Graph limits

Jan Hladký

SUMMARY

This thesis consists of the following papers:

- In *Part I: Topologies on the graphon space*,
  - [P1] Martin Doležal, Jan Hladký: *Cut-norm and entropy minimization over weak* limits*, Journal of Combinatorial Theory, Series B 137 (2019), 232–263.
  - [P2] Martin Doležal, Jan Grebík, Jan Hladký, Israel Rocha, Václav Rozhoň: *Relating the cut distance and the weak* topology for graphons*. Journal of Combinatorial Theory, Series B 147 (2021), 252–298.
  - [P3] Martin Doležal, Jan Grebík, Jan Hladký, Israel Rocha, Václav Rozhoň: *Cut distance identifying graphon parameters over weak* limits*. Journal of Combinatorial Theory, Series A 189 (2022), Paper No. 105615, 57 pp.
  - [P4] Frederik Garbe, Jan Hladký, Joonkyung Lee: *Two remarks on graph norms*. Discrete and Computational Geometry 67 (2022), no. 3, 919–929.
- In *Part II: Classical graph-theoretic concepts in graphons*,
  - [P5] Jan Hladký, Ping Hu, Diana Piguet: *Tilings in graphons*. European Journal of Combinatorics 93 (2021), Paper No. 103284, 23 pp.
  - [P6] Jan Hladký, Ping Hu, Diana Piguet: *Komlós's tiling theorem via graphon covers*. Journal of Graph Theory 90 (2019), no. 1, 24–45.
  - [P7] Martin Doležal, Jan Hladký: *Matching polytons*. Electronic Journal of Combinatorics 26 (2019), no. 4, Paper No. 4.38, 33 pp.
  - [P8] Jan Hladký, Israel Rocha: *Independent sets, cliques, and colorings in graphons*. European Journal of Combinatorics 88 (2020), 103108, 18 pp.
  - [P9] Jan Hladký, Eng Keat Hng: *Approximating fractionally isomorphic graphons.* European Journal of Combinatorics 113 (2023), 103751, 19 pp.
- In *Part III: Inhomogeneous random structures from graphons*,
  - [P10] Jan Hladký, Christos Pelekis, Matas Šileikis: *A limit theorem for small cliques in inhomogeneous random graphs*. Journal of Graph Theory 97 (2021), no. 4, 578–599.
  - [P11] Martin Doležal, Jan Hladký, András Máthé: *Cliques in dense inhomogeneous random graphs*. Random Structures and Algorithms 51 (2017), no. 2, 275–314.
  - [P12] Jan Hladký, Asaf Nachmias, Tuan Tran: *The local limit of the uniform spanning tree on dense graphs*. Journal of Statistical Physics 173 (2018), no. 3-4, 502–545.
  - [P13] Frederik Garbe, Jan Hladký, Matas Šileikis, Fiona Skerman: *From flip processes to dynamical systems on graphons*, Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques, accepted, available on journal webpage.
  - [P14] Pedro Araújo, Eng Keat Hng, Jan Hladký, Matas Šileikis: *Prominent examples of flip processes*, Random Structures and Algorithms, DOI: 10.1002/rsa.21192.
- In *Part IV: Limits of other discrete structures*,
  - [P15] Frederik Garbe, Robert Hancock, Jan Hladký, Maryam Sharifzadeh: *Limits of Latin squares,* Discrete Analysis 2023:8, 66pp.

[P16] Jan Hladký, András Máthé, Viresh Patel, Oleg Pikhurko: *Poset limits can be totally ordered*, Transactions of the American Mathematical Society, Volume 367 (2015), pages 4319-4337.

[P17] Frederik Garbe, Jan Hladký, Gábor Kun, Kristýna Pekárková: *On pattern-avoiding permutons*, Random Structures and Algorithms, DOI: 10.1002/rsa.21208.

All these papers relate and contribute to the theory of dense graph limits. The theory starts with work of Lovász and Szegedy [10] and of Borgs, Chayes, Lovász, Sós, and Vesztergombi [1], with preprints available in 2004 and 2007, respectively. In these two seminal papers, three fundamental concepts were introduced, namely, "graphons" which are measurable objects representing limits of sequences of finite graphs, the "cut distance" which is the metric underlying this limit notion, and "inhomogeneous random graphs", which in particular were used to show that the construction of the limit space is minimal possible, that is, that for every graphon there exists a sequence converging to it.

The main feature of the theory is that the cut distance is on one hand fine enough that many important combinatorial parameters (most importantly the so-called homomorphism densities) are continuous and on the other hand coarse enough that it is compact. This key compactness property was first proven by Lovász and Szegedy in their original paper [10] using Szemerédi's regularity lemma. Since then an alternative proof using nonstandard analysis was given in [2]. *Part I* of the thesis concerns the Lovász-Szegedy compactness theorem and related topological fundamentals of the theory of dense graph limits. In [P1] we found another proof of the Lovász-Szegedy compactness theorem based on the weak* topology. More specifically, we found a way of refining the weak* topology, which is known to be compact by the classical Banach-Alaoglu theorem but not fine enough for continuity of any interesting combinatorial parameters, to yield also compactness of the cut distance. The main tool in [P1] to this end was a certain construction which we called "entropy minimization". In [P2] and [P3] we extended this approach. In particular, [P2] contains a very abstract alternative proof of the Lovász-Szegedy compactness theorem. In [P3] we replaced entropy with the broadest possible class of parameters which fit in the proof scheme of [P1]. We call these parameters "cut distance identifying graphon parameters". We showed that these parameters are connected with the theory of the regularity lemma. Most importantly, essentially as a byproduct of our theory, we proved a conjecture of Kráľ, Martins, Pach and Wrochna [9] connecting graph norms and Sidorenko's conjecture. I will give more details of this result in the next paragraph.

The theory of dense graph limits has played a major role in several fields of graph theory. The most important applications appeared in extremal graph theory. The advantages of the limit view are severalfold. Firstly, the focus on the limit object allows to disregard lower order terms for certain parameters, allowing to express relations between them in a streamlined fashion. This feature is even more pronounced in Razborov's theory of "flag algebras" [12], which was developed in parallel to the theory of dense graph limits and which has led to breakthroughs on dozens of prominent extremal problems. The second advantage of dense graph limits is that the limit object automatically entails powers provided by the regularity lemma. Papers [P5] and [P6]

are an example of this benefit. It could be argued that while both above mentioned features are helpful, at least in theory proofs obtained by these methods could be “de-limitized”. That it, such proofs could be transferred to the language of finite graphs, at the cost of becoming more technical. Such as “de-limitization” is much more difficult if not impossible for the third advantage, namely the emergence of graph norms. Graph norms were introduced by Hatami in [5] as a class of norms which are defined on (a linear extension) of graphons.[1] Graph norms bring functional-analytic tools to extremal theory. Paper [P4] contains two results. The first one answers an open question from [5] about the moduli of convexity of graph norms, and the other is a small correction to the entire theory of graph norms as had been used since [5]. Already Hatami’s the first paper draws a clear potential of applicability of graph norms to Sidorenko’s conjecture. Sidorenko’s conjecture, one of the most notorious open problems in extremal graph theory, asserts that every bipartite graph satisfies a certain homomorphism density inequality. One way of proving Sidorenko’s conjecture for a particular bipartite graph is to prove that that graph has the norming property, as Hatami observed that each graph with the latter must also satisfy Sidorenko’s inequality. Indeed, one of Hatami’s main contributions was in proving that hypercubes have the norming property and are Sidorenko as a consequence, an addition of a new infinite family of bipartite graphs after previous 17 years of a stall. It is known that graph norms cannot answer Sidorenko’s conjecture (at least not in a straightforward fashion). That is, it is known that not all bipartite graphs are norming. However, answering a question of Kráľ, Martins, Pach and Wrochna from [9], in [P3] we almost characterized norming graphs as those that satisfy a stronger “step-Sidorenko inequality”.

Part II of the thesis is motivated by a very straightforward idea: If graphons extend finite graphs, then there should exist extensions to basic concepts from graph theory. Of course, this idea existed since the origins of the theory of dense graph limits. Homomorphism densities which were already mentioned are a notable example. However, many other concepts remained unexplored prior to our work. In [P7] and [P5] we investigated a counterpart to matchings and a generalization in which one asks for vertex-disjoint copies of a fixed graph F, often called F-tilings. In [P6] we used this theory to give an alternative proof of a theorem of Komlós [9] from extremal graph theory. The original proof of Komlós proceeds by an iterative applications of the regularity lemma, a proof scheme quite rare in extremal graph theory. By contrast, our proof in [P6] goes by a single application of LP duality. This difference in proofs reflects that the power of the regularity method is indeed contained in graphons, as mentioned earlier.

Paper [P8] deals with concepts of independent sets and colorings. Paper [P9] deals with fractional isomorphism. Fractional isomorphism in graphs is a prominent relaxation of isomorphism studied since the 1980s. Grebík and Rocha [3] developed a theory of fractional isomorphism for graphons paralleling that of graphs. An easy consequence of their theory is that if we have two sequences of mutually fractionally isomorphic graphs, that converge to two graphons then these graphons are fractionally isomorphic. In [P9] we prove the opposite direction of this relation: If we have two fractionally isomorphic graphons then there exist two sequences of mutually fractionally isomorphic graphs that converge to these two graphons.

1 In this simplified text we do not distinguish between *graph norms* and *weak graph norms*.

In *Part III* we study random structures coming from graphons. The simplest such random structure is the inhomogeneous random graph $\mathbf{G}(n, W)$, where W is a graphon. As we said, this random graph model appeared already in the first paper by Lovász and Szegedy [10] to show that the construction of the limit space is minimal possible. From purely a graph-limit perspective, this rather rudimentary result on $\mathbf{G}(n, W)$ is sufficient. However, further motivation comes from the theory of random graphs, which has been concerned for the last six decades with study of the Erdős-Rényi model $\mathbf{G}(n, p)$. It turns out that for p in the constant regime, the Erdős-Rényi model $\mathbf{G}(n, p)$ is just a special case of $\mathbf{G}(n, W)$. Note that $\mathbf{G}(n, p)$ for p constant is the most tractable one and most challenges emerge in sparser regimes such as $p=\log n/n$ or $p=C/n$. One of the main points I have been trying to make is that uncomplicatedness of $\mathbf{G}(n, p=\text{const})$ does not propagate to $\mathbf{G}(n, W)$, and that in fact $\mathbf{G}(n, W)$ can in itself encompass sparse random graphs. Papers [P10] and [P11] study properties of cliques in $\mathbf{G}(n, W)$. I believe that in particular [P11], which studies the size of the largest clique in $\mathbf{G}(n, W)$ and provides counterparts to results of Grimmett and McDiarmid from the 1970s, provides good evidence for the claim that $\mathbf{G}(n, W)$ is a much more complex model than $\mathbf{G}(n, p=\text{const})$.

Staying in *Part III*, papers [P12], [P13], and [P14] deal with different random objects than $\mathbf{G}(n, W)$. In particular, [P12] deals with the uniform spanning tree. The uniform spanning tree is a concept which plays a prominent role in modern probability theory, mostly because of the link between loop-erased random walk (which underlies the uniform spanning tree) and the Schramm-Loewner evolution. These connections have led to elegant theories for the uniform spanning trees on sparse graphs such as planar lattices or Cayley graphs. Our focus in [P12] is on dense graphs. Grimmett [4] showed in 1980 that from a local point of view, the uniform spanning tree of a large complete graph behaves like the Galton-Watson branching process with offspring distribution Poisson(1) conditioned on survival. In [P12] we obtain a generalization of this for any dense graph.

Papers [P13] and [P14] introduce and then study a new class of dynamics on graphs, which we call “flip processes”. Flip processes are given by a rule R of a given order k which is list of transformations of each k . The process starts with a given n-vertex graph G_0. In each step, the graph G_i is obtained by sampling k random vertices $v_1, \ldots, v_k$ of G_{i-1} and replacing the induced graph $F:=G_{i-1}[v_1,\ldots,v_k]$ by $R(F)$. This class contains several previously studied processes including the Erdős-Rényi random graph process and the triangle removal process. In [P13] we create a general theory of flip processes. For each flip process, we construct its “trajectories” which is a map having a graphon and a real time at input and a graphon at its output, formally $\Phi: \mathcal{W}\times[0,\infty) \to \mathcal{W}$. The main result asserts that if we start a flip process from a graph G_0 which is closed to a graphon U, then after $m=cn^2$ many steps, the random graph G_m will be typically close to the graphon $\Phi(U,c)$. This allows us to study evolution of flip processes using lenses of dynamical systems. Paper [P14] is a comprehensive study of several prominent processes.

The ideas that led to the theory of dense graph limits were picked to introduce limits of other discrete structures. *Part IV* of the thesis describes my contributions to these related theories. One of such theories, developed in [6], introduces “permutons” as limits of sequences of permutations. Combinatorics of permutation is a large field on its own. One of the most important lines of

research in combinatorics of permutations is the study of permutations avoiding a fixed pattern. The most celebrated enumerative result in this direction is a theorem of Marcus and Tardos [11] which asserts that the number of such permutations grows only exponentially. Our main contribution in [P17] is rather a structural one. It asserts that the structure of permutons avoiding a certain pattern is only one-dimensional, compered to the generally two-dimensional nature of permutons. Paper [P16] deals with posets. Their limit theory was developed by Janson [7]. Answering the main question from [7], in [P16] we showed that the limit object can be constructed in a more canonical way than Janson did. The last paper in Part *IV* is [P15]. Whereas [P16] and [P17] were contributions to already existing limit theories, [P15] works out a new theory, namely that of limits of Latin squares. The construction of "Latinons", limit objects associated to sequences of finite Latin squares, turns out to be much more complex than in the case of graphons and permutons.

I declare that my contribution to each of the papers [P1]-[P17] was substantial.

References

[1] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztergombi: *Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing*, Advances in Mathematics, Volume 219 (2008), pages 1801-1851.

[2] Gábor Elek, Balázs Szegedy: *A measure-theoretic approach to the theory of dense hypergraphs*, Advances in Mathematics, Volume 213 (2012), pages 1731-1772.

[3] Jan Grebík, Israel Rocha: *Fractional isomorphism of graphons*, Combinatorica, Volume 42 (2022), pages 365-404.

[4] G. R. Grimmett: *Random labelled trees and their branching networks*, Australian Mathematical Society. Journal. Series A. Pure Mathematics and Statistics, Volume 30, 1980/81, pages 229-237.

[5] Hamed Hatami: *Graph norms and Sidorenko's conjecture*, Israel Journal of Mathematics, Volume 175 (2010), pages 125-150.

[6] Carlos Hoppen, Yoshiharu Kohayakawa, Carlos Gustavo Moreira, Balázs Ráth, Rudini Menezes Sampaio: *Limits of permutation sequences*, Journal of Combinatorial Theory, Series B, Volume 103 (2013), pages 93-113.

[7] Svante Janson: *Poset limits and exchangeable random posets*, Combinatorica, Volume 31 (2011), pages 529-563.

[8] János Komlós: *Tiling Turán theorems*, Combinatorica, Volume 20 (2000), pages 203-218.

[9] Daniel Kráľ, Taísa L. Martins, Péter Pál Pach, Marcin Wrochna: *The step Sidorenko property and non-norming edge-transitive graphs*, Journal of Combinatorial Theory. Series A, Volume 162 (2019), pages 34-54.

[10] László Lovász, Balázs Szegedy: *Limits of dense graph sequences*, Journal of Combinatorial Theory. Series B, Volume 96 (2006), pages 933-957.

[11] Adam Marcus, Gábor Tardos: *Excluded permutation matrices and the Stanley-Wilf conjecture*, Journal of Combinatorial Theory, Series A, Volume 107 (2004), pages 153-160.

[12] Alexander A. Razborov: *Flag algebras*, Journal of Symbolic Logic, Volume 72 (2007), pages 1239-1282.

TABLE OF CONTENTS

Summary **i**

This table of contents **vii**

Part I: Topologies on the graphon space **1**
Cut-norm and entropy minimization over weak* limits 2
Relating the cut distance and the weak* topology for graphons 32
Cut distance identifying graphon parameters over weak* limits 77
Two remarks on graph norms 132

Part II: Classical graph-theoretic concepts in graphons **143**
Tilings in graphons 144
Komlós's tiling theorem via graphon covers 175
Matching polytons 200
Independent sets, cliques, and colorings in graphons 234
Approximating fractionally isomorphic graphons 258

Part III: Inhomogeneous random structures from graphons **283**
A limit theorem for small cliques in inhomogeneous random graphs 284
Cliques in dense inhomogeneous random graphs 309
The local limit of the uniform spanning tree on dense graphs 354
From flip processes to dynamical systems on graphons 406
Prominent examples of flip processes 471

Part IV: Limits of other discrete structures **525**
Limits of Latin squares 526
Poset limits can be totally ordered 594
On pattern-avoiding permutons 615

Part I

Topologies on the graphon space

CUT-NORM AND ENTROPY MINIMIZATION OVER WEAK* LIMITS

MARTIN DOLEŽAL AND JAN HLADKÝ

ABSTRACT. We prove that the accumulation points of a sequence of graphs $G_1, G_2, G_3, \dots$ with respect to the cut-distance are exactly the weak* limit points of subsequences of the adjacency matrices (when all possible orders of the vertices are considered) that minimize the entropy over all weak* limit points of the corresponding subsequence. In fact, the entropy can be replaced by any map $W \mapsto \iint f(W(x,y))$, where f is a continuous and strictly concave function. As a corollary, we obtain a new proof of compactness of the cut-distance topology.

1. INTRODUCTION

The theory of limits of dense graphs was developed in [19, 6] and has revolutionized graph theory since then. The key objects of the theory are so-called graphons. More precisely, a graphon is a symmetric Lebesgue measurable function from I^2 to $[0,1]$ where $I = [0,1]$ is the unit interval (equipped by the Lebesgue measure λ). In the heart of the theory is then the following statement.

Theorem 1 (Informally)**.** *Suppose that $G_1, G_2, G_3, \dots$ is a sequence of graphs. Then there exists a subsequence $G_{k_1}, G_{k_2}, G_{k_3}, \dots$ and a graphon $W : I^2 \to [0,1]$ such that $G_{k_1}, G_{k_2}, G_{k_3}, \dots$ converges to W.*

Roughly speaking, to obtain the graphon W one looks at the adjacency matrices of the graphs $(G_{k_n})_n$ from distance. One possible way an analyst might attempt to make this statement formal could be to take W as a weak* limit[1] of adjacency matrices of the graphs $(G_{k_n})_n$ represented as functions from I^2 to $\{0,1\}$. Such a version of Theorem 1 would be just an instance of the Banach–Alaoglu Theorem. However, the weak* topology turns out to be too coarse to provide the favorable properties that are available in the contemporary theory of graph limits.[2] A good toy example is the sequence of the complete balanced bipartite graphs $(K_{n,n})_{n=1}^{\infty}$. When considering adjacency matrices of these graphs with vertices grouped into the two parts of the bipartite graphs, the corresponding weak* limit is a 2×2-chessboard function with values 0 and 1, which we denote by $W_{\text{bipartite}}$. This turns out to be a desirable limit. On the other hand, one could consider adjacency matrices ordered differently. Ordering the vertices randomly, we get the constant $W_{\text{const}} \equiv \frac{1}{2}$ as the weak* limit (almost surely). We see that it is undesirable to get W_{const} as the limit object

Jan Hladký was supported by the Alexander von Humboldt Foundation. Research of Martin Doležal was supported by the GAČR project GA16-07378S..

[1] See the Appendix for basic information about the weak* topology.

[2] A primal example of such a favorable property is the continuity of subgraph densities.

as the only information carried by such an object is that the overall edge densities of the graphs along the sequence converge to $\frac{1}{2}$.

So, instead of the weak* topology one considers the so-called cut-norm topology, and this is also the topology to which "converges to W" in Theorem 1 refers. The cut-norm $\|\cdot\|_\square$ is a certain uniformization of the weak* topology. Indeed, recall that given symmetric measurable functions $\Gamma : I^2 \to [0,1]$ and $\Gamma_1, \Gamma_2, \Gamma_3, \ldots : I^2 \to [0,1]$, the two convergence notions compare as follows.

$$\Gamma_n \xrightarrow{\text{w}^*} \Gamma \qquad \Longleftrightarrow \qquad \sup_{B\subset I} \left\{ \limsup_n \left| \int_{x\in B} \int_{y\in B} \Gamma_n(x,y) - \Gamma(x,y) \right| \right\} = 0 \,,$$

$$\Gamma_n \xrightarrow{\|\cdot\|_\square} \Gamma \qquad \Longleftrightarrow \qquad \limsup_n \left\{ \sup_{B\subset I} \left| \int_{x\in B} \int_{y\in B} \Gamma_n(x,y) - \Gamma(x,y) \right| \right\} = 0 \,.$$

We shall state the formal version of Theorem 1 in a somewhat bigger generality for graphons. If $\Gamma, \Gamma' : I^2 \to [0,1]$ are two graphons then we say that they are *versions* of each other if they differ only by some measure-preserving transformation of I (see Section 2 for a precise definition).

Then the formal statement of Theorem 1 reads as follows.

Theorem 2. *Suppose that $\Gamma_1, \Gamma_2, \Gamma_3, \ldots : I^2 \to [0,1]$ is a sequence of graphons. Then there exists a sequence $k_1 < k_2 < k_3 < \cdots$ of natural numbers, versions $\Gamma'_{k_1}, \Gamma'_{k_2}, \Gamma'_{k_3}, \ldots$ of $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots$, and a graphon $W : I^2 \to [0,1]$ such that the sequence $\Gamma'_{k_1}, \Gamma'_{k_2}, \Gamma'_{k_3}, \ldots$ converges to W in the cut-norm.*

Prior to our work, there were three approaches to proving Theorem 2. One, taken in [19] and in [20], uses (variants of) the regularity lemma to group parts of I according to the structure of Γ_n. This way, one approximates the graphons by step-functions, and the limit graphon W is a limit of these step-functions.[3] A second approach, taken in [12], relies on ultraproduct techniques. This later approach is extremely technical, and was developed for the (more difficult) theory of limits of hypergraphs, where for some time the regularity approach was not available.[4] The third proof follows from the Aldous–Hoover theorem for exchangeable arrays ([1]). While the Aldous–Hoover theorem substantially precedes the theory of graph limits, the connection was realized substantially later by Diaconis and Janson, [9] and independently by Austin [4].

We present a fourth proof of Theorem 2. Our proof provides for the first time a characterization of the cut-norm convergence in terms of the weak* convergence. Namely, fixing any continuous and strictly concave function $f : [0,1] \to \mathbb{R}$, we prove that there is a

[3] A very general compactness result was given by Regts and Schrijver, [23]. This result in particular subsumes the compactness of the graphon space. Even when specialized to the space of graphons, there are differences of debatable significance between the proofs.

[4] A regularity approach to hypergraph limits was later found by Zhao, [27].

subsequence $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots$ such that the map $W \mapsto \iint f(W(x,y))$ attains its minimum on the space of all weak* accumulation points of versions of graphons $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots$, and that any such minimizer is an accumulation point of the sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ in the cut-distance. This result is consistent with our toy example above. Indeed, for any strictly concave function f we have $\iint f\left(W_{\text{bipartite}}(x,y)\right) < \iint f\left(W_{\text{const}}(x,y)\right)$ by Jensen's inequality. Jensen's inequality underlies the general proof of our result. This application of Jensen's inequality is in a sense analogous to the proof of the index-pumping lemma in proofs of the regularity lemma. We try to indicate this important link in Section 3.2.

1.1. **Statement of the main results.** Let $f : [0,1] \to \mathbb{R}$ be an arbitrary continuous and strictly concave function.[5] Given a graphon $\Gamma : I^2 \to [0,1]$, we write $\text{INT}_f(\Gamma) := \int_{x\in I} \int_{y\in I} f(\Gamma(x,y))$. When f is the binary entropy, the integration $\text{INT}_f(W)$ appears also in the work on large deviations in random graphs, [7] (which does not relate to the current work otherwise), and is called the *entropy of the graphon* W.[6]

For a sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots : I^2 \to [0,1]$ of graphons, we denote by $\mathbf{ACC}_{\text{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ the set of all functions $W : I^2 \to [0,1]$ for which there exist versions $\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots$ of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ such that W is a weak* accumulation point of the sequence $\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots$. We also denote by $\mathbf{LIM}_{\text{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ the set of all functions $W : I^2 \to [0,1]$ for which there exist versions $\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots$ of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ such that W is a weak* limit of the sequence $\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots$. We have $\mathbf{LIM}_{\text{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) \subset \mathbf{ACC}_{\text{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. Note that $\mathbf{LIM}_{\text{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ can be empty but $\mathbf{ACC}_{\text{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ cannot be empty by the sequential Banach–Alaoglu Theorem (see the Appendix for more details). Also, note that such weak* accumulation points (and thus also limits) are necessarily symmetric, Lebesgue measurable, $[0,1]$-valued, and thus graphons.

Our main result states that, given a sequence of graphons $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$, there is a subsequence $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots$ such that the minimum of $\text{INT}_f(\cdot)$ over the set $\mathbf{ACC}_{\text{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \ldots)$ is attained, and the graphon attaining this minimum is an accumulation point of the sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ in the cut-distance.

Theorem 3. *Suppose that $f : [0,1] \to \mathbb{R}$ is an arbitrary continuous and strictly concave function. Suppose that $\Gamma_1, \Gamma_2, \Gamma_3, \ldots : I^2 \to [0,1]$ is a sequence of graphons.*

(a) Suppose that $W \in \mathbf{ACC}_{\text{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ is not an accumulation point of the sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ in the cut-norm. Then there exists $\widetilde{W} \in \mathbf{ACC}_{\text{w}^*}(\Gamma_1, \Gamma_2, \ldots)$ such that $\text{INT}_f(\widetilde{W}) < \text{INT}_f(W)$.

[5] The additional assumption of the continuity of the concave function $f : [0,1] \to \mathbb{R}$ which we work with in this paper only means that f is continuous (from the appropriate sides) at 0 and at 1, so this is not a big extra restriction.

[6] As was pointed out to us by Svante Janson, Aldous ([2, p. 145]) worked with this quantity already in the 1980's in the context of exchangeability.

(b) There exist a subsequence $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots$ *and a graphon*

$$W_{\min} \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots)$$

such that

$$\mathrm{INT}_f(W_{\min}) = \inf\left\{\mathrm{INT}_f(W) \colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots)\right\} .$$

Clearly, Theorem 3 implies Theorem 2.

The proof of Theorem 3 is given in Sections 4 and 5.

To complete the "characterization of the cut-norm convergence in terms of the weak* convergence" advertised above, we prove that weak* limit points that do not minimize $\mathrm{INT}_f(\cdot)$ cannot be limit points in the cut-norm.

Proposition 4. *Suppose that* $f : [0,1] \to \mathbb{R}$ *is an arbitrary continuous and strictly concave function. Suppose that* $\Gamma_1, \Gamma_2, \Gamma_3, \ldots : I^2 \to [0,1]$ *is a sequence of graphons. If* $W \in \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ *is a cut-norm limit of versions of* $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ *then* W *is a minimizer of* $\mathrm{INT}_f(\cdot)$ *over the space* $\mathbf{LIM}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$.

In Section 4 we show that Proposition 4 is an easy consequence of a result of Borgs, Chayes, and Lovász [5] on uniqueness of graph limits. In addition, we give a self-contained proof.

2. Notation and tools

For every function $W : I^2 \to \mathbb{R}$, we define the *cut-norm* of W by

$$\|W\|_\square = \sup_A \left| \int_A \int_A W(x,y) \right| , \tag{1}$$

where A ranges over all measurable subsets of I. Another slightly different formula is also often used in the literature where one replaces the right-hand side of (1) by $\sup_{A,B} \left|\int_A \int_B W(x,y)\right|$ where two sets A and B range over all measurable subsets of I. However, it is easy to see that for every symmetric function W, we have

$$\sup_{A,B} \left| \int_A \int_B W(x,y) \right| \geq \sup_A \left| \int_A \int_A W(x,y) \right| \geq \frac{1}{2} \sup_{A,B} \left| \int_A \int_B W(x,y) \right| ,$$

and so the notion of convergence of sequences of graphons (which are symmetric) in the cut-norm is irrelevant to the choice between these two formulas.

We say that a graphon $\Gamma \colon I^2 \to [0,1]$ is a *step-graphon* with steps $I_1, I_2, \ldots, I_k \subset I$ if the sets $I_1, I_2, \ldots, I_k$ are pairwise disjoint, $I_1 \cup I_2 \cup \ldots \cup I_k = I$ and $W_{|I_i \times I_j}$ is constant (up to a null set) for every $i, j = 1, 2, \ldots, k$.

We say that a measurable function $\gamma : I \to I$ is an *almost-bijection* if there exist conull sets $J_1, J_2 \subset I$ such that $\gamma_{|J_1}$ is a bijection from J_1 onto J_2. When we talk about the inverse of such a function γ then we mean $(\gamma_{|J_1})^{-1}$ but we denote it only by γ^{-1}. Note

that this inverse γ^{-1} is not unique but that does not cause any problems as any two inverses of γ differ only on a null set.

If $\Gamma, \Gamma' : I^2 \to [0,1]$ are two graphons then we say that Γ' is a *version* of Γ if there exists a measure preserving almost-bijection $\gamma : I \to I$ such that $\Gamma'(x,y) = \Gamma(\gamma^{-1}(x), \gamma^{-1}(y))$ for almost every $(x,y) \in I^2$.

Related to versions, we recall that the *cut-distance* and L^1*-distance* between two graphons W_1, W_2 are defined as $\delta_\square(W_1, W_2) = \inf \|U_1 - W_2\|_\square$ and $\delta_1(W_1, W_2) = \inf \|U_1 - W_2\|_1$ where U_1 ranges over all versions of W_1.

By an *ordered partition* of I, we mean a partition of I with a fixed order of the sets from the partition. For an ordered partition $\mathcal{J}$ of I into finitely many sets $C_1, C_2, \ldots, C_k$, we define mappings $\alpha_{\mathcal{J},1}, \alpha_{\mathcal{J},2}, \ldots, \alpha_{\mathcal{J},k} : I \to I$, and a mapping $\gamma_\mathcal{J} : I \to I$ by

$$
\begin{aligned}
\alpha_{\mathcal{J},1}(x) &= \int_0^x \mathbf{1}_{C_1}(y)\,\mathrm{d}(y)\,, \\
\alpha_{\mathcal{J},2}(x) &= \alpha_{\mathcal{J},1}(1) + \int_0^x \mathbf{1}_{C_2}(y)\,\mathrm{d}(y)\,, \\
&\vdots \\
\alpha_{\mathcal{J},k}(x) &= \alpha_{\mathcal{J},1}(1) + \alpha_{\mathcal{J},2}(1) + \ldots + \alpha_{\mathcal{J},k-1}(1) + \int_0^x \mathbf{1}_{C_k}(y)\,\mathrm{d}(y)\,, \\
\gamma_\mathcal{J}(x) &= \alpha_{\mathcal{J},i}(x) \quad \text{if } x \in C_i, \quad i = 1, 2, \ldots, k\,.
\end{aligned} \tag{2}
$$

Informally, $\gamma_\mathcal{J}$ is defined in such a way that it maps the set C_1 to the left side of the interval I, the set C_2 next to it, and so on. Finally, the set C_k is mapped to the right side of the interval I. Clearly, $\gamma_\mathcal{J}$ is a measure preserving almost-bijection.

For a graphon $W : I^2 \to [0,1]$ and an ordered partition $\mathcal{J}$ of I into finitely many sets, we denote by ${}_\mathcal{J}W$ the version of W defined by ${}_\mathcal{J}W(x,y) = W(\gamma_\mathcal{J}^{-1}(x), \gamma_\mathcal{J}^{-1}(y))$ for every $(x,y) \in I^2$.

2.1. **Lebesgue points.** The Lebesgue density theorem asserts that given an integrable function $f : \mathbb{R}^n \to \mathbb{R}$, almost every point $x \in \mathbb{R}^n$ is a Lebesgue point of f, meaning that the value of $f(x)$ equals the limit of the averages of f on neighborhoods of x of diminishing sizes. There is some freedom in choosing the particular shapes of these neighborhoods. Below, we give a definition of Lebesgue points tailored to our purposes. Since we shall work with graphons, we state this definition for the domain I^2.

Definition 5. *Suppose that $W : I^2 \to \mathbb{R}$ is an integrable function. We say that $(x,y) \in I^2$ is a* Lebesgue point *of W if for every $\eta > 0$ there exists $\delta_0 > 0$ such that whenever $[p_1, p_2] \subset I$ and $[q_1, q_2] \subset I$ are intervals such that the length of the intervals is smaller or equal to δ_0, such that the ratio of the lengths of these intervals is at least $\frac{1}{2}$ and at most*

2, and such that $[p_1, p_2]$ *contains* x *and* $[q_1, q_2]$ *contains* y *then*

$$\left| W(x,y) - \frac{1}{(p_2-p_1)(q_2-q_1)} \int_{p_1}^{p_2} \int_{q_1}^{q_2} W(w,z)\,\mathrm{d}(w)\,\mathrm{d}(z) \right| < \eta\,. \tag{3}$$

We can now state the Lebesgue density theorem.

Theorem 6 (Lebesgue density theorem)**.** *Suppose that* $W : I^2 \to \mathbb{R}$ *is an integrable function. Then almost every point of* I^2 *is a Lebesgue point of* W*.*

2.2. **Stepping.** The next definition introduces graphons derived by an averaging of a given graphon W on a given partition of I. Here, we denote by $\lambda^{\oplus 2}$ the two-dimensional Lebesgue measure on I^2.

Definition 7. *Suppose that* $W : I^2 \to [0,1]$ *is a graphon. For a partition* $\mathcal{I}$ *of the unit interval into finitely many sets of positive measure,* $I = I_1 \sqcup I_2 \sqcup \ldots \sqcup I_k$*, we define a* stepping $W^{\ltimes \mathcal{I}}$ *which is defined on each rectangle* $I_i \times I_j$ *to be the constant* $\frac{1}{\lambda^{\oplus 2}(I_i \times I_j)} \int_{I_i} \int_{I_j} W(x,y)$*.*

The next lemma shows that we can replace any graphon W by its stepping (on some partition of I) without changing the value of $\mathrm{INT}_f(W)$ too much.

Lemma 8. *Let* $f : [0,1] \to \mathbb{R}$ *be an arbitrary continuous and strictly concave function, and let* $\mathcal{J}$ *be an arbitrary partition of* I *into finitely many intervals of positive measure. Suppose that* $W\colon I^2 \to [0,1]$ *is a graphon, and let* $\varepsilon > 0$*. Then there exists a partition* $\mathcal{I}$ *of* I *into finitely many intervals of positive measure such that* $\mathcal{I}$ *is a refinement of* $\mathcal{J}$ *and such that* $|\mathrm{INT}_f(W) - \mathrm{INT}_f(W^{\ltimes \mathcal{I}})| < \varepsilon$*.*

Proof. As f is continuous, there is $\eta > 0$ such that $|f(x) - f(y)| < \frac{1}{2}\varepsilon$ whenever $x, y \in [0,1]$ are such that $|x - y| < \eta$. Also, as W is an integrable function, almost every point $(x,y) \in I^2$ is a Lebesgue point of W. This implies that for a.e. $(x,y) \in I^2$ there is a natural number n such that whenever $[p_1, p_2] \subset I$ and $[q_1, q_2] \subset I$ are intervals of lengths smaller or equal to $\frac{2}{n}$ such that the ratio of the lengths is at least $\frac{1}{2}$ and at most 2, and such that $[p_1, p_2]$ contains x and $[q_1, q_2]$ contains y then inequality (3) holds. For every such (x,y), we denote by $n(x,y)$ the smallest n with this property. For every natural number n, we also put

$$D_n = \big\{(x,y) \in I^2 \colon n(x,y) > n\big\}\,.$$

Then it is easy to check that the sets $D_1 \supseteq D_2 \supseteq D_3 \supseteq \ldots$ are measurable and $\lambda^{\oplus 2}\left(\bigcap_{n=1}^{\infty} D_n\right) = 0$. So, after denoting $C := \max_{x \in [0,1]} |f(x)|$, we can find a natural number n_0 large enough such that

$$\lambda^{\oplus 2}\left(D_{n_0}\right) < \frac{1}{4C}\varepsilon\,, \tag{4}$$

and such that $\frac{1}{n_0}$ is smaller than the length of all intervals from the partition $\mathcal{J}$. Now let $\mathcal{I}$ be an arbitrary refinement of the partition $\mathcal{J}$ into finitely many intervals $I_1, I_2, \ldots, I_k$,

such that the length of each of these intervals is at least $\frac{1}{n_0}$ and at most $\frac{2}{n_0}$. For each $i,j = 1,2,\ldots,k$, we denote $C_{i,j} = \frac{1}{\lambda^{\oplus 2}(I_i \times I_j)} \int_{I_i}\int_{I_j} W(x,y)$. Inequality (3) then tells us that

$$|W(x,y) - C_{i,j}| < \eta \qquad \text{for every } (x,y) \in (I_i \times I_j) \setminus D_{n_0}, \quad i,j = 1,2,\ldots,k\,,$$

and so

$$(5)\ |f(W(x,y)) - f(C_{i,j})| < \tfrac{1}{2}\varepsilon \qquad \text{for every } (x,y) \in (I_i \times I_j) \setminus D_{n_0}, \quad i,j = 1,2,\ldots,k\,.$$

So we have

$$\begin{aligned}
&\left|\mathrm{INT}_f(W) - \mathrm{INT}_f(W^{\bowtie\mathcal{I}})\right| \\
&\leq \iint\limits_{D_{n_0}} \left|f(W(x,y)) - f(W^{\bowtie\mathcal{I}}(x,y))\right| + \sum_{i,j=1}^{k} \iint\limits_{(I_i\times I_j)\setminus D_{n_0}} \left|f(W(x,y)) - f(W^{\bowtie\mathcal{I}}(x,y))\right| \\
&\overset{(5)}{\leq} 2C \cdot \lambda^{\oplus 2}(D_{n_0}) + \frac{1}{2}\varepsilon \sum_{i,j=1}^{k} \lambda^{\oplus 2}\left((I_j \times I_j) \setminus D_{n_0}\right) \\
&\overset{(4)}{<} \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon\,,
\end{aligned}$$

as we wanted. □

The next lemma says that if a graphon is a weak* limit point then so is any graphon derived by an averaging of the original one on a given partition of I into intervals.

Lemma 9. *Suppose that $\Gamma_1, \Gamma_2, \Gamma_3, \ldots : I^2 \to [0,1]$ is a sequence of graphons. Suppose that $W \in \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ and that we have a partition $\mathcal{I}$ of I into finitely many intervals of positive measure. Then $W^{\bowtie\mathcal{I}} \in \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$.*

Moreover, whenever $\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots$ are versions of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ which converge to W in the weak topology then the versions $\Gamma''_1, \Gamma''_2, \Gamma''_3, \ldots$ of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ weak* converging to $W^{\bowtie\mathcal{I}}$ can be chosen in such a way that for every natural number j and for every intervals $K, L \in \mathcal{I}$ it holds*

$$\int_K \int_L \Gamma'_j(x,y) = \int_K \int_L \Gamma''_j(x,y)\,.$$

The proof of Lemma 9 follows a relatively standard probabilistic argument. Suppose for simplicity that $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ weak* converges to W. Then, for each n, we consider a version Γ'_n of Γ_n which is obtained by splitting each interval $A \in \mathcal{I}$ into n subsets of the same measure and then permuting these subsets of A at random. It can then be shown that $\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots$ converge to $W^{\bowtie\mathcal{I}}$ almost surely. The next two definitions are needed to make precise the notion of randomly permuting parts of the graphon within a given partition.

Definition 10. *Given a set $A \subset I$ of positive measure and a number $s \in \mathbb{N}$, we can consider a partition $A = [\![A]\!]^s_1 \sqcup [\![A]\!]^s_2 \sqcup \ldots \sqcup [\![A]\!]^s_s$, where each set $[\![A]\!]^s_i$ has measure $\frac{\lambda(A)}{s}$ and for each $1 \leq i < j \leq s$, the set $[\![A]\!]^s_i$ is entirely to the left of $[\![A]\!]^s_j$. These conditions define the partition $A = [\![A]\!]^s_1 \sqcup [\![A]\!]^s_2 \sqcup \ldots \sqcup [\![A]\!]^s_s$ uniquely, up to null sets. For each $i, j \in [s]$ there is a natural, uniquely defined (up to null sets), measure preserving almost-bijection $\chi^{A,s}_{i,j} : [\![A]\!]^s_i \to [\![A]\!]^s_j$ which preserves the order on the real line.*

Definition 11. *Suppose that $\Gamma : I^2 \to [0,1]$ is a graphon. For a partition $\mathcal{I}$ of I into finitely many sets of positive measure, $I = I_1 \sqcup I_2 \sqcup \ldots \sqcup I_k$, and for $s \in \mathbb{N}$, we define a discrete distribution $\mathbb{W}(\Gamma, \mathcal{I}, s)$ on graphons using the following procedure. We take $\pi_1, \ldots, \pi_k : [s] \to [s]$ independent uniformly random permutations. After these are fixed, we define a sample $W \sim \mathbb{W}(\Gamma, \mathcal{I}, s)$ by*

$$W(x,y) = \Gamma\left(\chi^{I_i,s}_{p,\pi_i(p)}(x), \chi^{I_j,s}_{q,\pi_j(q)}(y)\right) \quad \text{when } x \in [\![I_i]\!]^s_p,\ y \in [\![I_j]\!]^s_q,\ i,j \in [k],\ p,q \in [s]\ .$$

This defines the sample $W : I^2 \to [0,1]$ uniquely up to null sets, and thus defines the whole distribution $\mathbb{W}(\Gamma, \mathcal{I}, s)$. Observe that $\mathbb{W}(\Gamma, \mathcal{I}, s)$ is supported on (some) versions of Γ.

We call the sets $[\![I_j]\!]^s_q$ stripes.

Proof of Lemma 9. By considering suitable versions of the graphons Γ_n, we can without loss of generality assume that the sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ itself converges to W in the weak* topology. For each $n \in \mathbb{N}$, let us sample $U_n \sim \mathbb{W}(\Gamma_n, \mathcal{I}, n)$. We claim that the sequence $U_1, U_2, U_3, \ldots$ converges to $W^{\ltimes \mathcal{I}}$ in the weak* topology almost surely. As each U_n is a version of Γ_n, this will prove the lemma. So, let us now turn to proving the claim.

Let $i, j \in [k]$ be arbitrary. Further, let $0 \leq p_1 < p_2 \leq 1$ and $0 \leq r_1 < r_2 \leq 1$ be arbitrary rational numbers such that the rectangle $[p_1, p_2] \times [r_1, r_2]$ is contained (modulo a null set) in $I_i \times I_j$. Having fixed i, j, p_1, p_2, r_1, r_2, let us write c for the value of $W^{\ltimes \mathcal{I}}$ on $I_i \times I_j$. For each $n \in \mathbb{N}$, let E_n be the event that

$$\left| \iint_{[p_1,p_2]\times[r_1,r_2]} U_n \mathrm{d}(\lambda^{\oplus 2}) - c(p_2 - p_1)(r_2 - r_1) \right| > \sqrt[4]{1/n} + \tfrac{4}{n}\ .$$

Let us now bound the probability that E_n occurs. To this end, let Y_n be the value of

$$\iint_{[p_1,p_2]\times[r_1,r_2]} U_n \mathrm{d}(\lambda^{\oplus 2})\ .$$

We clearly have $\mathbb{E}[Y_n] = c(p_2 - p_1)(r_2 - r_1) \pm \frac{4}{n}$ (the error $\pm\frac{4}{n}$ comes from those products of pairs of stripes that intersect both $[p_1, p_2] \times [r_1, r_2]$ and its complement). Therefore, if E_n occurs then $|Y_n - \mathbb{E}[Y_n]| > \sqrt[4]{1/n}$. Suppose that we want to compute Y_n. From the k random permutations $\pi_1, \pi_2, \ldots, \pi_k : [n] \to [n]$ used in Definition 11 to define U_n, we only need to know the permutations π_i and π_j. To generate these, we toss in i.i.d. points $i_1, i_2, \ldots, i_n, j_1, j_2, \ldots, j_n$ into the unit interval I; the Euclidean order of the points

$i_1, i_2, \ldots, i_n$ naturally defines π_i and similarly the points $j_1, j_2, \ldots, j_n$ naturally define π_j.[7] So, we can view Y_n as a random variable on the probability space I^{2n}. Observe that if $\mathfrak{s} = (i_1, i_2, \ldots, i_n, j_1, j_2, \ldots, j_n)$ and $\mathfrak{s}' = (i'_1, i'_2, \ldots, i'_n, j'_1, j'_2, \ldots, j'_n)$ are two elements of I^{2n} that differ in only one coordinate, then $|Y_n(\mathfrak{s}) - Y_n(\mathfrak{s}')| \leq \frac{2}{n}$. Thus the Method of Bounded Differences (see [22]) tells us that

$$\mathbb{P}[E_n] \leq \mathbb{P}\left[|Y_n - \mathbb{E}[Y_n]| > \sqrt[4]{1/n}\right] \leq 2\exp\left(-\frac{2(\sqrt[4]{1/n})^2}{2n \cdot \left(\frac{2}{n}\right)^2}\right) = 2\exp\left(-\sqrt{n}/4\right) .$$

Because the sequence $(2\exp(-\sqrt{n}/4))_{n=1}^{\infty}$ is summable, the Borel–Cantelli lemma allows to conclude that only finitely many events E_n occur, almost surely. Thus, almost surely, for any weak* accumulation point U of the sequence $U_1, U_2, U_3, \ldots$, we have

$$\iint_{[p_1,p_2]\times[r_1,r_2]} U \mathrm{d}(\lambda^{\oplus 2}) = c(p_2 - p_1)(r_2 - r_1) . \tag{6}$$

By applying the union bound, we obtain that (6) holds for all (countably many) choices of i, j, p_1, p_2, r_1, r_2, almost surely. Since the elements of $\mathcal{I}$ are intervals, the above system of rectangles $[p_1, p_2] \times [r_1, r_2]$ generates the Borel σ-algebra on I^2. Consequently, we obtain that $U \equiv W^{\ltimes \mathcal{I}}$, almost surely.

The "moreover" part obviously follows from the proof. □

2.3. **Jensen's inequality and steppings.** Recall that one of the possible formulations of Jensen's inequality says that if (Ω, λ) is a measurable space with $\lambda(\Omega) > 0$, $g : \Omega \to \mathbb{R}$ is a measurable function and $f : \mathbb{R} \to \mathbb{R}$ is a concave function then

$$f\left(\frac{1}{\lambda(\Omega)} \int_\Omega g(x)\right) \geq \frac{1}{\lambda(\Omega)} \int_\Omega f(g(x)) . \tag{7}$$

We use this formulation of Jensen's inequality to prove the following simple lemma.

Lemma 12. *Let $f\colon [0,1] \to \mathbb{R}$ be a continuous and strictly concave function. Let $\Gamma\colon I^2 \to [0,1]$ be a step-graphon with steps $I_1, I_2, \ldots, I_k$, and let $W\colon I^2 \to [0,1]$ be another graphon such that $\int_{I_i \times I_j} W = \int_{I_i \times I_j} \Gamma$ for every $i, j = 1, 2, \ldots, k$. Then $\mathrm{INT}_f(W) \leq \mathrm{INT}_f(\Gamma)$.*

Proof. It clearly suffices to show that for every $i, j = 1, 2, \ldots, k$ it holds

$$\int_{I_i} \int_{I_j} f(W(x,y)) \leq \int_{I_i} \int_{I_j} f(\Gamma(x,y)) .$$

[7] The exception being when some of the points $i_1, i_2, \ldots, i_n$ or of the points $j_1, j_2, \ldots, j_n$ coincide, in which case the order of these points does not determine a permutation. This event however happens almost never.

So let us fix i, j, and let $C_{i,j}$ be the constant for which $\Gamma_{|I_i \times I_j} = C_{i,j}$ almost everywhere. Then we have

$$\begin{aligned}\int_{I_i}\int_{I_j} f(W(x,y)) &\overset{(7)}{\leq} \lambda^{\oplus 2}(I_i \times I_j) \cdot f\left(\frac{1}{\lambda^{\oplus 2}(I_i \times I_j)} \int_{I_i}\int_{I_j} W(x,y)\right) \\ &= \lambda^{\oplus 2}(I_i \times I_j) \cdot f\left(C_{i,j}\right) \\ &= \int_{I_i}\int_{I_j} f(\Gamma(x,y)) \,,\end{aligned}$$

as we wanted. □

3. Summaries of proofs

In this section, we give an overview of the proof of Theorem 3(a) in Section 3.1. Then, we explain in Section 3.2 that this proof can be viewed as an infinitesimal counterpart to the index-pumping lemma. Last, in Section 3.3 we give a detailed outline of Theorem 3(b).

3.1. **Overview of proof of Theorem 3(a).** Suppose for simplicity that the sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ converges to W in the weak* topology. The key step to the proof of Theorem 3(a) is Lemma 14. There we prove that whenever we fix a sequence $(B_n)_{n=1}^{\infty}$ of measurable subsets of I and define a new version Γ'_n of Γ_n (for every n) by "shifting the set B_n to the left side of the interval I", then any weak* accumulation point $\widetilde{W}$ of the sequence $\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots$ satisfies $\mathrm{INT}_f(\widetilde{W}) \leq \mathrm{INT}_f(W)$. As this result relies on Jensen's inequality, we actually get $\mathrm{INT}_f(\widetilde{W}) < \mathrm{INT}_f(W)$ when we choose the sets B_n carefully. "Carefully" means that each of the integrals $\int_{B_n}\int_{B_n} \Gamma_n(x,y)$ differs from the integral $\int_{B_n}\int_{B_n} W(x,y)$ at least by some given $\varepsilon > 0$. But observe that if the graphon W is not a cut-norm accumulation point of the sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ then it is always possible to choose the sets B_n.

3.2. **Connection between the proof of Theorem 3(a) and proofs of regularity lemmas.** Graphons could be regarded as "the ultimate regularization". Thus, it is instructive to see how our proof relates to the usual proofs of regularity lemmas (of which the weak regularity lemma of Frieze and Kannan [14] is the most relevant). Recall that in these proofs of regularity lemmas one keeps refining a partition of a graph until the partition is regular.

Let us give details. Let $f : [0,1] \to \mathbb{R}$ be an arbitrary continuous and strictly concave function. Suppose that G is an n-vertex graph, and let $\mathcal{P} = (P_i)_{i=1}^{k}$ be a partition of $V(G)$ into sets. Then for each $i, j \in [k]$, we define $d_{ij} := \frac{\sum_{u \in P_i, v \in P_j} \mathbf{1}_{uv \in E(G)}}{|P_i| \cdot |P_j|}$ (with the convention $\frac{0}{0} = 0$). If $i \neq j$ then d_{ij} corresponds to the bipartite density of the pair $G[P_i, P_j]$, and

otherwise this corresponds to the density of the graph $G[P_i]$. Then we write

$$\mathrm{INT}_f(G;\mathcal{P}) := \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{|P_i|\cdot|P_j|}{n^2}\cdot f(d_{ij})\ .$$

Note that we can express $\mathrm{INT}_f(G;\mathcal{P})$ as $\mathrm{INT}_f(W_{G;\mathcal{P}})$, where $W_{G;\mathcal{P}}$ is a graphon representation of densities of G according to the partition $\mathcal{P}$. The index-pumping lemma, which we state here in the setting of the weak regularity lemma, asserts that non-regular partitions can be refined while controlling the index. Let us recall that a partition $\mathcal{P} = (P_i)_{i=1}^k$ of $V(G)$ is *weak ϵ-regular* if for each $B \subset V(G)$ we have

$$e(G[B]) = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} d_{i,j}|B\cap P_i|\cdot|B\cap P_j| \pm \epsilon n^2\ .$$

Lemma 13 (Index-pumping lemma)**.** *Suppose that $\mathcal{C}$ is a partition of a graph G. If $B \subset V(G)$ is a witness that $\mathcal{C}$ is not ϵ-regular, then splitting each cell $C \in \mathcal{C}$ into $C\cap B$ and $C\setminus B$ yields a partition $\mathcal{D}$ for which*

$$\mathrm{INT}_{x\mapsto -x^2}(G;\mathcal{D}) < \mathrm{INT}_{x\mapsto -x^2}(G;\mathcal{C}) - \frac{\epsilon^2}{4}\ .$$

For completeness, let us recall the the proof of the weak regularity lemma, which states that for each $\epsilon > 0$, each graph has a weak ϵ-regular partition with at most $2^{\lceil \frac{4}{\epsilon^2}\rceil}$ parts. One starts with a singleton partition. At any stage, if the current partition is not weak ϵ-regular, then Lemma 13 allows to decrease the index by at least $\frac{\epsilon^2}{4}$ while doubling the number of cells in the partition. Since $\mathrm{INT}_{x\mapsto -x^2}(G;\cdot) \in [-1,0]$, we must terminate in at most $\lceil \frac{4}{\epsilon^2}\rceil$ steps.

Let us now draw the analogy between Lemma 13 and Theorem 3(a) and its proof. Firstly, note that Theorem 3 allows other functions than $x \mapsto -x^2$ used in Lemma 13. This is however not a serious restriction. Indeed, replacing the so-called "defect form of the Cauchy–Schwarz inequality" in the usual proof of Lemma 13 by Jensen's inequality, we could obtain a statement for general continuous strictly concave functions. In fact, this has already been used in [24], where a different choice of a concave function was necessary. So let us now move to the main analogy. Let us consider W as in Theorem 3(a). As in Section 3.1, let us assume that W is the weak* limit of $\Gamma_1, \Gamma_2, \Gamma_3 \ldots$. Suppose that W is not an accumulation point of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ with respect to the cut-norm, and let $B_1, B_2, B_3, \ldots \subset I$ be witnesses for this. Now, the "shifting B_n to the left" described in Section 3.1 can be viewed as splitting each interval $J \subset I$ into $J\cap B_n$ and $J\setminus B_n$ (we think of J as being very small, thus representing an "infinitesimally small cluster"), just as in Lemma 13.

We pose a conjecture which goes in this direction in Section 7.3.

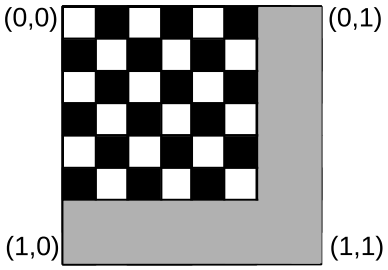


FIGURE 1. The graphon $\widetilde{W}_3$ from Section 3.3. Value 0 is white, value $\frac{1}{2}$ is gray, value 1 is black.

Last, let us mention that some sort of index-pumping arguments seem to be linked also with the Removal lemma, and in particular they appear in Fox's (regularity-lemma-free) quantitative improvement [13] (see also [8, Section 3.2]).

3.3. **Overview of proof of Theorem 3**(b)**.** Let us begin with the most straightforward attempt for a proof. For now, let us work with the simplifying assumption that all accumulation points are actually limits. As we shall see later, this simplifying assumption is a major cheat for which an extra patch will be needed. Then, let

$$m := \inf \{\mathrm{INT}_f(W) \colon W \in \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)\} \ .$$

For each $k \in \mathbb{N}$, let us fix a sequence $\Gamma_1^k, \Gamma_2^k, \Gamma_3^k, \ldots$ of versions of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ which converges in the weak* topology to a graphon $\widetilde{W}_k$ with $\mathrm{INT}_f(\widetilde{W}_k) < m + \frac{1}{k}$. Now, we might diagonalize and hope that any weak* accumulation point (whose existence is guaranteed by the Banach–Alaoglu Theorem) W^* of the sequence $\Gamma_1^1, \Gamma_2^2, \Gamma_3^3, \ldots$ satisfies $\mathrm{INT}_f(W^*) \leq m$. The reason for this hope being vain is the discontinuity of $\mathrm{INT}_f(\cdot)$ with respect to the weak* topology. As an example, let us take a situation when each $\widetilde{W}_k$ is a $2(k+2) \times 2(k+2)$-chessboard $\{0,1\}$-valued function, with the last two rows and columns having value $\frac{1}{2}$ (see Figure 1). In other words, most of each graphon $\widetilde{W}_k$ corresponds to a complete balanced bipartite graphon, to which an additional artificial subdivision to each of its parts to k subparts was introduced. These subparts were interlaced one after another, except that the vertices of the last subpart of each part were mixed together. (These graphons were clearly chosen nonoptimally in the sense that the mixing of the last two parts is undesired. We chose these graphons in this example here to have richer features to study.) All the graphons $\widetilde{W}_k$ have small values of $\mathrm{INT}_f(\cdot)$. On the other hand, the weak* limit of the sequence is the graphon $W_{\mathrm{const}} \equiv \frac{1}{2}$ whose value $\mathrm{INT}_f(\cdot)$ is bigger. There is a lesson to learn from this example. While for larger k, the versions in the sequence $\Gamma_1^k, \Gamma_2^k, \Gamma_3^k, \ldots$ will be aligned on I in a more optimal way locally, the global structure may get undesirably more convoluted as $k \to \infty$. To remedy this, we consider a sequence

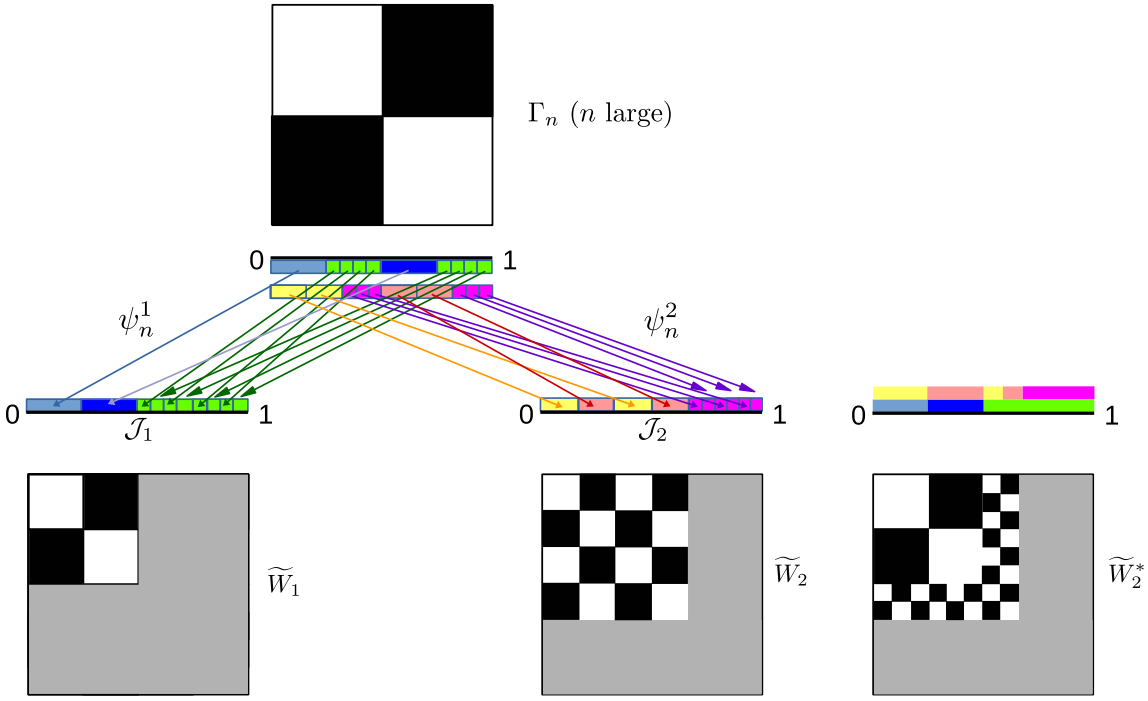


FIGURE 2. An example of reordering from Section 3.3. The top shows a graphon Γ_n, versions of which are close to $\widetilde{W}_1$ and $\widetilde{W}_2$ in the weak* topology. The two measure preserving transformations ψ_n^1 and ψ_n^2 which witness this closeness are shown with colors. The graphon $\widetilde{W}_2^*$ emerges by taking the partition whose global structure from $\widetilde{W}_1$ is refined according to the more local structure from $\widetilde{W}_2$. Iterating this process would lead to a sequence of graphons $(\widetilde{W}_k^*)_k$ which has the property that for any weak* accumulation point W^* we have $\mathrm{INT}_f(W^*) \leq \limsup_n \mathrm{INT}_f \widetilde{W}_n$.

of version of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ in which the structure of measure-preserving transformation on a rough level is inherited from measure preserving transformations leading to $\widetilde{W}_1$. Within each step corresponding to the step-graphon $\widetilde{W}_1$, the structure of the measure-preserving transformation is inherited from measure preserving transformations leading to $\widetilde{W}_2$, and so on. An example of this procedure is given in Figure 2. It can be shown that any weak* accumulation point W^* of these reordered graphons has the property that $\mathrm{INT}_f(W^*) \leq \limsup_n \mathrm{INT}_f \widetilde{W}_n$, as was needed.

Let us now explain why the assumption that all sequences converge weak* leaves a substantial gap in the proof. Recall that the information how the partition $\mathcal{J}^k$ of U_k interacts with the measure preserving almost-bijections on graphons $\mathfrak{s}_k \subset (\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ that converge to $\widetilde{W}_k$ gives us crucial directions as how to reorder and refine the subsequence

of graphons $\mathfrak{s}_{k+1}$ that converges to $\widetilde{W}_{k+1}$. Let us again stress that while the existence of the subsequences $\mathfrak{s}_j$ is guaranteed by weak* compactness, we have no control on their properties. So, it can be that $\mathfrak{s}_k$ is disjoint from $\mathfrak{s}_{k+1}$. In other words, we do not get the needed information how to reorder and refine the graphons in $\mathfrak{s}_{k+1}$. To remedy this problem, we prove a lemma (Lemma 16) which says that for every sequence $\Gamma_1, \Gamma_2, \Gamma_3, \dots$ of graphons there exists a subsequence $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \dots$ such that

$$\inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \dots)\} = \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \dots)\}\ .$$

Applying this lemma first, the arguments above become sound when we consider the subsequence $\Gamma_{k_1}, \Gamma_{k_2}, \dots$.

4. Proof of Theorem 3(a)

The following key lemma (or its subsequent corollary) is used in both proofs of Theorem 3(a) and Theorem 3(b).

Lemma 14. *Suppose that $f : [0,1] \to \mathbb{R}$ is an arbitrary continuous and strictly concave function. Suppose that $\Gamma_1, \Gamma_2, \Gamma_3, \dots : I^2 \to [0,1]$ is a sequence of graphons which converges to a graphon $W : I^2 \to [0,1]$ in the weak* topology. Suppose that $B_1, B_2, B_3, \dots$ is an arbitrary sequence of subsets of I. For each n, let $\mathcal{J}_n$ be the ordered partition of I into two sets B_n and $I \setminus B_n$ (in this order). Then every graphon $\widetilde{W}$ that is a weak* accumulation point of the sequence ${}_{\mathcal{J}_1}\Gamma_1, {}_{\mathcal{J}_2}\Gamma_2, {}_{\mathcal{J}_3}\Gamma_3, \dots$ satisfies $\mathrm{INT}_f(\widetilde{W}) \le \mathrm{INT}_f(W)$.*

Moreover, suppose that for the sequence $n_1 < n_2 < n_3 < \dots$ for which the sequence ${}_{\mathcal{J}_{n_1}}\Gamma_{n_1}, {}_{\mathcal{J}_{n_2}}\Gamma_{n_2}, {}_{\mathcal{J}_{n_3}}\Gamma_{n_3}, \dots$ weak converges to $\widetilde{W}$, we have that $\mathbf{1}_{B_{n_1}}, \mathbf{1}_{B_{n_2}}, \mathbf{1}_{B_{n_3}}, \dots$ converges to a function $\psi\colon I \to [0,1]$ in the weak* topology. Let $\theta\colon I \to I$ be defined by $\theta(x) = \int_0^x \psi(y)\,\mathrm{d}(y)$. If we have*

$$\lambda^{\otimes 2}\left(\left\{(x,y) \in I^2 \,:\, \psi(x) > 0, \psi(y) > 0, W(x,y) \neq \widetilde{W}(\theta(x),\theta(y))\right\}\right) > 0 \tag{8}$$

then $\mathrm{INT}_f(\widetilde{W}) < \mathrm{INT}_f(W)$.

Proof. By passing to a subsequence, we may assume that the sequence ${}_{\mathcal{J}_1}\Gamma_1, {}_{\mathcal{J}_2}\Gamma_2, {}_{\mathcal{J}_3}\Gamma_3, \dots$ is convergent to $\widetilde{W}$ in the weak* topology, and that the sequence $\mathbf{1}_{B_1}, \mathbf{1}_{B_2}, \mathbf{1}_{B_3}, \dots$ converges in the weak* topology to $\psi\colon I \to [0,1]$. We define $\xi\colon I \to I$ by $\xi(x) = \theta(1) + \int_0^x (1 - \psi(y))\,\mathrm{d}(y)$.

Claim 1. *For every two intervals $[p_1,p_2],[q_1,q_2]\subset I$ we have*

$$\begin{aligned}\int_{p_1}^{p_2}\int_{q_1}^{q_2} W(x,y) &= \int_{p_1}^{p_2}\int_{q_1}^{q_2} \widetilde{W}(\theta(x),\theta(y))\psi(x)\psi(y)\\ &\quad+\int_{p_1}^{p_2}\int_{q_1}^{q_2} \widetilde{W}(\theta(x),\xi(y))\psi(x)(1-\psi(y))\\ &\quad+\int_{p_1}^{p_2}\int_{q_1}^{q_2} \widetilde{W}(\xi(x),\theta(y))(1-\psi(x))\psi(y)\\ &\quad+\int_{p_1}^{p_2}\int_{q_1}^{q_2} \widetilde{W}(\xi(x),\xi(y))(1-\psi(x))(1-\psi(y))\,.\end{aligned} \tag{9}$$

Proof of Claim 1. By using the fact that $\Gamma_n \overset{w^*}{\to} W$ together with the identity $ab+a(1-b)+(1-a)b+(1-a)(1-b)=1$ we get that

$$\begin{aligned}\int_{p_1}^{p_2}\int_{q_1}^{q_2} W(x,y) &= \lim_{n\to\infty}\int_{p_1}^{p_2}\int_{q_1}^{q_2} \Gamma_n(x,y)\\ &= \lim_{n\to\infty}\int_{p_1}^{p_2}\int_{q_1}^{q_2} \Gamma_n(x,y)\mathbf{1}_{B_n}(x)\mathbf{1}_{B_n}(y)\\ &\quad+\lim_{n\to\infty}\int_{p_1}^{p_2}\int_{q_1}^{q_2} \Gamma_n(x,y)\mathbf{1}_{B_n}(x)(1-\mathbf{1}_{B_n}(y))\\ &\quad+\lim_{n\to\infty}\int_{p_1}^{p_2}\int_{q_1}^{q_2} \Gamma_n(x,y)(1-\mathbf{1}_{B_n}(x))\mathbf{1}_{B_n}(y)\\ &\quad+\lim_{n\to\infty}\int_{p_1}^{p_2}\int_{q_1}^{q_2} \Gamma_n(x,y)(1-\mathbf{1}_{B_n}(x))(1-\mathbf{1}_{B_n}(y))\,.\end{aligned} \tag{10}$$

Next we rewrite the integral following the first limit on the right-hand side of (10). To this end, we use the notation from (2) together with the obvious differentiation formula

$$(\alpha_{\mathcal{J}_n,1})'(x)=\mathbf{1}_{B_n}(x)\quad\text{for a.e. } x\in I \tag{11}$$

(and also, we use the fact that $\alpha_{\mathcal{J}_n,1|B_n}$ is an almost-bijection from B_n onto the interval $[0,\int_0^1\mathbf{1}_{B_n}(y)]$, and so it makes sense to talk about its inverse). We have

$$\begin{aligned}&\int_{p_1}^{p_2}\int_{q_1}^{q_2} \Gamma_n(x,y)\mathbf{1}_{B_n}(x)\mathbf{1}_{B_n}(y)\\ \boxed{\text{integration by substitution}}\quad&=\int_{\alpha_{\mathcal{J}_n,1}(p_1)}^{\alpha_{\mathcal{J}_n,1}(p_2)}\int_{\alpha_{\mathcal{J}_n,1}(q_1)}^{\alpha_{\mathcal{J}_n,1}(q_2)} \Gamma_n(\alpha^{-1}_{\mathcal{J}_n,1}(x),\alpha^{-1}_{\mathcal{J}_n,1}(y))\\ \boxed{\gamma_{\mathcal{J}_n}(x)=\alpha_{\mathcal{J}_n,1}(x)\text{ for every } x\in B_n}\quad&=\int_{\alpha_{\mathcal{J}_n,1}(p_1)}^{\alpha_{\mathcal{J}_n,1}(p_2)}\int_{\alpha_{\mathcal{J}_n,1}(q_1)}^{\alpha_{\mathcal{J}_n,1}(q_2)} \Gamma_n(\gamma^{-1}_{\mathcal{J}_n}(x),\gamma^{-1}_{\mathcal{J}_n}(y))\\ &=\int_{\alpha_{\mathcal{J}_n,1}(p_1)}^{\alpha_{\mathcal{J}_n,1}(p_2)}\int_{\alpha_{\mathcal{J}_n,1}(q_1)}^{\alpha_{\mathcal{J}_n,1}(q_2)} {}_{\mathcal{J}_n}\Gamma_n(x,y)\,.\end{aligned} \tag{12}$$

Therefore, we have

(13)
$$\begin{aligned}
&\left|\int_{p_1}^{p_2}\int_{q_1}^{q_2}\Gamma_n(x,y)\mathbf{1}_{B_n}(x)\mathbf{1}_{B_n}(y)-\int_{\theta(p_1)}^{\theta(p_2)}\int_{\theta(q_1)}^{\theta(q_2)}\mathcal{J}_n\Gamma_n(x,y)\right| \\
\overset{(12)}{=}&\left|\int_{\alpha_{\mathcal{J}_n,1}(p_1)}^{\alpha_{\mathcal{J}_n,1}(p_2)}\int_{\alpha_{\mathcal{J}_n,1}(q_1)}^{\alpha_{\mathcal{J}_n,1}(q_2)}\mathcal{J}_n\Gamma_n(x,y)-\int_{\theta(p_1)}^{\theta(p_2)}\int_{\theta(q_1)}^{\theta(q_2)}\mathcal{J}_n\Gamma_n(x,y)\right| \\
\leq&|\alpha_{\mathcal{J}_n,1}(p_1)-\theta(p_1)|+|\alpha_{\mathcal{J}_n,1}(p_2)-\theta(p_2)|+|\alpha_{\mathcal{J}_n,1}(q_1)-\theta(q_1)|+|\alpha_{\mathcal{J}_n,1}(q_2)-\theta(q_2)|\ .
\end{aligned}$$

The fact that $\mathbf{1}_{B_n}\overset{w^*}{\to}\psi$ immediately implies that $\alpha_{\mathcal{J}_n,1}(x)\to\theta(x)$ for every $x\in I$, and so we conclude that the right-hand side, and thus also the left-hand side, of (13), tends to 0. Therefore (note that the following limits exist as $\Gamma_n\overset{w^*}{\to}W$)

(14)
$$\begin{aligned}
\lim_{n\to\infty}\int_{p_1}^{p_2}\int_{q_1}^{q_2}\Gamma_n(x,y)\mathbf{1}_{B_n}(x)\mathbf{1}_{B_n}(y)&=\lim_{n\to\infty}\int_{\theta(p_1)}^{\theta(p_2)}\int_{\theta(q_1)}^{\theta(q_2)}\mathcal{J}_n\Gamma_n(x,y) \\
\boxed{\mathcal{J}_n\Gamma_n\overset{w^*}{\to}\widetilde{W}}\quad&=\int_{\theta(p_1)}^{\theta(p_2)}\int_{\theta(q_1)}^{\theta(q_2)}\widetilde{W}(x,y) \\
\boxed{\text{integration by substitution}}\quad&=\int_{p_1}^{p_2}\int_{q_1}^{q_2}\widetilde{W}(\theta(x),\theta(y))\psi(x)\psi(y)\ .
\end{aligned}$$

In a very analogous way as we derived (14), one can verify that

(15)
$$\lim_{n\to\infty}\int_{p_1}^{p_2}\int_{q_1}^{q_2}\Gamma_n(x,y)\mathbf{1}_{B_n}(x)(1-\mathbf{1}_{B_n}(y))=\int_{p_1}^{p_2}\int_{q_1}^{q_2}\widetilde{W}(\theta(x),\xi(y))\psi(x)(1-\psi(y))\ ,$$

(16)
$$\lim_{n\to\infty}\int_{p_1}^{p_2}\int_{q_1}^{q_2}\Gamma_n(x,y)(1-\mathbf{1}_{B_n}(x))\mathbf{1}_{B_n}(y)=\int_{p_1}^{p_2}\int_{q_1}^{q_2}\widetilde{W}(\xi(x),\theta(y))(1-\psi(x))\psi(y)\ ,$$

(17)
$$\lim_{n\to\infty}\int_{p_1}^{p_2}\int_{q_1}^{q_2}\Gamma_n(x,y)(1-\mathbf{1}_{B_n}(x))(1-\mathbf{1}_{B_n}(y))=\int_{p_1}^{p_2}\int_{q_1}^{q_2}\widetilde{W}(\xi(x),\xi(y))(1-\psi(x))(1-\psi(y))\ .$$

By putting (10), (14), (15), (16) and (17) together, we get (9). □

Since the sets of the form $[p_1, p_2] \times [q_1, q_2]$ generate the Borel σ-algebra on I^2, we conclude from Claim 1 that for almost every $(x, y) \in I^2$ we have that

$$\begin{aligned} W(x,y) =& \widetilde{W}(\theta(x),\theta(y))\psi(x)\psi(y) + \widetilde{W}(\theta(x),\xi(y))\psi(x)(1-\psi(y)) \\ &+ \widetilde{W}(\xi(x),\theta(y))(1-\psi(x))\psi(y) + \widetilde{W}(\xi(x),\xi(y))(1-\psi(x))(1-\psi(y)) \,. \end{aligned} \tag{18}$$

Note that the right-hand side of (18) is a convex combination of the four terms

$$\widetilde{W}(\theta(x),\theta(y)) \,, \quad \widetilde{W}(\theta(x),\xi(y)) \,, \quad \widetilde{W}(\xi(x),\theta(y)) \,, \quad \widetilde{W}(\xi(x),\xi(y)) \,. \tag{19}$$

Therefore we have

$$\begin{aligned} \mathrm{INT}_f(W) &= \int_0^1\int_0^1 f(W(x,y)) \\ \boxed{f \text{ is concave}} \quad &\overset{(18)}{\geq} \int_0^1\int_0^1 f\left(\widetilde{W}(\theta(x),\theta(y))\right)\psi(x)\psi(y) \\ &\quad + \int_0^1\int_0^1 f\left(\widetilde{W}(\theta(x),\xi(y))\right)\psi(x)(1-\psi(y)) \\ &\quad + \int_0^1\int_0^1 f\left(\widetilde{W}(\xi(x),\theta(y))\right)(1-\psi(x))\psi(y) \\ &\quad + \int_0^1\int_0^1 f\left(\widetilde{W}(\xi(x),\xi(y))\right)(1-\psi(x))(1-\psi(y)) \\ \boxed{\text{integration by substitution}} \quad &= \int_0^{\theta(1)}\int_0^{\theta(1)} f\left(\widetilde{W}(x,y)\right) + \int_0^{\theta(1)}\int_{\theta(1)}^1 f\left(\widetilde{W}(x,y)\right) \\ &\quad + \int_{\theta(1)}^1\int_0^{\theta(1)} f\left(\widetilde{W}(x,y)\right) + \int_{\theta(1)}^1\int_{\theta(1)}^1 f\left(\widetilde{W}(x,y)\right) \\ &= \int_0^1\int_0^1 f\left(\widetilde{W}(x,y)\right) = \mathrm{INT}_f(\widetilde{W}) \,. \end{aligned} \tag{20}$$

To prove the "moreover" part, suppose that we have (8). Then the convex combination (18) is not trivial on a set of positive measure. This is all we need as then we have a sharp inequality in (20) because f is strictly concave. □

We do not use the next corollary right now but we will need it in Section 5.

Corollary 15. *Suppose that $f : [0,1] \to \mathbb{R}$ is an arbitrary continuous and strictly concave function. Suppose that $\Gamma_1, \Gamma_2, \Gamma_3, \ldots : I^2 \to [0,1]$ is a sequence of graphons which converges to a graphon $W : I^2 \to [0,1]$ in the weak* topology. Suppose that ℓ is a fixed natural number and that for every n, $\mathcal{J}_n$ is an ordered partition of I into ℓ sets $B_1^n, B_2^n, \ldots, B_\ell^n$. Then for every graphon $\widetilde{W}$ that is a weak* accumulation point of the graphons ${}_{\mathcal{J}_1}\Gamma_1, {}_{\mathcal{J}_2}\Gamma_2, {}_{\mathcal{J}_3}\Gamma_3, \ldots$ we have $\mathrm{INT}_f(\widetilde{W}) \leq \mathrm{INT}_f(W)$.*

Proof. For every natural number n and every $i \in \{1, \dots, \ell\}$, we denote by $\mathcal{J}_n^i$ the ordered partition of I consisting of the sets $B_{\ell-i+1}^n, B_{\ell-i+2}^n, \dots, B_\ell^n$ and $I \setminus \bigcup_{j=\ell-i+1}^{\ell} B_j^n$ (in this order). Consider these $\ell+1$ sequences of graphons:

$$\begin{aligned}
\mathcal{S}_0 &\colon \Gamma_1, \Gamma_2, \Gamma_3, \dots \\
\mathcal{S}_1 &\colon {}_{\mathcal{J}_1^1}\Gamma_1, {}_{\mathcal{J}_2^1}\Gamma_2, {}_{\mathcal{J}_3^1}\Gamma_3, \dots \\
\mathcal{S}_2 &\colon {}_{\mathcal{J}_1^2}\Gamma_1, {}_{\mathcal{J}_2^2}\Gamma_2, {}_{\mathcal{J}_3^2}\Gamma_3, \dots \\
&\vdots \\
\mathcal{S}_\ell &\colon {}_{\mathcal{J}_1^\ell}\Gamma_1, {}_{\mathcal{J}_2^\ell}\Gamma_2, {}_{\mathcal{J}_3^\ell}\Gamma_3, \dots \ ,
\end{aligned}$$

so that the sequence $\mathcal{S}_\ell$ is precisely ${}_{\mathcal{J}_1}\Gamma_1, {}_{\mathcal{J}_2}\Gamma_2, {}_{\mathcal{J}_3}\Gamma_3, \dots$. Let us fix $\widetilde{W} \in \mathbf{ACC}_{\mathrm{w}^*}(\mathcal{S}_\ell)$. By passing to a subsequence, we may assume that the sequence $\mathcal{S}_i$ converges to some graphon W_i in the weak* topology for every $i = 1, 2, \dots, \ell-1$. It remains to apply Lemma 14 ℓ-times in a row. First, we apply it on the sequence $\mathcal{S}_0$ of graphons and on the sequence $B_\ell^1, B_\ell^2, B_\ell^3, \dots$ of subsets of I to conclude that $\mathrm{INT}_f(W_1) \le \mathrm{INT}_f(W)$. Next, we apply it on the sequence $\mathcal{S}_1$ of graphons and on the sequence $B_{\ell-1}^1, B_{\ell-1}^2, B_{\ell-1}^3, \dots$ of subsets of I to conclude that $\mathrm{INT}_f(W_2) \le \mathrm{INT}_f(W_1) \le \mathrm{INT}_f(W)$. In the last step, we apply it on the sequence $\mathcal{S}_{\ell-1}$ of graphons and on the sequence $B_1^1, B_1^2, B_1^3, \dots$ of subsets of I to conclude that $\mathrm{INT}_f(\widetilde{W}) \le \mathrm{INT}_f(W_{\ell-1}) \le \dots \le \mathrm{INT}_f(W_1) \le \mathrm{INT}_f(W)$. □

Now we can prove Theorem 3(a).

By passing to a subsequence, we may assume that the sequence $\Gamma_1, \Gamma_2, \Gamma_3, \dots$ converges to W in the weak* topology. As W is not an accumulation point of the sequence $\Gamma_1, \Gamma_2, \Gamma_3, \dots$ in the cut-norm, there is $\varepsilon > 0$ and a natural number n_0 such that $\|\Gamma_n - W\|_\square \ge \varepsilon$ for every $n \ge n_0$. By passing to a subsequence, we may suppose that $\|\Gamma_n - W\|_\square \ge \varepsilon$ for every natural number n. By the definition of the cut-norm, there is a sequence $B_1, B_2, B_3, \dots$ of subsets of I such that for every natural number n we have $\left|\int_{x\in B_n}\int_{y\in B_n}(\Gamma_n(x,y) - W(x,y))\right| \ge \varepsilon$. This means that either

$$\begin{aligned}
\int_{x\in B_n}\int_{y\in B_n}\Gamma_n(x,y) &\ge \int_{x\in B_n}\int_{y\in B_n}W(x,y) + \varepsilon \quad \text{or} \\
\int_{x\in B_n}\int_{y\in B_n}\Gamma_n(x,y) &\le \int_{x\in B_n}\int_{y\in B_n}W(x,y) - \varepsilon\ .
\end{aligned} \tag{21}$$

By passing to a subsequence, we may assume that only one of these two cases occurs. We stick to the case when (21) holds for every natural number n (the other case is analogous). By passing to a subsequence once again, we may assume that the sequence $\mathbf{1}_{B_1}, \mathbf{1}_{B_2}, \mathbf{1}_{B_3}, \dots$ converges in the weak* topology to some $\psi\colon I \to [0,1]$. For every natural number n, let $\mathcal{J}_n$ be the ordered partition of I into two sets B_n and $I \setminus B_n$ (in this order). This allows us to

define $\alpha_{\mathcal{J}_n,1}, \alpha_{\mathcal{J}_n,2}, \gamma_{\mathcal{J}_n} : I \to I$ as in (2), and versions ${}_{\mathcal{J}_1}\Gamma_1, {}_{\mathcal{J}_2}\Gamma_2, {}_{\mathcal{J}_3}\Gamma_3, \dots$ of $\Gamma_1, \Gamma_2, \Gamma_3, \dots$. We pass to a subsequence again to assure that the sequence ${}_{\mathcal{J}_1}\Gamma_1, {}_{\mathcal{J}_2}\Gamma_2, {}_{\mathcal{J}_3}\Gamma_3, \dots$ is convergent in the weak* topology, and we denote the weak* limit by $\widetilde{W}$. Now Lemma 14 tells us that $\mathrm{INT}_f(\widetilde{W}) \le \mathrm{INT}_f(W)$, and that to prove that this inequality is sharp we only need to verify (8). So to complete the proof, it suffices to prove the following claim.

Claim 2. *We have*

$$\int_0^1 \int_0^1 \widetilde{W}(\theta(x),\theta(y))\psi(x)\psi(y) \ge \int_0^1 \int_0^1 W(x,y)\psi(x)\psi(y) + \tfrac{1}{2}\varepsilon \,.$$

Proof of Claim 2. We have

$$\begin{aligned}
\int_0^1 \int_0^1 \widetilde{W}(\theta(x),\theta(y))\psi(x)\psi(y) &= \int_0^{\theta(1)} \int_0^{\theta(1)} \widetilde{W}(x,y) \\
\boxed{{}_{\mathcal{J}_n}\Gamma_n \overset{w^*}{\to} \widetilde{W}} \quad &= \lim_{n\to\infty} \int_0^{\theta(1)} \int_0^{\theta(1)} {}_{\mathcal{J}_n}\Gamma_n(x,y) \\
\boxed{\text{for large enough } n, \text{ as } \alpha_{\mathcal{J}_n,1}(1) \to \theta(1)} \quad &\ge \limsup_{n\to\infty} \int_0^{\alpha_{\mathcal{J}_n,1}(1)} \int_0^{\alpha_{\mathcal{J}_n,1}(1)} {}_{\mathcal{J}_n}\Gamma_n(x,y) - \tfrac{1}{2}\varepsilon \\
\boxed{\text{integration by substitution}} \quad &= \limsup_{n\to\infty} \int_0^1 \int_0^1 {}_{\mathcal{J}_n}\Gamma_n(\alpha_{\mathcal{J}_n,1}(x),\alpha_{\mathcal{J}_n,1}(y))\mathbf{1}_{B_n}(x)\mathbf{1}_{B_n}(y) - \tfrac{1}{2}\varepsilon \\
\boxed{\gamma_{\mathcal{J}_n}(x) = \alpha_{\mathcal{J}_n,1}(x) \text{ for every } x \in B_n} \quad &= \limsup_{n\to\infty} \int_{B_n} \int_{B_n} {}_{\mathcal{J}_n}\Gamma_n(\gamma_{\mathcal{J}_n}(x),\gamma_{\mathcal{J}_n}(y)) - \tfrac{1}{2}\varepsilon \\
&= \limsup_{n\to\infty} \int_{B_n} \int_{B_n} \Gamma_n(x,y) - \tfrac{1}{2}\varepsilon \\
&\overset{(21)}{\ge} \limsup_{n\to\infty} \int_{B_n} \int_{B_n} W(x,y) + \tfrac{1}{2}\varepsilon \\
\boxed{\mathbf{1}_{B_n} \overset{w^*}{\to} \psi} \quad &= \int_0^1 \int_0^1 W(x,y)\psi(x)\psi(y) + \tfrac{1}{2}\varepsilon \,.
\end{aligned}$$

□

Remark 1. *The initial step when we "shift the sets B_n to the left" crucially relies on the Euclidean order on I. This order is needless for the theory of graphons, i.e., graphons can be defined on a square of an arbitrary atomless separable probability space Ω. A linear order on Ω can be always introduced additionally, as Ω is measure-isomorphic to I. So, while our results work in full generality for an arbitrary Ω, we wonder if our argument can be modified so that the proof would naturally work without assuming a linear structure of the underlying probability space.*

5. Proof of Theorem 3(*b*)

The bulk of the proof is given after proving the following key lemma.

Lemma 16. *For every sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots : I^2 \to [0,1]$ of graphons there exists a subsequence $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots$ such that*

$$\inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \ldots)\} = \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \ldots)\} \ .$$

Proof. We start by finding countably many subsequences $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \ldots$ of the sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ such that for every natural number n we have:

(i) $\mathcal{S}_{n+1}$ is a subsequence of $\mathcal{S}_n$, and
(ii) there exists $W_{n+1} \in \mathbf{LIM}_{\mathrm{w}^*}(\mathcal{S}_{n+1})$ such that

$$\mathrm{INT}_f(W_{n+1}) < \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\mathcal{S}_n)\} + \tfrac{1}{n} \ . \tag{22}$$

This is done by induction. In the first step, we just define the sequence $\mathcal{S}_1$ to be the original sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$. Next suppose that we have already defined the subsequence $\mathcal{S}_n$ for some natural number n. Then there is a graphon $W_{n+1} \in \mathbf{ACC}_{\mathrm{w}^*}(\mathcal{S}_n)$ such that

$$\mathrm{INT}_f(W_{n+1}) < \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\mathcal{S}_n)\} + \tfrac{1}{n} \ .$$

Now we find a subsequence $\mathcal{S}_{n+1}$ of $\mathcal{S}_n$ such that some versions of the graphons from $\mathcal{S}_{n+1}$ converge to W_{n+1} in the weak* topology. This finishes the construction.

Now we use the diagonal method to define, for every natural number n, the graphon Γ_{k_n} to be the nth element of the sequence $\mathcal{S}_n$. Then we have for every n that

$$\begin{aligned}
&\inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots)\} \\
&\qquad \geq \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\mathcal{S}_n)\} \qquad \boxed{\Gamma_{k_n}, \Gamma_{k_{n+1}}, \Gamma_{k_{n+2}}, \ldots \text{ is a subsequence of } \mathcal{S}_n} \\
&\qquad \overset{(22)}{>} \mathrm{INT}_f(W_{n+1}) - \tfrac{1}{n} \\
&\qquad \geq \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \ldots)\} - \tfrac{1}{n} \ , \qquad \boxed{W_{n+1} \in \mathbf{LIM}_{\mathrm{w}^*}(\mathcal{S}_{n+1}) \subset \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \ldots)}
\end{aligned}$$

and so

$$\inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \ldots)\} \geq \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \ldots)\} \ .$$

The other inequality is trivial. □

We can now give the proof of Theorem 3(b).

By using Lemma 16 and by passing to a subsequence, we may assume that

$$\inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \ldots)\} = \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \ldots)\} \ .$$

We construct the desired subsequence $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots$ by the following construction.

In the first step, we find a graphon $W_1 \in \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ such that

$$\mathrm{INT}_f(W_1) < \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)\} + 1 \ .$$

By Lemma 8, there is a partition $\mathcal{J}_1$ of I into finitely many intervals of positive measure such that $|\mathrm{INT}_f(W_1) - \mathrm{INT}_f(W_1^{\bowtie\mathcal{J}_1})| < 1$. Then we clearly have

$$\mathrm{INT}_f(W_1^{\bowtie\mathcal{J}_1}) < \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)\} + 2\,.$$

By Lemma 9, the graphon $W_1^{\bowtie\mathcal{J}_1}$ is also an element of the set $\mathbf{LIM}_{\mathrm{w}^*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$, and so there is a sequence $\Gamma_1^1,\Gamma_2^1,\Gamma_3^1,\ldots$ of versions of $\Gamma_1,\Gamma_2,\Gamma_3,\ldots$ that converges to $\widetilde{W}_1 := W_1^{\bowtie\mathcal{J}_1}$ in the weak* topology. We define $\Gamma_{k_1} := \Gamma_1$, and we also define a sequence $q_1^1,q_2^1,q_3^1,\ldots$ to be the increasing sequence of all natural numbers.

Now fix a natural number n and suppose that we have already defined a finite subsequence $\Gamma_{k_1},\Gamma_{k_2},\ldots,\Gamma_{k_n}$ of $\Gamma_1,\Gamma_2,\Gamma_3,\ldots$. Suppose also that for every $1 \leq i \leq n$, we have already constructed

(i) a step-graphon $\widetilde{W}_i$ with steps given by some partition $\mathcal{J}_i$ of I into finitely many intervals of positive measure such that $\mathcal{J}_i$ is a refinement of $\mathcal{J}_{i-1}$ (if $i > 1$) and such that

$$\mathrm{INT}_f(\widetilde{W}_i) < \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)\} + \tfrac{2}{i}\,,$$

and

(ii) an increasing sequence $q_1^i,q_2^i,q_3^i,\ldots$ of natural numbers which is a subsequence of $q_1^{i-1},q_2^{i-1},q_3^{i-1},\ldots$ (if $i > 1$), together with a sequence $\Gamma^i_{q_1^i},\Gamma^i_{q_2^i},\Gamma^i_{q_3^i},\ldots$ of versions of $\Gamma_{q_1^i},\Gamma_{q_2^i},\Gamma_{q_3^i},\ldots$ which converges to $\widetilde{W}_i$ in the weak* topology and such that (if $i > 1$) for every natural number j and for every intervals $K,L \in \mathcal{J}_{i-1}$ it holds that

$$\int_K\int_L \Gamma^i_{q_j^i}(x,y) = \int_K\int_L \Gamma^{i-1}_{q_j^i}(x,y)\,.$$

Then we find a graphon $\overline{W}_{n+1} \in \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ such that

$$\mathrm{INT}_f(\overline{W}_{n+1}) < \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)\} + \tfrac{1}{n+1}\,.$$

Find a sequence $\overline{\Gamma}^{n+1}_{q_1^n},\overline{\Gamma}^{n+1}_{q_2^n},\overline{\Gamma}^{n+1}_{q_3^n},\ldots$ of versions of $\Gamma_{q_1^n},\Gamma_{q_2^n},\Gamma_{q_3^n},\ldots$ which converges to $\overline{W}_{n+1}$ in the weak* topology. For every natural number j, let $\phi_j : I \to I$ be the measure-preserving almost bijection satisfying $\overline{\Gamma}^{n+1}_{q_j^n}(x,y) = \Gamma^n_{q_j^n}(\phi_j^{-1}(x),\phi_j^{-1}(y))$ for a.e. $(x,y) \in I^2$ (such an almost-bijection exists as both $\overline{\Gamma}^{n+1}_{q_j^n}$ and $\Gamma^n_{q_j^n}$ are versions of the same graphon $\Gamma_{q_j^n}$). Let us fix some order of the sets from the partition $\mathcal{J}_n$. For every j, let $\mathcal{I}_j$ be the ordered partition of I consisting of the sets $\phi_j(K)$, $K \in \mathcal{J}_n$, with the order given by the order of the sets from $\mathcal{J}_n$. Let $r_1,r_2,r_3,\ldots$ be a subsequence of $q_1^n,q_2^n,q_3^n,\ldots$ such that for every $K \in \mathcal{J}_n$, the sequence $\mathbf{1}_{\phi_1(K)},\mathbf{1}_{\phi_2(K)},\mathbf{1}_{\phi_3(K)},\ldots$ is convergent in the weak* topology. Find an accumulation point W_{n+1} of the sequence ${}_{\mathcal{I}_1}\overline{\Gamma}^{n+1}_{r_1},{}_{\mathcal{I}_2}\overline{\Gamma}^{n+1}_{r_2},{}_{\mathcal{I}_3}\overline{\Gamma}^{n+1}_{r_3},\ldots$ (in the weak* topology). By Corollary 15, we have

$$\mathrm{INT}_f(W_{n+1}) \leq \mathrm{INT}_f(\overline{W}_{n+1}) < \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)\} + \tfrac{1}{n+1}\,.$$

Let $s_1, s_2, s_3, \ldots$ be a subsequence of $r_1, r_2, r_3, \ldots$ such that we have that the sequence ${}_{\mathcal{I}_1}\overline{\Gamma}^{n+1}_{s_1}, {}_{\mathcal{I}_2}\overline{\Gamma}^{n+1}_{s_2}, {}_{\mathcal{I}_3}\overline{\Gamma}^{n+1}_{s_3}, \ldots$ converges to W_{n+1} in the weak* topology. Note that for every natural number j and for every intervals $K, L \in \mathcal{J}_n$, it holds that

$$\begin{aligned}\int_K \int_L {}_{\mathcal{I}_j}\overline{\Gamma}^{n+1}_{s_j}(x,y) = \int_{\phi_J(K)} \int_{\phi_j(L)} \overline{\Gamma}^{n+1}_{s_j}(x,y) &= \int_{\phi_J(K)} \int_{\phi_j(L)} \Gamma^n_{s_j}(\phi_j^{-1}(x), \phi_j^{-1}(y)) \\ &= \int_K \int_L \Gamma^n_{s_j}(x,y)\ .\end{aligned} \tag{23}$$

By Lemma 8, there is a partition $\mathcal{J}_{n+1}$ of I into finitely many intervals of positive measure such that $\mathcal{J}_{n+1}$ is a refinement of $\mathcal{J}_n$ and such that $|\mathrm{INT}_f(W_{n+1}) - \mathrm{INT}_f(W_{n+1}^{\bowtie \mathcal{J}_{n+1}})| < \frac{1}{n+1}$. Then we clearly have

$$\mathrm{INT}_f(W_{n+1}^{\bowtie \mathcal{J}_{n+1}}) < \inf\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)\} + \tfrac{2}{n+1}\ .$$

By Lemma 9, the graphon $\widetilde{W}_{n+1} := W_{n+1}^{\bowtie \mathcal{J}_{n+1}}$ is a limit (in the weak* topology) of the sequence of some versions $\Gamma^{n+1}_{s_1}, \Gamma^{n+1}_{s_2}, \Gamma^{n+1}_{s_3}, \ldots$ of the graphons ${}_{\mathcal{I}_1}\overline{\Gamma}^{n+1}_{s_1}, {}_{\mathcal{I}_2}\overline{\Gamma}^{n+1}_{s_2}, {}_{\mathcal{I}_3}\overline{\Gamma}^{n+1}_{s_3}, \ldots$. By the "moreover" part of Lemma 9, we may further assume that for every natural number j and for every intervals $P, Q \in \mathcal{J}_{n+1}$, we have

$$\int_P \int_Q {}_{\mathcal{I}_j}\overline{\Gamma}^{n+1}_{s_j}(x,y) = \int_P \int_Q \Gamma^{n+1}_{s_j}(x,y)\ ,$$

which, together with (23), easily implies that for every natural number j and for every intervals $K, L \in \mathcal{J}_n$ it holds

$$\int_K \int_L \Gamma^{n+1}_{s_j}(x,y) = \int_K \int_L \Gamma^n_{s_j}(x,y)\ . \tag{24}$$

We define $\Gamma_{k_{n+1}} := \Gamma_{s^{n+1}_{n+1}}$, and we also define the sequence $q_1^{n+1}, q_2^{n+1}, q_3^{n+1}, \ldots$ to be the sequence $s_1, s_2, s_3, \ldots$. This completes the construction of the sequence $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots$.

Now let $W_{\min}$ be an arbitrary accumulation point (in the weak* topology) of the sequence $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots$, so that in particular $W_{\min} \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. It suffices to show that it holds for every n that $\mathrm{INT}_f(W_{\min}) \leq \mathrm{INT}_f(\widetilde{W}_n)$ as then we clearly have by our choice of the graphons $\widetilde{W}_1, \widetilde{W}_2, \widetilde{W}_3, \ldots$ that

$$\mathrm{INT}_f(W_{\min}) = \min\{\mathrm{INT}_f(W)\colon W \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)\}\ .$$

But for every three natural numbers $n < m$ and j and for every intervals $K, L \in \mathcal{J}_n$ it holds by (ii) that

$$\int_K \int_L \Gamma^m_{q^m_j}(x,y) = \int_K \int_L \Gamma^n_{q^m_j}(x,y)\ ,$$

and so (as $\Gamma^n_{q^n_j} \overset{w^*}{\to} \widetilde{W}_n$ as $j \to \infty$ for every n)

$$\int_K \int_L \widetilde{W}_m(x,y) = \int_K \int_L \widetilde{W}_n(x,y)\ .$$

It follows that for every n it holds

$$\int_K \int_L W_{\min}(x,y) = \int_K \int_L \widetilde{W}_n(x,y) \ .$$

The rest follows by Lemma 12.

6. Proof of Proposition 4

As promised, we give two proofs of Proposition 4. The first one is somewhat quicker, but uses a theorem of Borgs, Chayes, and Lovász [5] about uniqueness of graph limits. More precisely, the theorem states that if $U' : I^2 \to [0,1]$ and $U'' : I^2 \to [0,1]$ are two cut-norm limits of versions $\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots$ and $\Gamma''_1, \Gamma''_2, \Gamma''_3, \ldots$ of a graphon sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$, then there exists a graphon $U^* : I^2 \to [0,1]$ that is a cut-norm limit of versions of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$, and measure preserving transformations $\psi', \psi'' : I \to I$ such that for almost every $(x,y) \in I^2$, $U'(x,y) = U^*(\psi'(x), \psi'(y))$ and $U''(x,y) = U^*(\psi''(x), \psi''(y))$. Since then, the result was proven in several different ways, see [18, p.221]. Also, let us note that while all known proofs of the Borgs–Chayes–Lovász theorem are complicated, none uses the compactness of the space of graphons or the Regularity lemma. So, using this result as a blackbox, we still obtain a self-contained characterization of cut-norm limits in terms of weak* limits.

So, suppose that $W : I^2 \to [0,1]$ is a limit of versions of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ in the cut-norm. By Theorem 3 and by passing to a subsequence, we may assume that there exists a minimizer $W' : I^2 \to [0,1]$ of $\mathrm{INT}_f(\cdot)$ over $\mathbf{LIM}_{w^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ which is a limit of versions of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ in the cut-norm. Therefore, the Borgs–Chayes–Lovász theorem tells us that there exists a graphon $W^* : I^2 \to [0,1]$ and measure preserving maps $\psi, \psi' : I \to I$ such that $W(x,y) = W^*(\psi(x), \psi(y))$ and $W'(x,y) = W^*(\psi'(x), \psi'(y))$ for almost every $(x,y) \in I^2$. Since ψ and ψ' are measure preserving, we get $\mathrm{INT}_f(W) = \mathrm{INT}_f(W^*)$ and $\mathrm{INT}_f(W') = \mathrm{INT}_f(W^*)$. This finishes the proof. □

Let us now give a self-contained proof of Proposition 4. By Theorem 3 and by passing to a subsequence, we may assume that there exists a minimizer $W' : I^2 \to [0,1]$ of $\mathrm{INT}_f(\cdot)$ over $\mathbf{LIM}_{w^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ which is a limit of versions $\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots$ of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ in the cut-norm. Suppose that W is a graphon with $\mathrm{INT}_f(W) > \mathrm{INT}_f(W')$. This in particular means that there exists $\delta > 0$ so that

$$\|W' - U\|_1 > \delta \tag{25}$$

for any version U of W. We claim that there are no versions of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ that converge to W in the cut-norm. Indeed, suppose that such versions $\Gamma^*_1, \Gamma^*_2, \Gamma^*_3, \ldots$ exist. Observe that $\delta_1(\Gamma'_n, \Gamma^*_n) = 0$ for each n (in fact, the infimum in the definition of δ_1 is attained).

Now, [21, Lemma 2.11][8] tells us that

$$0 = \liminf_n 0 = \liminf_n \delta_1(\Gamma'_n, \Gamma^*_n) \geq \delta_1(W', W) \ ,$$

which is a contradiction to (25). □

7. Concluding remarks

7.1. **Specific concave and convex functions.** Perhaps the most natural choice of continuous concave function is the binary entropy H.

An equivalent characterization to our main result is that the limit graphons are the weak* limits that *maximize* INT_g for a strictly *convex* function g. The most interesting instance of this version of the statement is that the limit graphons are weak* limits maximizing the L^2-norm. Note that the L^2-norm is an infinitesimal counterpart to the notion of the "index" commonly used in proving the regularity lemma.

7.2. **Regularity lemmas as a corollary.** While the cut-distance is most tightly linked to the weak regularity lemma of Frieze and Kannan [14], a short reduction given in [20] shows that Theorem 2 implies also Szemerédi's regularity lemma [25], and its "superstrong" form, [3]. So, it is possible to obtain these regularity lemmas using the approach from this paper.[9]

The most remarkable difference of the current approach is that it does not use iterative index-pumping, as we explained in Section 3.2. That is, in our proof one refinement is sufficient for the argument. Such a shortcut is available only in the limit setting, it seems.

7.3. **A conjecture about finite graphs.** Let $f : [0,1] \to \mathbb{R}$ be an arbitrary continuous and strictly concave function. Suppose that G is an n-vertex graph, and let $\mathcal{P} = (P_i)_{i=1}^k$ be a partition of $V(G)$ into non-empty sets. Recall the notion of $\mathrm{INT}_f(G;\mathcal{P})$ and of densities d_{ij} defined in Section 3.2. We believe that a partition that minimizes $\mathrm{INT}_f(G;\cdot)$, when we range over all partitions $\mathcal{Q}$ of G with a given (but large) number of parts, provides a good approximation of G in the sense of the weak regularity lemma. To formulate this conjecture, let us say that a partition $\mathcal{Q}$ is *an* INT_f*-minimizing partition with* k *parts* if $\mathcal{Q}$ has k parts and for any partition $\mathcal{P}$ of $V(G)$ with k parts we have $\mathrm{INT}_f(G;\mathcal{Q}) \leq \mathrm{INT}_f(G;\mathcal{P})$.

Conjecture 17. *Suppose that $f : [0,1] \to \mathbb{R}$ is a continuous and strictly concave function, and that $\epsilon > 0$ is given. Then there exist numbers M, n_0 so that the following holds for each graph G of order at least n_0. If $\mathcal{Q}$ is an INT_f-minimizing partition of $V(G)$ with M parts then $\mathcal{Q}$ is also weak ϵ-regular.*

[8] Let us stress that [21, Lemma 2.11] does not rely on the Borgs–Chayes–Lovász theorem, and has a self-contained, one-page proof.

[9] With a notable drawback that we do not obtain any quantitative bounds.

This is a finite counterpart of our main result. Indeed, the space of weak* limits of graphons in Theorem 3 corresponds to an averaging over infinitesimally small sets, while in Conjecture 17 we range only over partitions with M parts. Of course, the much finer partitions considered in Theorem 3 provide an "$\epsilon = 0$ error".

If true, Conjecture 17 would provide a more direct link between regularity and index-like parameters than the index-pumping lemma.

While we were not able to prove Conjecture 17, let us present here a quick proof of a somewhat weaker statement.

Proposition 18. *Suppose that $f : [0,1] \to \mathbb{R}$ is a continuous and strictly concave function, and that $\epsilon > 0$ is given. Then there exist a finite set $X \subset \mathbb{N}$ so that for each graph G there exists $M \in X$ with the following property. If $\mathcal{Q}$ is an INT_f-minimizing partition of $V(G)$ with M parts then $\mathcal{Q}$ is also weak ϵ-regular.*

Actually to prove Proposition 18, one just needs to go through the proof of the weak regularity lemma. For simplicity, let us assume that $h : x \mapsto -x^2$ is the negative of the usual "index" used in the proof of the weak regularity lemma. Let us take $X := \{1, 2, 4, \ldots, 2^{\lceil \frac{4}{\epsilon^2} \rceil}\}$. Suppose for a contradiction that for each $i \in X$, there is an INT_f-minimizing partition $\mathcal{C}_i$ with i parts which is weak ϵ-irregular. Then Lemma 13 assert that there exists a partition $\mathcal{P}_{i+1}$ with $2i$ parts such that $\mathrm{INT}_h(G; \mathcal{P}_{i+1}) < \mathrm{INT}_h(G; \mathcal{C}_i) - \frac{\epsilon^2}{4}$. In particular, we have

$$\begin{aligned} 0 \geq \mathrm{INT}_h(G;\mathcal{C}_1) &> \mathrm{INT}_h(G;\mathcal{P}_2) + \frac{\epsilon^2}{4} \geq \mathrm{INT}_h(G;\mathcal{C}_2) + \frac{\epsilon^2}{4} \\ &> \mathrm{INT}_h(G;\mathcal{P}_3) + 2 \cdot \frac{\epsilon^2}{4} \geq \mathrm{INT}_h(G;\mathcal{C}_3) + 2 \cdot \frac{\epsilon^2}{4} > \ldots \\ &> \mathrm{INT}_h(G;\mathcal{P}_{i+1}) + i \cdot \frac{\epsilon^2}{4} \geq \mathrm{INT}_h(G;\mathcal{C}_{i+1}) + i \cdot \frac{\epsilon^2}{4} > \ldots \\ &> \mathrm{INT}_h(G;\mathcal{P}_{\lceil \frac{4}{\epsilon^2} \rceil + 1}) + \lceil \frac{4}{\epsilon^2} \rceil \cdot \frac{\epsilon^2}{4} \, . \end{aligned}$$

This is a contradiction to the fact that $\mathrm{INT}_h(\cdot;\cdot) \geq -1$.

One could consider even a "stability version" of Conjecture 17. That is, it may be that if $\mathrm{INT}_f(G;\mathcal{Q})$ is close to the minimum of $\mathrm{INT}_f(G;\mathcal{P})$ over partitions $\mathcal{P}$ with M parts, then $\mathcal{Q}$ is weak ϵ-regular. For example, repeating the proof of Proposition 18 for a set $X = \{1, 2, 4, \ldots, 2^{\lceil \frac{8}{\epsilon^2} \rceil}\}$, we get there exists $M \in X$ so that any partition $\mathcal{Q}$ with M parts for which

$$\mathrm{INT}_{x \mapsto -x^2}(G;\mathcal{Q}) \leq \frac{\epsilon^2}{8} + \min\left\{\mathrm{INT}_{x \mapsto -x^2}(G;\mathcal{P}) \, : \, \mathcal{P} \text{ has } M \text{ parts}\right\}$$

is weak ϵ-regular.

Also, Conjecture 17 could be asked for other versions of the regularity lemma.

7.4. **Attaining the infimum in Theorem 3(b).** Theorem 3(b) states that there exist a subsequence of graphons $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots$ such that the infimum of $\mathrm{INT}_f(\cdot)$ over $\mathbf{ACC}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots)$ is attained. Recently, Jon Noel showed us that passing to a subsequence is really needed. That is, taking f to be the binary entropy function, he constructed a sequence of graphons $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ such that $\inf\{\mathrm{INT}_f(\Gamma) : \Gamma \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)\} = 0$ but there exists no $\Gamma \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ with $\mathrm{INT}_f(\Gamma) = 0$. To this end, take $(W_\ell)_{\ell=1}^\infty$ to be rescaled adjacency matrices of a sequence of quasirandom graphs with edge density say 0.5, but replacing in each adjacency matrix one diagonal element (now represented by a square S_ℓ of size $\frac{1}{\ell} \times \frac{1}{\ell}$) by value say 0.7. Let $(\Gamma_n)_{n=1}^\infty$ be a sequence in which each graphon W_ℓ occurs infinitely many times.

Firstly, we claim that $\inf\{\mathrm{INT}_f(\Gamma) : \Gamma \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)\} = 0$. To see this, take ℓ large. Taking a subsequence $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots$ which consists only of copies of W_ℓ, we see that $W_\ell \in \mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. Now, $\mathrm{INT}_f(W_\ell) = \int_x \int_y f(W_\ell(x,y)) \leq \frac{1}{\ell^2}$, since the integrand is zero everywhere except S_ℓ.

Secondly, we claim that there is no graphon in $\mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ with zero entropy. Indeed, let us consider a weak* limit W of an arbitrary sequence of versions of $W_{\ell_1}, W_{\ell_2}, W_{\ell_3}, \ldots$. There are two cases. If the sequence $\ell_1, \ell_2, \ell_3, \ldots$ is unbounded then quasirandomness of the graphons implies that $W \equiv \frac{1}{2}$. The other case is when one index ℓ repeats infinitely many times. In that case, due to the value of 0.7 on S_ℓ, the graphon W cannot be $\{0,1\}$-valued, as the next lemma shows.

Lemma 19. *Suppose that Λ is an arbitrary probability measure space with a probability measure λ, and $\alpha > 0$. Suppose that $(A_s)_{s=1}^\infty$ is a sequence of functions, $A_s : \Lambda \to [0,1]$, which converges weak* to a function A. Suppose further that $\lambda(R_s) \geq \alpha$ for each $s \in \mathbb{N}$, where $R_s := \{x \in \Lambda : A_s = 0.7\}$. Then A is not $\{0,1\}$-valued.*

Proof. Suppose for a contradiction that A is $\{0,1\}$-valued. Let $X_0 = A^{-1}(0)$ and $X_1 = A^{-1}(1)$. Then for each $s \in \mathbb{N}$, we have $\lambda(R_s \cap X_0) \geq \alpha/2$ or $\lambda(R_s \cap X_1) \geq \alpha/2$. Let us consider the case that the set I_0 of indices s for which the former inequality occurs is infinite; the other case being analogous. For each $s \in I_0$ we have

$$\int_{X_0} A_s = \int_{X_0 \cap R_s} A_s + \int_{X_0 \setminus R_s} A_s \geq 0.7 \cdot \lambda(X_0 \cap R_s) + 0 \cdot \lambda(X_0 \setminus R_s) \geq 0.7 \cdot \frac{\alpha}{2} .$$

On the other hand, $\int_{X_0} A = 0$. So, the set X_0 witnesses that the functions $(A_s)_{s \in I_0}$ do not weak* converge to A, a contradiction. □

In either of the two cases above, W has positive entropy.

7.5. **Hypergraphs.** The theory of limits of dense hypergraphs of a fixed uniformity was worked out in [12] (using ultraproduct techniques) and in [27] (using hypergraph regularity lemma techniques), and is substantially more involved. It seems that the current approach may generalize to the hypergraph setting. This is currently work in progress.

7.6. **Role of weak* limits for other combinatorial structures.** In this paper, we have shown how to use weak* limits for sequences of graphs to obtain cut-distance limits. In the section above we indicated that a similar approach may lead to a construction of limits of hypergraphs of fixed uniformity. Of course, one can ask which other limit concepts can be approached by considering weak* limits as an intermediate step. Let us point out that limits of permutations (*permutons*) are particularly simple in this sense: Limits (in the "cut-distance" sense) of permutations arise simply by taking weak limits (here, it is weak rather than weak*, but the difference is not important) of certain objects associated directly to permutations. That is, no counterpart to our entropy minimization step is necessary, and every weak limit already has the desired combinatorial properties. See [17, Section 2]. These are, to the best of our knowledge, the only combinatorial structures for which weak/weak* convergence was used.

7.7. **Minimization with respect to different concave functions.** Suppose that f and g are two different strictly concave functions. Then for two graphons Γ_1 and Γ_2, we can have for example $\mathrm{INT}_f(\Gamma_1) < \mathrm{INT}_f(\Gamma_2)$ but $\mathrm{INT}_g(\Gamma_1) > \mathrm{INT}_g(\Gamma_2)$. As a (perhaps somewhat surprising) by-product of our main results, we cannot get such an inconsistency when searching global minima over the space of weak* limits. That is, Γ_1 achieves the minimum of INT_f on the space of weak* limits if and only if it achieves the minimum of INT_g. We do not know of a more direct proof of this fact.

7.8. **Recent developments.** After this paper was made available at arXiv in May 2017, the relation between the cut distance and the weak* topology was studied in more detail in [11] and [10]. The main novel feature in [11] is an abstract approach which allows to identify convergent subsequences and cut distance limits without minimization of any parameter over the space of weak* limits. The main two theorems in [11] which were inspired by the present paper are the following:

Theorem 20. *Suppose that* $\Gamma_1, \Gamma_2, \Gamma_3, \ldots : I^2 \to [0,1]$ *is a sequence of graphons. Then there exists a subsequence* $\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots$ *such that*

$$\mathbf{ACC}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots) = \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_{k_1}, \Gamma_{k_2}, \Gamma_{k_3}, \ldots) .$$

Theorem 21. *Suppose that* $\Gamma_1, \Gamma_2, \Gamma_3, \ldots : I^2 \to [0,1]$ *is a sequence of graphons. Then this sequence is cut-distance convergent if and only if*

$$\mathbf{ACC}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) = \mathbf{LIM}_{\mathrm{w}^*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) .$$

In particular, note that Theorem 20 substantially generalizes Lemma 16. Actually, investigating possible generalizations of Lemma 16 was the starting point for [11].

Besides this abstract approach, some further graphon parameters that can replace $\mathrm{INT}_f(\cdot)$ in Theorem 3 are found in [10]. These include, for example, the negative of the density of any even cycle, $-t(C_{2\ell},\cdot)$. On the other hand, in another very recent paper, Král', Martins, Pach, and Wrochna [16] identify a large class of (bipartite) graphs H for which $-t(H,\cdot)$ fails to identify cut distance limits. The problem of characterizing graphs H which this property is related to the Sidorenko conjecture and to norming graphs motivated by a question of Lovász and studied first in [15].

Also, the machinery introduced in [10] gives a short proof of a version of Theorem 3 which even allows to drop the requirement on the continuity of f.

Acknowledgements

This work was done while Jan Hladký was enjoying a lively atmosphere of the Institute for Geometry at TU Dresden, being hosted by Andreas Thom there.

We thank Dan Král and Oleg Pikhurko for encouraging conversations on the subject, Jon Noel for comments on an earlier version of the manuscript, and to Svante Janson and Guus Regts for bringing several important references to our attention. We also thank Jon Noel for his contribution included in Section 7.4.

Finally, we thank two anonymous referees for their comments, and in particular, for pointing out a gap in the proof of Lemma 14.

Appendix A. The weak* topology

Suppose that X is a Banach space and denote by X^* its dual. Then the weak* topology on X^* is the coarsest topology on X^* such that all mappings of the form $X^* \ni x^* \mapsto x^*(x)$, $x \in X$, are continuous. Recall that if the space X is separable then by the sequential Banach–Alaoglu Theorem (see e.g. [26, Theorem 1.9.14]), the unit ball of X^* is sequentially compact. This means that every bounded sequence of elements of the dual space X^* contains a weak*-convergent subsequence.

In this paper, we are interested in the case when X is the Banach space $L^1(\Omega)$ of all integrable functions on some probability space Ω. (Depending on our needs, the probability space Ω will be chosen to be either the unit interval I equipped with the one-dimensional Lebesgue measure or the unit square I^2 equipped with the two-dimensional Lebesgue measure). The space $L^1(\Omega)$ is equipped with the norm $\|f\|_1 = \int_\Omega |f(x)|$, $f \in L^1(\Omega)$. In this setting, the dual $X^* = (L^1(\Omega))^*$ is isometric to the space $L^\infty(\Omega)$ of all bounded measurable functions on Ω, equipped with the norm $\|g\|_\infty = \operatorname{ess\,sup}_{x\in\Omega} |g(x)|$. The duality between $L^1(\Omega)$ and $L^\infty(\Omega)$ is given by the formula $\langle g, f\rangle = \int_\Omega f(x)g(x)$ for $g \in L^\infty(\Omega)$

and $f \in L^1(\Omega)$. This means that a sequence $g_1, g_2, g_3, \ldots$ of elements of $L^\infty(\Omega)$ converges to $g \in L^\infty(\Omega)$ if and only if $\lim_{n\to\infty} \int_\Omega f(x)g_n(x) = \int_\Omega f(x)g(x)$ for every $f \in L^1(\Omega)$.

Now consider the Banach space $X = L^1(I^2)$ of all integrable functions defined on the unit square I^2 (which is equipped with the two-dimensional Lebesgue measure). Standard arguments show that the weak* topology on its dual space $L^\infty(I^2)$ can be equivalently generated by mappings of the form $L^\infty(I^2) \ni g \mapsto \int_A \int_B g(x,y)$ where A, B are measurable subsets of I. That is, the weak* topology can be equivalently generated only by characteristic functions of measurable rectangles (instead of all integrable functions on I^2). If we restrict this topology only to the space of all graphons $W : I^2 \to [0,1]$ defined on I^2 then it is easy to see that this restricted topology is generated only by mappings of the form $W \mapsto \int_A \int_A W(x,y)$ where A is a measurable subset of I (this is because each graphon is symmetric by the definition). This is the topology we refer to when we talk about convergence of graphons in the weak* topology. So this means that a sequence $W_1, W_2, W_3, \ldots$ of graphons defined on I^2 converges to a graphon W defined on I^2 if and only if $\lim_{n\to\infty} \int_A \int_A W_n(x,y) = \int_A \int_A W(x,y)$ for every measurable subset A of I. Note that the space of all graphons defined on I^2 is a weak* closed subset of the unit ball of $L^\infty(I^2)$, and so it is sequentially compact by the sequential Banach–Alaoglu Theorem (as the space $L^1(I^2)$ is separable).

While crucial to our arguments, it is worth noting that the Banach–Alaoglu Theorem is not a particularly deep statement and follows easily from Tychonoff's theorem for powers of compact spaces (and actually the version for countable powers is sufficient).

References

[1] D. J. Aldous. Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.*, 11(4):581–598, 1981.

[2] D. J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985.

[3] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[4] T. Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.*, 5:80–145, 2008.

[5] C. Borgs, J. Chayes, and L. Lovász. Moments of two-variable functions and the uniqueness of graph limits. *Geom. Funct. Anal.*, 19(6):1597–1619, 2010.

[6] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[7] S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011.

[8] D. Conlon and J. Fox. Graph removal lemmas. In *Surveys in combinatorics 2013*, volume 409 of *London Math. Soc. Lecture Note Ser.*, pages 1–49. Cambridge Univ. Press, Cambridge, 2013.

[9] P. Diaconis and S. Janson. Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)*, 28(1):33–61, 2008.

[10] M. Doležal, J. Grebík, J. Hladký, I. Rocha, and V. Rozhoň. Cut distance identifying graphon parameters over weak* limits. arXiv:1809.03797.

[11] M. Doležal, J. Grebík, J. Hladký, I. Rocha, and V. Rozhoň. Relating the cut distance and the weak* topology for graphons. arXiv:1809.03797.

[12] G. Elek and B. Szegedy. A measure-theoretic approach to the theory of dense hypergraphs. *Adv. Math.*, 231(3-4):1731–1772, 2012.

[13] J. Fox. A new proof of the graph removal lemma. *Ann. of Math. (2)*, 174(1):561–579, 2011.

[14] A. Frieze and R. Kannan. Quick Approximation to Matrices and Applications. *Combinatorica*, 19(2):175–220, 1999.

[15] H. Hatami. Graph norms and Sidorenko's conjecture. *Israel J. Math.*, 175:125–150, 2010.

[16] D. Král', T. Martins, P. P. Pach, and M. Wrochna. The step Sidorenko property and non-norming edge-transitive graphs. *J. Combin. Theory Ser. A*, 162:34–54, 2019.

[17] D. Král' and O. Pikhurko. Quasirandom permutations are characterized by 4-point densities. *Geom. Funct. Anal.*, 23(2):570–579, 2013.

[18] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[19] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[20] L. Lovász and B. Szegedy. Szemerédi's Lemma for the analyst. *J. Geom. and Func. Anal*, 17:252–270, 2007.

[21] L. Lovász and B. Szegedy. Testing properties of graphs and functions. *Israel J. Math.*, 178:113–156, 2010.

[22] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

[23] G. Regts and A. Schrijver. Compact orbit spaces in Hilbert spaces and limits of edge-colouring models. *European J. Combin.*, 52(part B):389–395, 2016.

[24] A. Scott. Szemerédi's regularity lemma for matrices and sparse graphs. *Combin. Probab. Comput.*, 20(3):455–466, 2011.

[25] E. Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

[26] T. Tao. *An epsilon of room, I: real analysis*, volume 117 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010. Pages from year three of a mathematical blog.

[27] Y. Zhao. Hypergraph limits: a regularity approach. *Random Structures Algorithms*, 47(2):205–226, 2015.

Institute of Mathematics, Czech Academy of Sciences. Žitná 25, 110 00, Praha, Czech Republic. The Institute of Mathematics of the Czech Academy of Sciences is supported by RVO:67985840.

RELATING THE CUT DISTANCE AND THE WEAK* TOPOLOGY FOR GRAPHONS

MARTIN DOLEŽAL, JAN GREBÍK, JAN HLADKÝ, ISRAEL ROCHA, VÁCLAV ROZHOŇ

Abstract. The theory of graphons is ultimately connected with the so-called cut norm. In this paper, we approach the cut norm topology via the weak* topology (when considering a predual of L^1-functions). We prove that a sequence $W_1, W_2, W_3, \dots$ of graphons converges in the cut distance if and only if we have equality of the sets of weak* accumulation points and of weak* limit points of all sequences of graphons $W_1', W_2', W_3', \dots$ that are weakly isomorphic to $W_1, W_2, W_3, \dots$. We further give a short descriptive set theoretic argument that each sequence of graphons contains a subsequence with the property above. This in particular provides an alternative proof of the theorem of Lovász and Szegedy about compactness of the space of graphons. We connect these results to «multiway cut» characterization of cut distance convergence from [Ann. of Math. (2) 176 (2012), no. 1, 151-219].

These results are more naturally phrased in the Vietoris hyperspace $K(\mathcal{W}_0)$ over graphons with the weak* topology. We show that graphons with the cut distance topology are homeomorphic to a closed subset of $K(\mathcal{W}_0)$, and deduce several consequences of this fact.

From these concepts a new order on the space of graphons emerges. This order allows to compare how structured two graphons are. We establish basic properties of this «structuredness order».

1. Introduction

Graphons emerged from the work of Borgs, Chayes, Lovász, Sós, Szegedy, and Vesztergombi [18, 4] on limits of sequences of finite graphs. We write $\mathcal{W}_0$ for the space of all *graphons*, i.e., all symmetric measurable functions from Ω^2 to $[0,1]$, after identifying graphons that are equal almost everywhere. Here as well as in the rest of the paper, Ω is an arbitrary separable atomless probability space with probability measure ν. While it is meaningful to investigate the space $\mathcal{W}_0$ with respect to several metrics and topologies, the two that relate the most to graph theory are the metrics $d_\square$ and $\delta_\square$ based on the so-called *cut norm* defined on the space $L^1(\Omega^2)$ by

$$\|Y\|_\square = \sup_{S,T\subseteq\Omega} \left| \int_{S\times T} Y \right| \quad \textit{for each } Y \in L^1(\Omega^2)\,.$$

Key words and phrases. graphon; graph limit; cut norm; weak* convergence.

Research of *Doležal* was supported by the GAČR project No. 17-27844S and RVO: 67985840. *Grebík* was supported by the GAČR project 17-33849L and RVO: 67985840. *Hladký* was supported by the Alexander von Humboldt Foundation and by the GAČR project No. 18-01472Y and RVO: 67985840. *Rocha* and *Rozhoň* were supported by the Czech Science Foundation, grant number GJ16-07822Y.

Given $U, W \in \mathcal{W}_0$ we set

$$d_\square(U,W) := \|U-W\|_\square = \sup_{S,T\subseteq\Omega}\left|\int_{S\times T} U - \int_{S\times T} W\right| \text{ , and}$$

$$\delta_\square(U,W) := \inf_{\varphi} d_\square(U, W^\varphi) \text{ ,} \tag{1.1}$$

where φ ranges over all measure preserving bijections of Ω and the graphon W^φ is defined by

$$W^\varphi(x,y) = W(\varphi(x),\varphi(y)) \text{ .} \tag{1.2}$$

We call $d_\square$ the *cut norm distance* and $\delta_\square$ the *cut distance*. We call graphons of the form W^φ *versions* of W. Passing to a version is an infinitesimal counterpart to considering another adjacency matrix of a graph, in which the vertices are reordered.

The key property of the space $\mathcal{W}_0$ is its compactness with respect to the cut distance $\delta_\square$. The result was first proven by Lovász and Szegedy [18] using the regularity lemma,[a] and then by Elek and Szegedy [10] using ultrafilter techniques, by Austin [1] and Diaconis and Janson [6] using the theory of exchangeable random graphs, and finally by Doležal and Hladký [8] by optimizing a suitable parameter over the set of weak* limits.

Theorem 1.1. *For every sequence $\Gamma_1, \Gamma_2, \Gamma_3, \dots$ of graphons there exists a subsequence $\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \dots$ and a graphon Γ such that $\delta_\square(\Gamma_{n_i}, \Gamma) \to 0$.*

1.1. **Overview of the results.** We view graphons as functions in the Banach space $L^\infty(\Omega^2)$, with a predual Banach space $L^1(\Omega^2)$ to which we associate the concept of weak* convergence. That means that a sequence of graphons $\Gamma_1, \Gamma_2, \Gamma_3, \dots$ *converges weak** to a graphon W if we have

$$\sup_{Q\subseteq\Omega^2} \lim_{n\to\infty} \int_Q \Gamma_n - \int_Q W = 0 \text{ .} \tag{1.3}$$

Since the sigma-algebra of all measurable sets $Q \subseteq \Omega^2$ is generated by sets of the form $S \times T$, where S and T are measurable subsets of Ω, we can equivalently rewrite (1.3) in a way which is more convenient for us,

$$\sup_{S,T\subseteq\Omega} \lim_{n\to\infty} \int_{S\times T} \Gamma_n - \int_{S\times T} W = 0 \text{ .} \tag{1.4}$$

The weak* topology is weaker than the topology generated by $d_\square$, of which the former can be viewed as a certain uniformization. Indeed, recall that a sequence of graphons $\Gamma_1, \Gamma_2, \Gamma_3, \dots$ converges to W in the cut norm if

$$\lim_{n\to\infty} \sup_{S,T\subseteq\Omega} \left\{\int_{S\times T} \Gamma_n - \int_{S\times T} W\right\} = 0 \text{ ,} \tag{1.5}$$

[a] see also [19] and [20] for variants of this approach

that is, (1.4) and (1.5) differ only in the order of the limit and the supremum.

In Section 3 and Section 4, which we consider the main contribution of the paper, we show that the interplay between the cut distance and weak* topology creates a rich theory.

In particular, we can make use of the weak* convergence to prove Theorem 1.1. To this end, we look at the set $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ of all weak* accumulation points of sequences

$$\left\{\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots : \Gamma'_n \text{ is a version of } \Gamma_n\right\} .$$

Similarly, denote by $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ the set of all graphons W for which there exist versions $\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots$ of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ such that W is a weak* limit of the sequence $\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots$. Note that equivalently, we could have required $\Gamma'_1, \Gamma'_2, \Gamma'_3, \ldots$ to be weakly isomorphic to $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$, rather than being versions of $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$. The set $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ is non-empty by the Banach–Alaoglu Theorem. In the set $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ we cleverly select one graphon Γ. The selection is done so that in addition to being a weak* accumulation point, Γ is also a cut distance accumulation point (the latter being clearly a stronger property). In [8], Doležal and Hladký carried out a similar program[b] where they showed that for the «clever selection» we can take Γ as the maximizer of an arbitrary graphon parameter of the form

$$\mathrm{INT}_f(W) := \int_x \int_y f\left(W(x,y)\right) \tag{1.6}$$

where $f : [0,1] \to \mathbb{R}$ is a fixed but arbitrary continuous strictly convex function.[c]

Our main result, Theorem 3.5, says that a sequence of graphons $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ is cut distance convergent if and only if $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) = \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. This is complemented by Theorem 3.3 which says that from any sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ of graphons, we can choose a subsequence $\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots$ such that

$$\mathbf{LIM}_{\mathrm{w}*}(\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots) = \mathbf{ACC}_{\mathrm{w}*}(\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots) .$$

In particular, this yields Theorem 1.1. In Section 7 we show that Theorem 3.5 is actually equivalent to one of the main results of [5] on so-called «multiway cuts».

It turns out that these results can be naturally phrased in terms of the so-called Vietoris hyperspace $K(\mathcal{W}_0)$ over graphons with the weak* topology (see Section 2.5 for definitions). To each graphon $W : \Omega^2 \to [0,1]$, we associate its *envelope* $\langle W \rangle = \mathbf{ACC}_{\mathrm{w}*}(W, W, W, \ldots)$ which is a subset of $L^\infty\left(\Omega^2\right)$. We show that cut distance convergence of graphons is equivalent to convergence of the corresponding envelopes in $K(\mathcal{W}_0)$, and that envelopes form a closed set in $K(\mathcal{W}_0)$. As $K(\mathcal{W}_0)$ is known to be compact, this connection in

[b] which in actuality is more complicated for reasons sketched in Section 3.2

[c] Most of [8] deals with *minimizing* $\mathrm{INT}_f(W)$ for a fixed continuous strictly *concave* function. This is obviously equivalent. Also, note that in [9] it is shown that the assumption of continuity is not needed.

particular provides an alternative proof of Theorem 1.1. However, the transference between the space of graphons and $K(\mathcal{W}_0)$ has other applications.

From these proofs a new partial order on the space of graphons naturally emerges. We say that U is *more structured* than W if $\langle U\rangle \supsetneq \langle W\rangle$, and write $U \succ W$. One illustrative example is to take U to be a complete balanced bipartite graphon and $W \equiv \frac{1}{2}$. Obviously, $\langle W\rangle = \{W\}$. By considering versions of U which create finer and finer chessboards, we have $\langle U\rangle \ni W$. Thus, $U \succeq W$. We establish basic properties of this «structuredness order». We investigate these properties in Section 8. In particular, in Proposition 8.5 we characterize minimal and maximal elements. In Section 4.4, we introduce the *range frequencies* of a graphon W as a pushforward probability measure on $[0,1]$ defined by

$$\mathbf{\Phi}_W(A) := \nu^{\otimes 2}\left(W^{-1}(A)\right),$$

for a measurable set $A \subseteq [0,1]$, and introduce a certain *flatness order* on these range frequencies. Roughly speaking, one probability measure on $[0,1]$ is flatter than another, if the former can be obtained by a certain sort of averaging of the latter. As we show in Proposition 4.17, this order is compatible with the structuredness order on the corresponding graphons. In [9] we use range frequencies to reprove in a very quick way the result of Doležal and Hladký (even for discontinuous functions, as mentioned in Footnote [c]).

The main motivation for introducing and studying the structuredness order is its connection to Theorem 1.1. Indeed, the «clever selection» in our proof of Theorem 1.1 is to take a maximal element (inside $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$) with respect to the structuredness order. However, the abstract theory of the weak* approach to cut norm convergence which we introduce in this paper has also applications in classical problems in extremal graph theory. More specifically, one of the main results which we prove in [9] says that if a graph is «step Sidorenko» then it is «weakly norming», thus answering an open question from [14]. Our proof in [9] relies heavily on the theory of the structuredness order introduced in the current paper, and we do not see an alternative proof which would avoid it.[d]

2. Preliminaries

2.1. **General notation.** We write $\overset{\varepsilon}{\approx}$ for equality up to ε. For example, $1 \overset{0.2}{\approx} 1.1 \overset{0.2}{\approx} 1.3$. We write P_k for a path on k vertices and C_k for a cycle on k vertices.

If A and B are measure spaces then we say that a map $f : A \to B$ is an *almost-bijection* if there exist measure zero sets $A_0 \subseteq A$ and $B_0 \subseteq B$ so that $f_{\restriction A\setminus A_0}$ is a bijection between $A \setminus A_0$ and $B \setminus B_0$. Note that in (1.1), we could have worked with measure preserving almost-bijections φ instead.

[d] See Remark 3.29 of [9].

2.2. **Graphon basics.** Our notation is mostly standard, following [17]. We write $\mathcal{W}_0$ for the space of all graphons, that is, symmetric measurable functions from Ω^2 to $[0,1]$, modulo differences on null-sets.

Graphons U and W are called *weakly isomorphic* if $\delta_\square(U,W)=0$. Note that in this case there is not necessarily a measure preserving bijection φ for which $d_\square(U,W^\varphi)=0$; see [17, Figure 7.1]. In other words, being versions and being weakly isomorphic are two slightly different notions. Let us denote the compact space of graphons after the weak isomorphism factorization as $\widetilde{\mathcal{W}_0}$. For every $W\in\mathcal{W}_0$ we denote its equivalence class $[\![W]\!]\in\widetilde{\mathcal{W}_0}$.

The probability measure underlying Ω is ν. We write $\nu^{\otimes k}$ for the product measure on Ω^k.

Remark 2.1. *Every separable atomless probability space is isomorphic to the unit interval with the Lebesgue measure. While most of our arguments are abstract and work with an arbitrary separable atomless probability space Ω, there are some other, where we will assume that graphons are defined on the square of the unit interval, and then will make use of the usual order on $[0,1]$.*

If $W\colon\Omega^2\to[0,1]$ is a graphon and φ,ψ are two measure preserving bijections of Ω then we use the short notation $W^{\psi\varphi}$ for the graphon $W^{\psi\circ\varphi}$, i.e.

$$W^{\psi\varphi}(x,y)=W(\psi(\varphi(x)),\psi(\varphi(y)))=W^\psi(\varphi(x),\varphi(y))=\left(W^\psi\right)^\varphi(x,y)$$

for $(x,y)\in\Omega^2$. Thus we have a right action of the group of all measure preserving isomorphisms of $[0,1]$ on $\mathcal{W}_0$.

The graphons that take values 0 or 1 almost everywhere are called *0-1 valued graphons*.

We call the quantity $\int_x\int_y W(x,y)$ the *edge density of* W. Recall also that for $x\in\Omega$, we have the *degree of* x in W defined as $\deg_W(x)=\int_y W(x,y)$. Recall that measurability of W gives that $\deg_W(x)$ exists for almost each $x\in\Omega$. We say that W is *p-regular* if for almost every $x\in\Omega$, $\deg_W(x)=p$.

2.2.1. *The stepping operator.* Suppose that $W:\Omega^2\to[0,1]$ is a graphon. We say that W is a *step graphon* if W is constant on each $\Omega_i\times\Omega_j$, for a suitable a finite partition $\mathcal{P}$ of Ω, $\mathcal{P}=\{\Omega_1,\Omega_2,\dots,\Omega_k\}$. We recall the definition of the stepping operator.

Definition 2.2. Suppose that $\Gamma:\Omega^2\to[0,1]$ is a graphon. For a finite partition $\mathcal{P}$ of Ω, $\mathcal{P}=\{\Omega_1,\Omega_2,\dots,\Omega_k\}$, we define a graphon $\Gamma^{\bowtie\mathcal{P}}$ by setting it on the rectangle $\Omega_i\times\Omega_j$ to be the constant $\frac{1}{\nu^{\otimes2}(\Omega_i\times\Omega_j)}\int_{\Omega_i}\int_{\Omega_j}\Gamma(x,y)$. We allow graphons to have not well-defined values on null sets which handles the cases $\nu(\Omega_i)=0$ or $\nu(\Omega_j)=0$.

In [17], a stepping is denoted by $\Gamma_\mathcal{P}$ rather than $\Gamma^{\bowtie\mathcal{P}}$.

Finally, we say that a graphon U *refines* a graphon W, if W is a step graphon for a suitable partition $\mathcal{P}$ of Ω, $\mathcal{P}=\{\Omega_1,\Omega_2,\dots,\Omega_k\}$, and $U^{\bowtie\mathcal{P}}=W$.

2.3. **Topologies on $\mathcal{W}_0$.** There are several natural topologies on $\mathcal{W}_0$. The $\|\cdot\|_1$ topology inherited from the normed space $L^1(\Omega^2)$, the topology given by the $\|\cdot\|_\square$ norm, and the weak* topology (when $\mathcal{W}_0$ is viewed as a subset of the dual space of $L^1(\Omega^2)$ i.e. a subset of $L^\infty(\Omega^2)$). Note that $\mathcal{W}_0$ is closed in L^1. We write $d_1(\cdot,\cdot)$ for the distance derived from the $\|\cdot\|_1$ norm. Recall also that by the Banach–Alaoglu Theorem, $\mathcal{W}_0$ equipped with the weak* topology is compact and that the weak* topology on $\mathcal{W}_0$ is metrizable. We shall denote by $d_{\mathrm{w}^*}(\cdot,\cdot)$ any metric compatible with this topology. For example, we can take some countable dense measure subalgebra $\{A_n\}_{n\in\mathbb{N}}$ of all measurable subsets of Ω, and define

$$d_{\mathrm{w}^*}(U,W) := \sum_{n,k\in\mathbb{N}} 2^{-(n+k)} \left| \int_{A_n\times A_k} (U-W)\,\mathrm{d}\nu \right|. \tag{2.1}$$

The following fact summarizes the relation of the above topologies.

Fact 2.3. *The following identity maps are continuous:* $(\mathcal{W}_0, d_1) \to (\mathcal{W}_0, d_\square) \to (\mathcal{W}_0, d_{\mathrm{w}^*})$.

Proof. The continuity of the first map is an easy consequence of the definitions and the second is explained in Subsection 1.1. □

2.4. **Auxiliary results about L^1-spaces.** We prove two auxiliary lemmas about L^1-spaces. Lemma 2.4 is an easy result about functions that do not converge in L^1.

Lemma 2.4. *Suppose that Λ is a probability measure space with measure λ. If we have functions $g, g_1, g_2, g_3, \ldots : \Lambda \to [0,1]$ for which $g_n \overset{\|\cdot\|_1}{\longnot\longrightarrow} g$, then there exists an interval $J \subseteq [0,1]$ and a number $c > 0$ such that for the interval $J^+ := \{x + d : x \in J, d \in [-c,c]\}$ we have $\lambda\left(g^{-1}(J) \setminus g_n^{-1}(J^+)\right) \not\longrightarrow 0$.*

Proof. By passing to a subsequence, we may assume that there exists a constant $\varepsilon > 0$ so that for each n, $\|g_n - g\|_1 > \varepsilon$. Take $k := \lceil 4/\varepsilon \rceil$ and a partition of $[0,1]$ into k intervals $J_1, J_2, \ldots, J_k$ of lengths at most $\frac{\varepsilon}{4}$ (ordered from left to right; it is not important if they are open, closed, or semiopen). For $j \in [k]$, $L_j := g^{-1}(J_j)$. For each n, there exists a number $j(n) \in [k]$ so that we have $\int_{L_{j(n)}} |g_n - g| > \varepsilon \cdot \lambda(L_{j(n)})$. Observe that the strict inequality forces that $\lambda(L_{j(n)}) > 0$. In particular, we then have that

$$|g_n(x) - g(x)| > \frac{\varepsilon}{2} \tag{2.2}$$

for a set of points $x \in L_{j(n)}$ of measure at least $\frac{\varepsilon}{2} \cdot \lambda\left(L_{j(n)}\right)$. Let j be a number that repeats infinitely often in the sequence $j(1), j(2), j(3), \ldots$. By passing to a subsequence once again, we can assume that $j = j(1) = j(2) = \ldots$. We set $J := J_j$, $c := \frac{\varepsilon}{4}$, and J^+ as in the statement of the lemma. Observe that whenever $x \in L_j$ satisfies (2.2), then $g_n(x) \notin J^+$. Therefore, we conclude that for each n, $\lambda\left(g^{-1}(J) \setminus g_n^{-1}(J^+)\right) \geq \frac{\varepsilon}{2} \cdot \lambda(L_j)$. This concludes the proof. □

We can now state the second lemma of this section.

Lemma 2.5. *For every graphon $\Gamma : \Omega^2 \to [0,1]$ and every $\varepsilon > 0$ there exists a finite partition $\mathcal{P}$ of Ω such that $\left\|\Gamma - \Gamma^{\ltimes\mathcal{P}}\right\|_1 < \varepsilon$.*

For the proof of Lemma 2.5, the following fact will be useful.

Fact 2.6. *Suppose that $f \in L^1(\Lambda)$ is an arbitrary function on a finite measure space Λ with measure λ. Set $a := \frac{1}{\lambda(\Lambda)} \cdot \int_\Lambda f$. Then for each $b \in \mathbb{R}$ we have that $\|f - a\|_1 \leq 2\|f - b\|_1$.*

Proof. We have

$$\begin{aligned}\|f-a\|_1 &= \int_\Lambda |f(x)-a| \leq \int_\Lambda |f(x)-b| + \int_\Lambda |a-b| = \|f-b\|_1 + \lambda(\Lambda)\cdot|a-b| \\ &= \|f-b\|_1 + \left|\int_\Lambda (f(x)-b)\right| \leq 2\|f-b\|_1 \ .\end{aligned}$$

□

Proof of Lemma 2.5. Since sets of the form $A \times B$, where $A, B \subseteq \Omega$ are measurable sets, generate the product sigma-algebra on Ω^2, there exists a finite partition $\mathcal{P}$ of Ω and a function $S : \Omega^2 \to \mathbb{R}$ such that S is constant on each rectangle of $\mathcal{P} \times \mathcal{P}$, and such that $\|\Gamma - S\|_1 < \frac{\varepsilon}{2}$. Now, for each rectangle $(A, B) \in \mathcal{P} \times \mathcal{P}$, we apply Fact 2.6 on the restricted function $\Gamma_{\restriction A\times B}$ and the constant $S_{\restriction A\times B}$. Summing up the contributions coming from these applications of Fact 2.6, we get that $\left\|\Gamma - \Gamma^{\ltimes\mathcal{P}}\right\|_1 \leq 2\|\Gamma - S\|_1 < \varepsilon$. □

We call $\Gamma^{\ltimes\mathcal{P}}$ with properties as in Lemma 2.5 *averaged L^1-approximation of Γ by a step-graphon for precision ε.*

2.5. **Hyperspace $K(\mathcal{W}_0)$.** Let X be a metrizable compact space. We denote as $K(X)$ the space of all compact subsets of X with the topology generated by sets of the form $\{L \in K(X) : L \subseteq U\}$ and $\{L \in K(X) : L \cap U \neq \emptyset\}$ where $U \subseteq X$ ranges over all open sets of X. Then $K(X)$ is called the *hyperspace of X with the Vietoris topology.*

Fact 2.7 ((4.22) and (4.26) in [13])**.** *Let X be a metrizable compact space with compatible metric ρ. Then $K(X)$ is metrizable compact (and hence separable). Furthermore, the Hausdorff metric on $K(X)$,*

$$d_\rho^{\mathrm{Hf}}(L,M) = \max\left\{\max_{x\in L}\{\rho(x,M)\}, \max_{y\in M}\{\rho(y,L)\}\right\} \tag{2.3}$$

is compatible with the Vietoris topology on $K(X)$.

Remark 2.8. *We will be interested in the situation where $X = \mathcal{W}_0$ is endowed with the weak* topology. By the discussion in Section 2.3, X is indeed metrizable compact.*

3. Weak* convergence and the cut distance

3.1. **Becoming familiar with $\mathbf{ACC}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ and $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$.** Let us observe some basic properties of the sets $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ and $\mathbf{ACC}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$. We have $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)\subseteq\mathbf{ACC}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$. The set $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ can be empty (for example when $\Gamma_1\equiv 0,\Gamma_2\equiv 1,\Gamma_3\equiv 0,\Gamma_4\equiv 1,\ldots$) but $\mathbf{ACC}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ is non-empty by the Banach–Alaoglu Theorem. Actually, we can describe some elements of $\mathbf{ACC}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ fairly easily. Let $T\subseteq[0,1]$ be the set of the accumulation points of the edge densities of the graphons $\Gamma_1,\Gamma_2,\Gamma_3,\ldots$, i.e., T is the set of the accumulation points of the sequence $\left(\int_x\int_y\Gamma_n(x,y)\right)_n$. Now, a constant $c\in[0,1]$ (viewed as a constant graphon) lies in $\mathbf{ACC}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ if and only if $c\in T$. The direction that $c\in\mathbf{ACC}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ implies $c\in T$ is obvious. Now, suppose that $c\in T$. That is, for some subsequence $\Gamma_{n_1},\Gamma_{n_2},\Gamma_{n_3},\ldots$ the densities converge to c. Partition each Γ_{n_i} into i sets of measure $\frac{1}{i}$ and consider a version $\widehat{\Gamma_{n_i}}$ of Γ_{n_i} obtained by a measure preserving bijection permuting these sets randomly. Then almost surely, $\widehat{\Gamma_{n_1}},\widehat{\Gamma_{n_2}},\widehat{\Gamma_{n_3}},\ldots$ weak* converge to c. This is included here just to get familiar with $\mathbf{ACC}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ and a proof is not needed at this point. However, the statement follows from Lemma 4.2(b).

The first non-trivial fact we will prove about the set $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ is that it is closed.

Lemma 3.1. *Let $\Gamma_1,\Gamma_2,\Gamma_3,\ldots\in\mathcal{W}_0$ be a sequence of graphons. Then the following holds for the set $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$.*

(a) $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ *is weak* closed in* $L^\infty\left(\Omega^2\right)$.
(b) $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ *is weak* compact in* $L^\infty\left(\Omega^2\right)$.
(c) $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ *is closed in* $L^1\left(\Omega^2\right)$.

Proof of Part (a). Suppose that $L_1,L_2,L_3,\ldots$ are elements of $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ such that $L_k\xrightarrow{w^*}L$ for $k\to\infty$. For every k let $\Gamma_1^k,\Gamma_2^k,\Gamma_3^k,\ldots$ be a sequence of versions of $\Gamma_1,\Gamma_2,\Gamma_3,\ldots$ converging to L_k. We find an increasing sequence $i_1,i_2,i_3,\ldots$ of integers such that for every k and for every $n\geq i_k$ we have $d_{w^*}(\Gamma_n^{k+1},L_{k+1})<\frac{1}{k}$. Then the following sequence of versions of $\Gamma_1,\Gamma_2,\Gamma_3,\ldots$ weak* converges to L:

$$\Gamma_1^1,\Gamma_2^1,\ldots,\Gamma_{i_1-1}^1,\Gamma_{i_1}^2,\Gamma_{i_1+1}^2,\ldots,\Gamma_{i_2-1}^2,\Gamma_{i_2}^3,\Gamma_{i_2+1}^3,\ldots,\Gamma_{i_3-1}^3,\ldots.$$

Proof of Part (b): Recall that the closed unit ball is compact in the weak* topology. Since $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ lies in this ball, it is weak* compact.

Proof of Part (c): The unit ball B of $L^\infty\left(\Omega^2\right)$ is closed in $L^1\left(\Omega^2\right)$. $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ is a weak* closed subset of B, and so it is also closed in B in the topology inherited from $L^1\left(\Omega^2\right)$ (by Fact 2.3). So, $\mathbf{LIM}_{w*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ is closed in $L^1\left(\Omega^2\right)$. □

Remark 3.2. *In* [8]*, a somewhat weaker closeness property of* $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ *than Lemma 3.1*(c) *was established and used, namely that the set*

$$\{\mathrm{INT}_f(W) : W \in \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)\}$$

attains it supremum (here, f *is a fixed continuous strictly convex function). Section 7.4 of* [8] *contains an example, due to Jon Noel, which shows that the set*

$$\{\mathrm{INT}_f(W) : W \in \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)\}$$

need not even achieve its supremum.

3.2. **Differentiating between $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ and $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ in** [8] **and in the present paper.** In the proof of Theorem 1.1 given in [8], which is in some sense a precursor of the current work, quite some work is put into zigzagging between $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ and $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$. Let us explain this in more detail. Let us fix a continuous strictly convex function f. The idea for finding the graphon Γ in Theorem 1.1 in [8] is as follows. Denoting by X either *(i)* $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ or *(ii)* $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$, we take $\Gamma \in X$ that maximizes $\mathrm{INT}_f(\Gamma)$. Using the definition of X, there exist versions $\Gamma'_{n_1},\Gamma'_{n_2},\Gamma'_{n_3},\ldots$ of $\Gamma_{n_1},\Gamma_{n_2},\Gamma_{n_3},\ldots$ that converge to Γ weak*.[e] The aim is to prove that $\Gamma'_{n_1},\Gamma'_{n_2},\Gamma'_{n_3},\ldots$ actually converge to Γ also in the cut norm — that would obviously prove Theorem 1.1. Now, the key step in [8] is to prove that if $\Gamma'_{n_1},\Gamma'_{n_2},\Gamma'_{n_3},\ldots$ do not converge to Γ in the cut norm, then there exist versions $\Gamma''_{n_{k_1}},\Gamma''_{n_{k_2}},\Gamma''_{n_{k_3}},\ldots$ of a suitable subsequence of $\Gamma'_{n_1},\Gamma'_{n_2},\Gamma'_{n_3},\ldots$ that weak* converge to a graphon Γ' with $\mathrm{INT}_f(\Gamma') > \mathrm{INT}_f(\Gamma)$. Since $\Gamma''_{n_{k_1}},\Gamma''_{n_{k_2}},\Gamma''_{n_{k_3}},\ldots$ witness that $\Gamma' \in X$, this is a contradiction. Now, let us explain why we need favorable properties of both (i) and (ii) for the proof. Firstly, note that in the sentence «Since $\Gamma''_{n_{k_1}},\Gamma''_{n_{k_2}},\Gamma''_{n_{k_3}},\ldots$ witness that $\Gamma' \in X$» we are referring to a subsequence, so this is a correct justification only in case $X = \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$. On the other hand, in the sentence «we take $\Gamma \in X$ that maximizes $\mathrm{INT}_f(\Gamma)$» we need the maximum to be achieved. Such a closeness property is enjoyed by $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ as we saw in Lemma 3.1(c), but not by $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ as we saw in Remark 3.2.

So, while differences between $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ and $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ were viewed in [8] as a nuisance that required a subtle and technical treatment, in this section we shall show that these differences capture the essence of the cut norm convergence. Namely, we shall prove in Theorem 3.3 that each sequence $\Gamma_1,\Gamma_2,\Gamma_3,\ldots$ of graphons contains a subsequence $\Gamma_{n_1},\Gamma_{n_2},\Gamma_{n_3},\ldots$ such that

$$\mathbf{LIM}_{\mathrm{w}*}\left(\Gamma_{n_1},\Gamma_{n_2},\Gamma_{n_3},\ldots\right) = \mathbf{ACC}_{\mathrm{w}*}\left(\Gamma_{n_1},\Gamma_{n_2},\Gamma_{n_3},\ldots\right)\ , \tag{3.1}$$

[e]Note that in variant (i), we actually have $n_1 = 1, n_2 = 2, n_3 = 3, \ldots$.

and in Theorem 3.5 we shall prove that (3.1) is equivalent to cut distance convergence of $\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots$. Of course, a proof of Theorem 1.1 then follows immediately.

3.3. **Main results: subsequences with $\mathbf{LIM}_{\mathrm{w}*} = \mathbf{ACC}_{\mathrm{w}*}$.** As we observed earlier we have $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) \subseteq \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ and equality usually does not hold. The next theorem however says that we can always achieve equality after passing to a subsequence.

Theorem 3.3. *Let $\mathcal{S} = (\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ be a sequence of graphons. Then there exists a subsequence $\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots$ such that $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots) = \mathbf{ACC}_{\mathrm{w}*}(\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots)$.*

The proof of Theorem 3.3 proceeds by transfinite induction. Crucially, we rely on a well-known fact from descriptive set theory, below referred to [13], which says that a strictly increasing transfinite sequence of closed sets in a second countable topological space must be of at most countable length. We shall apply this to the space $\left(L^\infty(\Omega^2), \mathrm{w}^*\right)$ which is second countable because it is metrizable and separable.

Proof. For two sequences[f] of graphons $\mathcal{U}$ and $\mathcal{T}$ we write $\mathcal{U} \leq^\star \mathcal{T}$ if deleting finitely many terms from $\mathcal{U}$ gives us a subsequence of $\mathcal{T}$. Note that the relation $\leq^\star$ is transitive. Note that if $\mathcal{U} \leq^\star \mathcal{T}$ then

$$\mathbf{LIM}_{\mathrm{w}*}(\mathcal{T}) \subseteq \mathbf{LIM}_{\mathrm{w}*}(\mathcal{U}) \ . \tag{3.2}$$

In the following, we construct a countable ordinal α_0 and a transfinite sequence $(\mathcal{S}_\alpha)_{\alpha \leq \alpha_0}$ of subsequences of $\mathcal{S}$ such that for every pair of ordinals $\gamma < \delta$ it holds that

$$\mathcal{S}_\delta \leq^\star \mathcal{S}_\gamma \ , \tag{3.3}$$

and also that $\mathbf{LIM}_{\mathrm{w}*}(\mathcal{S}_\gamma)$ is a proper subset of $\mathbf{LIM}_{\mathrm{w}*}(\mathcal{S}_\delta)$.

In the first step, we put $\mathcal{S}_0 = \mathcal{S}$. Now suppose that for some countable ordinal α, we have already constructed $\mathcal{S}_\beta$ for every $\beta < \alpha$. Either $\alpha = \beta + 1$ for some ordinal β or α is a limit ordinal. Suppose first that $\alpha = \beta + 1$ for some ordinal β. We distinguish two cases. If $\mathbf{LIM}_{\mathrm{w}*}(\mathcal{S}_\beta) = \mathbf{ACC}_{\mathrm{w}*}(\mathcal{S}_\beta)$ then we define $\alpha_0 = \beta$ and the construction is finished. Otherwise there is some graphon $W \in \mathbf{ACC}_{\mathrm{w}*}(\mathcal{S}_\beta) \setminus \mathbf{LIM}_{\mathrm{w}*}(\mathcal{S}_\beta)$. Then we proceed the construction by finding a subsequence $\mathcal{S}_{\beta+1}$ of $\mathcal{S}_\beta$ such that some versions of the graphons from $S_{\beta+1}$ converge to W. This way we have $\mathcal{S}_{\beta+1} \leq^\star \mathcal{S}_\beta$ and $W \in \mathbf{LIM}_{\mathrm{w}*}(\mathcal{S}_{\beta+1}) \setminus \mathbf{LIM}_{\mathrm{w}*}(\mathcal{S}_\beta)$. Now suppose that α is a countable limit ordinal. We find an increasing sequence $\beta_1, \beta_2, \beta_3, \ldots$ of ordinals such that $\beta_i \to \alpha$ for $i \to \infty$ (this is possible as α has countable cofinality). Now we use the diagonal method to define a sequence S_α such that $S_\alpha \leq^\star S_{\beta_i}$ for every i. Combined with (3.3) and with $\beta_i \to \alpha$, we get that $\mathcal{S}_\alpha \leq^\star \mathcal{S}_\beta$ for every $\beta < \alpha$. Plugging in (3.2), we conclude $\bigcup_{\beta<\alpha} \mathbf{LIM}_{\mathrm{w}*}(\mathcal{S}_\beta) \subseteq \mathbf{LIM}_{\mathrm{w}*}(\mathcal{S}_\alpha)$.

[f]By a *sequence*, we mean a system indexed by a countable initial segment of ordinals.

The obtained transfinite sequence $(\mathbf{LIM}_{w*}(\mathcal{S}_\alpha))_{\alpha\leq\alpha_0}$ is a strictly increasing sequence of subsets of unit ball in $L^\infty(\Omega^2)$. By Lemma 3.1(a), all of these subsets are weak* closed. By [13, Theorem 6.9], the sequence is at most countable, i.e. the previous construction stopped at some countable ordinal α_0. This means that $\mathbf{LIM}_{w*}(\mathcal{S}_{\alpha_0}) = \mathbf{ACC}_{w*}(\mathcal{S}_{\alpha_0})$. □

Remark 3.4. *Theorem 3.3 substantially extends the key Lemma 13 from* [8] *which states that any sequence of graphons* $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ *contains a subsequence* $\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots$ *such that*

$$\sup\{\mathrm{INT}_f(\Gamma) \,:\, \Gamma \in \mathbf{LIM}_{w*}(\Gamma_{n_1}, \Gamma_{n_2}, \ldots)\} = \sup\{\mathrm{INT}_f(\Gamma) \,:\, \Gamma \in \mathbf{ACC}_{w*}(\Gamma_{n_1}, \Gamma_{n_2}, \ldots)\}\ ,$$

for a continuous strictly convex function $f : [0,1] \to \mathbb{R}$. *Lemma 13 in* [8] *is proved by induction (over natural numbers) without any appeal to descriptive set theory.*

As promised, we shall now state that the property asserted in Theorem 3.3 is necessary and sufficient for cut distance convergence.

Theorem 3.5. *Let* $\Gamma_1, \Gamma_2, \Gamma_3, \ldots \in \mathcal{W}_0$. *The following are equivalent:*

(a) The sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ *is Cauchy with respect to the cut distance* $\delta_\square$,
(b) $\mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) = \mathbf{ACC}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$.

Furthermore, in case (a) (and hence also (b)) hold, we can take a maximal element W *in* $\mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ *with respect to the structuredness order (defined in Section 4 below) and then* $\Gamma_1, \Gamma_2, \Gamma_3, \ldots \xrightarrow{\delta_\square} W$.

We provide a proof of Theorem 3.5 in Section 5, after building key tools in Section 4. In Section 6 we state and prove Theorem 6.1 which extends Theorem 3.5 and relates cut distance convergence to convergence in the hyperspace $K(\mathcal{W}_0)$.

4. Envelopes and the structuredness order

Suppose that $W \in \mathcal{W}_0$ is a graphon. We call the set $\langle W\rangle := \mathbf{LIM}_{w*}(W, W, W, \ldots)$ the *envelope* of W. Envelopes allow us to introduce structuredness order on graphons. Intuitively, less-structured graphons have smaller envelopes. Extreme examples of this are constant graphons $W \equiv c$ (for some $c \in [0,1]$), which are obviously the only graphons for which $\langle W\rangle = \{W\}$. This leads us to say that a graphon U is *at most as structured as a graphon* W if $\langle U\rangle \subseteq \langle W\rangle$. We write $U \preceq W$ in this case. We write $U \prec W$ if $U \preceq W$ but it does not hold that $W \preceq U$. Observe that $\preceq$ is a quasiorder on the space of graphons and if $U \preceq W$ then also $U^\varphi \preceq W$ for every measure preserving bijection φ. As we shall see in Lemma 8.1, it is actually an order on the space of graphons modulo weak isomorphism. To prove these results we shall need several auxiliary results.

Lemma 4.1 (Lemma 7 in [8]). *Suppose that $\Gamma_1, \Gamma_2, \Gamma_3, \ldots : \Omega^2 \to [0,1]$ is a sequence of graphons. Suppose that $W \in \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ and that we have a partition $\mathcal{P}$ of Ω into finitely many sets. Then $W^{\ltimes \mathcal{P}} \in \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$.*

Lemma 4.2. *Suppose that $W \in \mathcal{W}_0$. Then*

(a) If $Q \subseteq \langle W \rangle$ then the weak* closure of Q is also contained in $\langle W \rangle$,
(b) $W^{\ltimes \mathcal{P}} \in \langle W \rangle$ for every finite partition $\mathcal{P}$ of Ω,
(c) $U \in \langle W \rangle$ if and only if $U \preceq W$,
(d) if $\delta_\square(W, U) = 0$ then $\langle W \rangle = \langle U \rangle$.

Proof. Item (a) follows from Lemma 3.1(a). Item (b) is a special case of Lemma 4.1.

Let us now turn to Item (c). If $U \preceq W$ then $U \in \langle W \rangle$ follows from the definition of $\preceq$ and the fact that $U \in \langle U \rangle$. To prove the opposite implication observe that if $(\varphi_n : \Omega \to \Omega)_n$ is a sequence of measure preserving bijections, $W^{\varphi_n} \xrightarrow{\mathrm{w}^*} U$ and $\psi : \Omega \to \Omega$ is a measure preserving bijection then $W^{\varphi_n \psi} \xrightarrow{\mathrm{w}^*} U^\psi$. Then we have that every version of U is in $\langle W \rangle$. Because $\langle U \rangle$ is exactly the weak* closure of the set of all versions of U we obtain that $U \preceq W$.

Let us now prove Item (d). If $\delta_\square(W, U) = 0$ then we have a sequence $W^{\varphi_n} \xrightarrow{\|\cdot\|_\square} U$ which by Fact 2.3 implies that $W^{\varphi_n} \xrightarrow{\mathrm{w}^*} U$. Therefore $U \preceq W$. A symmetric argument gives $W \preceq U$ and we may conclude that $\langle W \rangle = \langle U \rangle$. □

4.1. **Beyond Lemma 4.2(b).** In this short section we borrow results stated and proven further below to show, in Theorem 4.3, that Lemma 4.2(b) can be strengthened. We do not have applications of Theorem 4.3. On the other hand, together with the complement of Theorem 4.3, which we include as Problem 4.4, this would lead to a correspondence between the structuredness order and order on sub-sigma-algebras of the Borel sigma-algebra. Let us now give details.

Let $\mathcal{B}$ be the sigma-algebra of measurable sets on Ω. Suppose that $W : \Omega^2 \to [0,1]$ and $\mathcal{P}$ is a finite partition of Ω into measurable sets. Let $\mathcal{P}^*$ be the algebra generated by $\mathcal{P}$; clearly $\mathcal{P}^*$ is finite. We can express $W^{\ltimes \mathcal{P}}$ as a conditional expectation of W with respect to a sub-sigma-algebra, $W^{\ltimes P} = \mathbf{E}[W | \mathcal{P}^* \times \mathcal{P}^*]$. Then Lemma 4.2(b) tells us that this particular conditional expectation lies in $\langle W \rangle$. Here, the fact that we are taking a conditional expectation with respect to a finite sub-sigma-algebra is not needed.

Theorem 4.3. *Suppose $\mathcal{B}$ is the sigma-algebra of measurable sets on Ω, $\mathcal{A}$ is a sub-sigma-algebra of $\mathcal{B}$, and that $W : \Omega^2 \to [0,1]$ is a graphon. Then $\mathbf{E}[W | \mathcal{A} \times \mathcal{A}] \in \langle W \rangle$.*

Proof. Recall that we assume that Ω is a separable atomless measure space. This implies that there is an increasing sequence $\{\mathcal{A}_n\}_{n \in \mathbb{N}}$ of finite subalgebras of $\mathcal{A}$ such that $\bigcup_{n \in \mathbb{N}} \mathcal{A}_n$ generates $\mathcal{A}$, i.e., $\mathcal{A}$ is the smallest sub-sigma-algebra of $\mathcal{B}$ that contains $\mathcal{A}_n$ for every $n \in \mathbb{N}$.

Clearly, there are finite partitions $\mathcal{P}_n$ of Ω for every $n \in \mathbb{N}$ such that $\mathcal{P}_n^* = \mathcal{A}_n$, where we use the notation from the beginning of this subsection. It follows that $W^{\bowtie P_n} \in \left\langle W^{\bowtie P_{n+1}} \right\rangle$ and by Corollary 8.3 together with Lemma 4.5 we have that $W^{\bowtie P_n} \xrightarrow{\delta_\square} U$ where $U \in \langle W \rangle$. It follows from the Martingale convergence theorem [2, Theorem 35.5] that $W^{\bowtie P_n} \xrightarrow{\|\cdot\|_1} \mathbf{E}[W|\mathcal{A} \times \mathcal{A}]$ and as a consequence we get that $\delta_\square(\mathbf{E}[W|\mathcal{A} \times \mathcal{A}], U) = 0$. □

As advertised, we pose the complement of Theorem 4.3, which would give one-to-one correspondence between the structuredness order and containment of sub-sigma-algebras of $\mathcal{B}$, as an open problem.

Problem 4.4. Suppose $\mathcal{B}$ is the sigma-algebra of measurable sets on Ω, and that $U \preceq W$ are two graphons. Do there exist graphons U', W', and a sub-sigma-algebra $\mathcal{A}$ of $\mathcal{B}$ such that $\delta_\square(U, U') = \delta_\square(W, W') = 0$ and $U' = \mathbf{E}[W'|\mathcal{A} \times \mathcal{A}]$?

4.2. **$\preceq$-maximal elements in $\mathbf{LIM}_{\mathrm{w}*}$.** The main result of this section, Lemma 4.9, says that if

$$\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) = \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$$

then there exists a $\preceq$-maximal element in $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. Most of the work for the proof of Lemma 4.9 is done in Lemma 4.8 which is stated for step graphons only. To infer that certain favorable properties of a sequence of step graphons (on which Lemma 4.8 can be applied) can be transferred even to a graphon they approximate, Lemma 4.5 is introduced.

Lemma 4.5. *Suppose $U_1, U_2, U_3, \ldots$ is a sequence of graphons that converges weak* to U, and suppose that W is a graphon. Suppose that for each $n \in \mathbb{N}$ we have that $U_n \preceq W$. Then $U \preceq W$.*

Proof. Follows immediately from Lemma 4.2(a). □

For the key Lemma 4.8, we shall refine the structure of a graphon by «moving some parts to the left». To this end, it is convenient to work on $[0, 1]$ (see Remark 2.1). We introduce the following definitions.

Definition 4.6. By an *ordered partition* $\mathcal{P}$ of a set S, we mean a finite partition of S, $S = P_1 \sqcup P_2 \sqcup P_3 \sqcup \ldots \sqcup P_k$, $\mathcal{P} = (P_1, P_2, P_3, \ldots, P_k)$ in which the sets $P_1, P_2, P_3, \ldots, P_k$ are linearly ordered (in the way they are enumerated in $\mathcal{P}$).

Definition 4.7. For an ordered partition $\mathcal{J}$ of $I = [0,1]$ into finitely many sets $C_1, C_2, \ldots, C_k$, we define mappings $\alpha_{\mathcal{J},1}, \alpha_{\mathcal{J},2}, \ldots, \alpha_{\mathcal{J},k} : I \to I$, and a mapping $\gamma_{\mathcal{J}} : I \to I$ by

$$\alpha_{\mathcal{J},i}(x) = \lambda\left(\bigcup_{i=1}^{j-1} C_j\right) + \lambda\left(C_i \cap [0,x]\right), \quad i = 1,2,\ldots,k\,. \tag{4.1}$$

$$\gamma_{\mathcal{J}}(x) = \alpha_{\mathcal{J},i}(x) \quad \text{if } x \in C_i, \quad i = 1,2,\ldots,k\,.$$

Informally, $\gamma_{\mathcal{J}}$ is defined in such a way that it maps the set C_1 to the left side of the interval I, the set C_2 next to it, and so on. Finally, the set C_k is mapped to the right side of the interval I. Clearly, $\gamma_{\mathcal{J}}$ is a measure preserving almost-bijection.

Last, given a graphon $W : I^2 \to [0,1]$, we define a graphon ${}_{\mathcal{J}}W : I^2 \to [0,1]$ by

$${}_{\mathcal{J}}W(x,y) := W\left(\gamma_{\mathcal{J}}^{-1}(x), \gamma_{\mathcal{J}}^{-1}(y)\right)\,. \tag{4.2}$$

When $\mathcal{J}$ has only two parts, $\mathcal{J} = (A, I \setminus A)$, in order to simplify notation, we write

$${}_{A}W := {}_{\mathcal{J}}W\,. \tag{4.3}$$

Lemma 4.8. *Let $\Gamma_1, \Gamma_2, \Gamma_3, \ldots \in \mathcal{W}_0$ be a sequence of graphons on $[0,1]$. Suppose that $U, V \in \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ are two step graphons. Then there exists a step graphon $W \in \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \ldots)$ that refines U and such that $U, V \preceq W$.*

Proof. We at first assume that the ordered partition $\mathcal{P} = (P_1, \ldots, P_m)$ of U and the ordered partition $\mathcal{Q} = (Q_1, \ldots, Q_n)$ of V are composed of intervals, since each partition $\mathcal{J}$ can be reordered to intervals by the measure preserving almost-bijection $\gamma_{\mathcal{J}}$. We further assume without loss of generality that the sequence of measure preserving bijections certifying that $U \in \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \ldots)$ contains only identities, i.e., that $\Gamma_1, \Gamma_2, \ldots \xrightarrow{\mathrm{w}^*} U$. Let $\varphi_1, \varphi_2, \ldots$ be measure preserving bijections such that $\Gamma_1^{\varphi_1^{-1}}, \Gamma_2^{\varphi_2^{-1}}, \ldots \xrightarrow{\mathrm{w}^*} V$.

We now describe a sequence of measure preserving bijections $\psi_1, \psi_2, \ldots$ such that the weak star limit of $\Gamma_1^{\psi_1^{-1}}, \Gamma_2^{\psi_2^{-1}} \ldots$ gives the desired graphon W. For the ℓ-th graphon Γ_ℓ we define its partition $\mathcal{H}^{(\ell)} = \left(H_{1,1}^{(\ell)}, H_{1,2}^{(\ell)}, \ldots, H_{1,n}^{(\ell)}, H_{2,1}^{(\ell)} \ldots, H_{m,n}^{(\ell)}\right)$ where $H_{i,j}^{(\ell)} = P_i \cap \varphi_\ell^{-1}(Q_j)$ and set $\psi_\ell = \gamma_{\mathcal{H}^{(\ell)}}$ using Definition 4.7. The intuition behind ψ_ℓ is that it refines each block P_i of the partition $\mathcal{P}$ with the partition $\mathcal{Q}$; it can be, indeed, seen that for each i we have $\psi_\ell(P_i) = \psi_\ell\left(\bigcup_j H_{i,j}^{(\ell)}\right) = P_i$, where the equalities hold up to a null set.

We pass to a subsequence mn times to get that both endpoints of each of the intervals $\psi_\ell(H_{i,j}^{(\ell)})$ converge to some fixed numbers from $[0,1]$, thus giving us a limit partition $\mathcal{R} = (R_{1,1}, \ldots, R_{m,n})$ into intervals (some of them may be degenerate intervals of length 0). Note that as we know that for each i we have $\psi_\ell\left(\bigcup_j H_{i,j}^{(\ell)}\right) = P_i$, it is also true that $\bigcup_j R_{i,j} = P_i$. We also have for all j that $\nu\left(\bigcup_i H_{i,j}^{(\ell)}\right) = \nu(Q_j)$, hence $\nu\left(\bigcup_i R_{i,j}\right) = \nu(Q_j)$. Now we use the fact that the set of accumulation points of our sequence is non-empty due to Banach–Alaoglu theorem, thus after passing to a subsequence

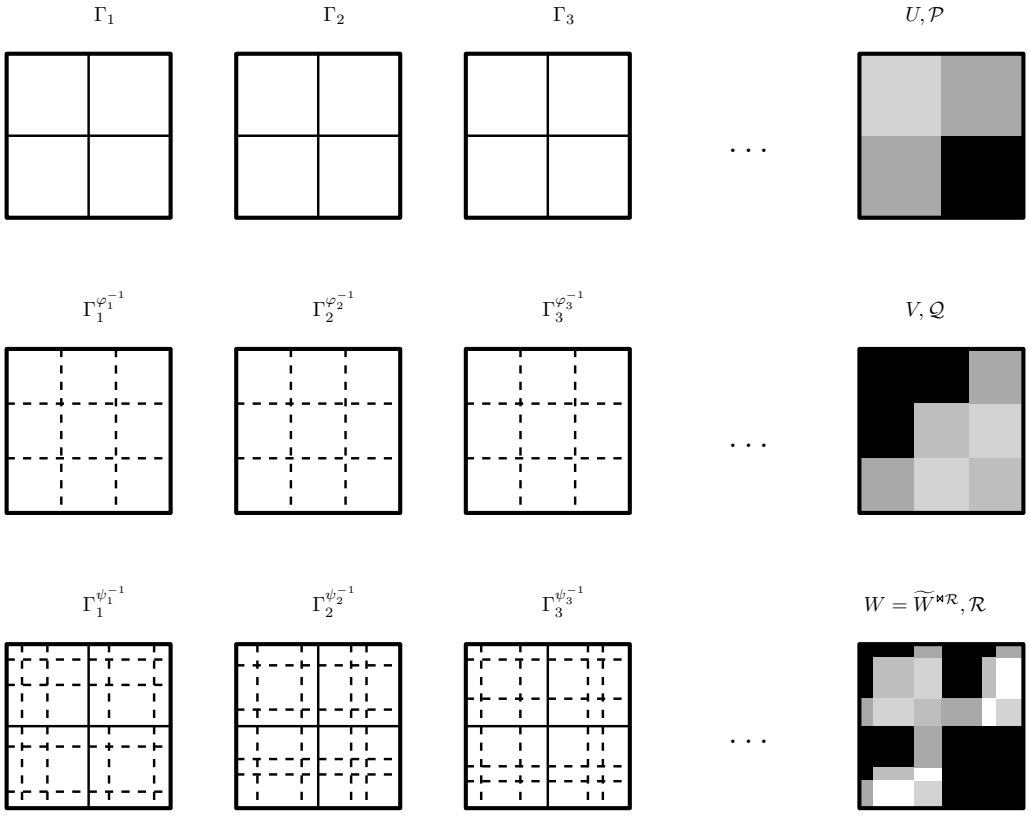


FIGURE 4.1. Two graphons U, V are step graphons with partitions $\mathcal{P}$ and $\mathcal{Q}$. A subsequence of $\Gamma_1^{\psi_1^{-1}}, \Gamma_2^{\psi_2^{-1}}, \dots$ converges to $\widetilde{W}$ and the corresponding partitions converge to the partition $\mathcal{R}$ that refines both $\mathcal{P}$ and $\mathcal{Q}$. The graphon $W = \widetilde{W}^{\wedge\mathcal{R}}$ is the desired step graphon that is structured more than both U and V.

yet again we get a subsequence $\Gamma_{k_1}^{\psi_{k_1}^{-1}}, \Gamma_{k_2}^{\psi_{k_2}^{-1}}, \dots \xrightarrow{\text{w}^*} \widetilde{W}$. Define W as $\widetilde{W}^{\wedge\mathcal{R}}$. We apply Lemma 3.1(a) to $\mathbf{LIM}_{\text{w}*}(\Gamma_{k_1}, \Gamma_{k_2}, \dots)$ and Lemma 4.2(b) to $\widetilde{W}$ and $\widetilde{W}^{\wedge\mathcal{R}}$ to get that $\widetilde{W}^{\wedge\mathcal{R}} \in \mathbf{ACC}_{\text{w}*}(\Gamma_1, \Gamma_2, \dots)$.

At first we prove that $\widetilde{W}^{\wedge\mathcal{R}}$ refines U. Since U is constant on each step $P_i \times P_j$, it suffices to prove that for any $\varepsilon > 0$ and any step $P_i \times P_j$ we have

$$\int_{P_i \times P_j} U = \int_{P_i \times P_j} \widetilde{W}^{\wedge\mathcal{R}}. \tag{4.4}$$

Take ℓ sufficiently large, so that

$$\left| \int_{P_i \times P_j} U - \int_{P_i \times P_j} \Gamma_{k_\ell} \right| < \varepsilon \tag{4.5}$$

and

$$\left| \int_{P_i \times P_j} \widetilde{W} - \int_{P_i \times P_j} \Gamma_{k_\ell}^{\psi_{k_\ell}^{-1}} \right| < \varepsilon. \tag{4.6}$$

Putting this together with the facts that $P_i = \psi_{k_\ell}(P_i)$ up to a null set and $R_{i,j} \subseteq P_i$ for all i, j, we get that

$$\begin{aligned}
&\int_{P_i \times P_j} U \\
\boxed{\text{Eq. (4.5)}} \quad &\overset{\varepsilon}{\approx} \int_{P_i \times P_j} \Gamma_{k_\ell} \\
\boxed{P_i = \psi_{k_\ell}(P_i)} \quad &= \int_{\psi_{k_\ell}(P_i) \times \psi_{k_\ell}(P_j)} \Gamma_{k_\ell} \\
&= \int_{P_i \times P_j} \Gamma_{k_\ell}^{\psi_{k_\ell}^{-1}} \\
\boxed{\text{Eq. (4.6)}} \quad &\overset{\varepsilon}{\approx} \int_{P_i \times P_j} \widetilde{W} \\
&= \int_{P_i \times P_j} \widetilde{W}^{\bowtie \mathcal{R}}.
\end{aligned}$$

Since this holds for every $\varepsilon > 0$, we get the desired Equation (4.4).

Now we prove that $\widetilde{W}^{\bowtie \mathcal{R}} \succeq V$. Since V is constant on each step $Q_i \times Q_j$, it suffices to prove that for any $\varepsilon > 0$ and any step $Q_i \times Q_j$ of V we have

$$\int_{Q_i \times Q_j} V = \sum_{\substack{1 \le g \le m \\ 1 \le h \le n}} \int_{R_{g,i} \times R_{h,j}} \widetilde{W}^{\bowtie \mathcal{R}}. \tag{4.7}$$

Take ℓ sufficiently large so that

$$\left| \int_{Q_i \times Q_j} V - \int_{Q_i \times Q_j} \Gamma_{k_\ell}^{\varphi_{k_\ell}^{-1}} \right| < \varepsilon, \tag{4.8}$$

$$\left| \sum_{g,h} \int_{R_{g,i} \times R_{h,j}} \widetilde{W}^{\bowtie \mathcal{R}} - \sum_{g,h} \int_{R_{g,i} \times R_{h,j}} \Gamma_{k_\ell}^{\psi_{k_\ell}^{-1}} \right| < \varepsilon \tag{4.9}$$

and, moreover, the length of each interval $\psi_{k_\ell}\left(H_{i,j}^{(k_\ell)}\right)$ differs from the length of interval $R_{i,j}$ by at most $\frac{\varepsilon}{4m^2n^2}$. Now we can bound the measure of overlap of each pair of rectangles $R_{g,i} \times R_{h,j}$ and $\psi_{k_\ell}\left(H_{g,i}^{(k_\ell)}\right) \times \psi_{k_\ell}\left(H_{h,j}^{(k_\ell)}\right)$. More precisely, we have

$$\nu^{\otimes 2}\left((R_{g,i} \times R_{h,j}) \triangle \left(\psi_{k_\ell}\left(H_{g,i}^{(k_\ell)}\right) \times \psi_{k_\ell}\left(H_{h,j}^{(k_\ell)}\right)\right)\right) < 4mn \cdot \frac{\varepsilon}{4m^2n^2} = \frac{\varepsilon}{mn}, \tag{4.10}$$

where $4mn$ comes from the facts that we bound the displacement of all four sides of the rectangles and that their displacement depends on the displacement of all preceding intervals. Putting all of this together, we get that

$$\begin{aligned}
&\sum_{g,h} \int_{R_{g,i}\times R_{h,j}} \widetilde{W}^{\bowtie\mathcal{R}} \\
&\qquad \boxed{\text{Eq. (4.9)}} \quad \overset{\varepsilon}{\approx} \sum_{g,h} \int_{R_{g,i}\times R_{h,j}} \Gamma_{k_\ell}^{\psi_{k_\ell}^{-1}} \\
&\qquad \boxed{\text{Eq. (4.10)}} \quad \overset{\varepsilon}{\approx} \sum_{g,h} \int_{\psi_{k_\ell}\left(H_{g,i}^{(k_\ell)}\right)\times\psi_{k_\ell}\left(H_{h,j}^{(k_\ell)}\right)} \Gamma_{k_\ell}^{\psi_{k_\ell}^{-1}} \\
&\qquad\qquad = \sum_{g,h} \int_{\left(P_g\cap\varphi_{k_\ell}^{-1}(Q_i)\right)\times\left(P_h\cap\varphi_{k_\ell}^{-1}(Q_j)\right)} \Gamma_{k_\ell} \\
&\qquad\qquad = \sum_{g,h} \int_{\left(\varphi_{k_\ell}(P_g)\cap Q_i\right)\times\left(\varphi_{k_\ell}(P_h)\cap Q_j\right)} \Gamma_{k_\ell}^{\varphi_{k_\ell}^{-1}} \\
&\qquad\qquad = \int_{Q_i\times Q_j} \Gamma_{k_\ell}^{\varphi_{k_\ell}^{-1}} \\
&\qquad \boxed{\text{Eq. (4.8)}} \quad \overset{\varepsilon}{\approx} \int_{Q_i\times Q_j} V \ .
\end{aligned}$$

This yields the desired equation (4.7). □

Lemma 4.9. *Let $\Gamma_1, \Gamma_2, \Gamma_3, \ldots \in \mathcal{W}_0$ be a sequence of graphons on $[0,1]$ for which*

$$\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) = \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) \ .$$

Then $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ contains a maximum element with respect to the structuredness order.

Proof. The space $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ is separable metrizable since the space $\mathcal{W}_0$ with the weak* topology is separable metrizable and therefore we may find a countable set

$$P \subseteq \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$$

such that its weak* closure is $\mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. For each $W \in P$ and $k \in \mathbb{N}$, consider a suitable graphon, denoted by $W(k)$, that is an averaged L^1-approximation of W by a step graphon for precision $\frac{1}{k}$. Such a graphon $W(k)$ exists by Lemma 2.5. Note also that if $W(k)$ is chosen as $W^{\ltimes\mathcal{P}}$ for some finite partition $\mathcal{P}$ (as in Lemma 2.5) then $W(k) \in \mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ by Lemma 4.1.

Let us now consider the set $Q := \{W(k) : W \in P, k \in \mathbb{N}\}$. The set Q is countable, contained in $\mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ and its weak* closure is $\mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. Let $U_1, U_2, U_3, \ldots$ be an enumeration of the elements of Q. Let $M_1 := U_1$ and define inductively a sequence $\{M_n\}_{n\in\mathbb{N}} \subseteq \mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ as follows. Suppose that we have $M_n \in \mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ and use Lemma 4.8 with M_n in place of U and U_{n+1} in place of V to find $M_{n+1} \in \mathbf{ACC}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. It follows from the assumption that $\mathbf{ACC}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) = \mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ that, in fact, $M_{n+1} \in \mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. Note that we have $U_n \preceq M_n$ for every $n \in \mathbb{N}$. Moreover, it follows from the construction that there is a sequence of finite partitions $\{\mathcal{P}_n\}_{n\in\mathbb{N}}$ such that $M_n = M_{n+1}^{\ltimes\mathcal{P}_n}$ and $\mathcal{P}_{n+1}$ is a refinement of $\mathcal{P}_n$ for every $n \in \mathbb{N}$.

By the Martingale convergence theorem [2, Theorem 35.5] we find M such that $M_n \xrightarrow{\|\cdot\|_1} M$ and $M^{\ltimes\mathcal{P}_n} = M_n$ for every $n \in \mathbb{N}$. Since $\mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ is weak* closed by Lemma 3.1(a) we have $M \in \mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ and $M \succeq U_n$ for every $n \in \mathbb{N}$.

Finally, we claim that M is a maximal element of $\mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. Indeed, let $\Gamma \in \mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ be arbitrary. Since Q is weak* dense in $\mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$, we can find a sequence $U_{n_1}, U_{n_2}, U_{n_3}, \ldots$ weak* converging to Γ. Note that we have $U_{n_i} \preceq M$ for each $i \in \mathbb{N}$. Lemma 4.5 gives $\Gamma \preceq M$, as was needed. □

4.3. **Cut distance identifying graphon parameters.** In this section, we mention the notion of cut distance identifying graphon parameters, which we then extensively study in a follow-up paper [9]. A *graphon parameter* is any function $\theta : \mathcal{W}_0 \to \mathbb{R}$ such that $\theta(W_1) = \theta(W_2)$ for any two graphons W_1 and W_2 with $\delta_\square(W_1, W_2) = 0$. We say that a graphon parameter $\theta(\cdot)$ is a *cut distance identifying graphon parameter* (*CDIP*) if we have that $W_1 \prec W_2$ implies $\theta(W_1) < \theta(W_2)$. In particular, the key role of $\prec$-maximal element, such as in Theorem (b), can be in this setting expressed as θ-maximal elements. Note that $\mathrm{INT}_f(\cdot)$, where f is a strictly convex convex, is CDIP by [9, Section 3.3]. Hence the proof of Theorem 1.1 in [8] is a particular instance of a CDIP-version of Theorem (b), the general version then being given in Theorem 3.3 and Theorem 3.4 in [9]. As we show in [9], CDIP have more applications: they can be used for «index-pumping» in the proof of the Frieze-Kannan regularity lemma, and they tell us quite a bit about graph norms.

4.4. **Values and degrees with respect to the structuredness order.** Given a graphon $W : \Omega^2 \to [0,1]$, we can define a pushforward probability measure on $[0,1]$ by

$$\mathbf{\Phi}_W(A) := \nu^{\otimes 2}\left(W^{-1}(A)\right) , \tag{4.11}$$

for a set $A \subseteq [0,1]$. The measure $\mathbf{\Phi}_W$ gives us the distribution of the values of W, and we call it *the range frequencies of* W. Similarly, we can take the pushforward measure of the degrees,

$$\mathbf{\Upsilon}_W(A) := \nu\left(\deg_W^{-1}(A)\right) , \tag{4.12}$$

for a set $A \subseteq [0,1]$. We call $\mathbf{\Upsilon}_W$ the *degree frequencies of* W. The measures $\mathbf{\Phi}_W$ and $\mathbf{\Upsilon}_W$ do not characterize W in general but certainly give us substantial information about W. Therefore, given two graphons $U \preceq W$ it is natural to ask how $\mathbf{\Phi}_U$ compares to $\mathbf{\Phi}_W$ and how $\mathbf{\Upsilon}_U$ compares to $\mathbf{\Upsilon}_W$. To this end, we introduce the following concept.

Definition 4.10. Suppose that Λ_1 and Λ_2 are two finite measures on $[0,1]$. We say that Λ_1 is at *least as flat* as Λ_2 if there exists a finite measure Ψ on $[0,1]^2$ such that Λ_1 is the marginal of Ψ on the first coordinate, Λ_2 is the marginal of Ψ on the second coordinate, and for each measurable $D \subseteq [0,1]$ we have

$$\int_{D\times[0,1]} x \, \mathrm{d}\Psi(x,y) = \int_{D\times[0,1]} y \, \mathrm{d}\Psi(x,y) . \tag{4.13}$$

In addition, we say that Λ_1 is *strictly flatter* than Λ_2 if $\Lambda_1 \neq \Lambda_2$.

The condition that the marginals of Ψ are Λ_1 and Λ_2 is well-known in probability theory and referred to as a *coupling of* Λ_1 *and* Λ_2. However, we are not aware of this concept being combined with the condition (4.13).

Example 4.11. We should understand Λ_1 as a certain averaging of Λ_2. For example, suppose that Λ_1 has an atom $a \in [0,1]$, say $\Lambda_1(\{a\}) = m > 0$. Then taking $D = \{a\}$, (4.13) tells us that by averaging y according to the normalized (by $\frac{1}{m}$) restriction of the measure Ψ to $\{a\} \times [0,1]$, we get a. As we show in Lemma 4.12, such a property extends also to non-atoms.

Let us prove two basic lemmas. In Lemma 4.14 we give a useful characterization of strictly flatter pairs of measures. In Lemma 4.15 we prove that the flatness relation is actually an order. Even though we do not need these lemmas, we believe that the theory we develop here would not be complete without them.

We say that a finite measure Ψ on $[0,1]^2$ is *diagonal* if we have

$$\Psi\left(\left\{(x,y) \in [0,1]^2 : x \neq y\right\}\right) = 0 .$$

Let us now recall the notion of disintegration of a measure. Suppose that (X,μ) is a probability Borel measure space on a Polish space X. Let $f : X \to Y$ be a Borel map

onto another Polish space Y and denote as $f^*\mu$ the push-forward measure via f. Then the Disintegration Theorem tells us that there is a system $\{F_y\}_{y\in Y}$ of probability Borel measures on X such that

(D1) $F_y\left(f^{-1}(y)\right) = 1$ for every $y \in Y$, and
(D2) $\int_X h(x)\,\mathrm{d}\mu(x) = \int_Y \left(\int_X h(x)\,\mathrm{d}F_y(x)\right)\mathrm{d}f^*\mu(y)$ for every Borel map $h : X \to [0,1]$.

We will use the disintegration exclusively in the situation where $X = [0,1]^2$, f is the projection on the i-th coordinate, $\mu = \Phi$ and $f^*\mu = \Lambda_i$ where Φ is a witness for the fact that Λ_1 is at least as flat as Λ_2. When we use the variable x for the first coordinate and y for the second coordinate then we obtain a disintegration $\left\{\Phi^1_x\right\}_{x\in[0,1]}$ and $\left\{\Phi^2_y\right\}_{y\in[0,1]}$. Moreover, to simplify notation we will always assume that each Φ^i_z lives on the interval $[0,1]$ instead of the corresponding (horizontal or vertical) strip. For example in case of disintegration on the second coordinate, the two conditions above then look like this:

(D1) $\Phi^2_y([0,1]) = 1$ for every $y \in [0,1]$, and
(D2) $\int_{[0,1]^2} h(x,y)\,\mathrm{d}\Phi(x,y) = \int_{[0,1]}\left(\int_{[0,1]} h(x,y)\,\mathrm{d}\Phi^2_y(x)\right) d\Lambda_2(y)$ for every Borel map $h : [0,1]^2 \to [0,1]$.

Lemma 4.12. *Suppose that Λ_1 is at least as flat as Λ_2, witnessed by a measure Φ. Then for Λ_1-almost every $x \in [0,1]$ we have*

$$x = \int_{[0,1]} y\,\mathrm{d}\Phi^1_x(y)\,. \tag{4.14}$$

Proof. Assume not, then we may assume without loss of generality that there is a set $D \subseteq [0,1]$ of positive Λ_1-measure such that $x > \int_{[0,1]} y\,\mathrm{d}\Phi^1_x(y)$ for each $x \in D$. Then we have

$$\int_{D\times[0,1]} x\,\mathrm{d}\Phi(x,y) = \int_D x\,\mathrm{d}\Lambda_1(x) > \int_D \left(\int_{[0,1]} y\,\mathrm{d}\Phi^1_x(y)\right)\mathrm{d}\Lambda_1(x) \overset{(D2)}{=} \int_{D\times[0,1]} y\,\mathrm{d}\Phi(x,y)$$

which contradicts (4.13). □

Lemma 4.13. *Suppose that Λ_1 is at least as flat as Λ_2. Then we have $\Lambda_1([0,1]) = \Lambda_2([0,1])$.*

Proof. Let Ψ be a witness that Λ_1 is at least as flat as Λ_2. Then the marginal condition of Definition 4.10 tells us that

$$\Lambda_1([0,1]) = \Psi([0,1]\times[0,1]) = \Lambda_2([0,1])\,.$$

□

Lemma 4.14. *Suppose that Λ_1 and Λ_2 are finite measures on $[0,1]$, and that Λ_1 is at least as flat as Λ_2. Then $\Lambda_1 = \Lambda_2$ if and only if the only measure which witnesses that Λ_1 is at least as flat as Λ_2 is diagonal.*

Proof. Suppose that we have a diagonal measure Φ whose marginals are Λ_1 and Λ_2. Then for every measurable set D we have

$$\Lambda_1(D) = \Phi(D \times [0,1]) = \Phi(D \times D) = \Phi([0,1] \times D) = \Lambda_2(D)\,,$$

and so $\Lambda_1 = \Lambda_2$.

On the other hand, suppose that we have a non-diagonal measure Φ whose marginals are Λ_1 and Λ_2. By Lemma 4.13 we may assume that both Λ_1 and Λ_2 are probability measures. Let us fix a strictly convex function $f : [0,1] \to [0,1]$. Let us consider the disintegration $\left\{\Phi^1_x\right\}_{x\in[0,1]}$ of Φ. Recall that by (D1), for each $x \in [0,1]$, Φ^1_x is a probability measure. In particular, Jensen's inequality gives us

$$f\left(\int y \,\mathrm{d}\Phi^1_x(y)\right) \leq \int f(y)\ \mathrm{d}\Phi^1_x(y)\,.$$

Observe that for a positive Λ_1-measure of x's, we have that Φ^1_x is not a Dirac measure. For each such x, the inequality above is strict. Then we have

$$\begin{aligned}\int f(x)\,\mathrm{d}\Lambda_1(x) &= \int f\left(\int y\,\mathrm{d}\Phi^1_x(y)\right)\,\mathrm{d}\Lambda_1(x)\\ \boxed{\text{Jensen's inequality as above}} \quad &< \int\left(\int f(y)\ \mathrm{d}\Phi^1_x(y)\right)\,\mathrm{d}\Lambda_1(x) = \int f(y)\,\mathrm{d}\Phi(x,y)\\ &= \int f(y)\ \mathrm{d}\Lambda_2(y).\end{aligned}$$

In particular, we can conclude that $\Lambda_1 \neq \Lambda_2$. □

Lemma 4.15. *Suppose that $\Lambda_\mathrm{A}, \Lambda_\mathrm{B}, \Lambda_\mathrm{C}$ are three finite measures on $[0,1]$. Suppose thatΛ_A is at least as flat as Λ_B and that Λ_B is at least as flat as Λ_C. Then Λ_A is at least as flat as Λ_C. If, in addition at least one of these flatness relations is strict, then Λ_A is strictly flatter than Λ_C.*

Proof. By Lemma 4.13, the measures Λ_A, Λ_B and Λ_C have the same total measure, say $\Lambda_\mathrm{A}([0,1]) = \Lambda_\mathrm{B}([0,1]) = \Lambda_\mathrm{C}([0,1]) = m$. By multiplying these measures, and all the corresponding measures witnessing the flatness relation by $\frac{1}{m}$, it is enough to restrict ourselves to the case of probability measures from now on.

Let $\widetilde{\Phi}$ be a witness that Λ_A is at least as flat as Λ_B, and let $\widehat{\Phi}$ be a witness that Λ_B is at least as flat as Λ_C. Let us disintegrate $\widetilde{\Phi}$ on the second coordinate, and $\widehat{\Phi}$ on the first coordinate. This gives us families $\left\{\widetilde{\Phi}^2_y\right\}_{y\in[0,1]}$ and $\left\{\widehat{\Phi}^1_y\right\}_{y\in[0,1]}$ of probability Borel measures on $[0,1]$. This way, for each measurable set $S \subseteq [0,1]^2$ we have

$$\widetilde{\Phi}(S) = \int_y \widetilde{\Phi}^2_y(\{x\,:\,(x,y)\in S\})\ \mathrm{d}\Lambda_\mathrm{B}(y) \text{ and } \widehat{\Phi}(S) = \int_y \widehat{\Phi}^1_y(\{z\,:\,(y,z)\in S\})\ \mathrm{d}\Lambda_\mathrm{B}(y)\,.$$

Now, for every set $S \subseteq [0,1]^2$ of the form $S = \bigsqcup_{i=1}^n A_i \times B_i$ where A_i and B_i are measurable subsets of $[0,1]$, we define

$$\Xi(S) := \sum_{i=1}^{n} \int_{y\in[0,1]} \widetilde{\Phi}_y^2(A_i) \cdot \widehat{\Phi}_y^1(B_i)\ \mathrm{d}\Lambda_{\mathrm{B}}(y)\ . \tag{4.15}$$

It is tedious but straightforward to verify that the value of $\Xi(S)$ does not depend on the choice of the decomposition $S = \bigsqcup_{i=1}^n A_i \times B_i$. Then Carathéodory's extension theorem allows us to extend Ξ to a Borel measure on $[0,1]^2$, which we still denote by Ξ. We claim that this measure witnesses that Λ_{A} is at least as flat as Λ_{C}. Firstly, let us check that the marginals of Ξ are Λ_{A} and Λ_{C}, respectively. For $D \subseteq [0,1]$, we have that

$$\begin{aligned}\Xi(D\times[0,1]) &\overset{4.15}{=} \int_{y\in[0,1]} \widetilde{\Phi}_y^2(D)\cdot\widehat{\Phi}_y^1([0,1])\ \mathrm{d}\Lambda_{\mathrm{B}}(y)\\ \boxed{\text{by (D2), } \widehat{\Phi}_y^1([0,1])=1} \quad &= \int_{y\in[0,1]} \widetilde{\Phi}_y^2(D)\ \mathrm{d}\Lambda_{\mathrm{B}}(y) = \widetilde{\Phi}(D\times[0,1]) = \Lambda_{\mathrm{A}}(D)\ .\end{aligned}$$

Similarly, one can verify that $\Lambda_C(D) = \Xi([0,1]\times D)$.

Let $D \subseteq [0,1]$. We have the following

$$\begin{aligned}\int_{D\times[0,1]} x\,\mathrm{d}\Xi(x,z) &= \int_D x\,\mathrm{d}\Lambda_{\mathrm{A}}(x)\\ &= \int_{D\times[0,1]} x\,\mathrm{d}\tilde{\Phi}(x,y)\\ &= \int_{D\times[0,1]} y\,\mathrm{d}\tilde{\Phi}(x,y) = \int_{[0,1]}\left(\int_{[0,1]} y\cdot\mathbf{1}_{D\times[0,1]}(x,y)\,\mathrm{d}\widetilde{\Phi}_y^2(x)\right)\mathrm{d}\Lambda_{\mathrm{B}}(y)\\ &= \int_{[0,1]} y\widetilde{\Phi}_y^2(D)\,\mathrm{d}\Lambda_B(y)\\ &= \int_{[0,1]}\left(\int_{[0,1]} z\,\mathrm{d}\widehat{\Phi}_y^1(z)\right)\widetilde{\Phi}_y^2(D)\,\mathrm{d}\Lambda_{\mathrm{B}}(y)\\ &= \int_{D\times[0,1]} z\,\mathrm{d}\Xi(x,z)\end{aligned}$$

where the last equality follows from the following claim.

Claim 4.16. Let $g:[0,1]\to[0,1]$ be a measurable function. Then

$$\int_{[0,1]}\left(\int_{[0,1]} g(z)\ \mathrm{d}\widehat{\Phi}_y^1(z)\right)\widetilde{\Phi}_y^2(D)\,\mathrm{d}\Lambda_{\mathrm{B}}(y) = \int_{D\times[0,1]} g(z)\ \mathrm{d}\Xi(x,z).$$

Proof. Let $A \subseteq [0,1]$ be a measurable set and consider its characteristic function $\mathbf{1}_A$. We have

$$\int_{[0,1]} \left(\int_{[0,1]} \mathbf{1}_A(z) \, \mathrm{d}\widehat{\Phi}^1_y(z) \right) \widetilde{\Phi}^2_y(D) \, \mathrm{d}\Lambda_{\mathrm{B}}(y) = \int_{[0,1]} \widehat{\Phi}^1_y(A) \, \widetilde{\Phi}^2_y(D) \, \mathrm{d}\Lambda_{\mathrm{B}}(y)$$
$$= \Xi(D \times A) = \int_{D \times [0,1]} \mathbf{1}_A(z) \, \mathrm{d}\Xi(x,z).$$

This implies that the claim holds for every step function. Assume now that $g_n \to g$ uniformly and $\{g_n\}_{n \in \mathbb{N}}$ are step functions. We have

$$\begin{aligned}
\int_{[0,1]} \left(\int_{[0,1]} g(z) \, \mathrm{d}\widehat{\Phi}^1_y(z) \right) \widetilde{\Phi}^2_y(D) \, \mathrm{d}\Lambda_{\mathrm{B}}(y) &= \int_{[0,1]} \left(\int_{[0,1]} \lim_n g_n(z) \, \mathrm{d}\widehat{\Phi}^1_y(z) \right) \widetilde{\Phi}^2_y(D) \, \mathrm{d}\Lambda_{\mathrm{B}}(y) \\
&= \int_{[0,1]} \lim_n \left(\int_{[0,1]} g_n(z) \, \mathrm{d}\widehat{\Phi}^1_y(z) \right) \widetilde{\Phi}^2_y(D) \, \mathrm{d}\Lambda_{\mathrm{B}}(y) \\
&= \lim_n \int_{[0,1]} \left(\int_{[0,1]} g_n(z) \, \mathrm{d}\widehat{\Phi}^1_y(z) \right) \widetilde{\Phi}^2_y(D) \, \mathrm{d}\Lambda_{\mathrm{B}}(y) \\
&= \lim_n \int_{D \times [0,1]} g_n(z) \, \mathrm{d}\Xi(x,z) \\
&= \int_{D \times [0,1]} g(z) \, \mathrm{d}\Xi(x,z)
\end{aligned}$$

(the second, third and fifth equalities follow by the uniform convergence) and that finishes the proof. □

It remains to prove the additional part. So suppose that either Λ_{A} is strictly flatter than Λ_{B} or Λ_{B} is strictly flatter than Λ_{C}. Fix a strictly convex function $f\colon [0,1] \to [0,1]$. In the very same way as in the proof of Lemma 4.14 it follows that

$$\int_{[0,1]} f(x) \, \mathrm{d}\Lambda_{\mathrm{A}}(x) \leq \int_{[0,1]} f(y) \, \mathrm{d}\Lambda_{\mathrm{B}}(y) \leq \int_{[0,1]} f(z) \, \mathrm{d}\Lambda_{\mathrm{C}}(z) \,,$$

and at least one of the inequalities above is strict. Therefore $\Lambda_{\mathrm{A}} \neq \Lambda_{\mathrm{C}}$. □

The next two propositions answer the question of relating $\mathbf{\Phi}_U$ to $\mathbf{\Phi}_W$ and $\mathbf{\Upsilon}_U$ to $\mathbf{\Upsilon}_W$ using the above concept of flatter measures. While we consider these result interesting per se, let us note that in [9] we give several quick and fairly powerful applications of these results. For example, we show that the result of Doležal and Hladký can be extended to discontinuous functions, as advertised in Footnote [c].

Proposition 4.17. *Suppose that we have two graphons $U \preceq W$. Then the measure $\mathbf{\Phi}_U$ is at least as flat as the measure $\mathbf{\Phi}_W$. Similarly, the measure $\mathbf{\Upsilon}_U$ is at least as flat as the measure $\mathbf{\Upsilon}_W$. Lastly, if $U \prec W$ then $\mathbf{\Phi}_U$ is strictly flatter than $\mathbf{\Phi}_W$.*

Example 4.18. We cannot conclude that $\mathbf{\Upsilon}_U$ is strictly flatter than $\mathbf{\Upsilon}_W$ if $U \prec W$. To this end, it is enough to take U the constant-p graphon (for some $p \in (0,1)$) and W some p-regular but non-constant graphon. Then $\mathbf{\Upsilon}_U$ and $\mathbf{\Upsilon}_W$ are both equal to the Dirac measure on p.

Example 4.19. A probabilist might say that the information inherited from $\mathbf{\Phi}_W$ to $\mathbf{\Phi}_U$ is only «annealed», and not «quenched». Let us explain this on an example. Suppose that U is the constant-$\frac{1}{2}$ graphon and W attains each of the values $0, \frac{1}{2}, 1$ on sets of measure $\frac{1}{3}$ each. Obviously, we have that $U \prec W$ and thus there is a sequence $W^{\pi_1}, W^{\pi_2}, W^{\pi_3}, \ldots \xrightarrow{\text{w}^*} U \equiv \frac{1}{2}$. Now, observe that there are many different scenarios where the values $\frac{1}{2}$ can arise in the limit, of which we give two extreme ones. The first possibility is that around each $(x,y) \in \Omega^2$, we have alternations of values 0's, $\frac{1}{2}$'s, and 1's in the graphons W^{π_n}, each with frequency $\frac{1}{3}$. The second possibility is that for the measure $\frac{1}{3}$ of (x,y)'s, the graphons W^{π_n} attain values only $\frac{1}{2}$ around (x,y), and for the remaining (x,y)'s of measure $\frac{2}{3}$, we have alternations of values 0's, and 1's, each with density $\frac{1}{2}$.

In the proof of Proposition 4.17, we will need some basic facts about the weak* convergence of measures on $[0,1]^2$ (this convergence is also often called weak convergence or narrow convergence in the literature) which we recall here. We say that a bounded sequence of finite positive measures $\Psi_1, \Psi_2, \Psi_3, \ldots$ on $[0,1]^2$ converges in the weak* topology to a finite positive measure Ψ if for every continuous real function f defined on $[0,1]^2$ we have

$$\lim_{n\to\infty} \int_{[0,1]^2} f(x,y)\,\mathrm{d}\Psi_n(x,y) = \int_{[0,1]^2} f(x,y)\,\mathrm{d}\Psi(x,y).$$

This definition has many equivalent reformulations but we will need only the following one: A sequence $\Psi_1, \Psi_2, \Psi_3, \ldots$ converges to Ψ in the weak* topology if and only if $\lim_{n\to\infty} \Psi_n(A) = \Psi(A)$ for every Borel subset A of $[0,1]^2$ which satisfies that the Ψ-measure of its boundary in $[0,1]^2$ is 0.[g] Recall also that every sequence of probability measures on a compact separable space has a weak* convergent subsequence.

The following lemma is the key ingredient of the proof of Lemma 4.21. However, it may be of independent interest as it connects our research with Choquet theory. (We will not use Choquet theory, and the rest of this paragraph is meant only to hint the connection.)

[g]The boundary of a set $A \subseteq [0,1]^2$ is defined as the set of all points in the closure of A which are not interior points of A. Note that the interior points are considered only from point of view of the topological space $[0,1]^2$ (not $\mathbb{R}^2$). So for example, the boundary of the closed set $A = [0, \frac{1}{2}] \times [0,1]$ is $\{\frac{1}{2}\} \times [0,1]$ as all other points from A are interior points of A in $[0,1]^2$.

Recall that from the point of view of Choquet theory, if Λ_1, Λ_2 are finite positive measures on some compact convex subset C of a normed space then we say that Λ_1 *is smaller than* Λ_2 if $\int_C f \, \mathrm{d}\Lambda_1 \leq \int_C f \, \mathrm{d}\Lambda_2$ for every continuous convex function $f\colon C \to \mathbb{R}$. The next lemma states that the relation «being flatter» is naturally embedded into this Choquet ordering of measures (when the compact convex set C is the unit interval $[0,1]$).

Lemma 4.20. *Let Λ_1, Λ_2 be finite measures on $[0,1]$ such that Λ_1 is at least as flat as Λ_2. Then for every continuous convex function $f\colon [0,1] \to \mathbb{R}$ it holds $\int_{[0,1]} f \, \mathrm{d}\Lambda_1 \leq \int_{[0,1]} f \, \mathrm{d}\Lambda_2$. Moreover, if Λ_1 is strictly flatter than Λ_2 then there is a continuous convex function $g\colon [0,1] \to \mathbb{R}$ such that $\int_{[0,1]} g \, \mathrm{d}\Lambda_1 < \int_{[0,1]} g \, \mathrm{d}\Lambda_2$.*

Proof. Let Ψ be a witness for Λ_1 being at least as flat as Λ_2, as in Definition 4.10. Fix a continuous convex function $f\colon [0,1] \to \mathbb{R}$ and $\varepsilon > 0$. Find a natural number n such that $f(x) - f(y) < \varepsilon$ whenever $x, y \in [0,1]$ are such that $|x-y| \leq \frac{2}{n}$. Let $[0,1] = I_1 \sqcup I_2 \sqcup \ldots \sqcup I_n$ be a partition of $[0,1]$ into pairwise disjoint intervals of lengths $\frac{1}{n}$. For every i fix a point $x_i \in I_i$. As for every i we have

$$\int_{(x,y)\in I_i\times[0,1]} x \, \mathrm{d}\Psi(x,y) = \int_{(x,y)\in I_i\times[0,1]} y \, \mathrm{d}\Psi(x,y) \,,$$

$$\left| x_i \Psi(I_i \times [0,1]) - \int_{(x,y)\in I_i\times[0,1]} x \, \mathrm{d}\Psi(x,y) \right| \leq \frac{1}{n} \cdot \Psi(I_i \times [0,1])$$

and similarly

$$\left| \sum_{j=1}^{n} y_j \Psi(I_i \times I_j) - \int_{(x,y)\in I_i\times[0,1]} y \, \mathrm{d}\Psi(x,y) \right| \leq \frac{1}{n} \cdot \Psi(I_i \times [0,1]) \,,$$

it follows that

$$\left| x_i - \sum_{j=1}^{n} \frac{\Psi(I_i \times I_j)}{\Psi(I_i \times [0,1])} y_j \right| \leq \frac{2}{n} \,.$$

So by the convexity of f we have for every i that

$$f(x_i) \overset{\varepsilon}{\approx} f\left(\sum_{j=1}^{n} \frac{\Psi(I_i \times I_j)}{\Psi(I_i \times [0,1])} y_j \right) \leq \sum_{j=1}^{n} \frac{\Psi(I_i \times I_j)}{\Psi(I_i \times [0,1])} f(y_j) \,.$$

Therefore

$$\begin{aligned}\int_{x\in[0,1]} f(x)\,\mathrm{d}\Lambda_1(x) &= \sum_{i=1}^{n}\int_{x\in I_i} f(x)\,\mathrm{d}\Lambda_1(x) \overset{\varepsilon\cdot\Psi([0,1]^2)}{\approx} \sum_{i=1}^{n} f(x_i)\Psi(I_i\times[0,1])\\ &\overset{\varepsilon\cdot\Psi([0,1]^2)}{\approx} \sum_{i=1}^{n} f\left(\sum_{j=1}^{n}\frac{\Psi(I_i\times I_j)}{\Psi(I_i\times[0,1])}y_j\right)\Psi(I_i\times[0,1])\\ &\le \sum_{i=1}^{n}\sum_{j=1}^{n} f(y_j)\Psi(I_i\times I_j)\\ &= \sum_{j=1}^{n} f(y_j)\Psi([0,1]\times I_j)\\ &\overset{\varepsilon\cdot\Psi([0,1]^2)}{\approx} \int_{y\in[0,1]} f(y)\,\mathrm{d}\Lambda_2(y)\,.\end{aligned}$$

As this is true for every $\varepsilon > 0$ we conclude that $\int_{[0,1]} f\,\mathrm{d}\Lambda_1 \le \int_{[0,1]} f\,\mathrm{d}\Lambda_2$.

Now suppose that Λ_1 is strictly flatter than Λ_2. Then there is a continuous function $h\colon [0,1]\to\mathbb{R}$ such that $\int_{[0,1]} h\,\mathrm{d}\Lambda_1 \ne \int_{[0,1]} h\,\mathrm{d}\Lambda_2$. Recall that functions of the form $g_1 - g_2$, where both $g_1, g_2\colon [0,1]\to\mathbb{R}$ are continuous and convex, are uniformly dense in the space of all continuous functions $h\colon [0,1]\to\mathbb{R}$. Indeed, this follows from the Stone–Weierstrass theorem as every polynomial function $p\colon [0,1]\to\mathbb{R}$ can be written as $p = g_1 - g_2$ for g_1, g_2 continuous convex (just define $g_1(x) := p(x) + Kx^2$ and $g_2(x) := Kx^2$ where K is a sufficiently large constant). So there is a continuous convex function $g\colon [0,1]\to\mathbb{R}$ such that $\int_{[0,1]} g\,\mathrm{d}\Lambda_1 \ne \int_{[0,1]} g\,\mathrm{d}\Lambda_2$. But then the previous part of the proof gives us that $\int_{[0,1]} g\,\mathrm{d}\Lambda_1 < \int_{[0,1]} g\,\mathrm{d}\Lambda_2$. □

Let us now give an intuitively clear lemma.

Lemma 4.21. *Let $\Theta^A, \Delta^A, \Theta^B, \Delta^B$ be four finite measures on $[0,1]$. Suppose that Θ^A is strictly flatter than Δ^A and that Θ^B is at least as flat as Δ^B. Then the measure $\Theta^A + \Delta^A$ is strictly flatter than $\Theta^B + \Delta^B$.*

Proof. Let Ψ^A (resp. Ψ^B) be a witness for Θ^A (resp. Θ^B) being as flat as Δ^A (resp. Δ^B), as in Definition 4.10. Then $\Psi^A + \Psi^B$ shows that $\Theta^A + \Delta^A$ is at least as flat as $\Theta^B + \Delta^B$. It remains to show that the relation is actually strict.

Let $g\colon [0,1]\to\mathbb{R}$ be a continuous convex function such that $\int_{[0,1]} g\,\mathrm{d}\Theta^A < \int_{[0,1]} g\,\mathrm{d}\Delta^A$. Such a function exists by Lemma 4.20. Then another application of Lemma 4.20 gives that $\int_{[0,1]} g\,\mathrm{d}(\Theta^A + \Delta^A) < \int_{[0,1]} g\,\mathrm{d}(\Theta^B + \Delta^B)$, and so the measures $\Theta^A + \Delta^A$ and $\Theta^B + \Delta^B$ are not the same. □

The proof of Proposition 4.17 relies on the following lemma.

Lemma 4.22. *Let B be a separable atomless finite measure space with the measure β, and let $(f_n : B \to [0,1])_n$ and $f : B \to [0,1]$ be measurable functions such that $\lim_{n\to\infty} \int_A f_n(x)\, d\beta(x) = \int_A f(x)\, d\beta(x)$ for every measurable $A \subseteq B$. Suppose that Δ_n and Δ are the pushforward measures of f_n and f, $\Delta_n(L) := \beta\left(f_n^{-1}(L)\right)$, $\Delta(L) := \beta\left(f^{-1}(L)\right)$. Suppose that the measures Δ_n weak* converge to a probability measure Δ^*. Then Δ is at least as flat as Δ^*.*

Proof. For every natural number n and every measurable subset A of $[0,1]^2$ we define

$$\Psi_n(A) = \beta\left(\{x \in B\colon\ (f(x), f_n(x)) \in A\}\right).$$

Clearly every Ψ_n is a measure on $[0,1]^2$. Let Ψ be some weak* accumulation point of the sequence $\Psi_1, \Psi_2, \Psi_3, \ldots$. Without loss of generality, we may assume that the sequence $\Psi_1, \Psi_2, \Psi_3, \ldots$ converges to Ψ. Let Z be the set consisting of all points $z \in (0,1)$ for which either $\Psi(\{z\} \times [0,1]) > 0$ or $\Psi([0,1] \times \{z\}) > 0$ or $\Delta^*(\{z\}) > 0$. Then Z is at most countable. So if $\mathcal{I}$ is the system of all intervals $I \subseteq [0,1]$ whose endpoints do not belong to Z then $\mathcal{I}$ is closed under taking finite intersections and it generates the sigma-algebra of all Borel subsets of $[0,1]$. Moreover, whenever $I, J \in \mathcal{I}$ then the boundary of $I \times J$ is of Ψ-measure 0 and of Δ^*-measure 0. Denote by Ψ^x and Ψ^y the marginals of Ψ on the first and on the second coordinate, respectively. Then for every $I \in \mathcal{I}$ we have

$$\begin{aligned}
\Psi^x(I) &= \Psi(I \times [0,1]) = \lim_{n\to\infty} \Psi_n(I \times [0,1]) \\
&= \lim_{n\to\infty} \beta\left(\{x \in B\colon\ (f(x), f_n(x)) \in I \times [0,1]\}\right) \\
&= \lim_{n\to\infty} \Delta(I) = \Delta(I).
\end{aligned}$$

As this is true for every $I \in \mathcal{I}$, it clearly follows that $\Psi^x = \Delta$. On the other hand, for every $I \in \mathcal{I}$ we have that

$$\begin{aligned}
\Psi^y(I) &= \Psi([0,1] \times I) = \lim_{n\to\infty} \Psi_n([0,1] \times I) \\
&= \lim_{n\to\infty} \beta\left(\{x \in B\colon\ (f(x), f_n(x)) \in [0,1] \times I\}\right) \\
&= \lim_{n\to\infty} \Delta_n(I) \\
\boxed{\Delta_n \xrightarrow{w^*} \Delta^*} &= \Delta^*(I).
\end{aligned}$$

So, again we have $\Psi^y = \Delta^*$. To finish the proof, it remains to show that Ψ satisfies (4.13). That is, we need to show that

$$\int_{(x,y)\in C\times[0,1]} x\, \mathrm{d}\Psi(x,y) = \int_{(x,y)\in C\times[0,1]} y\, \mathrm{d}\Psi(x,y) \tag{4.16}$$

for every Borel measurable subset C of $[0,1]$. Again, it is enough to show (4.16) only for every $C \in \mathcal{I}$. So fix $C \in \mathcal{I}$, $\varepsilon > 0$ and find some partition $\{I_1, I_2, \ldots, I_m\}$ of the interval

$[0,1]$ into intervals from $\mathcal{I}$ of lengths smaller than ε. Suppose first that the interval C is of length smaller than ε. Fix some points $x_0 \in C$ and $y_j \in I_j$ (for every j). Then we have

$$\begin{aligned}
\int_{(x,y)\in C\times[0,1]} y \,\mathrm{d}\Psi(x,y) \quad &= \sum_{j=1}^{m} \int_{(x,y)\in C\times I_j} y \,\mathrm{d}\Psi(x,y) \\
&\overset{\varepsilon\cdot\Psi(C\times[0,1])}{\approx} \sum_{j=1}^{m} y_j \Psi(C\times I_j) = \sum_{j=1}^{m} y_j \lim_{n\to\infty} \Psi_n(C\times I_j) \\
&= \sum_{j=1}^{m} y_j \lim_{n\to\infty} \beta\left(\{x\in B\colon\ (f(x), f_n(x)) \in C\times I_j\}\right) \\
&= \sum_{j=1}^{m} \lim_{n\to\infty} \int_{x\in f^{-1}(C)\cap f_n^{-1}(I_j)} y_j \,\mathrm{d}\beta \\
&\overset{\varepsilon\cdot\beta(f^{-1}(C))}{\approx} \lim_{n\to\infty} \int_{x\in f^{-1}(C)} f_n(x) \,\mathrm{d}\beta = \int_{x\in f^{-1}(C)} f(x) \,\mathrm{d}\beta \\
&\overset{\varepsilon\cdot\beta(f^{-1}(C))}{\approx} \int_{x\in f^{-1}(C)} x_0 \,\mathrm{d}\beta \\
&= x_0\cdot\beta\left(f^{-1}(C)\right) = x_0\cdot\Delta(C) = x_0\cdot\Psi^x(C) \\
&= \int_{(x,y)\in C\times[0,1]} x_0 \,\mathrm{d}\Psi(x,y) \overset{\varepsilon\cdot\Psi(C\times[0,1])}{\approx} \int_{(x,y)\in C\times[0,1]} x \,\mathrm{d}\Psi(x,y),
\end{aligned}$$

so

$$\int_{(x,y)\in C\times[0,1]} y \,\mathrm{d}\Psi(x,y) \overset{2\varepsilon\cdot(\Psi(C\times[0,1])+\beta(f^{-1}(C)))}{\approx} \int_{(x,y)\in C\times[0,1]} x \,\mathrm{d}\Psi(x,y).$$

In general, the interval C can be decomposed into subintervals belonging to $\mathcal{I}$ such that each of them is of length smaller than ε, and by additivity of integration we conclude that

$$\int_{(x,y)\in C\times[0,1]} y \,\mathrm{d}\Psi(x,y) \overset{2\varepsilon\cdot(\Psi([0,1]^2)+\beta(B))}{\approx} \int_{(x,y)\in C\times[0,1]} x \,\mathrm{d}\Psi(x,y).$$

As this is true for every $\varepsilon > 0$, we have verified (4.16). □

We are now ready to prove Proposition 4.17.

Proof of Proposition 4.17, non-strict part. Suppose that we have two graphons $U, W : \Omega^2 \to [0,1]$, where $U \preceq W$. Then there exist measure preserving bijections $\pi_1, \pi_2, \pi_3, \ldots$ on Ω so that

$$W^{\pi_n} \xrightarrow{\mathrm{w}^*} U\,. \tag{4.17}$$

Now, we can apply Lemma 4.22 with $B := \Omega^2$, $f_n := W^{\pi_n}$, $\Delta_n := \mathbf{\Phi}_{W^{\pi_n}} = \mathbf{\Phi}_W$, $f := U$, and $\Delta := \mathbf{\Phi}_U$. The lemma gives that $\mathbf{\Phi}_U$ is at least as flat as $\mathbf{\Phi}_W$.

Observe that (4.17) implies that for the degree functions $\deg_U : \Omega \to [0,1]$ and $\deg_{W^{\pi_n}} : \Omega \to [0,1]$ we have $\int_A \deg_{W^{\pi_n}} \xrightarrow{\text{w}^*} \int_A \deg_U$ for every measurable $A \subseteq \Omega$. Now, we can apply Lemma 4.22 with $B := \Omega$, $f_n := \deg_{W^{\pi_n}}$, $\Delta_n := \boldsymbol{\Upsilon}_{W^{\pi_n}} = \boldsymbol{\Upsilon}_W$, $f := \deg_U$, and $\Delta := \boldsymbol{\Upsilon}_U$. The lemma gives that $\boldsymbol{\Upsilon}_U$ is at least as flat as $\boldsymbol{\Upsilon}_W$. □

Proof of Proposition 4.17, strictly flatter part. Next, suppose that $U \prec W$. Then for the measure preserving bijections $\pi_1, \pi_2, \pi_3, \ldots$ as above, we have

$$W^{\pi_n} \overset{\|\cdot\|_1}{\not\longrightarrow} U\,. \tag{4.18}$$

Indeed, suppose that (4.18) is not true, that is, $W^{\pi_n} \xrightarrow{\|\cdot\|_1} U$. Then for measure preserving bijections $\psi_n := (\pi_n)^{-1}$ we have $U^{\psi_n} \xrightarrow{\|\cdot\|_1} W$, and in particular $W \preceq U$. This is a contradiction to the fact that $U \prec W$. Now, Lemma 2.4 implies that there exists an interval $J \subseteq [0,1]$ and a «strictly bigger» interval J^+ such that

$$\nu^{\otimes 2}\left(U^{-1}(J) \setminus (W^{\pi_n})^{-1}(J^+)\right) \not\longrightarrow 0\,.$$

By passing to a subsequence, let us assume that $\nu^{\otimes 2}\left(U^{-1}(J) \setminus (W^{\pi_n})^{-1}(J^+)\right) > \varepsilon$ for each n and for some $\varepsilon > 0$. We shall now apply Lemma 4.22 twice. To this end, we take $X := U^{-1}(J)$. Furthermore, we write the measures $\boldsymbol{\Phi}_U$ and $\boldsymbol{\Phi}_{W^{\pi_n}}$ as $\boldsymbol{\Phi}_U = \Phi_U^X + \Phi_U^{\Omega^2\setminus X}$ and $\boldsymbol{\Phi}_{W^{\pi_n}} = \Phi_{W^{\pi_n}}^X + \Phi_{W^{\pi_n}}^{\Omega^2\setminus X}$, where $\Phi_U^X(L) := \nu^{\otimes 2}\left(X \cap U^{-1}(L)\right)$, $\Phi_U^{\Omega^2\setminus X}(L) := \nu^{\otimes 2}\left(U^{-1}(L) \setminus X\right)$, and the measures $\Phi_{W^{\pi_n}}^X$ and $\Phi_{W^{\pi_n}}^{\Omega^2\setminus X}$ are defined analogously. Let $\left(\Delta^X, \Delta^{\Omega^2\setminus X}\right)$ be an arbitrary accumulation point of the sequence $\left(\Phi_{W^{\pi_n}}^X, \Phi_{W^{\pi_n}}^{\Omega^2\setminus X}\right)_n$ of pairs of measures with respect to the product of weak* topologies on measures on $[0,1]^2$. Crucially, note that $\boldsymbol{\Phi}_W = \Delta^X + \Delta^{\Omega^2\setminus X}$.

- We first apply Lemma 4.22 with $B := X$, $f_n := (W^{\pi_n})_{\restriction X}$, $\Delta_n := \Phi_{W^{\pi_n}}^X$, $f := U_{\restriction X}$, and $\Delta := \Phi_U^X$. The lemma gives that Φ_U^X is at least as flat as Δ^X. Recall that the support of the individual measures $\Phi_{W^{\pi_n}}^X$ uniformly exceeds the interval J^+. From this, we conclude that the support of their weak* accumulation point Δ^X does not lie in J. On the other hand, observe that the support of Φ_U^X lies inside J. Thus $\Phi_U^X \neq \Delta^X$. We conclude that Φ_U^X is strictly flatter than Δ^X.
- Next, we apply Lemma 4.22 with $B := \Omega^2 \setminus X$, $f_n := (W^{\pi_n})_{\restriction \Omega^2\setminus X}$, $\Delta_n := \Phi_{W^{\pi_n}}^{\Omega^2\setminus X}$, and $f := U_{\restriction \Omega^2\setminus X}$, $\Delta := \Phi_U^{\Omega^2\setminus X}$. The lemma gives that $\Phi_U^{\Omega^2\setminus X}$ is at least as flat as $\Delta^{\Omega^2\setminus X}$.

The proof now follows by Lemma 4.21 to $\boldsymbol{\Phi}_W = \Delta^X + \Delta^{\Omega^2\setminus X}$. □

Let us finish this section with an auxiliary result which will be applied in Section 8. The result states that probability measures supported on $\{0,1\}$ are maximal with respect to the flat order.

Lemma 4.23. *Suppose that Λ_1 and Λ_2 are two probability measures on $[0,1]$ such that Λ_1 is strictly flatter than Λ_2. Then Λ_1 is not supported on $\{0,1\}$.*

Proof. The proof is obvious. □

4.5. **Relationship between envelopes, the cut distance, and range frequencies.** Our last result states that two envelopes are equal if and only if the corresponding graphons are weakly isomorphic. This result relies on Theorem 3.5 which will be proven in Section 5, and is used later in the proof of Corollary 6.2, which puts into relation the cut distance and the Vietoris hyperspace $K(\mathcal{W}_0)$.

Corollary 4.24. *Let $U, W \in \mathcal{W}_0$. The following are equivalent:*

- $\langle U \rangle = \langle W \rangle$,
- $\delta_\square(U, W) = 0$.

Moreover, if $U \preceq W$ then the conditions above are equivalent to

- $\mathbf{\Phi}_U = \mathbf{\Phi}_W$.

Proof. Assume that $\langle U \rangle = \langle W \rangle$. Then the conditions in Theorem 3.5(b) are satisfied because we have $\langle U \rangle = \mathbf{LIM}_{\mathrm{w}*}(U, U, U \ldots) = \mathbf{ACC}_{\mathrm{w}*}(U, U, U, \ldots)$. By the «furthermore» part of Theorem 3.5, the constant sequence $U, U, U, \ldots$ converges in $\delta_\square$ to a maximal element of $\langle U \rangle$. By the assumption $\langle U \rangle = \langle W \rangle$, the graphon W is a maximal element of $\langle U \rangle$. The only possibility under which a constant sequence can converge in the cut distance to another graphon is when their cut distance is 0, that is $\delta_\square(U, W) = 0$.

The opposite implication is Lemma 4.2(d).

We now turn to the «moreover» part. Suppose that $\langle U \rangle = \langle W \rangle$. By Proposition 4.17 we have that $\mathbf{\Phi}_U$ is at least as flat as $\mathbf{\Phi}_W$ and vice-versa. By Lemma 4.20 this means that $\int f \, \mathrm{d}\mathbf{\Phi}_W = \int f \, \mathrm{d}\mathbf{\Phi}_U$ for every continuous convex function $f\colon [0,1] \to \mathbb{R}$. Applying the Stone–Weierstrass theorem in the same way as in the proof of Lemma 4.20, this is true for every continuous (not necessarily convex) function $f\colon [0,1] \to \mathbb{R}$. Therefore, $\mathbf{\Phi}_U = \mathbf{\Phi}_W$.

Assume finally that $U \preceq W$ and $\mathbf{\Phi}_U = \mathbf{\Phi}_W$. Then $\mathbf{\Phi}_U$ is not strictly flatter than $\mathbf{\Phi}_W$ and therefore by Proposition 4.17, it is not the case that $U \prec W$. This means that $\langle U \rangle = \langle W \rangle$. □

5. Proof of Theorem 3.5

The main idea of the proof of the implication (b)⇒(a) is the following. Let W be a $\preceq$-maximal element in $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. Such a W is guaranteed to exist by Lemma 4.9. Then the sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ converges to W in the cut distance. To make this argument precise we first recall several definitions and results from [8].

Lemma 5.1 (Claim 1 and Claim 2 in [8]). *Let $W, \Gamma_1, \Gamma_2, \ldots \in \mathcal{W}_0$ be graphons defined on $\Omega = [0,1]$ and assume that $\Gamma_n \xrightarrow{w^*} W$. Take some sequence $B_1, B_2, \ldots \subseteq [0,1]$ of measurable sets and a subsequence $(n_k)_k$ such that $1_{B_{n_k}} \xrightarrow{w^*} s$ and ${}_{B_{n_k}}\Gamma_{n_k} \xrightarrow{w^*} \tilde{W}$ (with the notation from (4.3)) for some function $s : [0,1] \to [0,1]$ and some graphon $\tilde{W}$. We define*

$$\psi(x) = \int_0^x s(y)\,\mathrm{d}y \qquad \text{and} \qquad \varphi(x) = \psi(1) + \int_0^x (1 - s(y))\,\mathrm{d}y. \tag{5.1}$$

Then for almost every $(x,y) \in [0,1]^2$ we have

$$\begin{aligned} W(x,y) =& \tilde{W}(\psi(x),\psi(y))s(x)s(y) + \tilde{W}(\psi(x),\varphi(y))s(x)(1-s(y)) \\ &+ \tilde{W}(\varphi(x),\psi(y))(1-s(x))s(y) + \tilde{W}(\varphi(x),\varphi(y))(1-s(x))(1-s(y)). \end{aligned} \tag{5.2}$$

Moreover, if W is not a cut-norm accumulation point of the sequence $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ and the sets $B_1, B_2, \ldots \subseteq [0,1]$ are chosen to witness this fact, i.e., such that $\int_{B_n \times B_n} (\Gamma_n - W) > \varepsilon$ or $\int_{B_n \times B_n} (\Gamma_n - W) < -\varepsilon$ for some $\varepsilon > 0$ (which does not depend on n), the convex combination (5.2) is proper[h] *on a set of positive $\nu^{\otimes 2}$ measure.*

Proof of Theorem 3.5, (b)⇒(a). Suppose that $\mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) = \mathbf{ACC}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ and assume that $W \in \mathbf{LIM}_{w*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ is a maximal element as given in Lemma 4.9. We may also assume that $\Gamma_n \xrightarrow{w^*} W$. We claim that this already implies that $\Gamma_n \xrightarrow{\|\cdot\|_\square} W$. Suppose not. Then by passing to a subsequence we may assume that there is an $\varepsilon > 0$ and a sequence $B_1, B_2, \ldots$ of Borel subsets of $[0,1]$ such that $\int_{B_n \times B_n} (\Gamma_n - W) > \varepsilon$ (or $\int_{B_n \times B_n} (\Gamma_n - W) < -\varepsilon$ which can be handled similarly). We will use versions ${}_{B_n}\Gamma_n$ of the graphons Γ_n that are defined via (4.3). We take a function $s : [0,1] \to [0,1]$ and a graphon $\tilde{W}$ such that $1_{B_n} \xrightarrow{w^*} s$ and ${}_{B_n}\Gamma_n \xrightarrow{w^*} \tilde{W}$. If 1_{B_n} or ${}_{B_n}\Gamma_n$ are not weak* convergent (that is, s or $\tilde{W}$ do not exist), then we pass to a suitable convergent subsequence (which we still index by $1, 2, \ldots$).

Claim. We have $\langle W \rangle = \langle \tilde{W} \rangle$.

Proof. For each $n \in \mathbb{N}$, let $\mathcal{I}_n = \{I_{n,0}, I_{n,1}, \ldots, I_{n,2^n-1}\}$ be a partition of $[0,1]$ into 2^n many equimeasurable intervals, i.e., $I_{n,k} = [\frac{k}{2^n}, \frac{k+1}{2^n})$. Define a measure preserving almost-bijection φ_n by

$$\varphi_n(x) = \begin{cases} \frac{k}{2^n} + x - \psi\left(\frac{k}{2^n}\right) & \text{if } x \in \left[\psi\left(\frac{k}{2^n}\right), \psi\left(\frac{k+1}{2^n}\right)\right) \text{ and} \\ \frac{k}{2^n} + \psi\left(\frac{k+1}{2^n}\right) - \psi\left(\frac{k}{2^n}\right) + x - \varphi\left(\frac{k}{2^n}\right) & \text{if } x \in \left[\varphi\left(\frac{k}{2^n}\right), \varphi\left(\frac{k+1}{2^n}\right)\right), \end{cases}$$

[h] that is, at least two summands on the right-hand side are positive

where ψ and φ are defined by (5.1). Define $\tilde{W}_n(x,y) = \tilde{W}\left(\varphi_n^{-1}(x), \varphi_n^{-1}(y)\right)$. We claim that $\tilde{W}_n \xrightarrow{\text{w}^*} W$. First note that if $m \in \mathbb{N}$ and $m \geq n$ and $k, l < 2^n$, then

$$\int_{I_{n,k}\times I_{n,l}} \tilde{W}_m = \sum_{i=2^{m-n}\cdot k}^{2^{m-n}\cdot(k+1)-1} \sum_{j=2^{m-n}\cdot l}^{2^{m-n}\cdot(l+1)-1} \int_{I_{m,i}\times I_{m,j}} \tilde{W}_m. \tag{5.3}$$

For each $n \in \mathbb{N}$ we have

$$\begin{aligned}
\int_{I_{n,k}\times I_{n,l}} \tilde{W}_n &= \int_{\psi(\frac{k}{2^n})}^{\psi(\frac{k+1}{2^n})} \int_{\psi(\frac{l}{2^n})}^{\psi(\frac{l+1}{2^n})} \tilde{W} + \int_{\psi(\frac{k}{2^n})}^{\psi(\frac{k+1}{2^n})} \int_{\varphi(\frac{l}{2^n})}^{\varphi(\frac{l+1}{2^n})} \tilde{W} \\
&+ \int_{\varphi(\frac{k}{2^n})}^{\varphi(\frac{k+1}{2^n})} \int_{\psi(\frac{l}{2^n})}^{\psi(\frac{l+1}{2^n})} \tilde{W} + \int_{\varphi(\frac{k}{2^n})}^{\varphi(\frac{k+1}{2^n})} \int_{\varphi(\frac{l}{2^n})}^{\varphi(\frac{l+1}{2^n})} \tilde{W} \\
&= \int_{I_{n,k}\times I_{n,l}} \tilde{W}(\psi(x),\psi(y))s(x)s(y) + \int_{I_{n,k}\times I_{n,l}} \tilde{W}(\psi(x),\varphi(y))s(x)(1-s(y)) \\
&+ \int_{I_{n,k}\times I_{n,l}} \tilde{W}(\varphi(x),\psi(y))(1-s(x))s(y) \\
&+ \int_{I_{n,k}\times I_{n,l}} \tilde{W}(\varphi(x),\varphi(y))(1-s(x))(1-s(y)) \\
&= \int_{I_{n,k}\times I_{n,l}} W.
\end{aligned}$$

Plugging this back to (5.3) gives

$$\int_{I_{n,k}\times I_{n,l}} \tilde{W}_m = \int_{I_{n,k}\times I_{n,l}} W\,.$$

This implies that for every $n \in \mathbb{N}$ and $k, l < 2^n$, we have $\int_{I_{n,k}\times I_{n,l}} \tilde{W}_m \to \int_{I_{n,k}\times I_{n,l}} W$ as $m \to \infty$. Consequently $\tilde{W}_m \xrightarrow{\text{w}^*} W$ and $W \in \langle \tilde{W} \rangle$. On the other hand, $\tilde{W} \in \langle W \rangle$ since $\tilde{W} \in \mathbf{LIM}_{\text{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$, and the maximality of W in $\mathbf{LIM}_{\text{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. □

Now Proposition 4.17 together with Lemma 4.20 give us that $\int f \, \mathrm{d}\mathbf{\Phi}_{\tilde{W}} = \int f \, \mathrm{d}\mathbf{\Phi}_W$ for every continuous convex function. So to get the desired contradiction it suffices to prove the following claim.

Claim. Let $f : [0,1] \to [0,1]$ be a strictly convex function. Then $\int f \, \mathrm{d}\mathbf{\Phi}_{\tilde{W}} > \int f \, \mathrm{d}\mathbf{\Phi}_W$.

Proof. Recall that the convex combination in (5.2) is proper on a set of positive $\nu^{\otimes 2}$ measure. Therefore the strict convexity of f gives us

$$\begin{aligned}
\int_0^1 f(x)\,\mathrm{d}\mathbf{\Phi}_W &= \int_0^1\int_0^1 f(W(x,y)) \\
&< \int_0^1\int_0^1 f\left(\tilde{W}(\psi(x),\psi(y))s(x)s(y)\right) + \int_0^1\int_0^1 f\left(\tilde{W}(\psi(x),\varphi(y))s(x)(1-s(y))\right) \\
&\quad + \int_0^1\int_0^1 f\left(\tilde{W}(\varphi(x),\psi(y))(1-s(x))s(y)\right) \\
&\quad + \int_0^1\int_0^1 f\left(\tilde{W}(\varphi(x),\varphi(y))(1-s(x))(1-s(y))\right) \\
&= \int_0^{\psi(1)}\int_0^{\psi(1)} f\left(\tilde{W}(x,y)\right) + \int_0^{\psi(1)}\int_{\psi(1)}^1 f\left(\tilde{W}(x,y)\right) \\
&\quad + \int_{\psi(1)}^1\int_0^{\psi(1)} f\left(\tilde{W}(x,y)\right) + \int_{\psi(1)}^1\int_{\psi(1)}^1 f\left(\tilde{W}(x,y)\right) \\
&= \int_0^1\int_0^1 f(\tilde{W}(x,y)) = \int_0^1 f(x)\,\mathrm{d}\mathbf{\Phi}_{\tilde{W}}\,.
\end{aligned}$$

□

Let $\Gamma_1,\Gamma_2,\Gamma_3,\ldots$ be a sequence of graphons which is Cauchy with respect to the cut distance. By Theorem 3.3 there is a subsequence $\Gamma_{n_1},\Gamma_{n_2},\Gamma_{n_3},\ldots$ such that

$$\mathbf{LIM}_{\mathrm{w}*}(\Gamma_{n_1},\Gamma_{n_2},\Gamma_{n_3},\ldots) = \mathbf{ACC}_{\mathrm{w}*}(\Gamma_{n_1},\Gamma_{n_2},\Gamma_{n_3},\ldots)\,.$$

By the proof of implication (b)⇒(a) this subsequence converges to some W in the cut distance. As the original sequence $\Gamma_1,\Gamma_2,\Gamma_3,\ldots$ is Cauchy with respect to the cut distance it follows that it converges to W as well. We may suppose that $\Gamma_n \xrightarrow{\|\cdot\|_\square} W$. To conclude that $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots) = \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ consider any $\Gamma_{n_k}^{\varphi_k} \xrightarrow{\mathrm{w}^*} U$. We claim that then also $W^{\varphi_k} \xrightarrow{\mathrm{w}^*} U$. To see this fix some Borel set $A\subseteq[0,1]$ then

$$\left|\int_{A\times A}(W^{\varphi_k}-U)\right| \le \left|\int_{\varphi_k(A)\times\varphi_k(A)}(W-\Gamma_{n_k})\right| + \left|\int_{A\times A}(\Gamma_{n_k}^{\varphi_k}-U)\right| \to 0$$

where the first term tends to 0 by $\|\cdot\|_\square$-convergence and the second by the weak* convergence. Then we use the same trick again:

$$\left|\int_{A\times A}(\Gamma_n^{\varphi_n}-U)\right| \le \left|\int_{\varphi_n(A)\times\varphi_n(A)}(\Gamma_n-W)\right| + \left|\int_{A\times A}(W^{\varphi_n}-U)\right| \to 0\,.$$

This gives us $\Gamma_n^{\varphi_n} \xrightarrow{\mathrm{w}^*} U$ which means that $U\in\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$. □

6. Relating the hyperspace $K(\mathcal{W}_0)$ and the cut distance

Lemma *3.1(b)* says that envelopes are members of the hyperspace $K(\mathcal{W}_0)$. First, we provide a proof of an extension of Theorem 3.5, stated in Theorem 6.1 below. In addition to the original statement, we include a characterization of cut distance convergence in terms of the hyperspace $K(\mathcal{W}_0)$, and also describe the limit graphon.

Theorem 6.1. *Let $W, \Gamma_1, \Gamma_2, \Gamma_3, \ldots \in \mathcal{W}_0$. The following are equivalent:*

(a) $\Gamma_n \xrightarrow{\delta_\square} W$,
(b) $\langle \Gamma_n \rangle \to \langle W \rangle$ *in* $K(\mathcal{W}_0)$,
(c) $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) = \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ *and* $W \in \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ *is the* $\preceq$*-maximal element of* $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$.

Proof. Condition (c) is equivalent to Condition (a) by Theorem 3.5. Before we prove that (b) is equivalent to (c) we recall that a basic open neighborhood of $K \in K(\mathcal{W}_0)$ is given by some finite open cover $\mathcal{U}$ of K, namely for every such cover $\mathcal{U}$ we define the open neighborhood of K as $\mathcal{O}_{\mathcal{U}} = \left\{L : L \subseteq \bigcup_{O \in \mathcal{U}} O \text{ and } \forall O \in \mathcal{U}\ L \cap O \neq \emptyset\right\}$.

(c) $\implies$ (b). First note that $\langle W \rangle = \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. Suppose that $\langle \Gamma_n \rangle \not\to \langle W \rangle$ in $K(\mathcal{W}_0)$. There must be a basic open neighborhood $\mathcal{O}_{\mathcal{U}}$ of $\langle W \rangle$ and a subsequence $\langle \Gamma_{k_1} \rangle, \langle \Gamma_{k_2} \rangle, \ldots$ such that either $\langle \Gamma_{k_j} \rangle \not\subseteq \bigcup_{O \in \mathcal{U}} O$ for every $j \in \mathbb{N}$ or there is $O \in \mathcal{U}$ such that $\langle \Gamma_{k_j} \rangle \cap O = \emptyset$ for every $j \in \mathbb{N}$. If the first possibility happens, then clearly $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_{k_1}, \Gamma_{k_2}, \ldots) \cap (\mathcal{W}_0 \setminus \bigcup_{O \in \mathcal{U}} O) \neq \emptyset$, which is a contradiction with $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_{k_1}, \Gamma_{k_2}, \ldots) = \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) \subseteq \bigcup_{O \in \mathcal{U}} O$. If the second possibility occurs, for some $O \in \mathcal{U}$ then clearly $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_{k_1}, \Gamma_{k_2}, \ldots) \cap O = \emptyset$. That is also a contradiction once $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_{k_1}, \Gamma_{k_2}, \ldots) = \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) = \langle W \rangle$ and $\langle W \rangle \cap O \neq \emptyset$ because $\langle W \rangle \in \mathcal{O}_{\mathcal{U}}$.

(b) $\implies$ (c). Let $U \in \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. We claim that $U \in \langle W \rangle$. Assume it is not true and take any open set $O \subseteq \mathcal{W}_0$ in the weak* topology such that $\langle W \rangle \subseteq O$ and U is not in the closure of O. Then by the convergence in $K(\mathcal{W}_0)$ we may find some $m \in \mathbb{N}$ such that for every $k \geq m$ we have $\langle \Gamma_k \rangle \subseteq O$. This implies that $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ is a subset of the closure of O and that is a contradiction because we assumed that U is not in that closure. Therefore, we have just proved that $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) \subseteq \langle W \rangle$.

Let $U \in \langle W \rangle$, then for every open set $O \subseteq \mathcal{W}_0$ in the weak* topology such that $U \in O$ we can find $m \in \mathbb{N}$ such that for every $k \geq m$ we have $\langle \Gamma_k \rangle \cap O \neq \emptyset$. Since $\mathcal{W}_0$ with the weak* topology is a metric space we may find a sequence of versions such that $\Gamma_n^{\varphi_n} \xrightarrow{\mathrm{w}^*} U$. Therefore, we have $\langle W \rangle \subseteq \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$, and so $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) = \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. By the first part of the argument it follows that

$$\langle W \rangle = \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$$

and thus W is the $\preceq$-maximal element of $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. □

We can now formulate Corollary 6.2, which is the final statement of this section. It allows us to transfer the space $\left(\widetilde{\mathcal{W}_0}, \delta_\square\right)$ into the hyperspace $K(\mathcal{W}_0)$. This transference will be useful later.

Prior to giving the statement, observe that for the envelope map $\langle\cdot\rangle$ defined on $\mathcal{W}_0$ we have $\langle W_1\rangle = \langle W_2\rangle$ for weakly isomorphic graphons W_1 and W_2. That allows us to define $\langle\cdot\rangle$ even on the factor-space $\widetilde{\mathcal{W}_0}$.

Corollary 6.2. *The envelope map $\langle\cdot\rangle : \left(\widetilde{\mathcal{W}_0}, \delta_\square\right) \to K(\mathcal{W}_0)$ is a continuous injection. We have that $\left(\widetilde{\mathcal{W}_0}, \delta_\square\right)$ is homeomorphic to some closed subspace X of $K(\mathcal{W}_0)$. Moreover, the metric $\delta_\square$ is equivalent*[i] *to the pullback χ of the hyperspace metric on $K(\mathcal{W}_0)$ defined in (2.3), that is,*

$$\chi\left(\llbracket U\rrbracket, \llbracket W\rrbracket\right) = \max\left\{\sup_\varphi\left\{\inf_\psi\{d_{\mathrm{w}^*}(U^\varphi, W^\psi)\}\right\}, \sup_\varphi\left\{\inf_\psi\{d_{\mathrm{w}^*}(U^\psi, W^\varphi)\}\right\}\right\},$$

for $\llbracket U\rrbracket, \llbracket W\rrbracket \in \widetilde{\mathcal{W}_0}$ and φ and ψ range through all measure preserving bijections on Ω.[j]

Finally, $\left(\widetilde{\mathcal{W}_0}, \delta_\square\right)$ is compact.

Proof. The map $\langle\cdot\rangle$ is well-defined and injective by Corollary 4.24 and it is continuous by Theorem 6.1. The set $X = \left\langle\widetilde{\mathcal{W}_0}\right\rangle$ is closed in the Vietoris topology. Indeed, suppose that $\langle\Gamma_n\rangle \to L$ in $K\left(\mathcal{W}_0\right)$ for some $L \in K\left(\mathcal{W}_0\right)$. Then it follows from [13, Exercise 4.21, Exercise 4.23] that $L = \mathbf{LIM}_{\mathrm{w}*}\left(\Gamma_1, \Gamma_2, \Gamma_3, ...\right) = \mathbf{ACC}_{\mathrm{w}*}\left(\Gamma_1, \Gamma_2, \Gamma_3, ...\right)$, see the definition of topological limit before [13, Exercise 4.23]. By Theorem 3.5 we know that there is some $W \in \mathcal{W}_0$ such that $\Gamma_n \xrightarrow{\delta_\square} W$. By the continuity of $\langle.\rangle$ we also have $\langle\Gamma_n\rangle \to \langle W\rangle$ in $K\left(\mathcal{W}_0\right)$.

$K(\mathcal{W}_0)$ is compact by Fact 2.7 and Remark 2.8. Since X is a closed subset of $K(\mathcal{W}_0)$, it is also compact. By Theorem 6.1, we have that the inverse map $\langle\cdot\rangle^{-1}$ is also continuous. Therefore $\widetilde{\mathcal{W}_0}$ is homeomorphic to X, and hence compact. Therefore, $\delta_\square$ and χ give the same compact topology on $\widetilde{\mathcal{W}_0}$. □

7. Relation to multiway cuts

In this section, we compare our view with [5]. The main result of [5] is a statement that several properties of graph(on) sequences are equivalent to cut distance convergence. These properties include so-called right-convergence and convergence of free energies. As

[i] Recall that two metrics on a topological space are *equivalent* if they give the same topology.

[j] Note that the definition of $\chi\left(\llbracket U\rrbracket, \llbracket W\rrbracket\right)$ does not depend on the particular representatives U and W. Also note that this statement holds for any metric d_{w^*} compatible with the weak* topology, not only the one given in (2.1).

we shall see, however, the one that relates to our approach involves so-called multiway cuts.

Let us give the key definition (see p. 178 in [5]). First, for two real matrices M and N of the same dimensions, their L_1-*distance* is defined as $\|M-N\|_1 := \sum_{i,j}|M_{ij}-N_{ij}|$. Suppose that $\mathbf{a}=(a_1,\ldots,a_q)$ is a vector with non-negative entries that sum up to 1. Given a graphon $W:\Omega^2\to[0,1]$, we write $\mathcal{S}_{\mathbf{a}}(W)$ for the set of all $q\times q$ matrices M that can be obtained in the following way. Partition suitably $\Omega=\Omega_1\sqcup\ldots\sqcup\Omega_q$ so that $\nu(\Omega_i)=a_i$. Then for each $i,j\in[q]$, define $M_{ij}:=\int_{\Omega_i}\int_{\Omega_j}W$. So, $\mathcal{S}_{\mathbf{a}}(W)\subseteq[0,1]^{q\times q}$. In particular, given two graphons U and W, we can talk about the *Hausdorff distance with respect to* ℓ_1 of the sets $\mathcal{S}_{\mathbf{a}}(U)$ and $\mathcal{S}_{\mathbf{a}}(W)$,

$$d_1^{\mathrm{Hf}}\left(\mathcal{S}_{\mathbf{a}}(U),\mathcal{S}_{\mathbf{a}}(W)\right):=\max\left\{\sup_{M\in\mathcal{S}_{\mathbf{a}}(U)}\inf_{N\in\mathcal{S}_{\mathbf{a}}(W)}\|M-N\|_1,\ \sup_{M\in\mathcal{S}_{\mathbf{a}}(W)}\inf_{N\in\mathcal{S}_{\mathbf{a}}(U)}\|M-N\|_1\right\}.$$

Then one of the main equivalences of [5], given there in Theorem 3.5 and stated here in a slightly tailored form[k], is the following.

Theorem 7.1 ([5])**.** *Let $\Gamma_1,\Gamma_2,\Gamma_3,\ldots\in\mathcal{W}_0$. The following are equivalent:*

(a) The sequence $\Gamma_1,\Gamma_2,\Gamma_3,\ldots$ is Cauchy with respect to the cut distance $\delta_\square$,

(b) For each $\mathbf{a}$ as above, the sequence $\mathcal{S}_{\mathbf{a}}(\Gamma_1),\mathcal{S}_{\mathbf{a}}(\Gamma_2),\mathcal{S}_{\mathbf{a}}(\Gamma_3),\ldots$ is Cauchy with respect to d_1^{Hf}.

To close the circle, we prove that the second conditions of Theorem 3.5 and of Theorem 7.1 are equivalent. We shall use the Hausdorff metric $d_{\mathrm{w}^*}^{\mathrm{Hf}}$ on $K(\mathcal{W}_0)$ derived as in (2.3) from the weak* metric defined, say, by (2.1) .

Proposition 7.2. *Let $\Gamma_1,\Gamma_2,\Gamma_3,\ldots\in\mathcal{W}_0$. The following are equivalent:*

(a) $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)=\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$,

(b) The sequence $\langle\Gamma_1\rangle,\langle\Gamma_2\rangle,\langle\Gamma_3\rangle,\ldots$ is Cauchy with respect to $d_{\mathrm{w}^*}^{\mathrm{Hf}}$,

(c) For each $\mathbf{a}$ as above, the sequence $\mathcal{S}_{\mathbf{a}}(\Gamma_1),\mathcal{S}_{\mathbf{a}}(\Gamma_2),\mathcal{S}_{\mathbf{a}}(\Gamma_3),\ldots$ is Cauchy with respect to d_1^{Hf}.

Proof. First we show that (c) implies (a). Suppose that $\Omega=[0,1]$ and let $\mathcal{I}_k=\{I_{k,1},\ldots,I_{k,k}\}$ be the canonical partition of $[0,1]$ into equimeasurable intervals, i.e., $\lambda(I_{k,i})=\frac{1}{k}$ and $I_{k,i}$ precedes $I_{k,i+1}$. If $\mathcal{P}=\{P_1,\ldots,P_k\}$ is another equitable partition of $[0,1]$, then we denote as $\pi_{\mathcal{P}}$ any fixed measure preserving bijection such that $\pi_{\mathcal{P}}(P_i)=I_{k,i}$ for every $i\in[k]$. Let $W\in\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$. We may suppose that there is a subsequence $\{n_j\}_{j\in\mathbb{N}}$ such that $\Gamma_{n_j}\xrightarrow{\mathrm{w}^*}W$. Fix $k\in\mathbb{N}$. It is easy to see that $\left(\Gamma_{n_j}^{\ltimes\mathcal{I}_k}\right)\xrightarrow{\mathrm{w}^*}W^{\ltimes\mathcal{I}_k}$. Define

[k]More precisely, Theorem 3.5 of [5] does not talk about $\mathcal{S}_{\mathbf{a}}(\cdot)$, but about a slightly different object $\mathcal{S}_q(\cdot)$. However, equivalence of these two approaches easily follows from Lemma 4.5 of [5].

$M_{n_j}(r,s) = \int_{I_{k,r}\times I_{k,s}} \Gamma_{n_j}$ and $N_k(r,s) = \int_{I_{k,r}\times I_{k,s}} W$ for every $r,s \in [k]$. Then we have $M_{n_j} \in \mathcal{S}_{\mathbf{a}}(\Gamma_{n_j})$ for every $j \in \mathbb{N}$ and $N_k \in \mathcal{S}_{\mathbf{a}}(W)$ where $\mathbf{a} = \left(\frac{1}{k},\ldots,\frac{1}{k}\right)$. Moreover, $M_{n_j} \xrightarrow{\|\cdot\|_1} N_k$. The assumption (c) allows to find an equitable partition $\mathcal{P}_n = \{P_i\}_{\in[k]}$ for every $n \notin \{n_j : j \in \mathbb{N}\}$ such that if we define $M_n(r,s) = \int_{P_r\times P_s} \Gamma_n$ for every $r,s \in [k]$, then we have $M_n \xrightarrow{\|\cdot\|_1} N_k$. One can easily verify that for $W_n := \left((\Gamma_n)^{\ltimes\mathcal{P}_n}\right)^{\pi_{\mathcal{P}_n}}$ we have $W_n = ((\Gamma_n)^{\pi_{\mathcal{P}_n}})^{\ltimes\mathcal{I}_k}$ and therefore $W_n \in \langle\Gamma_n\rangle$ by Lemma4.1. This implies that $W_n \xrightarrow{\text{w}^*} W^{\ltimes\mathcal{I}_k}$ because $W_n \restriction I_{k,r}\times I_{k,s} = k^{-2}\cdot M_n(r,s)$ holds for every $n \in \mathbb{N}$ and $r,s \in [k]$. Hence, $W^{\ltimes\mathcal{I}_k} \in \mathbf{LIM}_{\text{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$. Finally, since $W^{\ltimes\mathcal{I}_k} \xrightarrow{\text{w}^*} W$ and $\mathbf{LIM}_{\text{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$ is weak* closed by Lemma3.1 (a), we conclude that $W \in \mathbf{LIM}_{\text{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots)$.

The direction from (a) to (b) follows from Corollary 6.2. It remains to show that (b) implies (c). Let $\mathbf{a} \in [0,1]^q$ be a vector whose entries sum up to 1, $\epsilon > 0$ and $\mathcal{P} = \{P_i\}_{i=1}^q$ be a partition of Ω with the property that $\nu(P_i) = a_i$. Recall that we have fixed a collection $\{A_r\}_{r\in\mathbb{N}}$ of measurable subsets of $[0,1]$ that is dense in the measure algebra and that defines the Hausdorff metric $d_{\text{w}^*}^{\text{Hf}}$ on $K(\mathcal{W}_0)$ by (2.1). It follows that there is a sequence $\{A_{r_i}\}_{i=1}^q$ such that $\sum_{i=1}^q \lambda(P_i \triangle A_{r_i}) < \epsilon$. Let $W,U \in \mathcal{W}$ be such that $d_{\text{w}^*}^{\text{Hf}}(\langle U\rangle,\langle W\rangle) < \frac{\epsilon}{2^{2rq}}$.

Let $M \in \mathcal{S}_{\mathbf{a}}(W)$. Then there are measure preserving bijections π,ϕ such that $M(i,j) = \int_{P_i\times P_j} W^\phi = \int_{\phi^{-1}(P_i)\times\phi^{-1}(P_j)} W$ and $d_{w^*}\left(\left(W^\phi\right)^{\ltimes\mathcal{P}}, U^\pi\right) < \frac{\epsilon}{2^{2rq}}$. Define $N \in \mathcal{S}_{\mathbf{a}}(U)$ as $N(i,j) = \int_{P_i\times P_j} U^\pi = \int_{\pi^{-1}(P_i)\times\pi^{-1}(P_j)} U$. This gives

$$\begin{aligned}
\|M-N\|_1 &= \sum_{i,j=1}^q \left|\int_{P_i\times P_j} \left(W^\phi\right)^{\ltimes\mathcal{P}} - U^\pi\right| \le \sum_{i,j=1}^q \left|\int_{A_{r_i}\times A_{r_j}} \left(W^\phi\right)^{\ltimes\mathcal{P}} - U^\pi\right| + \epsilon \\
&\le 2^{2r_q} \sum_{i,j=1}^q 2^{-(r_i+r_j)} \left|\int_{A_{r_i}\times A_{r_j}} \left(W^\phi\right)^{\ltimes\mathcal{P}} - U^\pi\right| + \epsilon \le 2^{2r_q} d_{\text{w}^*}^{\text{Hf}}(\langle U\rangle,\langle W\rangle) + \epsilon \\
&\le 2\epsilon.
\end{aligned}$$

This shows that if $d_{\text{w}^*}^{\text{Hf}}(\langle U\rangle,\langle W\rangle) < \frac{\epsilon}{2^{2rq}}$, then $d_1^{\text{Hf}}(\mathcal{S}_{\mathbf{a}}(U),\mathcal{S}_{\mathbf{a}}(W)) \le 2\epsilon$. Consequently, (b) implies (c). □

8. Properties of the structuredness (quasi)order

Above, we obtained properties of the structuredness (quasi)order $\preceq$ that were needed for our abstract proof of Theorem 6.1. In this section, we establish further properties of $\preceq$. In Lemma 8.1 we prove that $\preceq$ is actually a closed order on $\widetilde{\mathcal{W}_0}$. In Corollary 8.3, we prove that $\preceq$-increasing/decreasing chains of graphons are cut distance convergent. In Corollary 8.4, we characterize the elements of $K(\mathcal{W}_0)$ that are envelopes of graphons. Finally, in Proposition 8.4, we characterize $\preceq$-minimal and $\preceq$-maximal elements. This

last-mentioned result is just starting point of investigating the structure of the poset $\preceq$, which we leave open.

Let us first prove that $\preceq$ is actually an order modulo weak isomorphism.

Lemma 8.1. *The relation $\preceq$ on the space $\widetilde{\mathcal{W}_0}$ is an order, and as a subset of $\widetilde{\mathcal{W}_0} \times \widetilde{\mathcal{W}_0}$ it is closed.*

Proof. Since by Corollary 6.2 the space $\left(\widetilde{\mathcal{W}_0}, \delta_\square\right)$ is homeomorphic to some closed subspace of $K(\mathcal{W}_0)$ and the relation $\preceq$ is interpreted as $\subseteq$ on $K(\mathcal{W}_0)$ it is enough to verify the properties for the relation $\subseteq$ on $K(\mathcal{W}_0)$. But both properties are trivially satisfied for the relation $\subseteq$. □

Next, we turn our attention to finding upper and lower bounds with respect to $\preceq$. Let us first give an auxiliary result, which is then utilized in Corollaries 8.3 and 8.4.

Proposition 8.2. *(a) Suppose that $P \subseteq \mathcal{W}_0$ is upper-directed in the structuredness order, i.e., for every $U_0, U_1 \in P$ there is $V \in P$ such that $U_0, U_1 \preceq V$. Then there is a graphon $W \in \mathcal{W}_0$ such that W is the supremum of P with respect to $\preceq$.*

(b) Suppose that $P \subseteq \mathcal{W}_0$ is down-directed in the structuredness order, i.e., for every $U_0, U_1 \in P$ there is $V \in P$ such that $V \preceq U_0, U_1$. Then there is a graphon $W \in \mathcal{W}_0$ such that W is the infimum of P with respect to $\preceq$.

Proof. Let us focus on (a). First of all consider the set $\langle P\rangle = \{\langle U\rangle : U \in P\}$. This set is upper-directed with respect to $\subseteq$ in $K(\mathcal{W}_0)$. Let K be the weak* closure of $\bigcup_{U\in P}\langle U\rangle$. Clearly, $K \in K(\mathcal{W}_0)$. Further, K is the supremum of $\langle P\rangle$ with respect to $\subseteq$ on $K(\mathcal{W}_0)$. To finish the proof, we only need to show that there exists $W \in \mathcal{W}_0$ such that $K = \langle W\rangle$. Consider a countable set $P_0 \subseteq P$ such that $\langle P_0\rangle$ is dense in $\langle P\rangle$. This can be done since $K(\mathcal{W}_0)$ is separable metrizable by Fact 2.7. Take some enumeration $U_1, U_2, \ldots$ of P_0. Define inductively an increasing chain $\Gamma_1, \Gamma_2, \ldots \in P$ such that for every $n \in \mathbb{N}$ we have that $U_1, \ldots, U_n, \Gamma_{n-1} \preceq \Gamma_n$. This can be done since P is upper-directed. Since $\Gamma_1 \preceq \Gamma_2 \preceq \Gamma_3 \preceq \ldots$, we have $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots) = \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. Indeed, whenever we take $\Gamma \in \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$, say $\Gamma_{n_1}^{\pi_{n_1}}, \Gamma_{n_2}^{\pi_{n_2}}, \Gamma_{n_3}^{\pi_{n_3}}, \ldots \xrightarrow{\|\cdot\|_\square} \Gamma$, then for each index i in the interval (n_k, n_{k+1}), we can use that $\Gamma_{n_k} \preceq \Gamma_i$ to approximate $\Gamma_{n_k}^{\pi_{n_k}}$ by some version $\Gamma_i^{\pi_i}$ of Γ_i (with a vanishing error as $i \to \infty$). With the gaps (n_k, n_{k+1}) filled-in this way, we have $\Gamma_{n_1}^{\pi_{n_1}}, \Gamma_{1+n_1}^{\pi_{1+n_1}}, \Gamma_{2+n_1}^{\pi_{2+n_1}}, \ldots \xrightarrow{\|\cdot\|_\square} \Gamma$, and consequently $\Gamma \in \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$, as we wanted. Moreover $K = \mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ because $\langle P_0\rangle$ is dense in $\langle P\rangle$. By Lemma 4.9, we can pick a $\preceq$-maximal element W of $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. Now, we have that $K = \langle W\rangle$.

The proof of (b) is similar. The only difference is that the desired infimum is of the form $K = \bigcap_{U\in P}\langle U\rangle$ and we inductively build a decreasing sequence. □

Along very similar lines, we can prove that $\preceq$-increasing/decreasing chains of graphons are cut distance convergent.

Corollary 8.3. *(a) Suppose that $W_1 \preceq W_2 \preceq W_3 \preceq \ldots$ is a sequence of graphons. Then this sequence is cut distance convergent.*

(b) Suppose that $W_1 \succeq W_2 \succeq W_3 \succeq \ldots$ is a sequence of graphons. Then this sequence is cut distance convergent.

Proof. Suppose that $W_1 \preceq W_2 \preceq W_3 \preceq \ldots$. Then we have $\mathbf{LIM}_{\mathrm{w}*}(W_1, W_2, W_3, \ldots) = \mathbf{ACC}_{\mathrm{w}*}(W_1, W_2, W_3, \ldots)$. By Lemma 4.9, we can pick a $\preceq$-maximal element W of the set $\mathbf{LIM}_{\mathrm{w}*}(W_1, W_2, W_3, \ldots)$. Now, by Theorem 6.1 we have that $W_n \xrightarrow{\delta_\square} W$.

The proof for a decreasing sequence is the same. □

Next, we characterize the elements of $K(\mathcal{W}_0)$ that are envelopes of graphons.

Corollary 8.4. *Let $K \in K(\mathcal{W}_0)$. Then there exists $W \in \mathcal{W}_0$ such that $K = \langle W \rangle$ if and only if K is upper-directed (for every $U_0, U_1 \in K$ there is $V \in K$ such that $U_0, U_1 \preceq V$) and downwards closed (for every $U \in K$ and $V \preceq U$ we have $V \in K$).*

Proof. Suppose first that $K = \langle W \rangle$. Then for every $U_0, U_1 \in K$, we have $U_0, U_1 \preceq W$. That is, K is upper-directed. Suppose next that $U \in K$ and $V \preceq U$. Let $\epsilon > 0$ be arbitrary. Since $V \preceq U$, there exists a version U^π of U such that $d_{\mathrm{w}^*}(V, U^\pi) < \frac{\epsilon}{2}$. Since $U \in K$, we have $U \preceq W$. Thus, also $U^\pi \preceq W$. Therefore, there exists a version W^θ of W such that $d_{\mathrm{w}^*}\left(U^\pi, W^\theta\right) < \frac{\epsilon}{2}$. We conclude that $d_{\mathrm{w}^*}\left(V, W^\theta\right) < \epsilon$. Since ϵ was arbitrarily small and since K is weak* closed, we conclude that $V \in K$.

Let us now turn to the other implication. As K is upper-directed then by Proposition 8.2 we can take its supremum W. Because K is downwards closed we have $\langle W \rangle = K$. □

Last, we characterize the minimal and maximal elements of the structuredness order; the latter part being suggested to us by László Miklós Lovász.

Proposition 8.5. *The minimal elements of the structuredness order are exactly constant graphons, and for every graphon W there is a minimal graphon $W_{\min} \preceq W$. The maximal elements of the structuredness order are exactly 0-1 valued graphons, and for every graphon W there is a maximal graphon $W_{\max} \succeq W$.*

Proof. The first part follows directly from the fact that an envelope of any graphon contains a constant graphon (see Lemma 4.2(b)), and also the graphon itself. For the second part we at first prove that only 0-1 valued graphons can be maximal.

Suppose that W is a graphon such that its value is neither 0 nor 1 on a set of positive measure. This implies that there is an $\varepsilon > 0$ such that W has values between ε and $1 - \varepsilon$ on a set of positive measure. Now consider a map $\varphi\colon [0,1] \to [0,1]$ such that $\varphi(x) = 2x$

for $0 \leq x \leq \frac{1}{2}$ and $\varphi(x) = 2x - 1$ for $\frac{1}{2} < x \leq 1$. The graphon W^φ contains four copies of W scaled by a factor of one half. Let $B \subseteq [0,1]^2$ be the set, on which W^φ takes values between ε and $1-\varepsilon$. Let $\widetilde{W}$ be a graphon such that $\widetilde{W} = W^\varphi + \varepsilon$ for $x, y \in \big([0,\frac{1}{2}]^2 \cup [\frac{1}{2},1]^2\big) \cap B$, $\widetilde{W} = W^\varphi - \varepsilon$ for $x, y \in \big([0,\frac{1}{2}] \times [\frac{1}{2},1] \cup [\frac{1}{2},1] \times [0,\frac{1}{2}]\big) \cap B$ and $\widetilde{W} = W$ otherwise. The values of $\widetilde{W}$ are bounded by 0 and 1 and W and $\widetilde{W}$ are not weakly isomorphic (compare $\mathrm{INT}_f(W)$ and $\mathrm{INT}_f(\widetilde{W})$ for any strictly convex function f). Moreover, we claim that $W \in \left\langle \widetilde{W} \right\rangle$. To see this, one can construct a sequence of measure preserving almost-bijections $\psi_1, \psi_2, \dots$, defined as $\psi_n(x) = \frac{\lfloor 2nx \rfloor}{2n} + x$ for $0 \leq x \leq \frac{1}{2}$ and $\psi_n(x) = \frac{\lfloor 2nx \rfloor - 2n + 1}{2n} + x$ for $\frac{1}{2} \leq x \leq 1$, that interlace the two intervals $[0,\frac{1}{2}]$ and $[\frac{1}{2},1]$ and thus serve as an approximation of φ. The fact that $\widetilde{W}^{\psi_1^{-1}}, \widetilde{W}^{\psi_2^{-1}}, \dots \xrightarrow{\text{w}^*} W$ can be seen directly from the definition of weak* convergence.

Next, we prove that all 0-1 graphons are maximal. Indeed, let W be a 0-1 valued graphon, and suppose that there exists some graphon U such that $U \succ W$. Then for the measures $\mathbf{\Phi}_U$ and $\mathbf{\Phi}_W$ we have that $\mathbf{\Phi}_W$ is strictly flatter than $\mathbf{\Phi}_U$ by Proposition 4.17. This contradicts Lemma 4.23.

Finally, let W be an arbitrary graphon. Consider the set $\mathcal{P}$ of all graphons $P \succeq W$. Then every chain in $\mathcal{P}$ has a supremum in the structuredness order by Proposition 8.2(a). Therefore, we can apply Zorn's lemma to conclude that there is a maximal graphon $W_{\max} \succeq W$. □

Remark 8.6. *Let us take $W \equiv \frac{1}{2}$ and a sequence W_n of graphons corresponding to Erdős–Rényi random graphs $\mathbb{G}(n, \frac{1}{2})$. By Proposition 8.5, W is a $\prec$-minimal elements, while each W_n is a $\prec$-maximal element. Yet, it is well-known that $W_n \xrightarrow{\delta_\square} W$ almost surely. This example shows that even a cut distance convergent sequence can be fairly «separated» from the limit point with respect to the structurdness order. (Note also, that while W is a $\prec$-minimal elements, and each W_n is a $\prec$-maximal element, W_n and W are typically incomparable in the structuredness order, since the density of W_n is typically not exactly $\frac{1}{2}$.)*

In the example below we show that the structuredness order does not have meet and joins in general.

Example 8.7. We shall construct graphons W_1, W_2, U_1, U_2 such that we have $U_1, U_2 \preceq W_1, W_2$, but there is no graphon V such that $U_1, U_2 \preceq V \preceq W_1, W_2$. For a fixed $\varepsilon > 0$ we define the four graphons (on the unit square, denoting the Lebesgue measure on $[0,1]$ as ν) as follows (the definitions should be clear from Figure 8.1).

(1) Define $W_1(x,y) = 1$ if and only if $(x,y) \in [\frac{1}{2},1] \times [0,\frac{1}{2}] \cup [0,\frac{1}{2}] \times [\frac{1}{2},1] \cup [\frac{1}{4} - \frac{\varepsilon}{2}, \frac{1}{4} + \frac{\varepsilon}{2}] \times [0,1] \cup [0,1] \times [\frac{1}{4} - \frac{\varepsilon}{2}, \frac{1}{4} + \frac{\varepsilon}{2}]$. Moreover, $W_1 = 4(\varepsilon - \varepsilon^2)$ for $(x,y) \in [\frac{1}{2},1]^2$ and is zero otherwise.

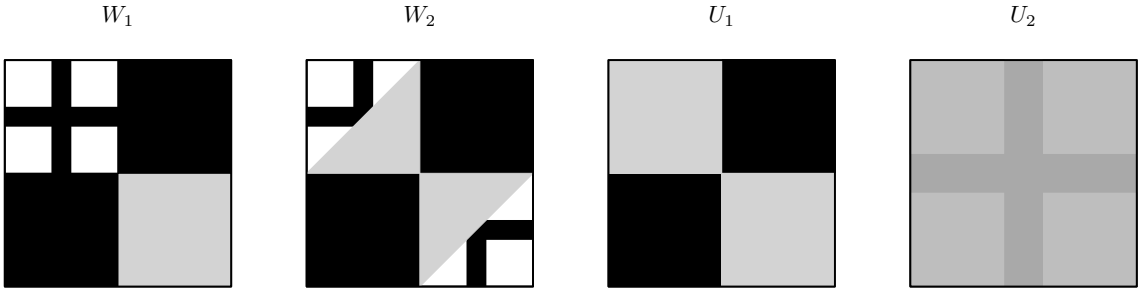


FIGURE 8.1. The four graphons W_1, W_2, U_1, U_2 witnessing that an intersection of two envelopes is not necessarily an envelope.

(2) Define W_2 as W_1 but switch its values on the triangle with vertices $[0, \frac{1}{2}], [\frac{1}{2}, \frac{1}{2}], [\frac{1}{2}, 0]$ with the triangle with vertices $[\frac{1}{2}, 1], [1, 1], [1, \frac{1}{2}]$. Note that $\int_{[0,\frac{1}{2}]^2} W_1 = \int_{[\frac{1}{2},1]^2} W_1 = \int_{[0,\frac{1}{2}]^2} W_2 = \int_{[\frac{1}{2},1]^2} W_2$.

(3) Define $U_1 = 4(\varepsilon - \varepsilon^2)$ for $(x, y) \in [0, \frac{1}{2}]^2 \cup [\frac{1}{2}, 1]^2$ and 1 otherwise. Note that $U_1 \preceq W_1, W_2$ (this follows by applying Lemma 4.2(b) twice just for the top-left and bottom-right subgraphons of the graphons W_1 and W_2).

(4) Finally, define U_2 to be such that

$$U_2(x,y) = \frac{1}{4}\left(W_1\left(\frac{x}{2},\frac{y}{2}\right) + W_1\left(\frac{x+1}{2},\frac{y}{2}\right) + W_1\left(\frac{x}{2},\frac{y+1}{2}\right) + W_1\left(\frac{x+1}{2},\frac{y+1}{2}\right)\right).$$

We have again $U_2 \preceq W_1, W_2$ (consider a sequence of bijections interlacing the two intervals $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$, as in the proof of Proposition 8.5).

We are now ready to get to the main two properties of this example.

Claim (Claim A). If V is a graphon such that $U_1 \preceq V$, then V contains a set J such that V is 1 almost everywhere on $J \times ([0,1] \setminus J)$.

Proof. The argument is similar to the proof of [7, Lemma 3] and [12, Lemma 2.3]. Observe that in U_1, the values on the rectangle $\left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]$ are 1 everywhere. Since there exists a sequence of measure preserving transformations $\pi_1, \pi_2, \ldots$ such that $V^{\pi_n} \xrightarrow{\text{w}^*} U_1$, taking $J_n := \pi_n\left(\left[0, \frac{1}{2}\right]\right)$ yields sets of measure $\frac{1}{2}$ with the property that

$$\lim_n \int_{J_n \times ([0,1]\setminus J_n)} V = \int_{[0,\frac{1}{2}]\times[\frac{1}{2},1]} U_1 = \frac{1}{4}\,. \tag{8.1}$$

Let us pass to a subsequence so that the sequence of indicators $(\mathbf{1}_{J_n})_n$ converges weak*, say to a function $g : [0,1] \to [0,1]$. Note that then the sequence $\left(\mathbf{1}_{[0,1]\setminus J_n}\right)_n$ converges weak* to $1 - g$.

We claim that V is 1 almost everywhere on $(\operatorname{supp} g) \times (\operatorname{supp}(1-g))$. Before proving this, let us explain, why this gives the desired set J. To this end, note that since $\nu(J_n) = \frac{1}{2}$,

weak* convergence implies $\int g = \frac{1}{2}$. In particular, the measure of the support of g is at least $\frac{1}{2}$ while at the same time the measure of points where $g \equiv 1$ is at most $\frac{1}{2}$. So, we take J to be an arbitrary set of measure exactly $\frac{1}{2}$ contained in the former set and containing the latter set. It is clear that all the required properties are satisfied.

So, let us prove that V is 1 almost everywhere on $(\operatorname{supp} g) \times (\operatorname{supp}(1-g))$. It suffices to prove that for each $a > 0$, the Lebesgue measure of those pairs $(x,y) \in g^{-1}((a,1]) \times g^{-1}([0,1-a))$ for which $V(x,y) < 1-a$ is zero. So, suppose that this fails for some $a > 0$. By basic properties of measure, we can then find two sets $X \subseteq g^{-1}((a,1])$, $Y \subseteq g^{-1}([0,1-a))$ of positive measure such that

$$\nu^{\otimes 2}(\{(x,y) \in X \times Y : V(x,y) \geq 1-a\}) < \frac{a^3}{100} \cdot \nu(X)\nu(Y)\,. \tag{8.2}$$

Now, observe that since g is a weak* limit of $(\mathbf{1}_{J_n})_n$ and $X \subseteq g^{-1}((a,1])$, we have that for all sufficiently large n, $\nu(X \cap J_n) > a \cdot \nu(X)$. By a similar reasoning, $\nu(Y \cap ([0,1] \setminus J_n)) > a \cdot \nu(Y)$ for all sufficiently large n. To say this in words, a substantial (non-vanishing) part of the set J_n is inside the set X and a substantial part of the set $[0,1] \setminus J_n$ is inside the set Y, and by (8.2), V is very far from being close to 1 on $X \times Y$. This contradicts (8.1), which is saying that V must be very close to 1 on most of $J_n \times ([0,1] \setminus J_n)$. □

Claim (Claim B). There does not exist any graphon V such that $U_1, U_2 \preceq V \preceq W_1, W_2$.

Proof. Assume that there is a graphon V such that $U_1, U_2 \preceq V \preceq W_1, W_2$. Let the set J be given from Claim A. Without loss of generality assume that $J = [0, \frac{1}{2}]$. Now, we turn our attention to the graphons W_i, $i = 1, 2$. From the fact that there is a sequence of measure preserving transformations $\varphi_1, \varphi_2, \ldots$ such that $W_1^{\varphi_1}, W_1^{\varphi_2}, \ldots \xrightarrow{\text{w}^*} V$ and, thus, $\lim_{n\to\infty} \int_{[0,\frac{1}{2}]\times[\frac{1}{2},1]} W_1^{\varphi_n} = \int_{[0,\frac{1}{2}]\times[\frac{1}{2},1]} V$, we obtain that for any $\delta > 0$ there is n sufficiently large such that either

$$\nu\left(\varphi_n\left([0,\frac{1}{2}]\right) \cap [0,\frac{1}{2}]\right) \leq \delta \quad \text{or} \quad \nu\left(\varphi_n\left([0,\frac{1}{2}]\right) \cap [0,\frac{1}{2}]\right) \geq \frac{1}{2} - \delta\,.$$

Indeed, suppose that $\frac{1}{2} - \delta > \nu\left(\varphi_n\left([0,\frac{1}{2}]\right) \cap [0,\frac{1}{2}]\right) > \delta$ and observe that the density of $W_1^{\varphi_n}$ is equal to $4(\varepsilon - \varepsilon^2)$ on a subset of $[0,\frac{1}{2}] \times [\frac{1}{2},1]$ of measure at least δ^2. Now it suffices to recall that $\int_{[0,\frac{1}{2}]\times[\frac{1}{2},1]} W_1^{\varphi_n} \to \nu\left([0,\frac{1}{2}] \times [\frac{1}{2},1]\right) = \frac{1}{4}$ which implies that the assumption is false for large enough n. From the fact that either $\nu\left(\varphi_n\left([0,\frac{1}{2}]\right) \cap [0,\frac{1}{2}]\right) \leq \delta$ or $\nu\left(\varphi_n\left([0,\frac{1}{2}]\right) \cap [0,\frac{1}{2}]\right) \geq \frac{1}{2} - \delta$ we conclude that actually (after passing to a subsequence) some versions of $W_1 \cap [0,\frac{1}{2}]^2, W_1 \cap [0,\frac{1}{2}]^2, \ldots$ converge weak* to either $V \cap [0,\frac{1}{2}]^2$ or $V \cap [\frac{1}{2},1]^2$ (redefine φ_n on a set of small measure such that it maps $[0,\frac{1}{2}]$ either onto $[0,\frac{1}{2}]$ or onto $[\frac{1}{2},1]$ to get the required measure preserving bijections). Without loss of generality assume that some versions of $W_1 \cap [\frac{1}{2},1]^2, W_1 \cap [\frac{1}{2},1]^2, \ldots$ converge weak* to $V \cap [\frac{1}{2},1]^2$. Similarly, we get that there are versions of $W_2 \cap [0,\frac{1}{2}]^2, W_2 \cap [0,\frac{1}{2}]^2, \ldots$ converging weak* to $V \cap [0,\frac{1}{2}]^2$

(the other case when the versions $W_2 \cap [\frac{1}{2}, 1]^2, W_2 \cap [\frac{1}{2}, 1]^2, \dots$ converge weak* to $V \cap [0, \frac{1}{2}]^2$ can be treated in the same way).

Notice that

$$\int_{[0,1]\times[\frac{1}{2}-\varepsilon,\frac{1}{2}+\varepsilon]} U_2 = (2\varepsilon)\cdot 1 \cdot \frac{3}{4} + o(\varepsilon) = \frac{3}{2}\varepsilon + o(\varepsilon)\,. \tag{8.3}$$

On the other hand,

$$\begin{aligned}
\sup_{\nu(C)=2\varepsilon} \int_{[0,1]\times C} V &= \sup_{A\subseteq[0,\frac{1}{2}],B\subseteq[\frac{1}{2},1],\nu(A\cup B)=2\varepsilon} \left(\int_{[0,1]\times A} V + \int_{[0,1]\times B} V \right) \\
&= 2\varepsilon\cdot\frac{1}{2} + \sup_{A\subseteq[0,\frac{1}{2}],B\subseteq[\frac{1}{2},1],\nu(A\cup B)=2\varepsilon} \left(\int_{[0,\frac{1}{2}]\times A} V + \int_{[\frac{1}{2},1]\times B} V \right) \\
&\leq \varepsilon + \sup_{A\subseteq[0,\frac{1}{2}],B\subseteq[\frac{1}{2},1],\nu(A\cup B)=2\varepsilon} \left(\int_{[0,\frac{1}{2}]\times A} W_2 + \int_{[\frac{1}{2},1]\times B} W_1 \right) \\
&= \varepsilon + \left(\frac{1}{4}\varepsilon + o(\varepsilon)\right) + o(\varepsilon) \\
&= \frac{5}{4}\varepsilon + o(\varepsilon), \qquad (8.4)
\end{aligned}$$

which, for ε small enough, is smaller than the appropriate value for U_2 appearing in (8.3). Now, we can conclude that $V \not\succeq U_2$. Indeed, suppose that $V \succeq U_2$. Then for every $a > 0$ there is a version V^π of V such that $d_{\mathrm{w}^*}(V^\pi, U_2) < a$. In particular, there exists a version V^π such that $\left|\int_{[0,1]\times[\frac{1}{2}-\varepsilon,\frac{1}{2}+\varepsilon]} V^\pi - U_2\right| < \frac{1}{100}$. Taking $C = \pi^{-1}\left([\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon]\right)$, (8.4) contradicts (8.3). □

9. Conclusion

Inspired by previous work [8], we created a comprehensive theory of approaching the cut distance convergence via the weak* topology. The main results, Theorem 3.5 and Theorem 3.3, say that in each sequence $\Gamma_1, \Gamma_2, \Gamma_3, \dots$ there exists a subsequence $\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \dots$ such that $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \dots) = \mathbf{ACC}_{\mathrm{w}*}(\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \dots)$, and that the latter property is equivalent to the cut distance convergence of the said subsequence. It follows from the equivalence established in Proposition 7.2 that Theorem 3.5 can be regarded as the functional-analytic formulation of Theorem 7.1, which was originally established in [5]. So, while each of the above mentioned theorems could have been obtained from previously existing results, we believe that the weak* approach we introduce provides an important perspective on the theory of graph limits. For example, Corollary 8.3 offers families of convergent graphon sequences, which are very natural with this perspective

but would be even difficult to define previously. It turns out that many of these concepts extend (in a nontrivial way) to hypergraphons, and this is currently work in progress.

Our initial motivation was to build a theory rather than to solve any particular problem, quite contrary to the usual perception of combinatorics (see [11]). That said, we already see that this abstract theory has its fruits on the problem-solving side: in [9] we use it to prove that a connected graph is weakly norming if and only if it is step Sidorenko, and that if a graph is norming then it is step forcing (see [9, Remark 3.20] for an explanation of the role of the the weak* theory in these results). We see a potential for more applications in the theory of hypergraphons.

The weak* approach may provide alternative proofs of various generalizations of Theorem 1.1, such as Banach space valued graphons [15], representations of limits of sparse graphs using measures [16], and an L^p-approach to limits of sparse graphs [3]. To indicate the plausibility of this approach in the setting L^p-limits (which seems the most feasible one), we recall that the corresponding notion of weak* convergence in this setting is based on tests against all L^q-functions (where $\frac{1}{p} + \frac{1}{q} = 1$). This would of course change the definition of $\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$, and all related definitions. Working out this program in detail is currently in progress.

Acknowledgments

We would like to thank László Miklós Lovász, Daniel Král and Oleg Pikhurko for their useful suggestions, and Frederik Garbe and Ondřej Kůrka for comments on a preliminary version of this paper. We thank a referee for his or her comments.

References

[1] T. Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.*, 5:80–145, 2008.

[2] P. Billingsley. *Probability and Measure*, volume 60 of *Wiley series in probability and mathematical statistics*. John Wiley & sons, New York, NY., 1995.

[3] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao. An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Trans. Amer. Math. Soc.*, 372(5):3019–3062, 2019.

[4] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[5] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math. (2)*, 176(1):151–219, 2012.

[6] P. Diaconis and S. Janson. Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)*, 28(1):33–61, 2008.

[7] M. Doležal and J. Hladký. Matching polytons. *Electron. J. Combin.*, 26(4):Paper 4.38, 33, 2019.

[8] J. Doležal, M. Hladký. Cut-norm and entropy minimization over weak* limits. *J. Combin. Theory Ser. B*, 137:232–263, 2019.

[9] M. Doležal, J. Grebík, J. Hladký, I. Rocha, and V. Rozhoň. Cut distance identifying graphon parameters over weak* limits. arXiv:1809.03797v4.

[10] G. Elek and B. Szegedy. A measure-theoretic approach to the theory of dense hypergraphs. *Adv. Math.*, 231(3-4):1731–1772, 2012.

[11] W. T. Gowers. The two cultures of mathematics. In *Mathematics: frontiers and perspectives*, pages 65–78. Amer. Math. Soc., Providence, RI, 2000.

[12] J. Hladký, P. Hu, and D. Piguet. Komlós's tiling theorem via graphon covers. *J. Graph Theory*, 90(1):24–45, 2019.

[13] A. S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[14] D. Král', T. Martins, P. P. Pach, and M. Wrochna. The step Sidorenko property and non-norming edge-transitive graphs. *J. Combin. Theory Ser. A*, 162:34–54, 2019.

[15] D. Kunszenti-Kovács, L. Lovász, and B. Szegedy. Multigraph limits, unbounded kernels, and banach space decorated graphs. arXiv:1406.7846.

[16] D. Kunszenti-Kovács, L. Lovász, and B. Szegedy. Measures on the square as sparse graph limits. *J. Combin. Theory Ser. B*, 138:1–40, 2019.

[17] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[18] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[19] L. Lovász and B. Szegedy. Szemerédi's Lemma for the analyst. *J. Geom. and Func. Anal*, 17:252–270, 2007.

[20] G. Regts and A. Schrijver. Compact orbit spaces in Hilbert spaces and limits of edge-colouring models. *European J. Combin.*, 52(part B):389–395, 2016.

Doležal: Institute of Mathematics, Czech Academy of Sciences. Žitná 25, 110 00, Praha, Czechia. With institutional support RVO:67985840, *Grebík:* Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK. *This work was done while affiliated with:* Institute of Mathematics, Czech Academy of Sciences. Žitná 25, 110 00, Praha, Czechia. With institutional support RVO:67985840, *Hladký:* Institute of Mathematics, Czech Academy of Sciences. Žitná 25, 110 00, Praha, Czechia. *Part of this work was done while affiliated with:* Institut für Geometrie, TU Dresden, 01062 Dresden, Germany, *Rocha*: Institute of Computer Science, Czech Academy of Sciences. Pod Vodárenskou věží 2, 182 07, Prague, Czechia. With institutional support RVO:67985807., *Rozhoň:* ETH Zürich, Switzerland. *This work was done while affiliated with:* Institute of Computer Science, Czech Academy of Sciences. Pod Vodárenskou věží 2, 182 07, Prague, Czechia. With institutional support RVO:67985807.

CUT DISTANCE IDENTIFYING GRAPHON PARAMETERS OVER WEAK* LIMITS

MARTIN DOLEŽAL, JAN GREBÍK, JAN HLADKÝ, ISRAEL ROCHA, VÁCLAV ROZHOŇ

Abstract. The theory of graphons comes with the so-called cut norm and the derived cut distance. The cut norm is finer than the weak* topology (when considering the predual of L^1-functions). Doležal and Hladký [J. Combin. Theory Ser. B 137 (2019), 232-263] showed, that given a sequence of graphons, a cut distance accumulation graphon can be pinpointed in the set of weak* accumulation points as a minimizer of the entropy. Motivated by this, we study graphon parameters with the property that their minimizers or maximizers identify cut distance accumulation points over the set of weak* accumulation points. We call such parameters *cut distance identifying*.

Of particular importance are cut distance identifying parameters coming from homomorphism densities, $t(H, \cdot)$. This concept is closely related to the emerging field of graph norms, and the notions of the step Sidorenko property and the step forcing property introduced by Kráľ, Martins, Pach and Wrochna [J. Combin. Theory Ser. A 162 (2019), 34-54]. We prove that a connected graph is weakly norming if and only if it is step Sidorenko, and that if a graph is norming then it is step forcing.

Further, we study convexity properties of cut distance identifying graphon parameters, and find a way to identify cut distance limits using spectra of graphons. We also show that continuous cut distance identifying graphon parameters have the «pumping property», and thus can be used in the proof of the Frieze–Kannan regularity lemma.

1. Introduction

The theory of graphons, initiated in [3, 30] and covered in depth in [29], provides a powerful formalism for handling large graphs that are dense, i.e., they contain a positive proportion of all potential edges. In this paper, we study the relation between the cut norm and the weak* topology on the space of graphons through various graphon parameters. Let us give basic definitions needed to explain our motivation and results.

We write $\mathcal{W}_0$ for the space of all *graphons*, i.e., all symmetric measurable functions from Ω^2 to $[0, 1]$. Here as well as in the rest of the paper, Ω is an arbitrary standard Borel space with an atomless probability measure ν. Given a graphon W and a measure preserving bijection (*m.p.b.*, for short) $\varphi : \Omega \to \Omega$, we define the *version of* W by

$$W^{\varphi}(x, y) = W(\varphi(x), \varphi(y)) \ .$$

Key words and phrases. graphon; graph limit; cut norm; weak* convergence; graph norms; Sidorenko's conjecture.

Hladký was supported by the Alexander von Humboldt Foundation. *Rocha* and *Rozhoň* were supported by the Czech Science Foundation, grant number GJ16-07822Y. *Doležal* and *Hladký* were supported by the Czech Science Foundation, grant number GJ18-01472Y. *Grebík* was supported by the Czech Science Foundation, grant number 17-33849L.

Let us recall that the *cut norm* is defined by[a]

$$\|Y\|_{\square} = \sup_{S,T\subset\Omega} \left| \int_{S\times T} Y \right| \quad \text{for each } Y \in L^1(\Omega^2)\ . \tag{1.1}$$

Given two graphons U and W we define in (1.2) their *cut norm distance* and in (1.3) their *cut distance*,

$$d_{\square}(U,W) := \|U-W\|_{\square}\ , \text{ and} \tag{1.2}$$

$$\delta_{\square}(U,W) := \inf_{\varphi:\Omega\to\Omega \text{ m.p.b.}} d_{\square}(U,W^{\varphi})\ . \tag{1.3}$$

Recall that the key property of the space $\mathcal{W}_0$, which makes the theory so powerful in applications in extremal graph theory, random graphs, property testing, and other areas, is its compactness with respect to the cut distance. This result was first proven by Lovász and Szegedy [30] using the regularity lemma,[b] and then by Elek and Szegedy [14] using ultrafilter techniques, by Austin [2] and Diaconis and Janson [11] using the theory of exchangeable random graphs, and finally by Doležal and Hladký [13] and by Doležal, Grebík, Hladký, Rocha, and Rozhoň [12] in a way explained below. For our later purposes, it is more convenient to state the compactness theorem in terms of the cut norm distance.

Theorem 1.1. *For every sequence $\Gamma_1,\Gamma_2,\Gamma_3,\dots$ of graphons exists is a subsequence $\Gamma_{n_1},\Gamma_{n_2},\Gamma_{n_3},\dots$, measure preserving bijections $\pi_{n_1},\pi_{n_2},\pi_{n_3},\dots:\Omega\to\Omega$ and a graphon Γ such that $d_{\square}\left(\Gamma_{n_i}^{\pi_{n_i}},\Gamma\right)\to 0$.*

Let us now explain the approach from [13] and from [12], which is based on the weak* topology. Throughout the paper, we regard graphons as functions in the Banach space $L^{\infty}(\Omega^2)$, to which we associate the concept of weak* convergence given by its predual Banach space $L^1(\Omega^2)$. Therefore a sequence of graphons $\Gamma_1,\Gamma_2,\Gamma_3,\dots$ *converges weak** to a graphon W if for every $Q\subset\Omega^2$ we have $\lim_{n\to\infty}(\int_Q \Gamma_n - \int_Q W)=0$. Since the sigma-algebra of measurable sets on Ω^2 is generated by sets of the form $S\times T$ where $S,T\subset\Omega$ we have that this is equivalent to requiring that $\lim_{n\to\infty}(\int_{S\times T}\Gamma_n - \int_{S\times T} W)=0$ for each $S,T\subset\Omega$. This latter perspective on the definition of weak* convergence is better suited for our purposes as we are ranging over the same space as in (1.1). In particular, we get that the weak* topology is weaker than the topology generated by $d_{\square}$, which can be viewed as a certain uniformization of the weak* topology.

So, the idea in [13] and [12], on a high level, is to look at the set $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\dots)$ of all weak* accumulation points of sequences,

$$\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\dots) = \bigcup_{\pi_1\pi_2,\pi_3,\dots:\Omega\to\Omega \text{ m.p.b.}} \text{weak* accumulation points of} \Gamma_1^{\pi_1},\Gamma_2^{\pi_2},\Gamma_3^{\pi_3},\dots\ ,$$

[a] All the sets and functions below are tacitly assumed to be measurable.

[b] See also [31] and [33] for variants of this approach.

and locate in the set $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ one graphon Γ that is an accumulation point not only with respect to the weak* topology but also with respect to the cut norm distance. In [13], this was done by choosing Γ as a maximizer[c] of an operator $\mathrm{INT}_f(\cdot)$ on $\mathcal{W}_0$, defined for a continuous strictly convex function $f : [0,1] \to \mathbb{R}$ by

$$\mathrm{INT}_f(W) := \int_x \int_y f\left(W(x,y)\right) \ . \tag{1.4}$$

In [12], we then approached Theorem 1.1 by more abstract means. Namely, we showed that Γ can be chosen as the element with the maximum «envelope» in $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$. We recall the notion of envelopes in Section 2.6. For now, it suffices to say that each envelope is a subset of $L^\infty(\Omega^2)$ and the notion of maximality is with respect to the set inclusion. In particular, envelopes are not numerical quantities.

The main focus of this paper is to return to the numerical program initiated in [13]. We provide a comprehensive study of graphon parameters where the maximization problem over $\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)$ pinpoints cut distance accumulation points. We call such parameters «cut distance identifying»; further we call parameters satisfying somewhat a weaker property «cut distance compatible» (definitions are given in Section 3.1). We introduce similar but more abstract (i.e., non-numerical) notions of «cut distance identifying graphon orders» and «cut distance compatible graphon orders». In Section 3.1 we sketch that each cut distance identifying parameter/order can indeed be used to prove Theorem 1.1. While we initially regarded cut distance identifying/compatible graphon parameters/orders merely as a tool to understanding the relation between the weak* and the cut norm topologies, as we shall see below, it naturally led to results regarding quasirandomness, the Regularity Lemma, and graph norms. Let us now highlight some of these results, following the order in the paper. In this presentation we are somewhat imprecise (in particular, we omit various continuity assumptions) and use notions that can be found in the main body of the text.

Relation to quasirandomness. Recall that the Chung–Graham–Wilson theorem provides several characterizations of quasirandom graph sequences. Two of these characterizations are minimization characterizations; dealing with the 4-cycle density and the spectrum of the adjacency matrix, respectively. In Section 3.1.1 we explain that each cut distance identifying parameter/order gives rise to such a minimization counterpart to the Chung–Graham–Wilson theorem. As we show, the 4-cycle density and the spectrum of the adjacency matrix (or, rather of the graphon), indeed possess these stronger properties (see Theorems 3.28 and 3.22) and can be used as cut distance identifying parameters/orders.

[c] In fact, the supremum of $\{\mathrm{INT}_f(W) : W \in \mathbf{ACC}_{\mathrm{w}*}(\Gamma_1, \Gamma_2, \Gamma_3, \ldots)\}$ need not be attained (see [13, Section 7.4]), so the rigorous treatment needs to be a bit more technical. Similarly, we simplify the presentation of the approach from [12] below. The correct way is shown in Theorems 3.4 and 3.5.

Index-pumping. Starting with Szemerédi's Regularity Lemma [39], the field of regularizations of graphs is now one of the most powerful areas of graph theory. In the heart of proofs of these regularity lemmas is a certain «index-pumping» parameter. Recall that the most common index-pumping parameter is the «mean-square density». Sometimes, another index-pumping parameter is more convenient. For example, Scott [34] used a slight modification of the mean-square density to get a better handle on sparse graphs. Also, Gowers [19] famously used «octahedral densities» for index-pumping in his hypergraph regularity lemma; projected to the 2-uniform case (i.e., graphs), this would correspond to using the C_4-density for index-pumping. In Section 3.2 we show that each cut distance identifying graphon parameter can be used for «index-pumping» in the Frieze-Kannan Regularity Lemma (see Proposition 3.11). Our results in particular imply that any norming graph can play the same role (by Theorem 3.28).

The parameter $\mathrm{INT}_f(\cdot)$. In Section 3.3 we reprove the result of Doležal and Hladký and show that the assumption of f being continuous in (1.4) is not really needed (see Theorem 3.14). This result is a short application of our concept of so-called «range frequencies» which we previously introduced in [12]. In particular, our current approach gives us a shorter proof of the results from [13], even when the necessary theory from [12] is counted.

Convex graphon parameters. In Section 3.4 we make a connection between graphon parameters that are convex on the space of graphons and cut distance compatible graphon parameters (see Theorem 3.17). A similar observation was made independently by Lee and Schülke [26] who derived from it that certain graphs are not norming/weakly norming (see Section 3.6.7).

Spectrum. In Section 3.5, we prove that a so-called «spectral quasiorder», which we define in Section 2.3.4 using the spectral properties of graphons, is a cut distance identifying graphon order (see Theorem 3.22). As was previously mentioned, this in particular strengthens the spectral part of the Chung–Graham–Wilson theorem.

Graph norms. Last, but most importantly, in Section 3.6 we study cut distance identifying and cut distance compatible graphon parameters of the form $t(H,\cdot)$, that is, densities of a fixed graph H. Such parameters are central in extremal graph theory. The famous «Sidorenko conjecture» (by Erdős–Simonovits [15] and independently by Sidorenko [35]) asserts that if H is a bipartite graph and W is a graphon of edge density p, then $t(H,W) \geq p^{e(H)}$, and the Forcing conjecture asserts that this inequality is strict unless $W \equiv p$. Kráľ, Martins, Pach and Wrochna [25] introduced a stronger concept. They say that H has the «step Sidorenko property» if for each graphon W and each finite partition $\mathcal{P}$ of Ω we have $t(H,W) \geq t\left(H, W^{\ltimes\mathcal{P}}\right)$, where $W^{\ltimes\mathcal{P}}$ is the stepping of W according to $\mathcal{P}$, that is, a graphon obtained by averaging W on the steps of $\mathcal{P}\times\mathcal{P}$. The «step forcing property» can be formulated similarly. These concepts are very much related to the main focus of

our paper. Indeed, as we show in Proposition 3.2, H has the step Sidorenko property if and only if $t(H,\cdot)$ is cut distance compatible. An analogous equivalence between the step forcing property and cut distance identifying parameters is the subject of Conjecture 3.3 where we expect that H has the step forcing property if and only if $t(H,\cdot)$ is cut distance identifying; let us note that the implication from right to left is trivial.

In Theorem 3.25 we prove that if for a connected graph H we have that $t(H,\cdot)$ is cut distance compatible, then H is weakly norming. This answers a question of Kráľ, Martins, Pach and Wrochna [25, Section 5]. The opposite implication is trivial. Combined with the equivalence between the step Sidorenko property and being cut distance compatible, we get a characterization of connected weakly norming graphs as those that have the step Sidorenko property.

Our another main result about graph norms, Theorem 3.28, states that for each norming graph H, the graphon parameter $t(H,\cdot)$ is cut distance identifying. Thus, by the trivial direction of Conjecture 3.3 mentioned above, we in particular obtain that each norming graph H has the step forcing property.

These implications are shown in Figure 3.2.

2. Preliminaries

In this section we introduce necessary notation and work up facts from real and functional analysis, probability theory and facts about graphons. Among these auxiliary results, Lemma 2.24 needs a good amount of preparation. We also recall results from [12] which we build on in this paper.

2.1. **General notation and basic analysis.** We write $\overset{\varepsilon}{\approx}$ for equality up to ε. For example, $1 \overset{0.2}{\approx} 1.1 \overset{0.2}{\approx} 1.3$. We write $\diamond$ for the symmetric difference of two sets. We write P_k for a path on k vertices and C_k for a cycle on k vertices.

If A and B are measure spaces then we say that a map $f : A \to B$ is an *almost-bijection* if there exist measure zero sets $A_0 \subset A$ and $B_0 \subset B$ such that $f_{\restriction A\setminus A_0}$ is a bijection between $A \setminus A_0$ and $B \setminus B_0$. Note that in (1.3), we could have worked with measure preserving almost-bijections φ instead.

2.1.1. *Moduli of convexity.* We recall the notion of the modulus of convexity. Suppose that Y is a linear space with a seminorm $\|\cdot\|_Y$. Then the *modulus of convexity of* $\|\cdot\|_Y$ is the function $\mathfrak{d}_Y : (0,+\infty) \to [0,+\infty)$ defined by

$$\mathfrak{d}_Y(\varepsilon) := \inf\left\{1 - \left\|\frac{x+y}{2}\right\|_Y \,:\, x,y \in Y, \|x-y\|_Y \geq \varepsilon, \|x\|_Y = \|y\|_Y = 1\right\}. \tag{2.1}$$

The seminorm $\|\cdot\|_Y$ is said to be *uniformly convex* if $\mathfrak{d}_Y(\varepsilon) > 0$ for each $\varepsilon > 0$.

For each $p \in (1, +\infty)$, the L^p-norm is known to be uniformly convex; the most streamlined argument to show this is due to Hanner [21].

Remark 2.1. *The modulus of convexity is a basic parameter in the theory of Banach spaces, but let us give some explanation for nonexperts. Any seminorm $\|\cdot\|_Y$ must satisfy the triangle inequality $\|x+y\|_Y \le \|x\|_Y + \|y\|_Y$, and this inequality is tight; we certainly have an equality if y is a nonnegative multiple of x. The modulus of convexity gives us a lower bound on the gap in the triangle inequality if we are guaranteed that x and y are far from being colinear.*

2.2. **Probability.** We write $\mathbf{E}$ and $\mathbf{P}$ for expectation and probability, respectively. We use two concentration inequalities, which we now recall. The first one is the Chernoff bound in the form that can be found in [1, Theorem A.1.16].

Lemma 2.2. *Suppose that $N \in \mathbb{N}$ and $X_1, X_2, \dots, X_N$ are mutually independent random variables with $\mathbf{E}[X_i] = 0$ and $|X_i| \le 1$ for each $i \in [N]$. Then for each $a > 0$ we have*

$$\mathbf{P}\left[\left|\sum X_i\right| > a\right] < 2\exp\left(\frac{-a^2}{2N}\right) .$$

Next, we give a tailored version of the Method of Bounded Differences [32]. For convenience, we give it in two versions, the former an easy consequence of the latter.

Lemma 2.3.

(a) Suppose that $r \in \mathbb{N}$, $a > 0$ and Z is a random variable on the probability space $[0,1]^r$. Suppose that for each two vectors $\mathbf{c}, \mathbf{c}' \in [0,1]^r$ that differ on at most one coordinate, we have $|Z(\mathbf{c}) - Z(\mathbf{c}')| \le a$. Then we have for each $d > 0$ that

$$\mathbf{P}\left[|Z - \mathbf{E}Z| > d\right] \le \exp\left(-\frac{2d^2}{ra^2}\right) .$$

(b) Suppose that $r \in \mathbb{N}$, $\mathbf{a} \in \mathbb{R}^r$ is a non-zero vector and Z is a random variable on the probability space $[0,1]^r$. Suppose that for each $i \in [r]$ and for each two vectors $\mathbf{c}, \mathbf{c}' \in [0,1]^r$ that differ only on the i-th coordinate, we have $|Z(\mathbf{c}) - Z(\mathbf{c}')| \le \mathbf{a}_i$. Then we have for each $d > 0$ that

$$\mathbf{P}\left[|Z - \mathbf{E}Z| > d\right] \le \exp\left(-\frac{2d^2}{\sum_i \mathbf{a}_i^2}\right) .$$

2.3. **Graphons.** Our notation is mostly standard, following [29]. Let us fix a standard Borel space Ω with an atomless probability measure ν. Let $\mathcal{W}$ denote the space of *kernels*, i.e. of all bounded symmetric measurable real functions defined on Ω^2. We always work modulo differences on null-sets. For example, if $U, W \in \mathcal{W}$ are such that $U \neq W$, then $\|U - W\|_1 > 0$. We write $\mathcal{W}_0 \subset \mathcal{W}$ for the space of all *graphons*, that is, symmetric

measurable functions from Ω^2 to $[0,1]$, and $\mathcal{W}^+ \subset \mathcal{W}$ for the space of all bounded symmetric measurable functions from Ω^2 to $[0,+\infty)$. The definitions of the cut norm and cut distance given in (1.2) and (1.3) extend to kernels verbatim. We write $\nu^{\otimes k}$ for the product measure on Ω^k.

For $p \in [0,1]$, we write $\mathcal{G}_p = \left\{W \in \mathcal{W}_0 : \int_x \int_y W(x,y) = p\right\}$ for all graphons with edge density p.

Remark 2.4. *It is a classical fact that there is a measure preserving bijection between each two standard atomless probability spaces. So, while most of the time we shall work with graphons on Ω^2, a graphon defined on a square of any other probability space as above can be represented (even though not in a unique way) on Ω^2.*

If $W\colon \Omega^2 \to [0,1]$ is a graphon and φ, ψ are two measure preserving bijections of Ω then we use the short notation $W^{\psi\varphi}$ for the graphon $W^{\psi\circ\varphi}$, i.e. $W^{\psi\varphi}(x,y) = W(\psi(\varphi(x)), \psi(\varphi(y))) = W^{\psi}(\varphi(x), \varphi(y)) = \left(W^{\psi}\right)^{\varphi}(x,y)$ for $(x,y) \in \Omega^2$.

The next well-known lemma says that one can define cut norm using just disjoint sets or squares, by losing just a constant factor.

Lemma 2.5 ([3, (7.1), (7.2)])**.** *Let $U, V \in \mathcal{W}_0$ be graphons with $\|U-V\|_\square = \delta$. Then there exist $A, B, X \subset \Omega$ with $A \cap B = \emptyset$ such that*

$$\left|\int_{A\times B} (U-V)\right| \geq \frac{\delta}{4}\,, \text{ and}$$
$$\left|\int_{X\times X} (U-V)\right| \geq \frac{\delta}{2}\,.$$

2.3.1. *Subgraphons.* Suppose that $W \in \mathcal{W}_0$ and $A, B \subset \Omega$ are disjoint sets of positive measures. Then we write $W[A,B]$ for the following bipartite graphon. The graphon is defined on $(A\cup B)^2$ where $A \cup B$ is equipped by the measure ν normalized by $\frac{1}{\nu(A\cup B)}$ (so that it is a probability measure), and it is given by $W[A,B](x,y) = W(x,y)$ if $(x,y) \in A\times B \cup B \times A$, and $W[A,B](x,y) = 0$ if $(x,y) \in A\times A \cup B \times B$.

Similarly, we write $W[A,A]$ for the following graphon. The graphon is defined on A^2 where A is equipped by the measure ν normalized by $\frac{1}{\nu(A)}$ (so that it is a probability measure), and it is given by $W[A,A](x,y) = W(x,y)$ for every $(x,y) \in A^2$.

2.3.2. *Homomorphism densities.* As usual, given a finite graph H on the vertex set $\{v_1, v_2, \ldots, v_n\}$ and a graphon W, we write

$$t(H,W) := \int_{x_1\in\Omega}\int_{x_2\in\Omega}\cdots\int_{x_n\in\Omega} \prod_{v_iv_j\in E(H)} W(x_i,x_j) \tag{2.2}$$

for the *homomorphism density* of H in W. Note that (2.2) extends to all $W \in \mathcal{W}$. The Counting Lemma allows to bound the difference between $t(H, W_1)$ and $t(H, W_2)$ in terms of the cut distance between W_1 and W_2.

Lemma 2.6 (Exercise 10.27 in [29])**.** *Suppose that H is a graph with m edges, and $W_1, W_2 \in \mathcal{W}$ satisfy $\|W_1\|_\infty, \|W_2\|_\infty \leq c$. Then $|t(H, W_1) - t(H, W_2)| \leq 4m \cdot c^{m-1} \cdot \delta_\square(W_1, W_2)$.*[d]

We call the quantity $t(P_2, W) = \int_x \int_y W(x, y)$ the *edge density of* W. Recall also that for $x \in \Omega$, we have the *degree of* x in W defined as $\deg_W(x) = \int_y W(x, y)$. Recall that measurability of W gives that $\deg_W(x)$ exists for almost each $x \in \Omega$. We say that W is *p-regular* if for almost every $x \in \Omega$, $\deg_W(x) = p$. Note that the notions of edge density, degree and regularity[e] extend to kernels. In particular, there exist non-trivial 0-regular kernels (for example the difference of the constant $\frac{1}{2}$-graphon and a complete balanced bipartite graphon).

We will need to generalize homomorphism densities to decorated graphs, as is done in [29, p. 120]. A *$\mathcal{W}$-decorated graph* is a finite simple graph H on the vertex set $\{v_1, v_2, \ldots, v_n\}$ in which each edge $v_i v_j \in E(H)$ is labelled by an element $W_{v_i v_j} \in \mathcal{W}$. We denote such a $\mathcal{W}$-decorated graph by (H, w), where $w = \left(W_{v_i v_j}\right)_{v_i v_j \in E(H)}$. For such a $\mathcal{W}$-decorated graph (H, w) we define

$$t(H, w) = \int_{x_1 \in \Omega} \int_{x_2 \in \Omega} \cdots \int_{x_n \in \Omega} \prod_{v_i v_j \in E(H)} W_{v_i v_j}(x_i, x_j) \, .$$

Analogous definitions can be formulated to introduce $\mathcal{W}_0$-decorated graphs and $\mathcal{W}^+$-decorated graphs.

2.3.3. *Tensor product.* Finally, we will need the definition of the tensor product of two graphons. Suppose that $U, V : \Omega^2 \to [0, 1]$ are two graphons. We define their *tensor product* as a $[0, 1]$-valued function $U \otimes V : \left(\Omega^2\right)^2 \to [0, 1]$ by $(U \otimes V)\left((x_1, x_2), (y_1, y_2)\right) = U(x_1, y_1) V(x_2, y_2)$.

Using Remark 2.4, we can think of $U \otimes V$ as a graphon in $\mathcal{W}_0$. Note that for every graph H we have

(2.3)

$$t(H, U \otimes V) = \int_{\Omega^{2v(H)}} \prod_{v_i v_j \in E(H)} U(x_i, x_j) \prod_{v_i v_j \in E(H)} V\left(x_{v(H)+i}, x_{v(H)+j}\right) = t(H, U) \cdot t(H, V) \, .$$

[d]Exercise 10.27 in [29] is stated with $\|W_1\|_\infty, \|W_2\|_\infty \leq 1$. To reduce our more general setting to that in [29], we divide the values of W_1 and of W_2 by c, hence decreasing $t(H, W_1)$ and $t(H, W_2)$ by a factor of c^m and $\delta_\square(W_1, W_2)$ by a factor of c.

[e]Here, by «regularity» we mean all the degrees having the same value, and not Szemerédi's concept of regularity.

One can deal with the generalised homomorphism density for decorations on a fixed finite graph H (where the tensor product $w_1 \otimes w_2$ is defined coordinatewise) in the same way and get that

$$t(H, w_1 \otimes w_2) = t(H, w_1) \cdot t(H, w_2) \; . \tag{2.4}$$

2.3.4. *Spectrum and the spectral quasiorder.* We recall the basic spectral theory for graphons, details and proofs can be found in [29, §7.5]. We shall work with the real Hilbert space $L^2(\Omega)$, inner product on which is denoted by $\langle \cdot, \cdot \rangle$. Given a graphon $W : \Omega^2 \to [0,1]$, we can associate to it an operator $T_W : L^2(\Omega) \to L^2(\Omega)$ given by

$$(T_W f)(x) := \int_y W(x,y) f(y) \; .$$

T_W is a Hilbert–Schmidt operator, and hence has a discrete spectrum of finitely or countably many non-zero eigenvalues (with possible multiplicities). All these eigenvalues are real, bounded in modulus by 1, and their only possible accumulation point is 0. Whenever there is no danger of confusion, we do not distinguish between the graphon W and the associated operator T_W in the following text. For a given graphon W we denote its eigenvalues, taking into account their multiplicities, by

$$\lambda_1^+(W) \geq \lambda_2^+(W) \geq \lambda_3^+(W) \geq \ldots \geq 0 \; ,$$
$$\lambda_1^-(W) \leq \lambda_2^-(W) \leq \lambda_3^-(W) \leq \ldots \leq 0 \; .$$

(We pad zeros if the spectrum has only finitely many positive or negative eigenvalues.)

We now introduce the notion of spectral quasiorder (it seems that this definition has not appeared in other literature). We write $W \overset{S}{\preceq} U$ if $\lambda_i^+(W) \leq \lambda_i^+(U)$ and $\lambda_i^-(W) \geq \lambda_i^-(U)$ for all $i = 1, 2, 3, \ldots$. Further we write $W \overset{S}{\prec} U$ if $W \overset{S}{\preceq} U$ and at least one of the above inequalities is strict. Then $\overset{S}{\preceq}$ is a quasiorder on $\mathcal{W}_0$, which we call the *spectral quasiorder*.

Recall that the eigenspaces are pairwise orthogonal. Recall also that (see e.g. [29, eq. (7.23)])

$$\|W\|_2^2 = \sum_i \lambda_i^+(W)^2 + \sum_i \lambda_i^-(W)^2 \; . \tag{2.5}$$

In Section 3.6 we shall use the following formula connecting eigenvalues and cycle densities. For any graphon W and for any $k \geq 3$, we have by [29, eq. (7.22)],

$$t(C_k, W) = \sum_i \lambda_i^+(W)^k + \sum_i \lambda_i^-(W)^k \; . \tag{2.6}$$

2.3.5. *The stepping operator.* Suppose that $W : \Omega^2 \to [0,1]$ is a graphon. We say that W is a *step graphon* if W is constant on each $\Omega_i \times \Omega_j$, for a suitable finite partition $\mathcal{P}$ of Ω, $\mathcal{P} = \{\Omega_1, \Omega_2, \ldots, \Omega_k\}$.

We recall the definition of the stepping operator.

Definition 2.7. Suppose that $\Gamma : \Omega^2 \to [0,1]$ is a graphon. For a finite partition $\mathcal{P}$ of Ω, $\mathcal{P} = \{\Omega_1, \Omega_2, \ldots, \Omega_k\}$, we define the graphon $\Gamma^{\ltimes\mathcal{P}}$ by setting it on the rectangle $\Omega_i \times \Omega_j$ to be the constant $\frac{1}{\nu^{\otimes 2}(\Omega_i \times \Omega_j)} \int_{\Omega_i} \int_{\Omega_j} \Gamma(x,y)$. We allow graphons to have not well-defined values on null sets which handles the cases $\nu(\Omega_i) = 0$ or $\nu(\Omega_j) = 0$.

In [29], the stepping is denoted by $\Gamma_{\mathcal{P}}$ rather than $\Gamma^{\ltimes\mathcal{P}}$.

We will need the following technical result which is Lemma 2.5 in [12].

Lemma 2.8. *Suppose that $\Gamma : \Omega^2 \to [0,1]$ is a graphon and ε is a positive number. Then there exists a finite partition $\mathcal{P}$ of Ω such that $\left\|\Gamma - \Gamma^{\ltimes\mathcal{P}}\right\|_1 < \varepsilon$.*

We call $\Gamma^{\ltimes\mathcal{P}}$ with properties as in Lemma 2.8 an *averaged L^1-approximation of Γ by a step-graphon for precision ε.*

The next easy lemma says that weak* convergence of graphons is preserved under stepping.

Lemma 2.9. *Suppose that $W : \Omega^2 \to [0,1]$ and $\big(U_n : \Omega^2 \to [0,1]\big)_n$ are graphons such that $U_n \xrightarrow{\text{w}^*} W$. Suppose that $\mathcal{R}$ is a finite partition of Ω. Then $(U_n)^{\ltimes\mathcal{R}} \xrightarrow{\text{w}^*} W^{\ltimes\mathcal{R}}$.*

Proof. We need to prove that for any $S, T \subset \Omega$ we have $\int_{S\times T} W^{\ltimes\mathcal{R}} = \lim_{n\to\infty} \int_{S\times T} (U_n)^{\ltimes\mathcal{R}}$. Let $\mathcal{R} = \{\Omega_1, \Omega_2, \ldots, \Omega_k\}$. Then we have

$$\int_{S\times T} W^{\ltimes\mathcal{R}} = \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{\nu(S\cap\Omega_i)\nu(T\cap\Omega_j)}{\nu(\Omega_i)\nu(\Omega_j)} \int_{\Omega_i\times\Omega_j} W\ , \tag{2.7}$$

and a similar decomposition formula for each U_n,

$$\int_{S\times T} (U_n)^{\ltimes\mathcal{R}} = \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{\nu(S\cap\Omega_i)\nu(T\cap\Omega_j)}{\nu(\Omega_i)\nu(\Omega_j)} \int_{\Omega_i\times\Omega_j} U_n\ . \tag{2.8}$$

Since $U_n \xrightarrow{\text{w}^*} W$, we have that for each i and j, $\int_{\Omega_i\times\Omega_j} W = \lim_{n\to\infty} \int_{\Omega_i\times\Omega_j} U_n$. Substituting this into (2.7) and (2.8), we get the claim. □

Finally, we say that a graphon U *refines* a graphon W, if W is a step graphon and for a suitable partition $\mathcal{P}$ of Ω we have $U^{\ltimes\mathcal{P}} = W$.

Recall that a partition $\mathcal{Q} = \{Q_1, Q_2, \ldots, Q_k\}$ of a finite measure space (Λ, λ) is an *equipartition* if $\lambda(Q_i) = \frac{1}{k}\cdot\lambda(\Lambda)$ for each $i = 1, \ldots, k$. The last lemma of this section asserts that given a step graphon $\Gamma^{\ltimes\mathcal{U}}$ such that the steps of $\mathcal{U}$ are refined by a certain finer equipartition $\mathcal{O}$, we can group cells of $\mathcal{O}$ into a coarser partition $\mathcal{R}$ but finer than $\mathcal{U}$ in such a way that $\Gamma^{\ltimes\mathcal{R}}$ is close to $\Gamma^{\ltimes\mathcal{U}}$ in L^1.

Lemma 2.10. *Suppose that $\Gamma : \Omega^2 \to [0,1]$ is a graphon. Let $\mathcal{O}$ and $\mathcal{U}$ be finite partitions of Ω such that $\mathcal{O}$ is an equipartition which refines $\mathcal{U}$, and the number b_U of cells of $\mathcal{O}$ in each part U of $\mathcal{U}$ is a multiple of a number $t \in \mathbb{N}$. Then for each $U \in \mathcal{U}$ there is a partition of the cells of $\mathcal{O}$ that lie in U into $d_U := \frac{b_U}{t}$ many groups $R_{U,1}, R_{U,2}, \ldots, R_{U,d_U}$ with t elements each such that the partition $\mathcal{R}$ of Ω defined as $\bigcup R_{U,\ell}$, where $U \in \mathcal{U}$, $\ell \in [d_U]$, satisfies $\left\|\Gamma^{\bowtie\mathcal{U}} - \Gamma^{\bowtie\mathcal{R}}\right\|_1 \leq 2t^{-1/4}$.*

Proof. For $A, B \subseteq \Omega$ define $m(A,B) = \frac{1}{\lambda(A)\lambda(B)} \int_{A\times B} \Gamma$. In the special case when $U, V \in \mathcal{U}$ we set $c_{U,V} = m(U,V)$. We equip the set of all t-partitions $\mathcal{R} = \{\bigcup R_{U,i}\}_{U\in\mathcal{U}, i\in[d_U]}$ that respect $\mathcal{O}$ with the uniform measure. For every $U, V \in \mathcal{U}$ and $i \in [d_U], j \in [d_V]$ we define a random variable $B_{U,V,i,j}$ as an indicator function of the event $|m(\bigcup R_{U,i}, \bigcup R_{V,j}) - c_{U,V}| > t^{-\frac{1}{4}}$. We will show that

$$\mathbf{E}\left[\sum B_{U,V,i,j}\right] \leq \left(\frac{|\mathcal{O}|}{t}\right)^2 \exp\left(-\sqrt{t}\right). \tag{2.9}$$

Once we have this then we can pick $\mathcal{R}$ that satisfies $\sum B_{U,V,i,j}(\mathcal{R}) \leq \left(\frac{|\mathcal{O}|}{t}\right)^2 \exp\left(-\sqrt{t}\right)$ and we have

$$\begin{aligned}\left\|\Gamma^{\bowtie\mathcal{U}} - \Gamma^{\bowtie\mathcal{R}}\right\|_1 &= \sum \left\|\left(\Gamma^{\bowtie\mathcal{U}} - \Gamma^{\bowtie\mathcal{R}}\right)_{\restriction \bigcup R_{U,i}\times\bigcup R_{V,j}}\right\|_1 \\ &\leq t^{-\frac{1}{4}} + \left(\frac{t}{|\mathcal{O}|}\right)^2 \left(\frac{|\mathcal{O}|}{t}\right)^2 \exp\left(-\sqrt{t}\right) \leq 2t^{-\frac{1}{4}}.\end{aligned}$$

This will finish the proof.

To show (2.9), it is clearly enough to show that $\mathbf{E}[B_{U,V,i,j}] \leq \exp\left(-\sqrt{t}\right)$ for each $U, V \in \mathcal{U}$ and $i \in [d_U], j \in [d_V]$. To prove that we use the method of bounded differences in three distinguished cases. Set $I_U = \{A \in \mathcal{O} : A \subseteq U\}$ for every $U \in \mathcal{U}$.

Case I: We have $U \neq V$.

Consider the space $\mathcal{X}_{U,V}$ of all pairs of unions of t-subsets of I_U and I_V endowed with the uniform measure, i.e., elements of $\mathcal{X}_{U,V}$ are of the form $(\bigcup I, \bigcup J)$ for some $I \subseteq I_U$ and $J \subseteq I_V$ such that $|I| = |J| = t$. It is easy to see that

$$\mathbf{E}[B_{U,V,i,j}] = \mathbf{P}\left[|m - c_{U,V}| > t^{-\frac{1}{4}}\right],$$

where we abuse the notation and view m as a random variable on $\mathcal{X}_{U,V}$. We can represent $\mathcal{X}_{U,V}$ as vectors in $[0,1]^{b_U} \times [0,1]^{b_V}$, endowed with the product Lebesgue measure, using the following procedure. Let $v \in [0,1]^{b_U}$ be a random vector taken with respect to the Lebesgue measure on $[0,1]^{b_U}$. Note that almost surely, the coordinates of v are pairwise distinct. Let $I \subset I_U$ be the set of the t indices of v with the smallest values. Note that this procedure is «Lipschitz» in the sense that if v and v' are two vectors that differ in only one coordinate then I and the corresponding I' differ in at most one index. Similarly

J can be selected from an independent random vector $w \in [0,1]^{b_V}$. Thus, we can think of m as a random variable on $[0,1]^{b_U} \times [0,1]^{b_V}$ with expectation $c_{U,V}$. Recall that the range of Γ is in $[0,1]$. Observe that if (v,w) changes in at most one index, then m changes by at most $\frac{1}{b_U}$ or $\frac{1}{b_V}$. Thus, Lemma 2.3(b) implies that

$$\mathbf{P}\left[|m - c_{U,V}| > d\right] \leq \exp\left(-\frac{2d^2}{b_U \cdot \left(\frac{1}{b_U}\right)^2 + b_V \cdot \left(\frac{1}{b_V}\right)^2}\right) \leq \exp\left(-d^2 t\right)$$

for every $d > 0$. The statement follows by taking $d := t^{-1/4}$.

Case II: We have $U = V$ and $i = j$.

First, note that if $b_U = t$, then there is nothing to prove, so we assume that $b_U \geq 2t$. Consider the space $\mathcal{X}_U$ of all unions of t-subsets of I_U endowed with the uniform measure, i.e., elements of $\mathcal{X}_U$ are of the form $\bigcup I$ for some $I \subseteq I_U$ such that $|I| = t$. Again, it is easy to see that

$$\mathbf{E}\left[B_{U,U,i,i}\right] = \mathbf{P}\left[|m - c_{U,U}| > t^{-\frac{1}{4}}\right],$$

where we abuse the notation and view m as a random variable on $\mathcal{X}_U$. We can represent $\mathcal{X}_U$ as vectors in $[0,1]^{b_U}$ as in Case I and use Lemma 2.3(b) to get

$$\mathbf{P}\left[|m - c_{U,U}| > d\right] \leq \exp\left(-\frac{2d^2}{b_U \cdot \left(\frac{2}{b_U}\right)^2}\right) \leq \exp\left(-d^2 t\right)$$

for every $d > 0$. The statement follows by taking $d := t^{-1/4}$.

Case III: We have $U = V$ and $i \neq j$.

Again, note that if $b_U = t$, then there is nothing to prove, so we assume that $b_U \geq 2t$. Consider the space $\mathcal{Y}_U$ of all pairs of unions of t-subsets of I_U that are disjoint endowed with the uniform measure, i.e., elements of $\mathcal{Y}_U$ are of the form $(\bigcup I, \bigcup I')$ for some $I, I' \subseteq I_U$ such that $|I| = |I'| = t$ and $I \cap I' = \emptyset$. Again, it is easy to see that

$$\mathbf{E}\left[B_{U,U,i,j}\right] = \mathbf{P}\left[|m - c_{U,U}| > t^{-\frac{1}{4}}\right],$$

where we view m as a random variable on $\mathcal{Y}_U$. We can represent $\mathcal{Y}_U$ as vectors in $[0,1]^{b_U}$ endowed with the Lebesgue measure. Namely, if $v \in [0,1]^{b_U}$, then we set $I \subseteq I_U$ to be the set of the t indices of v with the smallest values and $I' \subseteq I_U$ to be the set of the t indices of v with the biggest values. Observe that if v changes in at most one index, then m changes by at most $\frac{2}{b_U}$. Thus, Lemma 2.3(b) implies that

$$\mathbf{P}\left[|m - c_{U,V}| > d\right] \leq \exp\left(-\frac{2d^2}{b_U \cdot \left(\frac{2}{b_U}\right)^2}\right) \leq \exp\left(-d^2 t\right)$$

for every $d > 0$. The statement follows by taking $d := t^{-1/4}$. □

2.4. **Norms defined by graphs.** In this section we briefly recall how homomorphism densities $t(H, \cdot)$ induce norms on the space of graphons. More details can be found in [29, §14.1].

We now introduce the (semi)norming and weakly norming graphs and graphs with the (weak) Hölder property, concepts first introduced in [22]. We say that a graph H is *(semi)norming,* if the function

$$\|W\|_H := |t(H, W)|^{1/e(H)} \tag{2.10}$$

is a (semi)norm on $\mathcal{W}$. This means that we require that $\|\cdot\|_H$ is subadditive and homogeneous (i.e., $\|c \cdot W\|_H = |c| \cdot \|W\|_H$ for each $c \in \mathbb{R}$), and in the case of norming graphs we moreover assume that there does not exist a kernel W that is not identically zero, but $t(H, W) = 0$.

We list several properties of (semi)norming graphs.

Fact 2.11.

(a) [Exercise 14.7 in [29]*] Each seminorming graph is bipartite.*
(b) No tree is norming.

Proof of Fact 2.11(b). Theorem 2.10(ii) from [22] implies that a tree F is not norming, unless F is a star, say $K_{1,m}$. So, it remains to argue that a star $K_{1,m}$ cannot be norming. To this end, observe that for any 0-regular kernel U, we have that $\|U\|_{K_{1,m}} = 0$, while we can have $U \neq 0$.[f] □

We say that a graph H is *weakly norming,* if the function $\|W\|_H := t(H, |W|)^{1/e(H)}$ is a seminorm on $\mathcal{W}$. Note that by [29, Exercise 14.7 (a)], every weakly norming graph is bipartite. It follows that in this case the seminorm above is also a norm. Indeed, if $U \in \mathcal{W}$ is not zero almost everywhere, then U contains a Lebesgue point (a, b) such that $|U(a, b)| > 0$ and, denoting the bipartition of H as $\{u_1, \ldots, u_k\} \sqcup \{v_1, \ldots, v_\ell\}$, we can write

$$\|U\|_H^{e(H)} = t(H, |U|) = \int_{x_1} \cdots \int_{x_k} \int_{y_1} \cdots \int_{y_\ell} \prod_{u_i v_j \in E(H)} |U(x_i, y_j)| \, .$$

Now, if all x_i's are restricted to a sufficiently small neighborhood of a and all y_j's are in a small neighborhood of b, then the fact that (a, b) is a Lebesgue point at which $|U(a, b)| > 0$ tells us that $\prod_{u_i v_j \in E(H)} |U(x_i, y_j)|$ is positive, say at least 50% of the time. We conclude that $\|U\|_H^{e(H)} > 0$.

Since the homomorphism density $t(H, \cdot)$ satisfies $|t(H, cW)| = |c|^{e(H)} |t(H, W)|$, the only nontrivial requirement in the definition of seminorming or weakly norming graphs, respectively, is the subadditivity of the homomorphism density defined on the space $\mathcal{W}$ of

[f] Recall our remark about 0-regular kernels from Section 2.3.2.

kernels, or on the space $\mathcal{W}_0$ of graphons, respectively. In other words we ask that for each $W_1, W_2 \in \mathcal{W}$, or for each $W_1, W_2 \in \mathcal{W}_0$, respectively, we have

$$\|W_1 + W_2\|_H \leq \|W_1\|_H + \|W_2\|_H \ .$$

Complete bipartite graphs (in particular, stars), complete balanced bipartite graphs without a perfect matching, even cycles, and hypercubes are some of the most prominent examples of weakly norming graphs. All these classes fall within a much wider family of so-called «reflection graphs» which were shown to be weakly norming by Conlon and Lee [7].

A graph H has the *Hölder property*, if for every $\mathcal{W}$-decoration $w = (W_e)_{e\in E(H)}$ of H we have

$$t(H, w)^{e(H)} \leq \prod_{e\in E(H)} t(H, W_e) \ . \tag{2.11}$$

The graph H has the *weak Hölder property*, if (2.11) holds for every $\mathcal{W}_0$-decoration w of H.

Our next lemma says that for the weak Hölder property it is enough to test (2.11) over a slightly different set of decorations of H.

Lemma 2.12. *Suppose that H is graph which satisfies (2.11) for every $\mathcal{W}^+$-decoration $u = (U_e)_{e\in E(H)}$ with $t(H, U_e) = 1$ for every $e \in E(H)$. Then H has the weak Hölder property.*

Proof. Suppose that we need to check (2.11) for a given $\mathcal{W}_0$-decoration $w = (W_e)_{e\in E(H)}$ (or, actually, we will suppose, somewhat more generally, that w is a $\mathcal{W}^+$-decoration). Firstly, suppose that $t(H, W_e) > 0$ for every $e \in E(H)$. Then we define a $\mathcal{W}^+$-decoration $u = (U_e)_{e\in E(H)}$ by $U_e := t\,(H, W_e)^{-1/e(H)} \cdot W_e$. The decoration u satisfies (2.11) by the assumption of the lemma. Hence,

$$\left(\prod_{e\in E(H)} t\,(H, W_e)^{-1/e(H)} \ \cdot \ t(H, w)\right)^{e(H)} = t(H, u)^{e(H)} \leq \prod_{e\in E(H)} t(H, U_e) = \prod_{e\in E(H)} 1 = 1,$$

which is indeed equivalent to (2.11).

Secondly, suppose that w is a general $\mathcal{W}_0$-decoration. For $\alpha > 0$, let $w_\alpha = (W_{\alpha,e})_{e\in E(H)}$ be a $\mathcal{W}^+$-decoration where we add the constant α to each component, $W_{\alpha,e} = W_e + \alpha$. By the «firstly» part, we have $t(H, w_\alpha)^{e(H)} \leq \prod_{e\in E(H)} t(H, W_{\alpha,e})$. Since the graphons $W_{\alpha,e}$ converge to W_e in the cut norm as alpha tends to 0, and since the quantities $t(H, x)^{e(H)}$ and $\prod_{e\in E(H)} t\,(H, x_e)$ are cut norm continuous on the space of $\mathcal{W}^+$-decorations, we obtain the desired $t(H, w)^{e(H)} \leq \prod_{e\in E(H)} t(H, W_e)$. □

Literature concerning norming and weakly norming graphs seems to be imprecise when it comes to disconnected graphs. These impressions penetrated into previous versions of

this paper (up to version 3 at arXiv). The following recent result will help us to rescue the situation.

Fact 2.13 ([18])**.**

(a) Suppose that G is a disconnected norming graph. Then there exists a connected norming graph F such that each component of G is either isomorphic to F or is a vertex.

(b) Suppose that H is a disconnected weakly norming graph. Then there exists a connected weakly norming graph J such that each component of H is either isomorphic to J or is a vertex.

One of our main results in Section 3.6, Theorem 3.25, connects weakly norming graphs with the concept of the step Sidorenko property introduced below. To prove Theorem 3.25, we shall need the following characterization of weakly norming graphs from [22].

Theorem 2.14 (Theorem 2.8 in [22])**.** *A graph is seminorming if and only if it has the Hölder property. It is weakly norming if and only if it has the weak Hölder property.*

Another main result in Section 3.6, Theorem 3.28, connects norming graphs with the related step forcing property. To prove Theorem 3.28, we need Proposition 2.15 below. This proposition was proven in Section 5.2 of [7] in the discussion after equations (12) and (13). Formally, it was proven there only for the case $H = C_4$, but the authors noted that the same approach works in general. In particular, a straightforward generalization of their proof yields the following proposition.

Proposition 2.15 (Section 5.2 in [7])**.** *Suppose that H is a norming graph. Then for each kernel $W \in \mathcal{W}$ we have $t(H, W) \geq (\|W\|_\square)^{e(H)}$.*

2.4.1. *Moduli of convexity of seminorming graphs.* For every seminorming graph H, let $\|\cdot\|_H$ be the corresponding seminorm on $\mathcal{W}$. Hatami determined in Theorem 2.16 of [22] the moduli of convexity of the norms induced by connected seminorming graphs (note that Hatami wrongly claimed his result for disconnected graphs as well, see [18] for a discussion):

Theorem 2.16. *Let $m \in \mathbb{N}$ and let $\mathfrak{d}_m$ be the modulus of convexity of the L^m norm. There exists a constant $C_m > 0$ such that for any connected seminorming graph H with m edges we have the following. The modulus $\mathfrak{d}_H$ of convexity of the seminorm $\|\cdot\|_H$ satisfies*

$$C_m \mathfrak{d}_m \leq \mathfrak{d}_H \leq \mathfrak{d}_m \,.$$

In Section 2.1.1 we mentioned that the L^p-norm is uniformly convex, for $p \in (1, +\infty)$. Thus, when H is a connected seminorming graph which is not an edge, the seminorm $\|\cdot\|_H$ is uniformly convex.

2.5. **Topologies on $\mathcal{W}_0$.** There are several natural topologies on $\mathcal{W}_0$ and $\mathcal{W}$. The $\|\cdot\|_\infty$ topology inherited from the normed space $L^\infty(\Omega^2)$, the $\|\cdot\|_1$ topology inherited from the normed space $L^1(\Omega^2)$, the topology given by the $\|\cdot\|_\square$ norm, and the weak* topology inherited from the weak* topology of the dual Banach space $L^\infty(\Omega^2)$. Note that $\mathcal{W}_0$ is closed in both $L^1(\Omega^2)$ and $L^\infty(\Omega^2)$. We write $d_1(\cdot,\cdot)$ for the distance derived from the $\|\cdot\|_1$ norm and $d_\infty(\cdot,\cdot)$ for the distance derived from the $\|\cdot\|_\infty$ norm. The weak* topology of the dual Banach space $L^\infty(\Omega^2)$ is generated by elements of its predual $L^1(\Omega^2)$. That means that the weak* topology on $L^\infty(\Omega^2)$ is the smallest topology on $L^\infty(\Omega^2)$ such that all functionals of the form $g \in L^\infty(\Omega^2) \mapsto \int_{\Omega^2} fg$, where $f \in L^1(\Omega^2)$ is fixed, are continuous. Recall that by the Banach–Alaoglu theorem, $\mathcal{W}_0$ equipped with the weak* topology is compact. Recall also that the weak* topology on $\mathcal{W}_0$ is metrizable. We shall denote by $d_{\mathrm{w}^*}(\cdot,\cdot)$ any metric compatible with this topology. For example, we can take some countable family $\{A_n\}_{n\in\mathbb{N}}$ of measurable subsets of Ω which forms a dense set in the sigma-algebra of Ω, and define $d_{\mathrm{w}^*}(U,W) := \sum_{n,k\in\mathbb{N}} 2^{-(n+k)} \left|\int_{A_n\times A_k}(U-W)\right|$.

2.6. **Envelopes, the structuredness order, and the range and degree frequencies.** Here, we recall the key concepts from [12].

Suppose that $\Gamma_1, \Gamma_2, \Gamma_3, \ldots \in \mathcal{W}_0$ are graphons. Recall that

$$\mathbf{ACC}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots) = \bigcup_{\pi_1\pi_2,\pi_3,\ldots:\Omega\to\Omega\,\mathrm{m.p.b.}} \text{weak* accumulation points of} \Gamma_1^{\pi_1}, \Gamma_2^{\pi_2}, \Gamma_3^{\pi_3}, \ldots .$$

Similarly, we define

$$\mathbf{LIM}_{\mathrm{w}*}(\Gamma_1,\Gamma_2,\Gamma_3,\ldots) = \bigcup_{\pi_1\pi_2,\pi_3,\ldots:\Omega\to\Omega\,\mathrm{m.p.b.}} \text{weak* limit points of} \Gamma_1^{\pi_1}, \Gamma_2^{\pi_2}, \Gamma_3^{\pi_3}, \ldots .$$

For every graphon $W \in \mathcal{W}_0$ we define the set $\langle W\rangle := \mathbf{LIM}_{\mathrm{w}*}(W,W,W,\ldots)$. That is, a graphon $U \in \mathcal{W}_0$ belongs to $\langle W\rangle$ if and only if there are measure preserving bijections $\pi_1, \pi_2, \pi_3, \ldots$ of Ω such that the sequence $W^{\pi_1}, W^{\pi_2}, W^{\pi_3}, \ldots$ converges to U in the weak* topology. We call the set $\langle W\rangle$ the *envelope* of W.

We say that a graphon U is *at most as structured* as a graphon W if $\langle U\rangle \subset \langle W\rangle$. We write $U \preceq W$ in this case. We write $U \prec W$ if $U \preceq W$ but it does not hold that $W \preceq U$.

Fact 2.17 (Lemma 4.2(b) in [12])**.** *We have $W^{\ltimes\mathcal{P}} \in \langle W\rangle$ for every graphon W and every finite partition $\mathcal{P}$ of Ω.*

The «structurdness order» $\preceq$ fails to be antisymmetric on the space of graphons equipped with the cut norm. Indeed, if A and B are two graphons, $A = B^\pi$ for some measure preserving bijection π, then we can have $A \neq B$, $A \preceq B$ and $B \preceq A$. However, the structurdness order is a proper order on the factorspace given by the cut distance. This is stated below.

Fact 2.18 (Corollary 4.24 in [12]). *Suppose that $U \preceq W$. Then $W \preceq U$ if and only if $\delta_\square(U, W) = 0$.*

The next fact tells us that the relation $\preceq$ is (topogically) closed with respect to the cut distance topology.

Fact 2.19 (Lemma 8.1 in [12]). *Suppose that $A, A_1, A_2, A_3, \dots$ and $B, B_1, B_2, B_3, \dots$ are graphons, for each $n \in \mathbb{N}$ we have $B_n \preceq A_n$, $A_n \xrightarrow{\delta_\square} A$ and $B_n \xrightarrow{\delta_\square} B$. Then $B \preceq A$.*

It follows directly from the definition of the weak* topology that the edge density of a weak* limit of a sequence of graphons equals to the limit of the edge densities of the graphons in the sequence. Thus, we obtain the following.

Fact 2.20. *If two graphons have different edge densities then they are incomparable in the structuredness order.*

2.6.1. *The range frequency $\mathbf{\Phi}_W$, the degree frequency $\mathbf{\Upsilon}_W$, and the flatness order.* Given a graphon $W : \Omega^2 \to [0, 1]$, we can define a pushforward probability measure on $[0, 1]$ by

$$\mathbf{\Phi}_W(A) := \nu^{\otimes 2}\left(W^{-1}(A)\right), \tag{2.12}$$

for every Borel measurable set $A \subset [0, 1]$. The measure $\mathbf{\Phi}_W$ gives us the distribution of the values of W. In [12], $\mathbf{\Phi}_W$ is called the *range frequencies of* W. Similarly, we can take the pushforward measure of the degrees, which is called the *degree frequencies of* W,

$$\mathbf{\Upsilon}_W(A) := \nu\left(\deg_W^{-1}(A)\right), \tag{2.13}$$

for every Borel measurable set $A \subset [0, 1]$. The measures $\mathbf{\Phi}_W$ and $\mathbf{\Upsilon}_W$ provide substantial information about the graphon W. It is therefore natural to ask how these measures relate with respect to the structuredness order. To this end the following «flatness relation» on measures is introduced.

Definition 2.21. Suppose that Λ_1 and Λ_2 are two finite measures on $[0, 1]$. We say that Λ_1 is *at least as flat* as Λ_2 if there exists a finite measure Λ on $[0, 1]^2$ such that Λ_1 is the marginal of Λ on the first coordinate, Λ_2 is the marginal of Λ on the second coordinate, and for each $D \subset [0, 1]$ we have

$$\int_{(x,y)\in D\times[0,1]} x \,\mathrm{d}\Lambda = \int_{(x,y)\in D\times[0,1]} y \,\mathrm{d}\Lambda\,. \tag{2.14}$$

We say that Λ_1 is *strictly flatter* than Λ_2 if Λ_1 is at least as flat as Λ_2 and $\Lambda_1 \neq \Lambda_2$.

An illustration of the concept of flatness order is given in Figure 2.1. We can now state the main result of Section 4.4 of [12].

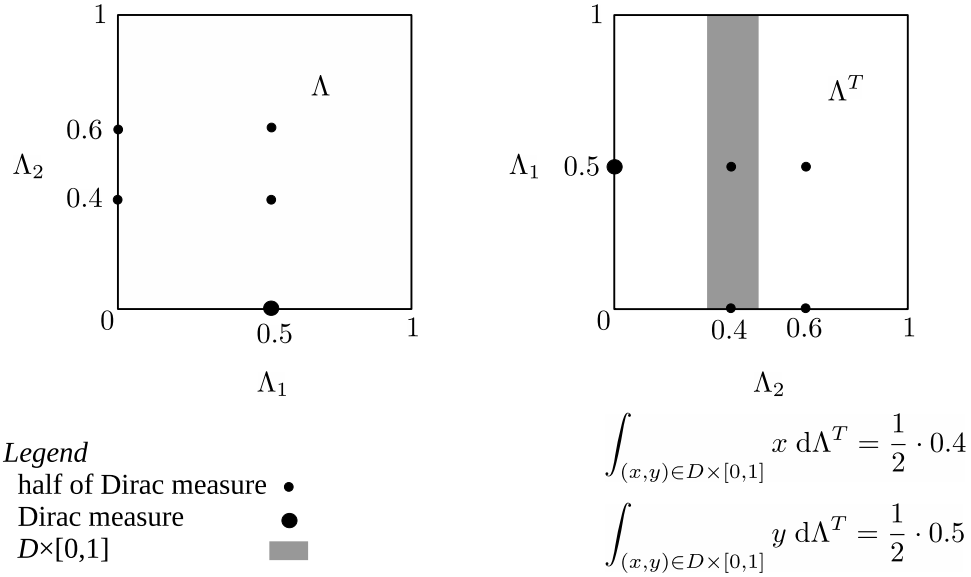


FIGURE 2.1. An illustration of the concept of flatness order. Let Λ_1 be the Dirac measure on 0.5. Let Λ_2 be a half of the sum of the Dirac measures on 0.4 and on 0.6. The left-hand side shows a measure Λ which witnesses that Λ_1 is at least as flat as Λ_2. The right-hand side gives an example of a set D which proves that the transposed measure Λ^T fails to witness that Λ_2 is at least as flat as Λ_1 (and indeed no such witness exists).

Proposition 2.22. *Suppose that we have two graphons $U \preceq W$. Then the measure $\mathbf{\Phi}_U$ is at least as flat as the measure $\mathbf{\Phi}_W$. Similarly, the measure $\mathbf{\Upsilon}_U$ is at least as flat as the measure $\mathbf{\Upsilon}_W$. Lastly, if $U \prec W$ then $\mathbf{\Phi}_U$ is strictly flatter than $\mathbf{\Phi}_W$.*

2.7. **Approximating a graphon by versions of a more structured graphon.** In this section we state and prove Lemma 2.24, which is the key technical step for one of our main results, Theorem 3.28. Since the proof of Lemma 2.24 is quite complex, we first state a simplified version in Lemma 2.23. We use this simplification to explain some key features of the proof and motivate some further notation.

Lemma 2.23 (Simplified version of Proposition 2.31)**.** *Suppose that $U, V \in \mathcal{W}_0$ and that $\mathcal{R}$ is a finite partition of Ω such that $V = U^{\ltimes\mathcal{R}}$. Then for each $\varepsilon > 0$ we can find a number N and measure preserving bijections $(\phi_i : \Omega \to \Omega)_{i=1}^N$ such that $\left\| V - \frac{\sum_{i=1}^N U^{\phi_i}}{N} \right\|_1 < \varepsilon$.*

While there are several possible proofs, the one which we need (and which we extend to prove Lemma 2.24) uses the probabilistic method. Let us sketch it now. Let $\mathcal{R} = \{\Omega_1, \Omega_2, \ldots, \Omega_k\}$. Suppose that we are given ε. We now take a large number s, and $N \gg s$. We partition each set Ω_j into sets $[\![\Omega_j]\!]^s_1 \sqcup [\![\Omega_j]\!]^s_2 \sqcup \ldots \sqcup [\![\Omega_j]\!]^s_s$ of the same measure

(Definition 2.26 below introduces this formally). For each $j \in [k]$, we can randomly shuffle the sets $\{[\![\Omega_j]\!]^s_1, [\![\Omega_j]\!]^s_2, \ldots, [\![\Omega_j]\!]^s_s\}$. Putting these shuffles together, we obtain a random measure preserving bijection $\phi : \Omega \to \Omega$ with the property that

$$\Omega_j = \phi(\Omega_j) \text{ for each } j \in [k], \tag{2.15}$$

and hence a random version U^ϕ of U (Definition 2.27 below introduces this formally). Such a random version typically blurs whatever structure there was in each rectangle $\Omega_j \times \Omega_\ell$ $(j, \ell \in [k])$.[g] Hence, a rather straightforward application of the Law of Large Numbers gives that with high probability, the mean of independent random versions $U^{\phi_1}, U^{\phi_2} \ldots, U^{\phi_N}$ approximates V in the L^1-distance.

Our actual Lemma 2.24 strengthens Lemma 2.23 in two ways. Firstly, it assumes that $V \preceq U$ which is more general than $V = U^{\ltimes\mathcal{R}}$ for some finite partition $\mathcal{R}$. This represents only a minor complication in the proof as these properties are almost the same (see for example Lemma 2.28). So, we describe the second (and main) strengthening under the notationally more convenient assumption that $V = U^{\ltimes\mathcal{R}}$ for $\mathcal{R} = \{\Omega_1, \Omega_2, \ldots, \Omega_k\}$. In addition to the approximation property as in Lemma 2.23, we require that many of the pairs $U^{\phi_{2i-1}}$ and $U^{\phi_{2i}}$ are at least as far apart in the cut norm distance, as a constant multiple of $\delta_\square(U, V)$. (Note that this statement is void when $\delta_\square(U, V) = 0$. Indeed in that case there is no way we could hope for such a property.) Let us explain why we expect this to occur for two independent random versions $U^{\phi_{2i-1}}$ and $U^{\phi_{2i}}$ with high probability. To this end, let us fix $S, T \subset \Omega$ for which we have $\left|\int_{S\times T}(U - V)\right| > \frac{\delta_\square(U,V)}{2}$. Without loss of generality, let us assume that $\int_{S\times T}(U - V) > \frac{\delta_\square(U,V)}{2}$. Let us now look at $U^{\phi_{2i-1}}$. We clearly have $\int_{S\times T} U = \int_{\phi^{-1}_{2i-1}(S)\times\phi^{-1}_{2i-1}(T)} U^{\phi_{2i-1}}$. Since V is a step-function on $\mathcal{R}\times\mathcal{R}$, for any measure preserving bijection ϕ satisfying (2.15) we have $\int_{S\times T} V = \int_{\phi^{-1}(S)\times\phi^{-1}(T)} V$, and hence

$$\int_{\phi^{-1}_{2i-1}(S)\times\phi^{-1}_{2i-1}(T)} (U^{\phi_{2i-1}} - V) > \frac{\delta_\square(U, V)}{2}\,. \tag{2.16}$$

Let us now look at $\int_{\phi^{-1}_{2i-1}(S)\times\phi^{-1}_{2i-1}(T)} U^{\phi_{2i}}$. As we said earlier (recall Footnote [g]), the version $U^{\phi_{2i}}$ with high probability blurs any structure on each rectangle $\Omega_j \times \Omega_\ell$. Thus, with high probability, $\int_{\phi^{-1}_{2i-1}(S)\times\phi^{-1}_{2i-1}(T)} U^{\phi_{2i}} \approx \int_{\phi^{-1}_{2i-1}(S)\times\phi^{-1}_{2i-1}(T)} V$. Combined with (2.16), this proves that $U^{\phi_{2i-1}}$ and $U^{\phi_{2i}}$ are far apart in the cut norm distance. In the actual proof, we need to deal with several technical difficulties.

[g]For example, if $\Omega_j \times \Omega_\ell$ consisted of two parts of equal measure, U being 0.1 on one part and 0.7 on the other, then U^ϕ on $\Omega_j \times \Omega_\ell$ will consist with high probability of a checkerboard with random-like alternation of 0.1's and 0.7's. In particular, for each fixed $X \subset \Omega_j \times \Omega_\ell$, we will have with high probability that $\int_X U^\phi \approx \frac{1}{2}(0.1 + 0.7)\nu^{\otimes 2}(X)$.

Lemma 2.24. *Suppose that $U, V \in \mathcal{W}_0$ and $V \prec U$. Then for any $\varepsilon > 0$ we can find an even number N and measure preserving bijections $(\phi_i : \Omega \to \Omega)_{i=1}^N$ such that*

$$\left\| V - \frac{\sum_{i=1}^N U^{\phi_i}}{N} \right\|_1 < \varepsilon \,. \tag{2.17}$$

Moreover, for at least half of the indices $i \in \left\{1, 2, \ldots, \frac{N}{2}\right\}$ we have

$$\left\| U^{\phi_{2i-1}} - U^{\phi_{2i}} \right\|_\square > \frac{\delta_\square(U, V)}{32} \,. \tag{2.18}$$

Remark 2.25. *Obviously, by rescaling, Lemma 2.24 can be extended to $U, V \in \mathcal{W}^+$.*

The rest of this section is devoted to proving Lemma 2.24. The key construction in the proof of Lemma 2.24 is very similar to the proof of Lemma 9 in [13] (however, our proof is substantially more complex due to the additional property (2.18)). We borrow the following two definitions from [13].

Definition 2.26. Given a set $A \subset [0,1]$ of positive measure and a number $s \in \mathbb{N}$, we can consider a partition $A = [\![A]\!]_1^s \sqcup [\![A]\!]_2^s \sqcup \ldots \sqcup [\![A]\!]_s^s$, where each set $[\![A]\!]_i^s$ has measure $\frac{\lambda(A)}{s}$ and for each $1 \le i < j \le s$, the set $[\![A]\!]_i^s$ is entirely to the left of $[\![A]\!]_j^s$. These conditions define the partition $A = [\![A]\!]_1^s \sqcup [\![A]\!]_2^s \sqcup \ldots \sqcup [\![A]\!]_s^s$ uniquely, up to null sets. For each $i, j \in [s]$ there is a natural, uniquely defined (up to null sets), measure preserving almost-bijection $\chi_{i,j}^{A,s} : [\![A]\!]_i^s \to [\![A]\!]_j^s$ which preserves the order on the real line.

We can now give a definition of the model of random versions of a given graphon which we need. An illustration is given in Figure 2.2.

Definition 2.27. Suppose that $\Gamma : [0,1]^2 \to [0,1]$ is a graphon. For a finite partition $\mathcal{R} = \{\Omega_1, \Omega_2, \ldots, \Omega_k\}$ of $[0,1]$ and for $s \in \mathbb{N}$, we define a discrete probability distribution $\mathbb{W}(\Gamma, \mathcal{R}, s)$ on versions of Γ as follows. We take $\pi_1, \ldots, \pi_k : [s] \to [s]$ independent uniformly random permutations. After these are fixed, we define a sample $W \sim \mathbb{W}(\Gamma, \mathcal{R}, s)$ by

$$W(x,y) = \Gamma\left(\chi_{p,\pi_i(p)}^{\Omega_i,s}(x), \chi_{q,\pi_j(q)}^{\Omega_j,s}(y)\right) \quad \text{when } x \in [\![\Omega_i]\!]_p^s,\ y \in [\![\Omega_j]\!]_q^s,\ i,j \in [k],\ p,q \in [s] \,.$$

This defines the sample $W : [0,1]^2 \to [0,1]$ uniquely up to null sets, and thus defines the whole distribution $\mathbb{W}(\Gamma, \mathcal{R}, s)$. Observe that $\mathbb{W}(\Gamma, \mathcal{R}, s)$ is supported on (some) versions of Γ. We call the sets $[\![\Omega_j]\!]_q^s$ *stripes*.

We shall also need the following technical lemmas.

Lemma 2.28. *Let $V \preceq U$ be two graphons on Ω^2 and $\varepsilon > 0$. Then there is a measure preserving bijection $\varphi : \Omega \to \Omega$ and a finite partition $\mathcal{R}$ of Ω such that*

$$\left\| V - (U^\varphi)^{\ltimes \mathcal{R}} \right\|_1 < \varepsilon.$$

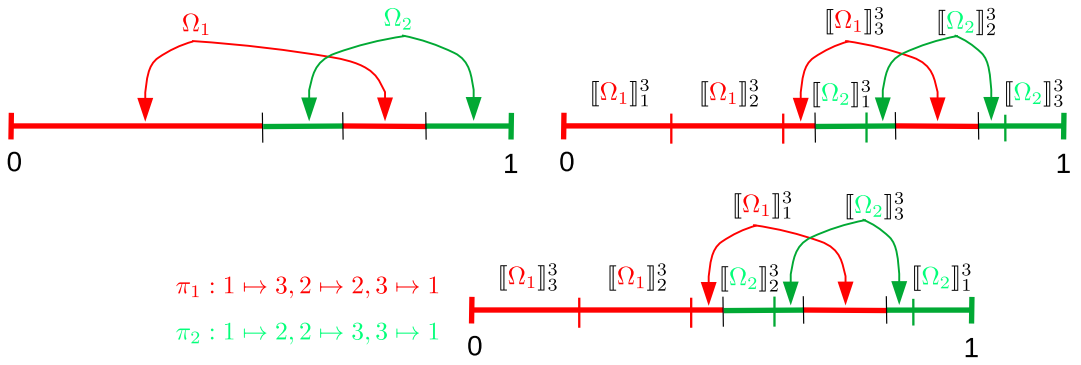


FIGURE 2.2. An illustration to Definitions 2.26 and 2.27. A sample partition $\mathcal{R} = \{\Omega_1, \Omega_2\}$ of $[0,1]$ in the top left figure. To create $\mathbb{W}(\Gamma, \mathcal{R}, s = 3)$, the unit interval is partitioned as in the top right figure. Finally, for randomly chosen permutations π_1, π_2, the unit interval is effectively reshuffled as in the bottom figure.

Proof. Use Lemma 2.8 to find a finite partition $\mathcal{R}$ such that $\left\|V - V^{\bowtie\mathcal{R}}\right\|_1 < \frac{\varepsilon}{2}$. By the definition of $V \preceq U$ we may find a sequence $\{U^{\varphi_i}\}_{i\in\mathbb{N}}$ of versions of U such that $U^{\varphi_i} \xrightarrow{\text{w}^*} V$. By Lemma 2.9, we have $(U^{\varphi_i})^{\bowtie\mathcal{R}} \xrightarrow{\text{w}^*} V^{\bowtie\mathcal{R}}$. Since on the left-hand side, we have step-functions on the same grid $\mathcal{R}\times\mathcal{R}$, the weak* convergence is in this case equivalent to the convergence in $\|\cdot\|_1$, $(U^{\varphi_i})^{\bowtie\mathcal{R}} \xrightarrow{\|\cdot\|_1} V^{\bowtie\mathcal{R}}$. Now, we can find $i \in \mathbb{N}$ such that $\left\|V - (U^{\varphi_i})^{\bowtie\mathcal{R}}\right\|_1 \leq \left\|V - V^{\bowtie\mathcal{R}}\right\|_1 + \left\|V^{\bowtie\mathcal{R}} - (U^{\varphi_i})^{\bowtie\mathcal{R}}\right\|_1 < \varepsilon$. □

Lemma 2.29. *Let $A \subset [0,1]$ be a measurable set, $\mathcal{R} = \{\Omega_1, \dots, \Omega_k\}$ be a finite partition of $[0,1]$ and $\varepsilon > 0$. Then there is $s_0 \in \mathbb{N}$ such that for every $s \geq s_0$ we may find a collection $\mathcal{I}_{\mathcal{A}} \subset [k]\times[s]$ such that for the symmetric difference of A and $\bigcup_{(i,x)\in\mathcal{I}_A} [\![\Omega_i]\!]^s_x$ we have*

$$\nu\left(A \diamond \bigcup_{(i,x)\in\mathcal{I}_A} [\![\Omega_i]\!]^s_x\right) < \varepsilon.$$

Proof. First we demonstrate that it is enough to show the lemma for the special case when $\mathcal{R}$ is an interval partition. Suppose that $A \subset [0,1]$ and $\varepsilon > 0$ is given. We may find a measurable almost-bijection φ such that the restriction of φ to each Ω_j preserves the order of the real line and such that $\varphi(\mathcal{R})$ is an interval partition. Then we find the correct $s_0 \in \mathbb{N}$ when applied for $\varphi(A)$. It is clear that this $s_0 \in \mathbb{N}$ works because $\varphi([\![\Omega_i]\!]^s_x) = [\![\varphi(\Omega_i)]\!]^s_x$.

If $\mathcal{R}$ is a finite interval partition then we can work in each interval separately. This implies that we may restrict ourselves to the case where $\mathcal{R} = \{[0,1]\}$. The latter is a basic fact about the Lebesgue measure. □

Lemma 2.30. *Let $U \in \mathcal{W}_0$, $\mathcal{R} = \{\Omega_1, \dots, \Omega_k\}$ be a finite partition and $\varepsilon > 0$. Then there is $s_0 \in \mathbb{N}$ such that for every $s \geq s_0$ we have $\left\| U - U^{\ltimes [\![\mathcal{R}]\!]^s} \right\|_1 < \varepsilon$ where*

$$[\![\mathcal{R}]\!]^s = \{[\![\Omega_i]\!]^s_x : (i,x) \in [k] \times [s]\}\,.$$

Proof. Using Lemma 2.29, it is straightforward to show that the assertion is true if there is a finite partition $\mathcal{P} = \{P_1, \dots, P_\ell\}$ of Ω such that U is constant on each $P_i \times P_j$. To prove the general case, use Lemma 2.8 to find a suitable partition $\mathcal{P}$ of Ω, then apply the previous observation to $U^{\ltimes \mathcal{P}}$ and use the inequality

$$\|U - U^{\ltimes [\![\mathcal{R}]\!]^s}\|_1 \leq \|U - U^{\ltimes \mathcal{P}}\|_1 + \|U^{\ltimes \mathcal{P}} - (U^{\ltimes \mathcal{P}})^{\ltimes [\![\mathcal{R}]\!]^s}\|_1 + \|(U^{\ltimes \mathcal{P}})^{\ltimes [\![\mathcal{R}]\!]^s} - U^{\ltimes [\![\mathcal{R}]\!]^s}\|_1\,.$$

□

Proposition 2.31. *Let U be a graphon, $\mathcal{R} = \{\Omega_1, \dots, \Omega_k\}$ a finite partition of $[0,1]$ and $\varepsilon > 0$. Then there is $s_0 \in \mathbb{N}$ such that for every $s \geq s_0$ there is $N_0 \in \mathbb{N}$ such that for every $N \geq N_0$, an independently chosen random N-tuple $(W_i \sim \mathbb{W}(U, \mathcal{R}, s))_{i=1}^N$ satisfies*

$$\left\| U^{\ltimes \mathcal{R}} - \frac{\sum_{i=1}^N W_i}{N} \right\|_1 < \varepsilon \tag{2.19}$$

with probability at least 0.9.

Proof. We will set s_0, s, N_0, N later, and start with some bounds that hold for all s and N, which we from now on suppose to be fixed. The idea is to split (2.19) to the contributions of individual parts $\left([\![\Omega_i]\!]^s_x \times [\![\Omega_j]\!]^s_y\right)_{i,j \in [k], x,y \in [s]}$.

We call a quadruple $(i,j,x,y) \in [k] \times [k] \times [s] \times [s]$ a *diagonal quadruple* if $i = j$ and $x = y$. A quadruple $(i,j,p,q) \in [k] \times [k] \times [s] \times [s]$ is *off-diagonal of Type I* if $i \neq j$, and it is *off-diagonal of Type II* if $i = j$ and $x \neq y$.

Claim (Claim A). Suppose that $(i,j,x,y) \in [k] \times [k] \times [s] \times [s]$ is an off-diagonal quadruple. Then for each $a > 0$ we have with probability at least $1 - 2\exp\left(-\frac{a^2 N}{2}\right)$ that

$$\left| \frac{1}{s^2} \int_{\Omega_i \times \Omega_j} U \mathrm{d}\nu^{\otimes 2} - \frac{1}{N} \sum_{\ell=1}^N \int_{[\![\Omega_i]\!]^s_x \times [\![\Omega_j]\!]^s_y} W_\ell \mathrm{d}\nu^{\otimes 2} \right| \leq a + \frac{2\nu^{\otimes 2}(\Omega_i \times \Omega_j)}{s^2\,(s-1)}\,.$$

Proof of Claim A. For each $\ell \in [N]$, we define

$$Y_\ell := \int_{[\![\Omega_i]\!]^s_x \times [\![\Omega_j]\!]^s_y} W_\ell \mathrm{d}\nu^{\otimes 2}\,.$$

Then $Y_\ell : \mathbb{W}(U, \mathcal{R}, s) \to [0,1]$ is a random variable with the expectation

$$e_1 := \frac{1}{s^2} \cdot \int_{\Omega_i \times \Omega_j} U \mathrm{d}\nu^{\otimes 2}$$

if (i,j,x,y) if off-diagonal of Type I, or

$$e_2 := \frac{1}{s\,(s-1)} \int_{\Omega_i\times\Omega_j\setminus\cup_{d=1}^{s}[\![\Omega_i]\!]_d^s\times[\![\Omega_j]\!]_d^s} U\mathrm{d}\nu^{\otimes 2} = \left(\frac{1}{s^2}\int_{\Omega_i\times\Omega_j} U\mathrm{d}\nu^{\otimes 2}\right) \pm \frac{2\nu^{\otimes 2}\left(\Omega_i\times\Omega_j\right)}{s^2\,(s-1)}$$

if (i,j,x,y) if off-diagonal of Type II. To see this, consider first the case of Type I. In that case, using the notation from Definition 2.27, we have a random permutation π_i permuting stripes of Ω_i and a different random permutation π_j permuting stripes of Ω_j. Such a pair of random permutations induces a permutation of the grid on $\Omega_i\times\Omega_j$ such that the probability that any given cell is placed onto the cell $[\![\Omega_i]\!]_x^s\times[\![\Omega_j]\!]_y^s$ is $\frac{1}{s^2}$, which justifies that the average is e_1 in this case. Similarly, in the case of Type II (i.e., $i=j$), we have one permutation π_i which permutes simultaneously rows and columns of the grid on $\Omega_i\times\Omega_i$. In that case, the probability that any given off-diagonal cell is placed onto the cell $[\![\Omega_i]\!]_x^s\times[\![\Omega_i]\!]_y^s$ is $\frac{1}{s(s-1)}$.

Observe that $(Y_\ell)_{\ell=1}^N$ are independent random variables. Thus, the Chernoff bound (Lemma 2.2) gives us that for each $a>0$ we have with probability at least $1-2\exp\left(-\frac{a^2N^2}{2N}\right) = 1-2\exp\left(-\frac{a^2N}{2}\right)$ that

$$\left|\frac{1}{s^2}\int_{\Omega_i\times\Omega_j} U\mathrm{d}\nu^{\otimes 2} - \frac{1}{N}\sum_{\ell=1}^{N}\int_{[\![\Omega_i]\!]_x^s\times[\![\Omega_j]\!]_y^s} W_\ell\mathrm{d}\nu^{\otimes 2}\right| = \left|e_1 - \left(\frac{1}{N}\sum_{\ell=1}^{N} Y_\ell\right)\right| \le a \tag{2.20}$$

in the case of off-diagonal quadruple of Type I and

$$\begin{aligned}\left|\frac{1}{s^2}\int_{\Omega_i\times\Omega_j} U\mathrm{d}\nu^{\otimes 2} - \frac{1}{N}\sum_{\ell=1}^{N}\int_{[\![\Omega_i]\!]_x^s\times[\![\Omega_j]\!]_y^s} W_\ell\mathrm{d}\nu^{\otimes 2}\right| \le \\ \le \left|e_2 - \left(\frac{1}{N}\sum_{\ell=1}^{N} Y_\ell\right)\right| + \frac{2\nu^{\otimes 2}\left(\Omega_i\times\Omega_j\right)}{s^2\,(s-1)} \le a + \frac{2\nu^{\otimes 2}\left(\Omega_i\times\Omega_j\right)}{s^2\,(s-1)}\end{aligned} \tag{2.21}$$

in the case of off-diagonal quadruple of Type II. Hence, (2.20) and (2.21) give the statement of the claim. □

Take $s_0\ge 3$ that satisfies Lemma 2.30 with $\frac{\varepsilon}{2}$, and such that $\frac{k^2+3}{s_0-1} < \frac{\varepsilon}{2}$. Then for every $s\ge s_0$ we may find N_0 such that for every $N\ge N_0$ we have

$$1-2\exp\left(-\frac{\left(\frac{1}{s^3}\right)^2 N}{2}\right) > 1-\frac{1}{10k^2s^2}\,.$$

We have

$$\left\|U^{\bowtie\mathcal{R}}-\frac{1}{N}\sum_{\ell=1}^{N}W_\ell\right\|_1\leq\left\|U^{\bowtie\mathcal{R}}-\frac{1}{N}\sum_{\ell=1}^{N}(W_\ell)^{\bowtie[\![\mathcal{R}]\!]^s}\right\|_1+\left\|\frac{1}{N}\sum_{\ell=1}^{N}(W_\ell)^{\bowtie[\![\mathcal{R}]\!]^s}-\frac{1}{N}\sum_{\ell=1}^{N}W_\ell\right\|_1$$
$$\leq\left\|U^{\bowtie\mathcal{R}}-\frac{1}{N}\sum_{\ell=1}^{N}(W_\ell)^{\bowtie[\![\mathcal{R}]\!]^s}\right\|_1+\frac{\varepsilon}{2}.$$

The following simple observation allows us to use Claim A:

$$\begin{aligned}&\left\|U^{\bowtie\mathcal{R}}-\frac{1}{N}\sum_{\ell=1}^{N}(W_\ell)^{\bowtie[\![\mathcal{R}]\!]^s}\right\|_1\\&=\sum_{(i,j,x,y)\in[k]\times[k]\times[s]\times[s]}\int_{[\![\Omega_i]\!]^s_x\times[\![\Omega_j]\!]^s_y}\\&\quad\left|\frac{1}{\nu^{\otimes2}(\Omega_i\times\Omega_j)}\int_{\Omega_i\times\Omega_j}U\mathrm{d}\nu^{\otimes2}-\frac{1}{N}\sum_{\ell=1}^{N}\frac{s^2}{\nu^{\otimes2}(\Omega_i\times\Omega_j)}\int_{[\![\Omega_i]\!]^s_x\times[\![\Omega_j]\!]^s_y}W_\ell\mathrm{d}\nu^{\otimes2}\right|\mathrm{d}\nu^{\otimes2}\\&=\sum_{(i,j,x,y)\in[k]\times[k]\times[s]\times[s]}\left|\frac{1}{s^2}\int_{\Omega_i\times\Omega_j}U\mathrm{d}\nu^{\otimes2}-\frac{1}{N}\sum_{\ell=1}^{N}\int_{[\![\Omega_i]\!]^s_x\times[\![\Omega_j]\!]^s_y}W_\ell\mathrm{d}\nu^{\otimes2}\right|.\end{aligned}$$

Claim A applied with $a=\frac{1}{s^3}$ gives that with probability at least

$$1-\left(\sum_{i,j\in[k]}\sum_{x,y\in[s]}\frac{1}{10k^2s^2}\right)=0.9$$

we have

$$\begin{aligned}&\sum_{(i,j,x,y)\in[k]\times[k]\times[s]\times[s]}\left|\frac{1}{s^2}\int_{\Omega_i\times\Omega_j}U\mathrm{d}\nu^{\otimes2}-\frac{1}{N}\sum_{\ell=1}^{N}\int_{[\![\Omega_i]\!]^s_x\times[\![\Omega_j]\!]^s_y}W_\ell\mathrm{d}\nu^{\otimes2}\right|\\&\leq\sum_{i,j\in[k]\times[k]}s^2\left(a+\frac{2\nu^{\otimes2}\left(\Omega_i\times\Omega_j\right)}{s^2\left(s-1\right)}\right)+\sum_{(i,x)\in[k]\times[s]}\nu^{\otimes2}\left([\![\Omega_i]\!]^s_x\times[\![\Omega_i]\!]^s_x\right)\\&\leq k^2s^2a+\frac{2}{s-1}+\frac{1}{s}\leq\frac{k^2+3}{s-1}<\frac{\varepsilon}{2}\,.\end{aligned}$$

□

Proposition 2.32. *Let U be a graphon, $\mathcal{R}=\{\Omega_1,\ldots,\Omega_k\}$ be a finite partition of $[0,1]$. Then there is $s_0\in\mathbb{N}$ such that for every $s\geq s_0$, a random graphon $W\sim\mathbb{W}\left(U,\mathcal{R},s\right)$ satisfies*

$$\|U-W\|_\square>\frac{\delta_\square\left(U,U^{\bowtie\mathcal{R}}\right)}{8}-\frac{2}{s^{\frac{1}{4}}}$$

with probability at least $1-\exp\left(-\frac{\sqrt{s}}{8k}\right)$.

Proof. Put $V = U^{\bowtie\mathcal{R}}$ and write $\delta = \delta_\square(U,V)$. By Lemma 2.5 we may find $S_0, T_0 \subset \Omega$ such that $S_0 \cap T_0 = \emptyset$ and

$$\left|\int_{S_0\times T_0}(U-V)\right| \geq \frac{\delta}{4}\,.$$

By a slight modification of Lemma 2.29 there is $s_0 \geq 3$ such that for every $s \geq s_0$ there are $S, T \subset \Omega$ that are unions of stripes such that

$$\left|\int_{S\times T}(U-V)\right| \geq \frac{\delta}{8}$$

and $S \cap T = \emptyset$. Now fix $s \geq s_0$ and denote $\mathcal{I}_S, \mathcal{I}_T \subset [k]\times[s]$ the sets of indices giving the corresponding stripes, i.e.,

$$S = \bigcup_{(i,x)\in\mathcal{I}_S} [\![\Omega_i]\!]^s_x$$

and similarly for T. Define $\mathcal{I}^i_S = \{x \in [s] : (i,x) \in \mathcal{I}_S\}$, $\mathcal{C}_S = \{[\![\Omega_i]\!]^s_x\}_{(i,x)\in\mathcal{I}_S}$ and $\mathcal{C}^i_S = \{[\![\Omega_i]\!]^s_x\}_{x\in\mathcal{I}^i_S}$, similarly define $\mathcal{I}^i_T$, $\mathcal{C}_T$ and $\mathcal{C}^i_T$.

We may assume that $S_i := \bigcup \mathcal{C}^i_S$ is on the left side of Ω_i and that $T_i := \bigcup \mathcal{C}^i_T$ is exactly next to it. To see this note that if $\mathcal{C}^i_S$ and $\mathcal{C}^i_T$ are in some general position, then we may find a measure preserving bijection φ that is invariant on each Ω_i and permutes the stripes accordingly. Note that this is possible because S, T are disjoint. Then for $\mathcal{C}^i_S, \mathcal{C}^i_T$ in the general position use the same argument with conjugation by φ as in Lemma 2.29.

Define the random variable $Z : \mathbb{W}(U,\mathcal{R},s) \to \mathbb{R}$ as

$$Z(W) = \int_{S\times T} W \mathrm{d}\nu^{\otimes 2}.$$

Claim (Claim B). We have $\mathbf{E}[Z] = \int_{S\times T} V\mathrm{d}\nu^{\otimes 2} \pm \frac{1}{s-1}$.

Proof of Claim B. For fixed $i,j \in [k]$ define $Z_{i,j}(W) = \int_{S_i\times T_j} W\mathrm{d}\nu^{\otimes 2}$. Then $Z = \sum_{i,j\in[k]} Z_{i,j}$ and also $\mathbf{E}[Z] = \sum_{i,j\in[k]} \mathbf{E}[Z_{i,j}]$. It suffices to show that $\mathbf{E}[Z_{i,j}] = \int_{S_i\times T_j} V\mathrm{d}\nu^{\otimes 2}$ if $i \neq j$ and $\mathbf{E}[Z_{i,i}] = \int_{S_i\times T_i} V\mathrm{d}\nu^{\otimes 2} \pm \frac{\nu^{\otimes 2}(\Omega_i\times\Omega_i)}{s-1}$ if $i = j$.

We use the notation from the proof of Proposition 2.31. Take an off-diagonal quadruple $(i,j,x,y) \in [k]\times[k]\times[s]\times[s]$ and define

$$Y_{i,j,x,y} = \int_{[\![\Omega_i]\!]^s_x \times [\![\Omega_j]\!]^s_y} W\mathrm{d}\nu^{\otimes 2}\,.$$

There are two cases depending on the type of (i,j,x,y). Suppose that (i,j,x,y) is of Type I. Then we have

$$\mathbf{E}[Y_{i,j,x,y}] = \frac{1}{s^2}\cdot\int_{\Omega_i\times\Omega_j} U\mathrm{d}\nu^{\otimes 2} = \frac{1}{s^2}\cdot\int_{\Omega_i\times\Omega_j} V\mathrm{d}\nu^{\otimes 2}\,,$$

and summing over all $(x,y) \in \mathcal{I}_S^i \times \mathcal{I}_T^j$ we get for $i \neq j$ that

$$\mathbf{E}\left[Z_{i,j}\right] = \frac{\left|\mathcal{I}_S^i\right| \times \left|\mathcal{I}_T^j\right|}{s^2} \int_{\Omega_i \times \Omega_j} V \mathrm{d}\nu^{\otimes 2} = \int_{S_i \times T_j} V \mathrm{d}\nu^{\otimes 2}.$$

Suppose that (i,i,x,y) is of Type II. Then we have

$$\mathbf{E}\left[Y_{i,i,x,y}\right] = \left(\frac{1}{s^2} \int_{\Omega_i \times \Omega_i} V \mathrm{d}\nu^{\otimes 2}\right) \pm \frac{2\nu^{\otimes 2}\left(\Omega_i \times \Omega_i\right)}{s^2\left(s-1\right)},$$

and summing over all $(x,y) \in \mathcal{I}_S^i \times \mathcal{I}_T^i$ (with x and y distinct) we get for every i that

$$\begin{aligned}\mathbf{E}\left[Z_{i,i}\right] &= \frac{\left|\mathcal{I}_S^i\right| \times \left|\mathcal{I}_T^i\right|}{s^2} \left(\int_{\Omega_i \times \Omega_i} V \mathrm{d}\nu^{\otimes 2} \pm \frac{2\nu^{\otimes 2}\left(\Omega_i \times \Omega_i\right)}{\left(s-1\right)}\right) \\ &= \int_{S_i \times T_i} V \mathrm{d}\nu^{\otimes 2} \pm \frac{\nu^{\otimes 2}\left(\Omega_i \times \Omega_i\right)}{s-1}\end{aligned}$$

because $\frac{\left|\mathcal{I}_S^i\right| \times \left|\mathcal{I}_T^i\right|}{s^2} \leq \frac{1}{2}$. □

In order to use the Method of Bounded Differences we introduce the following correspondence between permutations that induce $\mathbb{W}\left(U, \mathcal{R}, s\right)$ and $[0,1]^{k \times s}$ (which we view as functions from $[k] \times [s]$ to $[0,1]$). Namely, for each $f \in [k] \times [s] \to [0,1]$ which is injective we define a permutation α_f of $[k] \times [s]$ such that $\alpha_f(\{i\} \times [s]) = \{i\} \times [s]$ for each i, and such that the relative position of $\alpha_f(i,x)$ inside the block $\{i\} \times [s]$ is the same as the relative position of $f\left(i,x\right)$ inside the set of numbers $\left\{f\left(i,y\right)\right\}_{y \in [s]}$. We leave α_f undefined for non-injective functions, which form a nullset on $[0,1]^{k \times s}$. One can verify that the assignment $f \mapsto \left(\alpha_f\right)^{-1}$, which maps each f to the inverse of α_f, is measure preserving where we have the Lebesgue measure on $[0,1]^{k \times s}$ and the uniform measure on the permutations of $[k] \times [s]$ that fix the first coordinate. Note that these are exactly the permutations that naturally induce $\mathbb{W}\left(U, \mathcal{R}, s\right)$. Hence, we may consider the random variable Z to be defined on $[0,1]^{k \times s}$ with values in $\mathbb{R}$.

We show that Z satisfies the assumptions of Lemma 2.3(a). Recall that we assume that each S_i is concentrated on the left-most part of the interval Ω_i and T_i is exactly next to it. Suppose that $f, f' \in [0,1]^{k \times s}$ differ in at most one coordinate in the i-th block. Then

$$\left|\left(\alpha_f\right)^{-1}\left(\mathcal{I}_S^i\right) \diamond \left(\alpha_{f'}\right)^{-1}\left(\mathcal{I}_S^i\right)\right| \leq 2 \quad \text{and} \quad \left|\left(\alpha_f\right)^{-1}\left(\mathcal{I}_T^i\right) \diamond \left(\alpha_{f'}\right)^{-1}\left(\mathcal{I}_T^i\right)\right| \leq 2.$$

Then we may compute

$$\begin{aligned}
\left|Z(f)-Z\left(f'\right)\right| &= \left|\int_{S\times T} U^{(\alpha_f)^{-1}}\mathrm{d}\nu^{\otimes 2}-\int_{S\times T} U^{(\alpha_{f'})^{-1}}\mathrm{d}\nu^{\otimes 2}\right| \\
&= \left|\int_{(\alpha_f)^{-1}(S)\times(\alpha_f)^{-1}(T)} U\mathrm{d}\nu^{\otimes 2}-\int_{(\alpha_{f'})^{-1}(S)\times(\alpha_{f'})^{-1}(T)} U\mathrm{d}\nu^{\otimes 2}\right| \\
&\leq \nu^{\otimes 2}\left(\left((\alpha_f)^{-1}(S)\times(\alpha_f)^{-1}(T)\right)\diamond\left((\alpha_{f'})^{-1}(S)\times(\alpha_{f'})^{-1}(T)\right)\right) \\
&\leq 4\frac{\nu(\Omega_i)}{s}\leq\frac{4}{s}.
\end{aligned}$$

By Lemma 2.3(a) we have

$$\mathbf{P}\left[|Z-\mathbf{E}Z|>d\right]\leq\exp\left(-\frac{d^2 s}{8k}\right),\text{for any } d>0.$$

In particular, taking $d=s^{-\frac{1}{4}}$ we have

$$\mathbf{P}\left[|Z-\mathbf{E}Z|>s^{-\frac{1}{4}}\right]\leq\exp\left(-\frac{\sqrt{s}}{8k}\right)$$

and therefore with probability at least $1-\exp\left(-\frac{\sqrt{s}}{8k}\right)$ we have that

$$|Z(W)-\mathbf{E}Z|\leq s^{-\frac{1}{4}} \tag{2.22}$$

for $W\in\mathbb{W}(U,\mathcal{R},s)$. We conclude that with probability at least $1-\exp\left(-\frac{\sqrt{s}}{8k}\right)$ we have that

$$\begin{aligned}
\|W-U\|_\square &\geq \left|\int_{S\times T} U\mathrm{d}\nu^{\otimes 2}-\int_{S\times T} W\mathrm{d}\nu^{\otimes 2}\right| \\
&\geq \left|\int_{S\times T} U\mathrm{d}\nu^{\otimes 2}-\int_{S\times T} V\mathrm{d}\nu^{\otimes 2}\right|-\left|\int_{S\times T} V\mathrm{d}\nu^{\otimes 2}-\int_{S\times T} W\mathrm{d}\nu^{\otimes 2}\right| \\
&\geq \frac{\delta}{8}-\left|Z(W)-\left(\mathbf{E}[Z]\pm\frac{1}{s-1}\right)\right| \\
\boxed{\text{by (2.22)}} \quad &\geq \frac{\delta}{8}-\frac{1}{s-1}-\frac{1}{s^{\frac{1}{4}}}>\frac{\delta}{8}-\frac{2}{s^{\frac{1}{4}}},
\end{aligned}$$

as was needed (we used that $s\geq s_0\geq 3$ in the last inequality). □

Now we are ready to prove Lemma 2.24.

Proof of Lemma 2.24. Let $\delta:=\delta_\square(V,U)$. By Fact 2.18, we have that $\delta>0$. First use Lemma 2.28 to approximate V by some $(U^\varphi)^{\ltimes\mathcal{R}}$ such that $\left\|V-(U^\varphi)^{\ltimes\mathcal{R}}\right\|_1<\min\left(\frac{\varepsilon}{2},\frac{\delta}{2}\right)$. We may assume without loss of generality that φ is the identity and therefore work with

$U^{\bowtie\mathcal{R}}$ instead of $(U^{\varphi})^{\bowtie\mathcal{R}}$. We have

$$\delta_{\square}\left(U,U^{\bowtie\mathcal{R}}\right)\geq\delta_{\square}\left(V,U\right)-\delta_{\square}\left(V,U^{\bowtie\mathcal{R}}\right)\geq\delta_{\square}\left(V,U\right)-\left\|V-U^{\bowtie\mathcal{R}}\right\|_{1}>\frac{\delta}{2}\,.$$

We use Proposition 2.31 and Proposition 2.32 to find $s\in\mathbb{N}$ and an even number N such that $\|U-W\|_{\square}>\frac{\delta}{32}$ for $W\in\mathbb{W}\left(U,\mathcal{R},s\right)$ with probability at least 0.9, and also $\left\|U^{\bowtie\mathcal{R}}-\frac{\sum_{k=1}^{N}W_k}{N}\right\|_1<\frac{\varepsilon}{2}$ for $(W_k)_{k=1}^{N}\in\left(\mathbb{W}\left(U,\mathcal{R},s\right)\right)^N$ with probability at least 0.9.

Define a random variable $Q:\left(\mathbb{W}\left(U,\mathcal{R},s\right)\right)^N\to\mathbb{R}$,

$$Q:(W_k)_{k=1}^{N}\mapsto\frac{2}{N}\cdot\left|\left\{k\in[N/2]:\|W_{2k-1}-W_{2k}\|_{\square}\leq\frac{\delta}{32}\right\}\right|.$$

Note that for any $U^*\in\mathbb{W}(U,\mathcal{R},s)$, the distributions $\mathbb{W}(U,\mathcal{R},s)$ and $\mathbb{W}(U^*,\mathcal{R},s)$ on versions of U coincide with probability 1. So for every $k\in[N/2]$, we can equivalently first sample $W_{2k-1}\sim\mathbb{W}(U,\mathcal{R},s)$, and then sample $W_{2k}\sim\mathbb{W}(W_{2k-1},\mathcal{R},s)$. Thus, for a fixed $k\in[N/2]$ the probability that $\|W_{2k-1}-W_{2k}\|_{\square}>\frac{\delta}{32}$ is at least 0.9 due to Proposition 2.32. Hence, $\mathbf{E}\left[Q\right]\leq 0.1$. By Markov's inequality, $\mathbf{P}\left[Q\geq 0.4\right]\leq 0.25$. By the union bound, with probability at most 0.35 we have that $Q\geq 0.4$ or that $\left\|U^{\bowtie\mathcal{R}}-\frac{\sum_{k=1}^{N}W_k}{N}\right\|_1\geq\frac{\varepsilon}{2}$. In particular, there exists a choice $\left(U^{\phi_k}\right)_{k=1}^{N}=(W_k)_{k=1}^{N}$ of an N-tuple of versions of U which does not have any of these two «bad» properties. Such an N-tuple $\left(U^{\phi_k}\right)_{k=1}^{N}$ satisfies (2.18) since $Q<0.4<0.5$. It also satisfies (2.17) since

$$\left\|V-\frac{\sum_{i=1}^{N}U^{\phi_i}}{N}\right\|_1\leq\left\|V-U^{\bowtie\mathcal{R}}\right\|_1+\left\|U^{\bowtie\mathcal{R}}-\frac{\sum_{i=1}^{N}U^{\phi_i}}{N}\right\|_1<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon\,.$$

□

3. Cut distance identifying graphon parameters

3.1. **Basics.** In [12], we based our treatment of the cut distance on $\mathbf{ACC}_{\mathrm{w}*}\left(W_1,\ldots\right)$ and $\mathbf{LIM}_{\mathrm{w}*}\left(W_1,\ldots\right)$, which are sets of functions. In contrast, the key objects in [13] are the sets of numerical values

$$\left\{\mathrm{INT}_f(W):W\in\mathbf{ACC}_{\mathrm{w}*}\left(W_1,W_2,\ldots\right)\right\}\text{ and }\left\{\mathrm{INT}_f(W):W\in\mathbf{LIM}_{\mathrm{w}*}\left(W_1,W_2,\ldots\right)\right\},$$

with notation taken from (1.4). In this section, we introduce an abstract framework to approaching the cut distance via similar optimization problems. Our key definitions of cut distance identifying graphon parameters and cut distance compatible graphon parameters use $\mathbb{R}^n$ together with lexicographical ordering and Euclidean metric, and $\mathbb{R}^{\mathbb{N}}$ together with lexicographical ordering which we denote just $\leq$.

By a *graphon parameter* we mean any function $\theta:\mathcal{W}_0\to\mathbb{R}$, $\theta:\mathcal{W}_0\to\mathbb{R}^n$ (for some $n\in\mathbb{N}$), or $\theta:\mathcal{W}_0\to\mathbb{R}^{\mathbb{N}}$, such that $\theta(W_1)=\theta(W_2)$ for any two graphons W_1 and W_2 with

$\delta_\square(W_1, W_2) = 0$. By a *graphon order*, we mean a preorder $\lhd$ on $\mathcal{W}_0$ which does not change within the weak isomorphism classes, i.e., for each $W_1, W_2 \in \mathcal{W}_0$ with $\delta_\square(W_1, W_2) = 0$ we have $W_1 \lhd W_2$ and $W_2 \lhd W_1$. With these preliminary definitions, we can introduce the central concept of this paper, which we do in four variants, the distinction being whether we require strict monotonicity or not, and whether we work in the setting of graphon parameters or graphon orders.

Definition 3.1.

- We say that a graphon parameter θ is a *cut distance identifying graphon parameter* if we have that $W_1 \prec W_2$ implies $\theta(W_1) < \theta(W_2)$ (here, by $<$ we understand the usual Euclidean order on $\mathbb{R}$ in case $\theta : \mathcal{W}_0 \to \mathbb{R}$ and the lexicographic order in case $\theta : \mathcal{W}_0 \to \mathbb{R}^n$ or $\theta : \mathcal{W}_0 \to \mathbb{R}^{\mathbb{N}}$).
- We say that a graphon parameter ψ is a *cut distance compatible graphon parameter* if we have that $W_1 \preceq W_2$ implies $\psi(W_1) \leq \psi(W_2)$.
- We say that a graphon order $\lhd$ is a *cut distance identifying graphon order* if $W_1 \prec W_2$ implies $W_1 \lhd W_2$ and $W_2 \not\lhd W_1$.
- We say that a graphon order $\lessdot$ is a *cut distance compatible graphon order* if $W_1 \preceq W_2$ implies $W_1 \lessdot W_2$.

Cut distance identifying/compatible orders are abstract versions of their parameter counterparts. Indeed, given a graphon parameter θ, we have that θ is cut distance identifying if and only if a graphon order $\lhd_\theta$ defined by

$$U \lhd_\theta W \quad \text{if and only if } \theta(U) < \theta(W)$$

is cut distance identifying. Similarly, a graphon parameter ψ is cut distance compatible if and only if a graphon order $\lessdot_\psi$ defined by

$$U \lessdot_\psi W \quad \text{if and only if } \psi(U) \leq \psi(W) \tag{3.1}$$

is cut distance compatible. However, there exist cut distance identifying and cut distance compatible graphon orders that do not arise from graphon parameters. Indeed, Proposition 2.22 tells us that the at-least-as-flat relation on degree frequencies induces a cut distance compatible graphon order and that the strictly-flatter relation on range frequencies induces a cut distance identifying graphon order.[h]

The following proposition provides a useful criterion for cut distance compatible graphon parameters. In this criterion, we restrict ourselves to L^1-continuous parameters. This is only a mild restriction. Indeed, many prominent graphon parameters such as homomorphism densities are even continuous with respect to the cut norm (which is a coarser

[h]For this argument to make sense, we need the flatness relation to be transitive. This follows from Lemma 4.13 in [12].

topology). As another important example, the graphon parameter $\mathrm{INT}_f(\cdot)$ is L^1-continuous if f is a continuous function.

Proposition 3.2. *Suppose that θ is a graphon parameter that is continuous with respect to the L^1 norm. Then θ is cut distance compatible if and only if for each graphon $W : \Omega^2 \to [0,1]$ and each finite partition $\mathcal{P}$ of Ω we have $\theta\left(W^{\bowtie\mathcal{P}}\right) \leq \theta(W)$.*

Proof. The $\Rightarrow$ direction is obvious, since $W^{\bowtie\mathcal{P}} \preceq W$ by Fact 2.17 (L^1continuity is not needed for this direction). For the reverse direction, suppose that θ is not cut distance compatible. That is, there exist two graphons $U \preceq W$ so that $\theta(U) > \theta(W)$. Since θ is L^1-continuous at U we can use Lemma 2.8 to find a finite partition $\mathcal{Q}$ such that

$$\theta\left(U^{\bowtie\mathcal{Q}}\right) > \theta(W). \tag{3.2}$$

As $U \preceq W$, there exist measure preserving bijections $\pi_1, \pi_2, \pi_3, \ldots$ so that $W^{\pi_n} \xrightarrow{\text{w}^*} U$. In particular, the sequence $\left((W^{\pi_n})^{\bowtie\mathcal{Q}}\right)_n$ converges to $U^{\bowtie\mathcal{Q}}$ in L^1. Thus the L^1-continuity of θ at $U^{\bowtie\mathcal{Q}}$ gives us that for some n, $\theta\left((W^{\pi_n})^{\bowtie\mathcal{Q}}\right)$ is nearly as big as $\theta\left(U^{\bowtie\mathcal{Q}}\right)$. In particular, using (3.2) we have that $\theta\left((W^{\pi_n})^{\bowtie\mathcal{Q}}\right) > \theta(W)$. We let π_n act on the partition $\mathcal{Q}$, $\mathcal{P} := \pi_n(\mathcal{Q})$. Obviously, $(W^{\pi_n})^{\bowtie\mathcal{Q}}$ is a version of $W^{\bowtie\mathcal{P}}$, and thus $\theta\left(W^{\bowtie\mathcal{P}}\right) = \theta\left((W^{\pi_n})^{\bowtie\mathcal{Q}}\right) > \theta(W)$, as was needed. □

It is natural to believe that there is a similar characterization for cut distance identifying parameters. We were however unable to prove it, so we leave it as a conjecture.

Conjecture 3.3. *Suppose that θ is a graphon parameter that is continuous with respect to the L^1 norm. Then θ is cut distance identifying if and only if for each graphon W and each finite partition $\mathcal{P}$ of Ω for which $W^{\bowtie\mathcal{P}} \neq W$ we have $\theta\left(W^{\bowtie\mathcal{P}}\right) < \theta(W)$.*

Note that the $\Rightarrow$ direction is obvious as in Proposition 3.2.

Cut distance identifying graphon parameters/orders can be used to prove compactness of the graphon space. This is stated in the next two theorems.

Theorem 3.4. *Let $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ be a sequence of graphons.*

For orders: *Suppose that $\lhd$ is a cut distance compatible graphon order. Then there exists a subsequence $\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots$ such that $\mathbf{ACC}_{\text{w}*}(\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots)$ contains an element Γ with $W \lhd \Gamma$ for each $W \in \mathbf{ACC}_{\text{w}*}(\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots)$.*

For parameters: *Suppose that θ is a cut distance compatible graphon parameter. Then there exists a subsequence $\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots$ such that $\mathbf{ACC}_{\text{w}*}(\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots)$ contains an element Γ with $\theta(\Gamma) = \sup\{\theta(W) : W \in \mathbf{ACC}_{\text{w}*}(\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots)\}$.*

In both cases this follows immediately from [12, Theorem 3.3] and [12, Lemma 4.9]. Note that the version for orders is more general, since the parameter version can be

reduced by (3.1). Let us note that one could use the ideas from the proof of Lemma 16 from [13] to obtain an alternative proof of the parameter version of Theorem 3.4. This latter proof is more elementary and does not need transfinite induction or any appeal to the Vietoris topology, which the machinery from [12] does. However, one needs to be a little careful while doing so because not every subset of $\mathbb{R}^N$ (or $\mathbb{R}^{\mathbb{N}}$) has a supremum in the lexicographical ordering. On the other hand, the parameter version of Theorem 3.4 implicitly says that the supremum of the set $\{\theta(W) : W \in \mathbf{ACC}_{\mathrm{w}*}(\Gamma_{n_1}, \Gamma_{n_2}, \Gamma_{n_3}, \ldots)\}$ exists.

Theorem 3.5. *Let $W_1, W_2, W_3, \ldots$ be a sequence of graphons.*

For orders: *Suppose that $\lhd$ is a cut distance identifying graphon order. Suppose that $\Gamma \in \mathbf{LIM}_{\mathrm{w}*}(W_1, W_2, W_3, \ldots)$ is such that $W \lhd \Gamma$ for each $W \in \mathbf{ACC}_{\mathrm{w}*}(W_1, W_2, W_3, \ldots)$. Then $W_1, W_2, W_3, \ldots$ converges to Γ in the cut distance.*

For parameters: *Suppose that θ is a cut distance identifying graphon parameter. Suppose that a graphon $\Gamma \in \mathbf{LIM}_{\mathrm{w}*}(W_1, W_2, W_3, \ldots)$ is such that*

$$\theta(\Gamma) = \sup\{\theta(W) : W \in \mathbf{ACC}_{\mathrm{w}*}(W_1, W_2, W_3, \ldots)\} \ .$$

Then $W_1, W_2, W_3, \ldots$ converges to Γ in the cut distance.

Proof. As the first step, we show that $\langle\Gamma\rangle = \mathbf{ACC}_{\mathrm{w}*}(W_1, W_2, \ldots) = \mathbf{LIM}_{\mathrm{w}*}(W_1, W_2, \ldots)$. Let $U \in \mathbf{ACC}_{\mathrm{w}*}(W_1, W_2, \ldots)$. By Theorem 3.3 from [12] we can find a subsequence $W_{n_1}, W_{n_2}, \ldots$ such that

$$\mathbf{LIM}_{\mathrm{w}*}(W_{n_1}, W_{n_2}, \ldots) = \mathbf{ACC}_{\mathrm{w}*}(W_{n_1}, W_{n_2}, \ldots)$$

and $U \in \mathbf{LIM}_{\mathrm{w}*}(W_{n_1}, W_{n_2}, \ldots)$. Note that we have $\Gamma \in \mathbf{LIM}_{\mathrm{w}*}(W_{n_1}, W_{n_2}, \ldots)$. Using Lemma 4.9 from [12], we can find a maximum element $W \in \mathbf{LIM}_{\mathrm{w}*}(W_{n_1}, W_{n_2}, \ldots)$ with respect to the structuredness order. It follows that $\Gamma \preceq W$. Therefore $\Gamma \lhd W$ or $\theta(\Gamma) \leq \theta(W)$, respectively. Using our assumption on Γ and the fact that $\lhd$ is a cut distance identifying graphon order or that θ is a cut distance identifying graphon parameter, respectively, we must have $\langle\Gamma\rangle = \langle W\rangle$. This implies that $U \in \langle W\rangle = \langle\Gamma\rangle \subset \mathbf{LIM}_{\mathrm{w}*}(W_1, W_2, \ldots)$ where we used the fact that $\mathbf{LIM}_{\mathrm{w}*}(W_1, W_2, \ldots)$ is weak* closed (see [12, Lemma 3.1]). This immediately finishes the first step.

We may suppose that $W_n \xrightarrow{\mathrm{w}^*} \Gamma$. To show that in fact $W_n \xrightarrow{\delta_\square} \Gamma$, we can mimic the proof of Theorem 3.5 (b) $\Longrightarrow$ (a) from [12]. □

So, while the concepts of cut distance identifying graphon parameters or orders do not bring any new tools compared to the structuredness order, knowing that a particular parameter or order is cut distance identifying allows calculations that are often more direct than working with the structuredness order.

3.1.1. *Relation to quasirandomness.* Recall that dense quasi-random finite graphs correspond to constant graphons. Thus, the key question in the area of quasirandomness is which graphon parameters can be used to characterize constant graphons.[i]

The Chung–Graham–Wilson Theorem [4], a version of which we state below, provides the most classical parameters whose minimizer in $\mathcal{G}_p$ is the constant-p graphon.

Theorem 3.6. *Let $p \in [0,1]$. Then the constant-p graphon is the only graphon U in the family $\mathcal{G}_p$ satisfying any of the following conditions.*

(a) We have $t(C_{2\ell}, U) \leq p^{2\ell}$ for a fixed $\ell \in \{2,3,4,\dots\}$.
(b) The largest eigenvalue of U is at most p and all other eigenvalues are zero.

Such characterizations of quasirandomness fit very nicely our framework of cut distance identifying graphon parameters. Indeed, constant graphons are exactly the minimal elements in the structuredness order; we refer to [12, Proposition 8.5] for an easy proof. Thus, each cut distance identifying graphon parameter can be used to characterize constant graphons.

In the opposite direction, we show in Sections 3.5 and 3.6 that the graphon parameters considered in Theorem 3.6 are actually cut distance identifying. Such a strengthening is not automatic (even for reasonable graphon parameters); for example the parameter $t(C_4^+, \cdot)$ (here, C_4^+ is a 4-cycle with a pendant edge) is shown in [25, Section 2] to be minimized on constant graphons but not to be cut distance identifying.[j]

3.1.2. *Uniformity of cut distance identifying graphon parameters.* If θ is a cut distance identifying graphon parameter and $U \prec W$ are two graphons then we know that $\theta(U) < \theta(W)$. In Proposition 3.7 below we prove that this relation can be made uniform (if θ is assumed to be cut distance continuous). That is, if $\delta_\square(U,W) \geq \varepsilon$ then $\theta(U) \leq \theta(W) - b_\varepsilon$, where $b_\varepsilon > 0$ depends only on θ and ε.

We shall make use of Proposition 3.7 in Section 3.2.

Proposition 3.7. *Suppose that $\theta : \mathcal{W}_0 \to \mathbb{R}$ is an arbitrary cut distance identifying graphon parameter that is continuous with respect to the cut distance. For every $\varepsilon > 0$ there exists a $b_\varepsilon > 0$ such that the following holds. Suppose that $U, W : \Omega^2 \to [0,1]$ are graphons such that $U \preceq W$ and $\delta_\square(U,W) \geq \varepsilon$. Then $\theta(U) \leq \theta(W) - b_\varepsilon$.*

[i] Strictly speaking, only parameters that are continuous with respect to the cut distance are relevant for characterizing sequences of quasi-random graphs. Indeed, the assumption of continuity is used to transfer between finite graphs and their limits. The two main parameters we treat below — homomorphism densities $t(H,\cdot)$ and spectrum — are indeed well-known to be cut distance continuous (see Theorems 11.3 and 11.53 in [29]). The parameter $\mathrm{INT}_f(\cdot)$ is not cut distance continuous, and hence does not admit such a transference.

[j] See Remark 3.27 for a more general result.

Proof. Suppose that the claim fails for some $\varepsilon > 0$. That is, for each $n \in \mathbb{N}$, there exist graphons $U_n, W_n : \Omega^2 \to [0,1]$, $U_n \preceq W_n$,

$$\delta_\square (U_n, W_n) \geq \varepsilon \,, \tag{3.3}$$

and yet

$$\theta (W_n) \leq \theta (U_n) + \frac{1}{n} \,. \tag{3.4}$$

As the square of the metric space $(\mathcal{W}_0, \delta_\square)$ is compact, there exists a pair (U, W) of graphons and a sequence $i_1 < i_2 < i_3 < \dots$ so that $U_{i_\ell} \xrightarrow{\delta_\square} U$, $W_{i_\ell} \xrightarrow{\delta_\square} W$. By the continuity of θ, we get from (3.4) that $\theta(W) \leq \theta(U)$. Also, by (3.3) we get that

$$\delta_\square (U, W) \geq \varepsilon > 0 \,. \tag{3.5}$$

Further, using Fact 2.19, we infer that

$$U \preceq W \,. \tag{3.6}$$

Combined with (3.5), we get that $U \prec W$. Since θ is cut distance identifying, we should have $\theta(U) < \theta(W)$, a contradiction. □

3.2. **Using cut distance identifying graphon parameters for index-pumping.** In this section, we show that any cut distance identifying graphon parameter that is continuous with respect to the cut distance can replace the «index», in the Frieze–Kannan regularity lemma. In particular, by Theorem 3.28 below, any norming graph can be used for index-pumping.

We state a graphon version of the Frieze–Kannan regularity lemma [17] in Theorem 3.8 below.

Theorem 3.8 ([29, Corollary 9.13])**.** *For every $\varepsilon > 0$ there exists a number $M \in \mathbb{N}$ so that for each graphon $W : \Omega^2 \to [0,1]$ there exists a partition $\mathcal{P}$ of Ω with at most M parts so that $\|W - W^{\ltimes \mathcal{P}}\|_\square \leq \varepsilon$.*

The number M in Theorem 3.8 can be taken as $M = 2^{O(1/\varepsilon^2)}$ and this is essentially optimal, [9]. Let us recall the main steps of the proof of Theorem 3.8.

[FK1] We start with the trivial partition $\mathcal{P}_1 = \{\Omega\}$.

[FK2] At any given step $i = 1, 2, \dots$, if $\left\|W - W^{\ltimes \mathcal{P}_i}\right\|_\square \leq \varepsilon$, then we output the partition $\mathcal{P}_i$, but …

[FK3] … if $\left\|W - W^{\ltimes \mathcal{P}_i}\right\|_\square > \varepsilon$ let us take a set $X \subset \Omega$ which is a witness for this (c.f. Lemma 2.5), that is, $\left|\int_{X \times X} (W - W^{\ltimes \mathcal{P}_i})\right| > \frac{\varepsilon}{4}$. The so-called index pumping lemma asserts that defining a new partition $\mathcal{P}_{i+1} := \{C \cap X, C \setminus X : C \in \mathcal{P}_i\}$ we

have

$$\mathrm{INT}_{x\mapsto x^2}(W^{\bowtie\mathcal{P}_{i+1}}) > \mathrm{INT}_{x\mapsto x^2}(W^{\bowtie\mathcal{P}_i}) + c_\varepsilon \ ,$$

where $c_\varepsilon > 0$ depends on ε only.

[FK4] Since the mapping $\mathrm{INT}_{x\mapsto x^2}(\cdot)$ takes values in the interval $[0,1]$, we conclude that the above iteration in [FK3] cannot occur more than $\frac{1}{c_\varepsilon}$-many times. Since $|\mathcal{P}_{i+1}| \leq 2\,|\mathcal{P}_i|$, we conclude that the theorem holds with $M := 2^{\lceil 1/c_\varepsilon \rceil}$.

Our approach is as follows, in the first step, we replace the index $\mathrm{INT}_{x\mapsto x^2}(\cdot)$ by the C_4-density, and in the second step, using Proposition 3.7, we obtain a general result for any cut distance identifying graphon parameter θ that is continuous with respect to the cut distance.[k] Let us state the result about the C_4-density first.

Proposition 3.9. *Suppose that $\varepsilon > 0$, $W : \Omega^2 \to [0,1]$ is a graphon, $\mathcal{P}$ is a finite partition of Ω, and $X \subset \Omega$ is such that*

$$\left| \int_{X\times X} (W - W^{\bowtie\mathcal{P}}) \right| > \varepsilon \ . \tag{3.7}$$

Define $\mathcal{P}^* := \{C \cap X, C \setminus X : C \in \mathcal{P}\}$. Then $t\left(C_4, W^{\bowtie\mathcal{P}^*}\right) > t\left(C_4, W^{\bowtie\mathcal{P}}\right) + \frac{\varepsilon^4}{100}$.

Our proof of Proposition 3.9 is based on an extension of an auxiliary but technical result from [10], which we now state.

Lemma 3.10 (Lemma 11 in [10])**.** *Suppose that Q and R are step graphons with respect to equipartitions $\mathcal{Q}$ and $\mathcal{R}$, respectively. Suppose further that $\mathcal{Q}$ refines $\mathcal{R}$ and that $R = Q^{\bowtie\mathcal{R}}$. Then $t\,(C_4, Q) \geq t\,(C_4, R) + \frac{\|Q-R\|_\square^4}{8}$.*

Proof of Proposition 3.9. Suppose first that there exists an equipartition partition $\mathcal{Q} = \{Q_1, Q_2, \ldots, Q_k\}$ of $[0,1]$ that refines $\mathcal{P}^*$ and $N \in \mathbb{N}$ such that $|I_A|$ is a multiple of N for every $A \in \mathcal{P}^*$ where $I_A = \{i \in [k] : Q_i \subseteq A\}$. Note that if such an equipartition $\mathcal{Q}$ and $N \in \mathbb{N}$ exists, then we may assume that N is arbitrarily big and $\sqrt{N} \in \mathbb{N}$. Let $k_A = \frac{|I_A|}{\sqrt{N}}$ and partition each I_A into $I_{A,1} \ldots, I_{A,k_A}$ with $\sqrt{N}$-elements each as in Lemma 2.10. Denote as $\mathcal{D}$ the partition of Ω with pieces $\bigcup I_{A,i}$ where $A \in \mathcal{P}^*$ and $i \in [k_A]$. Note that $\mathcal{D}$ is an equipartition that refines $\mathcal{P}^*$. Then we have

$$\left\| W^{\bowtie\mathcal{P}^*} - W^{\bowtie\mathcal{D}} \right\|_1 \leq 2N^{-\frac{1}{8}} \tag{3.8}$$

by Lemma 2.10. For each $B \in \mathcal{P}$ we denote as $A_1, A_2 \in \mathcal{P}^*$ the unique elements such that $B = A_1 \cup A_2$. We define $J_B = \{X \in \mathcal{D} : X \subseteq A_1 \vee X \subseteq A_2\}$. Note that it follows from the assumption on N that for the number $r_B := \frac{|J_B|}{\sqrt{N}}$ we have $r_B \in \mathbb{N}$. Partition each J_B

[k] Note that a tempting shortcut in which we would deduce the pumping-up property of θ directly from the pumping-up property of $\mathrm{INT}_{x\mapsto x^2}(\cdot)$ does not work. The reason for this is that $\mathrm{INT}_{x\mapsto x^2}(\cdot)$ is not cut distance continuous.

into $J_{B,1}\dots,J_{B,r_B}$ groups with $\sqrt{N}$-elements each as in Lemma 2.10 and define $\mathcal{T}$ as the partition of Ω with pieces $\bigcup J_{B,i}$ where $B\in\mathcal{P}$ and $i\in[r_B]$. We have

$$\left\|W^{\bowtie\mathcal{P}}-W^{\bowtie\mathcal{T}}\right\|_1\leq 2N^{-\frac{1}{8}} \tag{3.9}$$

by Lemma 2.10. Observe that $\mathcal{D}$ is an equipartition that refines $\mathcal{T}$. Using (3.8), (3.9) and the fact that $\|W^{\bowtie\mathcal{P}}-W^{\bowtie\mathcal{P}^*}\|_\square>\varepsilon$ (by 3.7) we get

$$\left\|W^{\bowtie\mathcal{D}}-W^{\bowtie\mathcal{T}}\right\|_\square>\varepsilon-4N^{-\frac{1}{8}}.$$

By Lemma 3.10, we have $t\left(C_4,W^{\bowtie\mathcal{D}}\right)\geq t\left(C_4,W^{\bowtie\mathcal{T}}\right)+\frac{\left(\varepsilon-4N^{-\frac{1}{8}}\right)^4}{8}$. Further, by the Lemma 2.6 (using (3.8) and (3.9)), we have $t\left(C_4,W^{\bowtie\mathcal{D}}\right)=t\left(C_4,W^{\bowtie\mathcal{P}^*}\right)\pm 32N^{-\frac{1}{8}}$ and $t\left(C_4,W^{\bowtie\mathcal{T}}\right)=t\left(C_4,W^{\bowtie\mathcal{P}}\right)\pm 32N^{-\frac{1}{8}}$. Taking N large enough finishes the proof in the special case.

In the general case we assign to each $A\in\mathcal{P}^*$ a measurable set A' such that $\lambda(A')\in\mathbb{Q}$, $\lambda(A\diamond A')$ is arbitrarily small and the collection $\mathcal{R}=\{A':A\in\mathcal{P}^*\}$ is a partition of Ω. We use $\mathcal{R}$ in an obvious way to build $\mathcal{S}$ that approximate $\mathcal{P}$, i.e., if $B\in\mathcal{P}$ and $A_1\cup A_2=B$ where $A_1,A_2\in\mathcal{P}^*$, then define $B'=A_1'\cup A_2'$. It is easy to see that since W is a bounded function we can always find $\mathcal{R}$ such that $\left\|W^{\bowtie\mathcal{P}^*}-W^{\bowtie\mathcal{R}}\right\|_1$ and $\left\|W^{\bowtie\mathcal{P}}-W^{\bowtie\mathcal{S}}\right\|_1$ are arbitrary small. The rest is an easy application of the triangle inequality. □

We now show how to extend Proposition 3.9 to all continuous cut distance identifying graphon parameters.

Proposition 3.11. *Suppose that $\theta:\mathcal{W}_0\to\mathbb{R}$ is a cut distance identifying graphon parameter that is continuous with respect to the cut distance. For every $\varepsilon>0$ there exists a $b_\varepsilon>0$ such that the following holds. Suppose that $W:\Omega^2\to[0,1]$ is a graphon, $\mathcal{P}$ is a finite partition of Ω, and $X\subset\Omega$ is such that $\left|\int_{X\times X}(W-W^{\bowtie\mathcal{P}})\right|>\frac{\varepsilon}{4}$. Define $\mathcal{P}^*:=\{C\cap X,C\setminus X:C\in\mathcal{P}\}$. Then $\theta\left(W^{\bowtie\mathcal{P}^*}\right)>\theta\left(W^{\bowtie\mathcal{P}}\right)+b_\varepsilon$.*

Proof. By Proposition 3.9, we have $t\left(C_4,W^{\bowtie\mathcal{P}^*}\right)>t\left(C_4,W^{\bowtie\mathcal{P}}\right)+\frac{\varepsilon^4}{25600}$. By Lemma 2.6, we have

$$\delta_\square\left(W^{\bowtie\mathcal{P}^*},W^{\bowtie\mathcal{P}}\right)\geq\frac{\varepsilon^4}{409600}\,.$$

Finally, Proposition 3.7 gives $\theta\left(W^{\bowtie\mathcal{P}^*}\right)>\theta\left(W^{\bowtie\mathcal{P}}\right)+b_\varepsilon$, for some b_ε that depends only on ε and θ. □

Remark 3.12. *We would like to emphasize that in this section we showed that any continuous cut distance identitifying graphon parameter has a similar «pumping property» as the index, but did not obtain any new self-contained proof of the Frieze–Kannan regularity lemma.*

- *Firstly, for our proof, we need to borrow Lemma 3.9 which readily says that* some parameter *($t(C_4,\cdot)$, in this case) has the pumping property, and the existence of*

any one such parameter already allows to run the proof scheme [FK1]-[FK4]. This step was needed to infer Proposition 3.11, and it would be interesting to have a direct argument for this.

- *Secondly, we used the compactness of the space $(\mathcal{W}_0, \delta_\square)$, which is actually known to be equivalent to the Frieze–Kannan regularity lemma,* [31].

We believe that the same setting can be used in the setting of the Szemerédi regularity lemma. We pose this as a problem.

Conjecture 3.13. *Each cut distance identifying graphon parameter that is continuous with respect to the cut distance can be used as an «index» in the Szemerédi regularity lemma.*

The difficulty here is to provide a counterpart to Proposition 3.11 in the setting of the Szemerédi regularity lemma. That is, (without explaining all the notation) we do not have a single set X witnessing large cut norm but rather many witnesses of irregularity on individual pairs of clusters, none of them being substantial in the global sense of the cut norm.

3.3. **Revising the parameter $\mathrm{INT}_f(\cdot)$.** Recall that in [13], the parameter $\mathrm{INT}_f(\cdot)$ (for a strictly convex continuous function $f : [0,1] \to \mathbb{R}$) was used to identify cut distance limits of sequences of graphons (thus providing a new proof of Theorem 1.1). One of the key steps in [13] was to show that a certain refinement of a graphon leads to an increase of $\mathrm{INT}_f(\cdot)$. While not approached this way in [13], this hints that $\mathrm{INT}_f(\cdot)$ is cut distance identifying. We prove this statement in the current section, as a quick application of the results from [12, Section 4.4]. Also, here we show that the requirement of continuity of f was just an artifact of the proof in [13].

Theorem 3.14. *(a) Suppose that $f : [0,1] \to \mathbb{R}$ is a convex function. Then $\mathrm{INT}_f(\cdot)$ is cut distance compatible.*

(b) Suppose that $f : [0,1] \to \mathbb{R}$ is a strictly convex function. Then $\mathrm{INT}_f(\cdot)$ is cut distance identifying.

Proof of Part (a). Recall that every convex function admits left and right derivatives which are both increasing functions. The key is to observe that for a graphon Γ, we have $\mathrm{INT}_f(\Gamma) = \int_{x\in[0,1]} f(x)\, \mathrm{d}\mathbf{\Phi}_\Gamma$, where $\mathbf{\Phi}_\Gamma$ is defined by (2.12). Suppose that $U \preceq W$. By Proposition 2.22, we have that $\mathbf{\Phi}_U$ is at least as flat as $\mathbf{\Phi}_W$. Let Λ be a measure on $[0,1]^2$ as in Definition 2.21 that witnesses this fact. If Λ is carried by the diagonal of $[0,1]^2$ then $\mathbf{\Phi}_U = \mathbf{\Phi}_W$. In that case $U \not\prec W$ by Proposition 2.22. In other words, $\langle U\rangle = \langle W\rangle$. By Fact 2.18, we have $\delta_\square(U,W) = 0$. Since θ is a graphon parameter, we conclude that $\theta(U) = \theta(W)$. So it remains to consider the case when Λ is not carried by the diagonal.

Then there are intervals $[a,b],[c,d] \subset [0,1]$ with $\Lambda([a,b]\times[c,d]) > 0$ and $b < c$ (the other case when $d < a$ is similar).

Fix $\varepsilon > 0$ and note that f is continuous on the open interval $(0,1)$ by convexity, thus the points 0 and 1 are the only possible points of discontinuity of f. So for every $x \in (0,1)$ there is an interval $J_x \subset (0,1)$ containing x such that every two values of f on J_x differ by at most ε. Take a covering of $(0,1)$ consisting of at most countably many such intervals, add the singletons $\{0\}$ and $\{1\}$, and then refine the resulting family to a countable disjoint covering $\{J_1, J_2, \ldots\}$ of $[0,1]$. Then for every i and for every $x \in J_i$ we have $|f(x) - f(x_i)| \leq \varepsilon$ where x_i is the $\mathbf{\Phi}_U$-mean value of x on J_i, i.e., (by (2.14))

(3.10)
$$x_i = \frac{1}{\mathbf{\Phi}_U(J_i)} \int_{J_i} x \,\mathrm{d}\mathbf{\Phi}_U = \frac{1}{\Lambda(J_i\times[0,1])} \int_{J_i\times[0,1]} x \,\mathrm{d}\Lambda = \frac{1}{\Lambda(J_i\times[0,1])} \int_{J_i\times[0,1]} y \,\mathrm{d}\Lambda$$

(if for some i we have $\Phi_U(J_i) = 0$ then we can define x_i to be an arbitrary element of J_i). We may moreover assume that for every i either $J_i \subset [a,b]$ or $J_i \cap [a,b] = \emptyset$, then $x_i \in [a,b]$ whenever $J_i \subset [a,b]$. Note that convexity of f implies that

$$f(y) \geq f'_+(x_i)\cdot y + (f(x_i) - f'_+(x_i)\cdot x_i) \tag{3.11}$$

for every $y \in [c,d]$ and every i with $J_i \subset [a,b]$.

We have

$$\mathrm{INT}_f(U) = \int_{x\in[0,1]} f(x)\,\mathrm{d}\mathbf{\Phi}_U = \sum_i \int_{x\in J_i} f(x)\,\mathrm{d}\mathbf{\Phi}_U \overset{\varepsilon}{\approx} \sum_i f(x_i)\mathbf{\Phi}_U(J_i) = \sum_i f(x_i)\Lambda(J_i\times[0,1])\,.$$

We continue by employing Jensen's inequality and (3.10),

$$\begin{aligned}\sum_i f(x_i)\Lambda(J_i\times[0,1]) &\leq \sum_i \int_{(x,y)\in J_i\times[0,1]} f(y)\,\mathrm{d}\Lambda = \int_{(x,y)\in[0,1]^2} f(y)\,\mathrm{d}\Lambda \\ &= \int_{y\in[0,1]} f(y)\,\mathrm{d}\mathbf{\Phi}_W = \mathrm{INT}_f(W)\ .\end{aligned}$$

As this is true for every $\varepsilon > 0$ we conclude that $\mathrm{INT}_f(U) \leq \mathrm{INT}_f(W)$.

Proof of Part (b). Suppose that $U \prec W$ (then $\mathbf{\Phi}_U$ is strictly flatter than $\mathbf{\Phi}_W$, and so the witnessing measure Λ cannot be carried by the diagonal of $[0,1]^2$). In that case both one-sided derivatives of f are strictly increasing, and so it is easy to see that there is $\delta > 0$ such that Equation (3.11) holds in the stronger form

$$f(y) \geq f'_+(x_i)\cdot y + (f(x_i) - f'_+(x_i)\cdot x_i) + \delta \tag{3.12}$$

for every $y \in [c,d]$ and every i with $J_i \subset [a,b]$. We show that then the application of Jensen's inequality above ensures that $\mathrm{INT}_f(U) < \mathrm{INT}_f(W)$. To this end it suffices to show that there is a constant $K > 0$ not depending on ε such that

$$\sum_{i\colon J_i\subset[a,b]} f(x_i)\Lambda(J_i\times[0,1]) \le \sum_{i\colon J_i\subset[a,b]} \int_{(x,y)\in J_i\times[0,1]} f(y)\,\mathrm{d}\Lambda - K\,.$$

For every i denote $g_i(y) := f'_+(x_i)\cdot y + (f(x_i) - f'_+(x_i)\cdot x_i)$. Then we have

$$\begin{aligned}
&\sum_{i\colon J_i\subset[a,b]} \int_{(x,y)\in J_i\times[0,1]} f(y)\,\mathrm{d}\Lambda \\
=\ &\sum_{i\colon J_i\subset[a,b]} \int_{(x,y)\in J_i\times[c,d]} f(y)\,\mathrm{d}\Lambda + \sum_{i\colon J_i\subset[a,b]} \int_{(x,y)\in J_i\times([0,1]\setminus[c,d])} f(y)\,\mathrm{d}\Lambda \\
\boxed{\text{(3.12) and convexity}}\quad \ge\ &\sum_{i\colon J_i\subset[a,b]} \int_{(x,y)\in J_i\times[c,d]} (g_i(y)+\delta)\,\mathrm{d}\Lambda \\
&+ \sum_{i\colon J_i\subset[a,b]} \int_{(x,y)\in J_i\times([0,1]\setminus[c,d])} g_i(y)\,\mathrm{d}\Lambda \\
=\ &\sum_{i\colon J_i\subset[a,b]} \int_{(x,y)\in J_i\times[0,1]} g_i(y)\,\mathrm{d}\Lambda + \delta\cdot\Lambda([a,b]\times[c,d]) \\
\overset{(3.10)}{=}\ &\sum_{i\colon J_i\subset[a,b]} f(x_i)\Lambda(J_i\times[0,1]) + \delta\cdot\Lambda([a,b]\times[c,d])\,.
\end{aligned}$$

So it suffices to set $K := \delta\cdot\Lambda([a,b]\times[c,d])$. □

For a later reference, let us apply Theorem 3.14 to the strictly convex function $x \mapsto x^2$, for which $\mathrm{INT}_{x\mapsto x^2}(\cdot) = \|\cdot\|_2^2$.

Corollary 3.15. *Suppose that U and W are two graphons with $U \prec W$. Then $\|U\|_2 < \|W\|_2$.*

3.4. **Convex graphon parameters.** In Definition 3.16 we introduce convex graphon parameters. In Theorem 3.17 we prove that such parameters are cut distance compatible if they are also L^1-continuous. In Example 3.18 we observe that the opposite implication is not true.

Definition 3.16. A graphon parameter $g : \mathcal{W}_0 \to \mathbb{R}$ is *convex* if for every $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_k \in [0,1]$ with $\sum_i \alpha_i = 1$ and graphons $W, W_1, W_2, \ldots, W_k \in \mathcal{W}_0$ with $W = \sum_i \alpha_i W_i$ we have $f(W) \le \sum_i \alpha_i g(W_i)$.

Theorem 3.17. *Let $g : \mathcal{W}_0 \to \mathbb{R}$ be a graphon parameter that is convex and continuous in L^1. Then g is cut distance compatible.*

Theorem 3.17 can be used to give a third proof of a weaker version of the first part of Theorem 3.14, in which — just like the version in [13] — it is needed to require that $f : [0,1] \to \mathbb{R}$ is continuous. Indeed, the continuity of f easily implies that the graphon parameter INT_f is continuous in L^1, and the convexity of INT_f is also clear.

Now we prove Theorem 3.17.

Proof of Theorem 3.17. Suppose that $U, V : \Omega^2 \to [0,1]$ are arbitrary graphons such that $V \prec U$. Suppose that $\varepsilon > 0$ is arbitrary. Let $N(\varepsilon) \in \mathbb{N}$ and $(\phi_{\varepsilon,i})_{i=1}^N$ satisfy (2.17) for U, V and error ε (we will not use the feature (2.18) in this application of Lemma 2.24). For every $i \in [N]$ we denote the version $U^{\phi_{\varepsilon,i}}$ of U by $U_{\varepsilon,i}$. Then we have

$$\begin{aligned} g(V) &= g\left(\sum_{i=1}^{N(\varepsilon)} \frac{1}{N(\varepsilon)} U_{\varepsilon,i}\right) + \left(g(V) - g\left(\sum_{i=1}^{N(\varepsilon)} \frac{1}{N(\varepsilon)} U_{\varepsilon,i}\right)\right) \\ &\boxed{\text{convexity}} \quad \leq \sum_{i=1}^{N(\varepsilon)} \frac{1}{N(\varepsilon)} g(U_{\varepsilon,i}) + \left(g(V) - g\left(\sum_{i=1}^{N(\varepsilon)} \frac{1}{N(\varepsilon)} U_{\varepsilon,i}\right)\right) \\ &\boxed{g(U_{\varepsilon,i}) = g(U)} \quad = g(U) + \left(g(V) - g\left(\sum_{i=1}^{N(\varepsilon)} \frac{1}{N(\varepsilon)} U_{\varepsilon,i}\right)\right) . \end{aligned} \tag{3.13}$$

Now, as ε goes to 0, the graphon $\sum_{i=1}^{N(\varepsilon)} \frac{1}{N(\varepsilon)} U_{\varepsilon,i}$ goes to V in $L^1(\Omega^2)$. Thus, the L^1-continuity of g tells us that the last term in (3.13) vanishes, and thus $g(V) \leq g(U)$. Thus g is cut distance compatible. □

Example 3.18. In this example we first construct two graphons U and V such that V is a convex combination of versions of U but $V \not\preceq U$. We then use this to construct a cut distance compatible graphon parameter f^* that is not convex. The graphons U and V are shown in Figure 3.1. The graphon U is defined as $U(x,y) = 1$ if and only if $(x,y) \in [0, \frac{1}{2}]^2$ and $U(x,y) = 0$ otherwise, while $V(x,y) = \frac{1}{2}$ if and only if $(x,y) \in [0, \frac{1}{2}]^2 \cup [\frac{1}{2}, 1]^2$ and $V(x,y) = 0$ otherwise. If we set $\varphi(x) = 1 - x$, then clearly $V = \frac{U + U^\varphi}{2}$. Let us now argue that $V \not\preceq U$. For any measure preserving bijection π we have

$$\int_{[0,\frac{1}{2}]\times[\frac{1}{2},1]} U^\pi = \nu\left(\pi\left([0,\frac{1}{2}]\right) \cap [0,\frac{1}{2}]\right) \cdot \nu\left(\pi\left([0,\frac{1}{2}]\right) \cap [\frac{1}{2},1]\right) .$$

Thus, for any sequence of measure preserving bijections $\pi_1, \pi_2, \ldots$ such that $U^{\pi_n} \overset{w*}{\to} V$ we have (after passing to a subsequence if necessary) either

$$\nu\left(\pi_n\left([0,\frac{1}{2}]\right) \cap [0,\frac{1}{2}]\right) \to 0$$

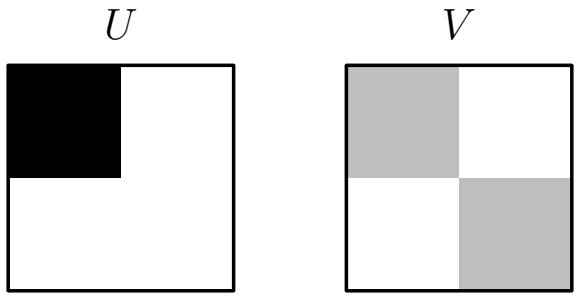


FIGURE 3.1. Graphons U and V from Example 3.18

or

$$\nu\left(\pi_n\left([0,\frac{1}{2}]\right)\cap[0,\frac{1}{2}]\right)\to\frac{1}{2}\,.$$

This is clearly a contradiction.

Now, take any cut distance compatible parameter f and suppose that it is convex. In particular, we have that $\frac{1}{2}f\left(U\right)+\frac{1}{2}f\left(U^{\varphi}\right)\geq f\left(V\right)$ for the two graphons U and V defined above. We can now define

$$f^{*}\left(W\right)=f\left(W\right)+\left(\frac{1}{2}f\left(U\right)+\frac{1}{2}f\left(U^{\varphi}\right)-f\left(V\right)+1\right)$$

for each graphon W such that $W\succeq V$ and

$$f^{*}(W)=f(W)$$

otherwise. The graphon parameter f^{*} is clearly cut distance compatible, but no longer convex, since

$$f^{*}\left(V\right)=\frac{1}{2}f\left(U\right)+\frac{1}{2}f\left(U^{\varphi}\right)+1>\frac{1}{2}f^{*}(U)+\frac{1}{2}f^{*}(U^{\varphi})\,.$$

This example works even if we restrict ourselves to graphons lying in the envelope of a certain fixed graphon W, since if we set $W\left(x,y\right)=1$ if and only if $(x,y)\in[0,\frac{1}{4}]^{2}\cup[\frac{1}{4},\frac{1}{2}]^{2}$ and $W(x,y)=0$ otherwise, and set $U'=\frac{U}{2},V'=\frac{V}{2}$, then we have three graphons U',V',W such that $U',V'\preceq W$, $V'=\frac{U'+U'^{\varphi}}{2}$, but $V'\not\preceq U'$.

The function f^{*} from Example 3.18 is, however, very unnatural since it is not continuous with respect to L^{1} (for a continuous parameter f at least). We leave it as an open problem, whether there is a continuous example.

Problem 3.19. Is there a graphon parameter $f:\mathcal{W}_{0}\to\mathbb{R}$ that is not convex, but is continuous in L^{1} and cut distance compatible?

Remark 3.20. *For homomorphism densities $t(H,\cdot) : \mathcal{W}_0 \to \mathbb{R}$, Theorem 3.17 can be reversed, under the additional assumption that H is a connected graph: $t(H,\cdot)$ is cut distance compatible if and only if it is convex. Let us give details of the direction not covered by Theorem 3.17. Suppose that $t(H,\cdot)$ is cut distance compatible, and H is connected. By Theorem 3.25 below, H is weakly norming. In particular, for the function $f : \mathcal{W}^+ \to \mathbb{R}$, $f(U) := t(H,U)^{1/e(H)}$ we have that $f(U_1+U_2) \leq f(U_1)+f(U_2)$. Now, for every $\alpha_1,\alpha_2,\alpha_3,\ldots,\alpha_k \in [0,1]$ with $\sum_i \alpha_i = 1$ and every graphons $W, W_1, W_2, \ldots, W_k \in \mathcal{W}_0$ with $W = \sum_i \alpha_i W_i$, we have*

$$t(H,W) = f\left(\sum_i \alpha_i W_i\right)^{e(H)} \leq \left(\sum_i f(\alpha_i W_i)\right)^{e(H)}$$
$$= \left(\sum_i \alpha_i f(W_i)\right)^{e(H)} \leq \sum_i \alpha_i \left(f(W_i)\right)^{e(H)} = \sum_i \alpha_i t(H,W_i) \ ,$$

as required.

3.5. **Spectrum.** The main result in this section, Theorem 3.22, asserts that the spectral quasiorder defined in Section 2.3.4 is a cut distance identifying graphon order. But first we need an easy lemma.

Lemma 3.21. *Let $(W_n)_n$ be a sequence of graphons on Ω^2 such that $W_n \xrightarrow{w^*} U$ for some graphon U. Let $u, v \in L^2(\Omega)$. Then we have $\langle W_n u, v\rangle \to \langle Uu, v\rangle$.*

Proof. Since step functions are dense in $L^2(\Omega)$, and since the forms $\langle W_n\cdot,\cdot\rangle$ and $\langle U\cdot,\cdot\rangle$ are obviously bilinear, it suffices to prove the statement for indicator functions of sets, $u = \mathbf{1}_A$, $v = \mathbf{1}_B$ (where $A, B \subset \Omega$). But in that case $\langle W_n u, v\rangle = \int_{A\times B} W_n$ and $\langle Uu, v\rangle = \int_{A\times B} U$. The statement follows since $W_n \xrightarrow{w^*} U$. □

We are now ready to prove the main result of this section. Let us note that the arguments that we use to prove this result also turned out to be useful in the setting of finitely forcible graphs; in particular Kráľ, Lovász, Noel, and Sosnovec [24] used our arguments in the final step of their proof that for each graphon and each $\varepsilon > 0$, there exists a finitely forcible graphon that differs from the original one only on a set of measure at most ε.

Theorem 3.22. *The spectral quasiorder is a cut distance identifying graphon order. That is, given two graphons $U, W \in \mathcal{W}_0$,*

(a) if $\delta_\square(U,W) = 0$, then the spectra of U and W are the same, and
(b) if $U \prec W$, then $U \overset{S}{\prec} W$.

Proof. Part (a) follows from [29, Theorem 11.54].

So, the main work is to prove (b). Consider the sequence $(W^{\pi_n})_n$ of versions of W such that $W^{\pi_n} \xrightarrow{\text{w}^*} U$. Let $\lambda_1^+ \geq \lambda_2^+ \geq \lambda_3^+ \geq \ldots \geq 0$ be the positive eigenvalues of U with associated pairwise orthogonal unit eigenvectors $u_1, u_2, u_3, \ldots$, and let $\beta_1^+ \geq \beta_2^+ \geq \beta_3^+ \geq \ldots \geq 0$ be the positive eigenvalues of W. First, we will prove that for any given $\varepsilon > 0$ and k, we have $\beta_k^+ \geq \lambda_k^+ - \varepsilon$. By the maxmin characterization of eigenvalues we have

$$\beta_k^+ = \max_{\substack{H \text{ subspace of } L^2(\Omega) \\ \dim(H)=k}} \min_{\substack{g \in H \\ \|g\|_2=1}} \langle Wg, g\rangle \ . \tag{3.14}$$

Given n, consider the space $\widetilde{H}_n = \text{span}\left\{u_1^{\pi_n^{-1}}, u_2^{\pi_n^{-1}}, \ldots, u_k^{\pi_n^{-1}}\right\}$, where $u_i^{\pi_n^{-1}}(x) = u_i(\pi_n^{-1}(x))$. Then (3.14) gives

$$\beta_k^+ \geq \min_{\substack{g \in \widetilde{H}_n \\ \|g\|_2=1}} \langle Wg, g\rangle \, . \tag{3.15}$$

Furthermore, by Lemma 3.21 we can find n large enough so that for all $i, j \leq k$ we have

$$|\langle W^{\pi_n} u_i, u_j\rangle - \langle Uu_i, u_j\rangle| < \frac{\varepsilon}{k^2}. \tag{3.16}$$

Now, for $g \in \widetilde{H}_n$ that realizes the minimum in (3.15), we can write its orthogonal decomposition as $g = \sum_{i=1}^k c_i u_i^{\pi_n^{-1}}$, where $\sum_{i=1}^k c_i^2 = 1$. Thus, we obtain

$$\langle Wg, g\rangle = \langle W^{\pi_n} g^{\pi_n}, g^{\pi_n}\rangle = \left\langle W^{\pi_n} \sum_{i=1}^k c_i u_i, \sum_{i=1}^k c_i u_i \right\rangle = \sum_{i,j=1}^k c_i c_j \langle W^{\pi_n} u_i, u_j\rangle \ .$$

We can now use (3.16) to replace the terms $\langle W^{\pi_n} u_i, u_j\rangle$ by the terms $\langle Uu_i, u_j\rangle$,

$$\begin{aligned}\langle Wg, g\rangle &= \sum_{i=1}^k c_i^2 \left(\langle Uu_i, u_i\rangle \pm \frac{\varepsilon}{k^2}\right) + \sum_{\substack{i,j=1 \\ i\neq j}}^k c_i c_j \left(\langle Uu_i, u_j\rangle \pm \frac{\varepsilon}{k^2}\right) \\ &\geq \sum_{i=1}^k c_i^2 \left(\lambda_i^+ - \frac{\varepsilon}{k^2}\right) - \sum_{\substack{i,j=1 \\ i\neq j}}^k |c_i c_j| \frac{\varepsilon}{k^2} \geq \lambda_k^+ - \varepsilon \ .\end{aligned}$$

Thus (3.15) implies $\beta_k^+ \geq \lambda_k^+ - \varepsilon$.

A similar argument can be used for the negative eigenvalues $\lambda_1^- \leq \lambda_2^- \leq \lambda_3^- \leq \ldots \leq 0$ of U and $\beta_1^- \leq \beta_2^- \leq \beta_3^- \leq \ldots \leq 0$ of W to show that $\beta_k^- \leq \lambda_k^- + \varepsilon$. That implies $U \overset{S}{\preceq} W$.

To show that for at least one eigenvalue the corresponding inequality is strict, assume for contradiction that the eigenvalues of U and W are all the same. Then a double application of (2.5) gives

$$\|W\|_2^2 = \sum \left(\beta_i^+\right)^2 + \sum \left(\beta_i^-\right)^2 = \sum \left(\lambda_i^+\right)^2 + \sum \left(\lambda_i^-\right)^2 = \|U\|_2^2 \, .$$

But this is a contradiction with Corollary 3.15. This finishes the proof. □

3.6. **Homomorphism densities.** In this section, we address the following problem.

Problem 3.23. Characterize graphs H for which $t(H,\cdot) : \mathcal{W}_0 \to \mathbb{R}$ is a cut distance compatible (respectively a cut distance identifying) graphon parameter.

Observe that thanks to Proposition 3.2, for the case of compatible graphon parameters, Problem 3.23 reduces to characterizing graphs H for which we have

$$t\left(H, W^{\bowtie\mathcal{P}}\right) \leq t\left(H, W\right) \text{ for each } W \in \mathcal{W}_0 \text{ and each finite partition } \mathcal{P}. \tag{3.17}$$

Similarly, if true, our Conjecture 3.3 implies that for the case of identifying graphon parameters, Problem 3.23 reduces to characterizing graphs H for which we have

(3.18)
$$t\left(H, W^{\bowtie\mathcal{P}}\right) < t\left(H, W\right) \text{ for each } W \in \mathcal{W}_0 \text{ and each finite partition } \mathcal{P} \text{ for which } W \neq W^{\bowtie\mathcal{P}}.$$

This is closely related to Sidorenko's conjecture (which was asked independently by Simonovits, and by Sidorenko, [36, 35]) and the Forcing conjecture (first hinted in [37, Section 5]). Indeed, these conjectures — when stated in the language of graphons — ask to characterize graphs H for which we have

$$t\left(H, W^{\bowtie\{\Omega\}}\right) \leq t\left(H, W\right) \text{ for each } W \in \mathcal{W}_0 \tag{3.19}$$
(Sidorenko's conjecture), and
$$t\left(H, W^{\bowtie\{\Omega\}}\right) < t\left(H, W\right) \text{ for each nonconstant } W \in \mathcal{W}_0 \tag{3.20}$$
(Forcing conjecture).

Recall that Sidorenko's conjecture asserts that H satisfies (3.19) if and only if H is bipartite. Similarly, the Forcing conjecture asserts that H satisfies (3.20) if and only if H is bipartite and contains a cycle. In both cases, the $\Rightarrow$ direction is easy. Let us recall that the reason why at least one cycle is required for the Forcing conjecture is that the homomorphism density of any forest H in any p-regular graphon (whether constant-p, or not) is $p^{e(H)}$. The other direction in both conjectures is open, despite being known in many special cases, see [7, 28, 23, 5, 22, 27, 38, 8, 6].

Because all the properties we investigate in this section strengthen (3.19), we are concerned only with bipartite graphs throughout. The only exception is Remark 3.31 which addresses a possible «converse» definition of cut distance identifying properties.

Graphs satisfying (3.17) were investigated in [25] where these graphs are said to have the *step Sidorenko property*. Similarly, graphs satisfying (3.18) are said to have the *step forcing property*. Clearly, these properties imply (3.19) and (3.20), respectively. These stronger

«step» properties do not follow automatically from (3.19) and (3.20); in [25, Section 2] it is shown that the 4-cycle with a pendant edge C_4^+ has the Sidorenko property but not the step Sidorenko property. Thus, every graph having the step Sidorenko property must be bipartite and every graph having the step forcing property must be bipartite with a cycle. The focus of [25] was in providing negative examples. For example, it was shown in [25] that a Cartesian product of cycles does not have the step Sidorenko property, unless all the cycles have length 4.

The connection to our running Problem 3.23 comes from Proposition 14.13 of [29] which implies that each weakly norming graph has the step Sidorenko property.

Corollary 3.24. *For each weakly norming graph H the function $t(H,\cdot)$ is cut distance compatible (or, equivalently, H has the step Sidorenko property).*

Corollary 3.24 also directly follows from Theorem 3.17. We recall the proof from [29] in Section 3.6.1.

In Section 3.6.2 we prove Theorem 3.25 which states that among connected graphs, the graphs with the step Sidorenko property are exactly the weakly norming graphs (thus answering a question of Kráľ, Martins, Pach and Wrochna [25, Section 5]).

Theorem 3.25. *Suppose that H is a connected graph. If the function $t(H,\cdot)$ is cut distance compatible (or, equivalently, if H has the step Sidorenko property), then H is weakly Hölder.*

Remark 3.26. *For disconnected graphs, the statement of Corollary 3.24 actually does not require the graph to be weakly norming, and can be strengthened as follows.* If each component H_i of a graph $H = H_1 \sqcup \ldots \sqcup H_k$ is weakly norming then $t(H,\cdot)$ is cut distance compatible. *Indeed, suppose that $U \preceq W$. Then Corollary 3.24 tells us that for each component, $t(H_i,U) \le t(H_i,W)$. Thus, $t(H,U) = \prod t(H_i,U) \le \prod t(H_i,W) = t(H,W)$, as was needed. We remark, that this relation between weakly norming disconnected graphs and cut distance compatibility might perhaps be an equivalence.*

Remark 3.27. *Two nontrivial necessary conditions for a graph H to be weakly Hölder are established in* [22, Theorem 2.10]. *One of them basically says that H does not contain a subgraph denser than itself. The other condition says that if $V(H) = A_1 \sqcup A_2$ is a bipartition of H and $u,v \in A_i$ are two vertices from the same part, then* $\deg(u) = \deg(v)$. *Thus, Theorem 3.25 restricts quite substantially the class of graphs having the step Sidorenko property, compared to the class of all bipartite graphs which are conjectured to have the Sidorenko property. In particular, we see directly that C_4^+ does not have the step Sidorenko property.*

The next theorem, which we prove in Section 3.6.3, is our another main result.

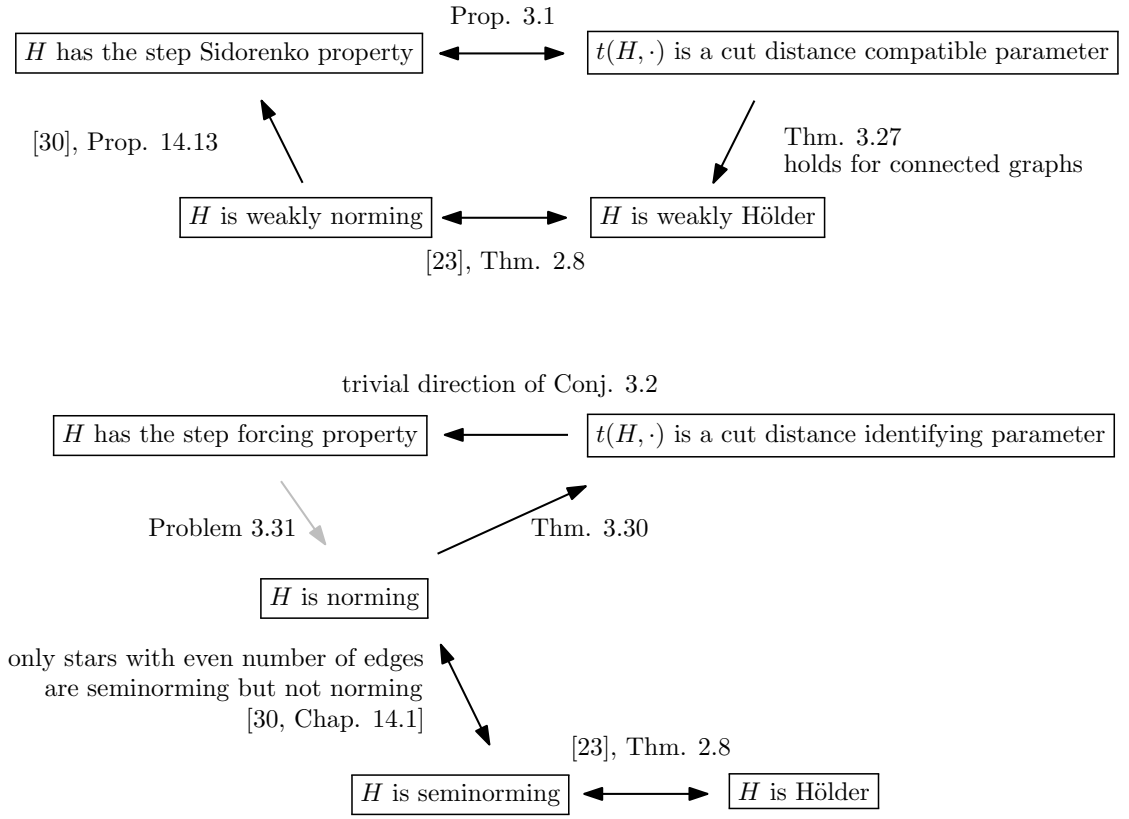


FIGURE 3.2. Diagram of notions regarding homomorphism densities used in this paper and their relations.

Theorem 3.28. *Suppose that H is a norming graph. Then the parameter $t(H,\cdot)$ is cut distance identifying. In particular, by the trivial direction of Conjecture 3.3, H is step forcing.*

Note that Theorem 3.28 is an implication only. It is reasonable to ask about the converse (for connected graphs, for the same reasons as in Remark 3.26).

Problem 3.29. Is it true that if a connected graph H has the step forcing property (or if $t(H,\cdot)$ is cut distance identifying, which may be a more restrictive assumption), we also have that H is norming?

Before showing the proofs of Theorem 3.25 and Theorem 3.28, we summarize in Figure 3.2 the known and conjectured relations for weakly norming graphs, norming graphs, graphs with the step Sidorenko or the step forcing property, and graphs that give cut distance compatible or cut distance identifying parameters.

3.6.1. *Proof of Corollary 3.24.* Here, we prove that each weakly norming graph H has the step Sidorenko property. Our argument is a tailored version of the proof of Proposition 14.13 of [29] (where the statement is proven in bigger generality, for so-called smooth invariant norms). The reason why we recall this argument is that it will allow us to understand the strategy for proving Theorem 3.28, as we explain at the end of this section.

So, suppose that H is a weakly norming graph, $W : \Omega^2 \to [0,1]$ is a graphon, $\mathcal{P}$ is a finite partition of Ω. We need to prove that $t(H,W) \geq t\left(H, W^{\ltimes\mathcal{P}}\right)$. Without loss of generality, we can assume that $\Omega = [0,1)$ (see Remark 2.4), and that $\mathcal{P}$ is a partition into intervals $\{[a_\ell, a_{\ell+1})\}_{\ell=1}^{|\mathcal{P}|}$. Fix a number η which is irrational with respect to the lengths of all the intervals $[a_\ell, a_{\ell+1})$. Consider the map $\gamma : [0,1) \to [0,1)$ that maps each number $x \in [0,1)$, say $x \in [a_i, a_{i+1})$, to $((x - a_i + \eta)) \mod (a_{i+1} - a_i) + a_i$. Clearly, the map γ is a measure preserving bijection on $[0,1)$, where each interval $[a_\ell, a_{\ell+1})$ is γ-invariant, and γ restricted to each $[a_\ell, a_{\ell+1})$ is ergodic. It follows that the map $(x,y) \mapsto (\gamma(x), \gamma(y))$ is ergodic when restricted on each set of the form $[a_\ell, a_{\ell+1}) \times [a_k, a_{k+1})$.

For $n \in \mathbb{N}$, let U_n be the version of W obtained using the n-th iteration of γ, $U_n := W^{\gamma^n}$. For $n \in \mathbb{N}$, let $S_n := \frac{1}{n}\sum_{k=1}^n U_k$. The Pointwise Ergodic Theorem tells us that the graphons S_n converge pointwise to $W^{\ltimes\mathcal{P}}$. Hence,

$$\sqrt[e(H)]{t\left(H, W^{\ltimes\mathcal{P}}\right)} = \lim_{n\to\infty} \sqrt[e(H)]{t\left(H, S_n\right)} = \lim_{n\to\infty} \sqrt[e(H)]{t\left(H, \frac{1}{n}\sum_{k=1}^n U_k\right)}$$

$$\boxed{\sqrt[e(H)]{t\left(H, |\cdot|\right)} \text{ is a seminorm}} \quad \leq \lim_{n\to\infty} \frac{1}{n} \cdot \sum_{k=1}^n \sqrt[e(H)]{t\left(H, U_k\right)}$$

$$\boxed{t\left(H, U_k\right) = t\left(H, W\right)} \quad = \sqrt[e(H)]{t\left(H, W\right)},$$

as was needed. This finishes the proof of Corollary 3.24.

Let us now look back at the argument to see what needs to be strengthened to give Theorem 3.28. Actually, in this informal sketch, we only want to show that H is step forcing, rather than $t(H,\cdot)$ being cut distance identifying.

That is, we have a norming graph H, a graphon $W : \Omega^2 \to [0,1]$, a finite partition $\mathcal{P}$ of Ω, such that $W \neq W^{\ltimes\mathcal{P}}$. We need to prove that $t(H,W) > t\left(H, W^{\ltimes\mathcal{P}}\right)$. The only space for getting the needed strict inequality in the calculation above is in the triangle inequality on the second line. In view of Remark 2.1, and using the fact that H is norming and hence $\|\cdot\|_H$ uniformly convex by Theorem 2.16, it only remains to argue that many of the graphons U_k are far from colinear, where «far from colinear» is measured in the

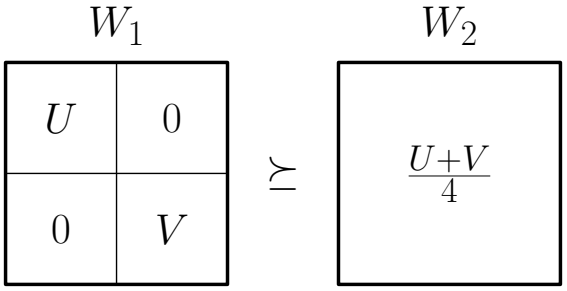


FIGURE 3.3. Graphons U, V, W_1 and W_2 from the proof of Theorem 3.25

$\|\cdot\|_H$-norm.[l] This is indeed plausible: since $W \neq W^{\ltimes\mathcal{P}}$, the graphons U_k must indeed be different.

As we shall see in Section 3.6.3, there are several difficulties in the actual proof of Theorem 3.28. In particular, we were not able to show that this approach works with $U_k = W^{\gamma^k}$, and needed to choose the approximating graphons U_k with the help of (rather technical) Lemma 2.24.

3.6.2. *Proof of Theorem 3.25.* Let H be a connected graph such that $t(H,\cdot)$ is cut distance compatible. Suppose that H has m edges and n vertices. We prove that H is weakly Hölder. By Theorem 2.14 we already know that weakly Hölder graphs are exactly weakly norming graphs. We divide the proof of the theorem into two parts. At first we prove that $t(H,\cdot)$ is subadditive up to a constant loss, specifically, we show that

$$t(H,U)^{1/m} + t(H,V)^{1/m} \geq \frac{1}{4}\cdot t(H,U+V)^{1/m}\,. \tag{3.21}$$

Then we use this inequality to prove that H is weakly norming using the tensoring technique in the same way as it is used in the proof of Theorem 2.8 from [22].

Let U and V be two arbitrary graphons and let W_1 be a graphon containing a copy of U scaled by the factor of one half in its top-left corner (i.e., $W_1(x,y) = U(2x,2y)$ for $(x,y) \in [0,\frac{1}{2}]^2$), a copy of V in its bottom-right corner (i.e., $W_1(x,y) = V\left(2(x-\frac{1}{2}),2(y-\frac{1}{2})\right)$ for $(x,y) \in [\frac{1}{2},1]^2$), and zero otherwise (see Figure 3.3). Note that for the homomorphism density $t(H,W_1)$ we have

$$t(H,W_1) = \frac{t(H,U)+t(H,V)}{2^n}\,.$$

This is because H is connected and, thus, homomorphisms that map a positive number of vertices of H to $[0,\frac{1}{2}]$, and a positive number of vertices to $[\frac{1}{2},1]$ do not contribute

[l]Note that we need to be careful about quantification: for example having just one graphon U_k to be «somewhat far» from the others, could result in a strict triangle inequality that would disappear when taking $\lim_{n\to\infty}$. So, we really need «many» graphons that are «uniformly far from colinear».

to the value of the integral $t(H, W_1)$. Now consider the graphon $W_2 = \frac{U+V}{4}$. By [12, Lemma 4.2] we have $W_1 \succeq W_2$. It follows that $t(H, W_1) \geq t(H, W_2)$. Observe that $t(H, W_2) = t\left(H, \frac{U+V}{4}\right) = \frac{t(H,U+V)}{4^m}$, hence we get

$$\frac{t(H,U) + t(H,V)}{2^n} \geq \frac{t(H,U+V)}{4^m} .$$

We are actually interested in the quantity $t(H,U)^{1/m}$, so we rewrite this as

$$(t(H,U) + t(H,V))^{1/m} \geq \frac{2^{n/m}}{4} \cdot t(H,U+V)^{1/m} \geq \frac{1}{4} \cdot t(H,U+V)^{1/m} .$$

Finally note that $t(H,U)^{1/m} + t(H,V)^{1/m} \geq (t(H,U) + t(H,V))^{1/m}$, as can be verified by raising the inequality to the m-th power. This yields the desired inequality (3.21).

Now we need to improve the constant on the right-hand side of (3.21). To this end we use the tensor power trick in the same way as was used in [22, Theorem 2.8] and in [29, Theorem 14.1].

Note that the inequality (3.21) can be inductively generalised to yield that for a sequence of graphons $U_1, \ldots, U_\ell$ we have

$$\sum_{i=1}^{\ell} t(H,U_i)^{1/m} \geq \left(\frac{1}{4}\right)^{\ell-1} \cdot t\left(H, \sum_{i=1}^{\ell} U_i\right)^{1/m} . \tag{3.22}$$

Let (H,w) be a $\mathcal{W}^+$-decoration of H. By Lemma 2.12 we may assume that $t(H, W_e) = 1$ for every W_e. We want to prove that $t(H,w) \leq 1$, but at first we prove the weaker inequality $t(H,w) \leq 4^{m(m-1)} \cdot m^m$. We have

$$t(H,w) \leq t\left(H, \sum_{e\in E(H)} W_e\right) \leq \left(4^{m-1} \cdot \sum_{e\in E(H)} t(H,W_e)^{1/m}\right)^m = 4^{m(m-1)} \cdot m^m , \tag{3.23}$$

where in the first inequality we replaced each W_e by $\sum_{e\in E(H)} W_e$, while the second inequality is due to the bound (3.22). Now suppose that we decorate each edge e of H by $W_e^{\otimes k}$ for $k \geq 1$. As we observed in (2.4) and (2.3), we then have $t\left(H, w^{\otimes k}\right) = t(H,w)^k$ and $t\left(H, W_e^{\otimes k}\right) = t\left(H, W_e\right)^k = 1$. Thus inequality (3.23) gives that $t(H,w)^k = t(H, w^{\otimes k}) \leq 4^{m(m-1)} \cdot m^m$, thus $t(H,w) \leq \left(4^{m(m-1)} \cdot m^m\right)^{1/k}$. Since this holds for any $k \geq 1$, we conclude that $t(H,w) \leq 1$.

3.6.3. *Proof of Theorem 3.28.* Let H be a norming graph. By Fact 2.13, it is enough to consider the case that H is connected. Suppose that H has m edges. Since $m > 1$ (cf. Fact 2.11(b)), we know from Section 2.4.1 that the modulus of convexity $\mathfrak{d}_H$ of the norm $\|\cdot\|_H$ defined by (2.10) is strictly positive.

Suppose that $U, V \in \mathcal{W}^+$ are such that $V \prec U$. We want to prove that $t(H,U) > t(H,V)$. Clearly, we may assume that $t(H,V) > 0$. By rescaling, we may moreover

assume that $t(H,U)=1$. Let us define $\delta := \frac{\delta_\square(U,V)}{32}$. We now set

$$\xi := \min\left(\mathfrak{d}_H(\delta), \|V\|_H\right) \quad \text{and} \quad \varepsilon := \frac{1}{4m\cdot\|U\|_\infty^{m-1}}\cdot\left(\frac{1}{10}\cdot\xi\right)^m .$$

Let an even number N and measure preserving bijections $(\phi_i)_{i=1}^N$ be given by Lemma 2.24 (see also Remark 2.25) for the inputs U and V. Since $V \preceq U$, we have $\|V\|_\infty \le \|U\|_\infty$. Since $\left\|V - \frac{\sum_{i=1}^N U^{\phi_i}}{N}\right\|_1 < \varepsilon$, we have

$$\begin{aligned}(3.24)\qquad \left|t\left(H, \frac{\sum_{i=1}^N U^{\phi_i}}{N}\right) - t(H,V)\right| &\overset{L2.6}{\le} 4m\cdot\|U\|_\infty^{m-1}\cdot\delta_\square\left(\frac{\sum_{i=1}^N U^{\phi_i}}{N}, V\right)\\ &\le 4m\cdot\|U\|_\infty^{m-1}\cdot\left\|V - \frac{\sum_{i=1}^N U^{\phi_i}}{N}\right\|_1 < \left(\frac{\xi}{10}\right)^m .\end{aligned}$$

Observe that for each index $i \in \left\{1,2,\dots,\frac{N}{2}\right\}$ that satisfies (2.18), an application of Proposition 2.15 to the kernel $X := U^{\phi_{2i-1}} - U^{\phi_{2i}}$ gives that

$$\|U^{\phi_{2i-1}} - U^{\phi_{2i}}\|_H \ge \|U^{\phi_{2i-1}} - U^{\phi_{2i}}\|_\square \ge \delta .$$

Since $\|U\|_H = 1$, any two versions of U that are far apart can play the role of x and y in (2.1). Thus, if we have an index $i \in \left\{1,2,\dots,\frac{N}{2}\right\}$ that satisfies (2.18), then we get that

$$1 - \sqrt[m]{t\left(H, \frac{U^{\phi_{2i-1}} + U^{\phi_{2i}}}{2}\right)} \ge \mathfrak{d}_H(\delta) \ge \xi.$$

Since $t(H,U)=1$, we may equivalently write

$$(3.25)\qquad \sqrt[m]{t(H,U^{\phi_{2i-1}})} + \sqrt[m]{t(H,U^{\phi_{2i}})} \ge \sqrt[m]{t(H,U^{\phi_{2i-1}} + U^{\phi_{2i}})} + 2\xi .$$

We are now in a position to do the the final calculation. We have

$$(3.26)\qquad \sqrt[m]{t(H,U)} = \frac{1}{N}\left(\sqrt[m]{t\left(H,U^{\phi_1}\right)} + \sqrt[m]{t\left(H,U^{\phi_2}\right)} + \cdots + \sqrt[m]{t\left(H,U^{\phi_N}\right)}\right) .$$

Let us now group the summands on the right-hand side of (3.26) into $\frac{N}{2}$ pairs $\sqrt[m]{t\left(H,U^{\phi_{2i-1}}\right)} + \sqrt[m]{t\left(H,U^{\phi_{2i}}\right)}$. Either a given pair satisfies (3.25), or, if not, the subadditivity of $\|\cdot\|_H$ gives us somewhat weaker $\sqrt[m]{t(H,U^{\phi_{2i-1}})} + \sqrt[m]{t(H,U^{\phi_{2i}})} \ge \sqrt[m]{t(H,U^{\phi_{2i-1}} + U^{\phi_{2i}})}$. Recall that

at least $\frac{N}{4}$ pairs satisfy (3.25). Thus,

$$\begin{aligned}
\sqrt[m]{t(H,U)} &\geq \frac{1}{N}\left(\sqrt[m]{t\left(H,U^{\phi_1}+U^{\phi_2}\right)}+\cdots+\sqrt[m]{t\left(H,U^{\phi_{N-1}}+U^{\phi_N}\right)}+\frac{N}{4}\cdot 2\xi\right)\\
\boxed{\text{subadditivity of } \|\cdot\|_H} \quad &\geq \frac{1}{N}\sqrt[m]{t\left(H,U^{\phi_1}+U^{\phi_2}+\cdots+U^{\phi_N}\right)}+\frac{\xi}{2}\\
&= \sqrt[m]{t\left(H,\frac{\sum_{i=1}^N U^{\phi_i}}{N}\right)}+\frac{\xi}{2}\\
\boxed{\text{by (3.24); } t(H,V)>(\tfrac{\xi}{10})^m} \quad &\geq \sqrt[m]{t(H,V)-\left(\frac{\xi}{10}\right)^m}+\frac{\xi}{2}\\
&\geq \sqrt[m]{t(H,V)}-\frac{\xi}{10}+\frac{\xi}{2}\\
&> \sqrt[m]{t(H,V)}\,,
\end{aligned}$$

as was needed.

3.6.4. *Discussion.*

Remark 3.30. *Let us comment on the role of the abstract weak* approach we introduced in* [12] *in our proofs of Theorem 3.25 and Theorem 3.28.*

- *In Theorem 3.25 we deduced the property of being weakly Hölder by using inequalities of the form $t(H,W_1)\geq t(H,W_2)$, where W_1 and W_2 are constructed from graphons U and V as in Figure 3.3. These inequalities follow from the fact that $W_1\succeq W_2$; but we do not have that W_2 is a stepping of W_1. In other words, the immediate property of H we use is that $t(H,\cdot)$ is cut distance compatible, rather than H having the step Sidorenko property. (Of course, the two properties are equivalent, by Proposition 3.2.) So, the weak* approach and the notion of the structuredness order were instrumental here.*
- *Theorem 3.28 cannot be even stated without the notion of the structuredness order (until the validity of Conjecture 3.3 is confirmed). On the other hand, a version of Theorem 3.28 which would not use the structuredness order, «Suppose that H is a norming graph. Then H is step forcing.» is either equivalent (if Conjecture 3.3 holds), or only a tiny bit weaker. Our proof of Theorem 3.28 could then be easily modified so that it avoids any notions introduced in* [12]*, that is using only traditional technology available after Hatami* [22]*. Actually, in this setting it would be enough to prove Lemma 2.24 for $V=U^{\ltimes\mathcal{P}}$ rather than $V\prec U$, which would result in a proof of that lemma shorter by one or two pages.*

Remark 3.31. *Note that the definition of cut distance compatible (resp. identifying) parameters given at the beginning of Section 3.1 was somewhat arbitrary. That is, instead*

of requiring that $W_1 \preceq W_2$ implies $\theta(W_1) \le \theta(W_2)$ (resp. that $W_1 \prec W_2$ implies $\theta(W_1) < \theta(W_2)$), we could have reversed the inequalities to $\theta(W_1) \ge \theta(W_2)$ (resp. $\theta(W_1) > \theta(W_2)$). However, among graphon parameters induced by graph densities, there are only trivial examples of cut distance compatible parameters in this sense. These correspond to the graphs that are disjoint unions of cliques on 1 and 2 vertices. For these graphs the homomorphism densities are either always constant 1 (if the graph is a disjoint union of vertices), or the power of the edge density of the graph (otherwise). Since we know that $U \succeq V$ implies that the edge densities of the two graphons are the same (Fact 2.20), these examples are cut distance compatible parameters in both senses for a trivial reason, and, in particular, they are not cut distance identifying parameters in this reverse sense. To see that there are no other examples of cut distance compatible parameters in the reverse sense, consider the two following graphons: a graphon W_{clique} consisting of a clique of measure 0.5 ($W_{\text{clique}}(x,y) = 1$ if and only if $0 \le x, y \le \frac{1}{2}$ and $W_{\text{clique}}(x,y) = 0$ otherwise), and the constant graphon $W_{\text{const}} \equiv \frac{1}{4}$. Now let H be a graph that is not a disjoint union of cliques of order one or two. Without loss of generality we may assume that H does not contain any component consisting of a single vertex. Hence $2e(H) > v(H)$. Now we have $W_{\text{const}} \preceq W_{\text{clique}}$, but $t(H, W_{\text{const}}) = \left(\frac{1}{4}\right)^{e(H)} < \left(\frac{1}{2}\right)^{v(H)} = t(H, W_{\text{clique}})$.

3.6.5. *Local Sidorenko's conjecture.* An interesting weakening of Sidorenko's conjecture is to require (3.19) only for graphons W that are close to a constant graphon. More precisely, we say that a graph H has the *local Sidorenko property with respect to the L^1-norm* (resp. *with respect to the cut norm* or *with respect to the L^∞-norm*) if for each $p \in [0,1]$ there exists an $\varepsilon > 0$ such that for each graphon W of edge density p and with $\|W - p\|_1 < \varepsilon$ (resp. with $\|W - p\|_\square < \varepsilon$ or with $\|W - p\|_\infty < \varepsilon$) we have that $t(H, W) \ge p^{e(H)}$. This weakening was first considered by Lovász [28] who proved that bipartite graphs are indeed locally Sidorenko (even with respect to the cut norm, which is the strongest of the results). Recently a full characterization of graphs with the local Sidorenko was announced by Fox and Wei [16]: a graph is locally Sidorenko if and only if it is a forest or has even girth.

We can combine the «step» and the «local» features in an obvious way. We say that a graph H has the *local step Sidorenko property* if for each partition $\mathcal{P} = (\Omega_i)_{i=1}^k$ of Ω and each template of densities $(p_{ij} \in [0,1])_{i,j \in [k]}$ there exists $\varepsilon > 0$ such that for each graphon W for which the average of W on each $\Omega_i \times \Omega_j$ equals p_{ij}, and for which $W^{\ltimes \mathcal{P}}$ is ε-close to W in some fixed norm as above, we have $t\left(H, W^{\ltimes \mathcal{P}}\right) \le t(H, W)$. Locally step forcing graphs can be defined analogously.

Problem 3.32. Characterize locally step Sidorenko and locally step forcing graphs (with respect to the norms $\|\cdot\|_1$, $\|\cdot\|_\square$, or $\|\cdot\|_\infty$).

3.6.6. *Two positive results directly.* We conclude the treatment of Problem 3.23 by two positive results, namely that stars are step Sidorenko and that even cycles are step forcing.

Propositions 3.33 and 3.34 in the case $\ell = 2$ are not new and follow from the results on weakly norming and Hölder graphs above. Yet, the short proofs given here nicely employ other parts of the theory established in this paper.

Proposition 3.33. *For each $\ell \in \mathbb{N}$, the graphon parameter $t(K_{1,\ell}, \cdot) : \mathcal{W}_0 \to \mathbb{R}$ is cut distance compatible.*

Proof. The key is to observe that for a graphon Γ, we have $t(K_{1,\ell}, \Gamma) = \int_{x\in[0,1]} x^\ell \mathrm{d}\boldsymbol{\Upsilon}_\Gamma$, where $\boldsymbol{\Upsilon}_\Gamma$ is defined by (2.13). So, suppose that $U \preceq W$. By Proposition 2.22, we have that $\boldsymbol{\Upsilon}_U$ is at least as flat as $\boldsymbol{\Upsilon}_W$. Let Λ be the measure on $[0,1]^2$ as in Definition 2.21 that witnesses this. In the following inequality, the measures Λ^1_x, $x \in [0,1]$, are given by the disintegration of the measure Λ on the first coordinate; see [12, p. 272] for a formal definition. We have

$$\begin{aligned}
t(K_{1,\ell}, U) &= \int_{x\in[0,1]} x^\ell \mathrm{d}\boldsymbol{\Upsilon}_U \\
\boxed{\text{by Lemma 4.12 from [12]}} \quad &= \int_{x\in[0,1]} \left(\int_{y\in[0,1]} y \,\mathrm{d}\Lambda^1_x \right)^\ell \mathrm{d}\boldsymbol{\Upsilon}_U \\
\boxed{\text{Jensen's inequality}} \quad &\leq \int_{x\in[0,1]} \int_{y\in[0,1]} y^\ell \,\mathrm{d}\Lambda^1_x \,\mathrm{d}\boldsymbol{\Upsilon}_U \\
\boxed{\text{definition of disintegration}} \quad &= \int_{(x,y)\in[0,1]^2} y^\ell \,\mathrm{d}\Lambda \\
&= \int_{y\in[0,1]} y^\ell \,\mathrm{d}\boldsymbol{\Upsilon}_W \\
&= t(K_{1,\ell}, W) \ .
\end{aligned}$$

□

Proposition 3.34. *For each $\ell \in \{2, 3, 4, \ldots\}$, the graphon parameter $t(C_{2\ell}, \cdot) : \mathcal{W}_0 \to \mathbb{R}$ is cut distance identifying.*

Before giving a proof, let us note that Lemma 11 in [10] is equivalent to the case $\ell = 2$ of the proposition. However, the proof in [10] does not seem to generalize to any higher ℓ, in which case Proposition 3.34 seems to be new.

Proof of Proposition 3.34. To prove the proposition, suppose that ℓ is fixed and $W_1 \prec W_2$ are two graphons. Theorem 3.22 tells us that $W_1 \overset{S}{\prec} W_2$. That is, the sum of the (2ℓ)-th powers of eigenvalues of W_1 is strictly smaller than that of W_2. The statement now follows from Equation (2.6). □

3.6.7. *A recent result of Lee and Schülke.* Combining Theorem 3.17 and Remark 3.20 (see also Figure 3.2), we get that for a connected graph H the function $t(H, \cdot)$ is convex on $\mathcal{W}_0$

if and only if H is weakly norming. A direct proof of this equivalence, together with a counterpart equivalence of the convexity of $t(H,\cdot)$ on $\mathcal{W}$ and H being norming was proven in [26]. These equivalences are then used in [26] to argue that $K_{5,5}\setminus C_{10}$ (one of the smallest graphs where Sidorenko's conjecture is open) is not weakly norming, and that $K_{t,t}$ minus a perfect matching is not norming.

4. Conclusion and possible further directions

In this paper, we studied cut distance identifying and cut distance compatible graphon parameters and graphon orders. This was based on the structuredness order $\preceq$ introduced in [12]. The basic theory of the structurdness order and the key fact that $\preceq$-maximal elements in the space of weak* limits are actually cut-distance limits readily translate to some other combinatorial-analytic objects such as kernels (that is, we allow even negative values), or digraphons (limits of directed graphons, i.e., not necessarily symmetric measurable functions $D:\Omega^2\to[0,1]$), and so does the main feature of the pushforward measures expressed in Proposition 2.22.[m] We think it would be interesting to investigate cut distance identifying/compatible parameters for these structures. For example, for digraph(on)s, there is a reasonable theory of quasirandomness (see [20] and references therein), and, as we saw, characterizing quasirandomness is in a sense dual to characterizing $\preceq$-maximal elements in the space of weak* limits.

The same program could be attempted for limits of k-uniform hypergraphs. However, already the basic theory of the weak* approach to hypergraphons seems to be substantially more involved (work in progress). Also, note that the transition from graphon parameters to hypergraphon parameters will not be automatic at all; for example, the Sidorenko conjecture does not have a reasonable counterpart for hypergraphons (see [38]).

Acknowledgments

We would like to thank David Conlon, Frederik Garbe, Dan Kráľ, Dávid Kunszenti-Kovács, Joonkyung Lee. We also thank two anonymous referees for their useful suggestions. This in particular concern a substantial simplification related to Proposition 2.15.

References

[1] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., Hoboken, NJ, third edition, 2008. With an appendix on the life and work of Paul Erdős.

[2] T. Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.*, 5:80–145, 2008.

[m] Even though these modifications have not been explicitly worked out in [12].

[3] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[4] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[5] D. Conlon, J. Fox, and B. Sudakov. An approximate version of Sidorenko's conjecture. *Geom. Funct. Anal.*, 20(6):1354–1366, 2010.

[6] D. Conlon, J. H. Kim, C. Lee, and J. Lee. Some advances on Sidorenko's conjecture. *J. Lond. Math. Soc. (2)*, 98(3):593–608, 2018.

[7] D. Conlon and J. Lee. Finite reflection groups and graph norms. *Adv. Math.*, 315:130–165, 2017.

[8] D. Conlon and J. Lee. Sidorenko's conjecture for blow-ups. *Discrete Anal.*, pages 1–13 (paper no. 2), 2021.

[9] J. W. Cooper, T. Kaiser, D. Král', and J. A. Noel. Weak regularity and finitely forcible graph limits. *Trans. Amer. Math. Soc.*, 370(6):3833–3864, 2018.

[10] J. W. Cooper, D. Kral, and T. L. Martins. Finitely forcible graph limits are universal. *Adv. Math.*, 340:819–854, 2018.

[11] P. Diaconis and S. Janson. Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)*, 28(1):33–61, 2008.

[12] M. Doležal, J. Grebík, J. Hladký, I. Rocha, and V. Rozhoň. Relating the cut distance and the weak* topology for graphons. *J. Combin. Theory Ser. B*, 147:252–298, 2021.

[13] J. Doležal, M. Hladký. Cut-norm and entropy minimization over weak* limits. *J. Combin. Theory Ser. B*, 137:232–263, 2019.

[14] G. Elek and B. Szegedy. A measure-theoretic approach to the theory of dense hypergraphs. *Adv. Math.*, 231(3-4):1731–1772, 2012.

[15] P. Erdős and M. Simonovits. Supersaturated graphs and hypergraphs. *Combinatorica*, 3(2):181–192, 1983.

[16] J. Fox and F. Wei. On the local approach to Sidorenko's conjecture. In *European Conference on Combinatorics, Graph Theory and Applications (EuroComb 2017)*, volume 61 of *Electron. Notes Discrete Math.*, pages 459–465. Elsevier Sci. B. V., Amsterdam, 2017.

[17] A. Frieze and R. Kannan. Quick Approximation to Matrices and Applications. *Combinatorica*, 19(2):175–220, 1999.

[18] F. Garbe, J. Hladký, and J. Lee. Two remarks on graph norms. Discrete Comput. Geom. (2021), https://doi.org/10.1007/s00454-021-00280-w.

[19] W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Ann. of Math. (2)*, 166(3):897–946, 2007.

[20] S. Griffiths. Quasi-random oriented graphs. *J. Graph Theory*, 74(2):198–209, 2013.

[21] O. Hanner. On the uniform convexity of L^p and l^p. *Ark. Mat.*, 3:239–244, 1956.

[22] H. Hatami. Graph norms and Sidorenko's conjecture. *Israel J. Math.*, 175:125–150, 2010.

[23] J. H. Kim, C. Lee, and J. Lee. Two approaches to Sidorenko's conjecture. *Trans. Amer. Math. Soc.*, 368(7):5057–5074, 2016.

[24] D. Král', L. M. Lovász, J. A. Noel, and J. Sosnovec. Finitely forcible graphons with an almost arbitrary structure. *Discrete Anal.*, pages 1–36 (paper no. 9), 2020.

[25] D. Král', T. Martins, P. P. Pach, and M. Wrochna. The step Sidorenko property and non-norming edge-transitive graphs. *J. Combin. Theory Ser. A*, 162:34–54, 2019.

[26] J. Lee and B. Schülke. Convex graphon parameters and graph norms. *Israel J. Math.*, 242(2):549–563, 2021.

[27] J. L. X. Li and B. Szegedy. On the logarithimic calculus and Sidorenko's conjecture. arXiv:1107.1153.

[28] L. Lovász. Subgraph densities in signed graphons and the local Simonovits-Sidorenko conjecture. *Electron. J. Combin.*, 18(1):Paper 127, 21, 2011.

[29] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[30] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[31] L. Lovász and B. Szegedy. Szemerédi's Lemma for the analyst. *J. Geom. and Func. Anal*, 17:252–270, 2007.

[32] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

[33] G. Regts and A. Schrijver. Compact orbit spaces in Hilbert spaces and limits of edge-colouring models. *European J. Combin.*, 52(part B):389–395, 2016.

[34] A. Scott. Szemerédi's regularity lemma for matrices and sparse graphs. *Combin. Probab. Comput.*, 20(3):455–466, 2011.

[35] A. Sidorenko. A correlation inequality for bipartite graphs. *Graphs Combin.*, 9(2):201–204, 1993.

[36] M. Simonovits. Extremal graph problems, degenerate extremal problems, and supersaturated graphs. In *Progress in graph theory (Waterloo, Ont., 1982)*, pages 419–437. Academic Press, Toronto, ON, 1984.

[37] J. Skokan and L. Thoma. Bipartite subgraphs and quasi-randomness. *Graphs Combin.*, 20(2):255–262, 2004.

[38] B. Szegedy. An information theoretic approach to Sidorenko's conjecture. arXiv:1406.6738.

[39] E. Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

Doležal: Institute of Mathematics, Czech Academy of Sciences. Žitná 25, 115 67, Praha 1, Czechia. With institutional support RVO:67985840, *Grebík:* Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK. *Most of this work was done while the author worked* at the Institute of Mathematics, Czech Academy of Sciences., *Hladký:* Institute of Computer Science, Czech Academy of Sciences. Pod Vodárenskou věží 2, 182 07, Prague, Czechia. With institutional support RVO:67985807. *Most of this work was done while the author worked* at TU Dresden and the Institute of Mathematics, Czech Academy of Sciences., *Rocha*: *This work was done while the author worked* at Institute of Computer Science, Czech Academy of Sciences. Pod Vodárenskou věží 2, 182 07, Prague, Czechia. With institutional support RVO:67985807., *Rozhoň:* ETH Zürich, Switzerland. *Most of this work was done while the author worked* at the Institute of Computer Science, Czech Academy of Sciences. With institutional support RVO:67985807.

TWO REMARKS ON GRAPH NORMS

FREDERIK GARBE, JAN HLADKÝ, AND JOONKYUNG LEE

Abstract. For a graph H, its homomorphism density in graphs naturally extends to the space of two-variable symmetric functions W in L^p, $p \geqslant e(H)$, denoted by $t(H, W)$. One may then define corresponding functionals $\|W\|_H := |t(H, W)|^{1/e(H)}$ and $\|W\|_{r(H)} := t(H, |W|)^{1/e(H)}$ and say that H is (semi-)norming if $\|.\|_H$ is a (semi-)norm and that H is weakly norming if $\|.\|_{r(H)}$ is a norm.

We obtain two results that contribute to the theory of (weakly) norming graphs. Firstly, answering a question of Hatami, who estimated the modulus of convexity and smoothness of $\|.\|_H$, we prove that $\|.\|_{r(H)}$ is neither uniformly convex nor uniformly smooth, provided that H is weakly norming. Secondly, we prove that every graph H without isolated vertices is (weakly) norming if and only if each component is an isomorphic copy of a (weakly) norming graph. This strong factorisation result allows us to assume connectivity of H when studying graph norms. In particular, we correct a negligence in the original statement of the aforementioned theorem by Hatami.

1. Introduction

One of the cornerstones of the theory of quasirandomness, due to Chung–Graham–Wilson [1] and to Thomason [10], is that a graph is quasirandom if and only if it admits a random-like count for any even cycle. A modern interpretation of this phenomenon is that the even cycle counts are essentially equivalent to the Schatten–von Neumann norms on the space of two variable symmetric functions, which are the natural limit object of large dense graphs. Indeed, Lovász [7] asked the natural question whether other graph counts can also induce a similar norm, which motivated Hatami's pioneering work [5] in the area. Since then, graph norms have been an important concept in the theory of graph limits and received considerable attention. For instance, Conlon and the third author [3] obtained a large class of graph norms, Kráľ, Martins, Pach, and Wrochna [6] proved that edge-transitive non-norming graphs exist, and very recently, the first author with Doležal, Grebík, Rocha, and Rozhoň [4] linked graph norms to the so-called step Sidorenko property.

The current note contributes further to this emerging theory of graph norms. We recall the basic definitions given in Hatami's work [5] with slight modifications taken from [8]. Let Ω be an arbitrary standard Borel space with an atomless probability measure ν. Whenever we consider a subset of Ω, we tacitly assume that it is measurable. We denote by $\mathcal{W}$ the linear space of all bounded symmetric measurable functions $W : \Omega^2 \to \mathbb{R}$. Also let $\mathcal{W}_{\geqslant 0} \subseteq \mathcal{W}$ be the set of non-negative functions in $\mathcal{W}$. Recall that functions in $\mathcal{W}_{\geqslant 0}$ that are bounded above by 1 are called *graphons*, and arise as limits of graph sequences [9].

Let H be a graph on the vertex set $\{v_1, \dots, v_n\}$. Given a symmetric measurable real-valued function W on Ω^2, set

$$t(H, W) := \int_{x_1 \in \Omega} \dots \int_{x_n \in \Omega} \prod_{\{v_i, v_j\} \in E(H)} W(x_i, x_j) \, d\nu^{\otimes n} \, . \tag{1}$$

Let $\mathcal{W}_H$ (resp. $\mathcal{W}_{r(H)}$) be the set of those symmetric measurable functions $W : \Omega^2 \to \mathbb{R}$ for which $t(H, W)$ (resp. $t(H, |W|)$) is defined and is finite. Obviously, $\mathcal{W}_H$ is a subspace of $\mathcal{W}_{r(H)}$, and Hölder's inequality immediately proves that $L^p(\Omega^2)$ is contained in $\mathcal{W}_{r(H)}$ whenever $p \geqslant e(H)$.

We then say that H is *(semi-)norming* if $\| \cdot \|_H := |t(H, \cdot)|^{1/e(H)}$ is a (semi-)norm on $\mathcal{W}_H$. Likewise, we say that H is *weakly norming* if $\| \cdot \|_{r(H)} := t(H, | \cdot |)^{1/e(H)}$ is a norm on $\mathcal{W}_{r(H)}$. Since $\mathcal{W}$ is a dense subset of the Banach space[1] $(\mathcal{W}_H, \| \cdot \|_H)$, this definition does not depend on whether we work in the Banach space $(\mathcal{W}, \| \cdot \|_H)$ or $(\mathcal{W}_H, \| \cdot \|_H)$. Analogously, in the definition of weakly norming property, $\mathcal{W}_{r(H)}$ can be replaced by $\mathcal{W}$. Note that, as the names suggest, norming graphs are seminorming and seminorming graphs are weakly norming.

Prominent examples of norming graphs are even cycles C_{2k} and complete bipartite graphs $K_{2n,2m}$ with an even number of vertices per partite set. Seminorming graphs that are not norming are stars with an even number of edges. Examples of weakly norming graphs that are not (semi)norming are complete bipartite graphs $K_{n,m}$ with $m > 1$ being odd. We refer the reader to [3, 8, 5] for more details and examples.

In what follows, we shall give short proofs of two results concerning (weakly) norming graphs. Firstly, we study basic geometric properties of the space $(\mathcal{W}_{r(H)}, \| \cdot \|_{r(H)})$. The definitions of uniform smoothness and uniform convexity will be precisely given in the next section.

Theorem 1.1. *Let H be a weakly norming graph. Then the normed space $(\mathcal{W}_{r(H)}, \| \cdot \|_{r(H)})$ is neither uniformly smooth nor uniformly convex.*

This answers a question of Hatami, who proved that $(\mathcal{W}, \| \cdot \|_H)$ *is* uniformly smooth and uniformly convex whenever H is seminorming and asked for a counterpart of his theorem for weakly norming graphs.

Theorem 1.1 not only answers a natural question arising from a functional-analytic perspective, but is also meaningful in the theory of quasirandomness. In [4], Hatami's theorem about uniform convexity and smoothness (see Theorem 2.2 for a precise statement) is the key ingredient in proving that every norming graph has the 'step forcing property'. By inspecting the proof in [4], one may see that the same conclusion for weakly norming

[1] By the topological equivalence between the cut norm and graph norms (see, for instance, Section 5.2 in [3]) and completeness of $\mathcal{W}$ under the cut norm, $\| \cdot \|_{r(H)}$ and $\| \cdot \|_H$ also define Banach spaces.

graphs H (except forests) could also be obtained if $\|\cdot\|_{r(H)}$ defined a uniformly convex space. However, Theorem 1.1 proves that such a modification is impossible.

Secondly, we prove a strong 'factorisation' result for disconnected weakly norming graphs.

Theorem 1.2. *A graph H is weakly norming if and only if all its non-singleton connected components are isomorphic and weakly norming. The same statement with weakly norming replaced by either seminorming or norming also holds.*

The 'if' direction is obvious, since $|t(H,W)|^{1/e(H)} = |t(H',W)|^{1/e(H')}$ whenever $W \in \mathcal{W}$ and H is a vertex-disjoint union of copies of H' and an arbitrary number of isolated vertices, but the converse is non-trivial.

Theorem 1.2 corrects a negligence that assumes connectivity of graphs without stating it, which in fact appeared in Hatami's work [5] and Lovász's book [8] that study graph norms. We also remark that for Sidorenko's conjecture, a major open problem in extremal combinatorics, even a weak factorisation result — such as each component of a graph satisfying the conjecture again satisfies it — is unknown, even though weakly norming graphs satisfy the conjecture. In fact, Conlon and the third author [2, Corollary 1.3] proved that the weak factorisation result, if it exists, implies the full conjecture.

2. Moduli of convexity and smoothness

We begin by recalling the definitions of moduli of convexity and moduli of smoothness of a normed space.

Definition 2.1. Let $(X, \|\cdot\|)$ be a normed space. The *modulus of convexity of* X is a function $\mathfrak{d}_X : (0,2] \to \mathbb{R}$ defined by

$$\mathfrak{d}_X(\varepsilon) := \inf\left\{1 - \left\|\frac{x+y}{2}\right\| \,:\, x,y \in X, \|x-y\| \geqslant \varepsilon, \|x\| = \|y\| = 1\right\}. \tag{2}$$

The *modulus of smoothness of* X is a function $\mathfrak{s}_X : (0,\infty) \to \mathbb{R}$ defined by

$$\mathfrak{s}_X(\varepsilon) := \sup\left\{\frac{1}{2}\left(\|x+y\| + \|x-y\| - 2\right) \,:\, x,y \in X, \|x\| = 1, \|y\| = \varepsilon\right\}. \tag{3}$$

The normed space $(X, \|\cdot\|)$ is *uniformly convex* if $\mathfrak{d}_X(\varepsilon) > 0$ for each $\varepsilon > 0$ and is *uniformly smooth* if $\lim_{\varepsilon \searrow 0} \frac{\mathfrak{s}_X(\varepsilon)}{\varepsilon} = 0$. For convenience, we write $\mathfrak{d}_H$, $\mathfrak{s}_H$, $\mathfrak{d}_{r(H)}$ and $\mathfrak{s}_{r(H)}$ instead of $\mathfrak{d}_{\mathcal{W}_H}$, $\mathfrak{s}_{\mathcal{W}_H}$, $\mathfrak{d}_{\mathcal{W}_{r(H)}}$ and $\mathfrak{s}_{\mathcal{W}_{r(H)}}$, respectively.

Hatami [5] determined $\mathfrak{d}_H$ and $\mathfrak{s}_H$ for connected norming graphs H up to a multiplicative constant by relating them to the moduli of convexity and of smoothness of ℓ^p-spaces, which are well-understood.

Theorem 2.2 (Theorem 2.16 in [5]). *For each $m \in \mathbb{N}$, there exist constants $C_m, C'_m > 0$ such that the following holds: let H be a connected seminorming graph with m edges. Then the Banach space $(\mathcal{W}_H, \|\cdot\|_H)$ satisfies $C_m \cdot \mathfrak{d}_{\ell^m} \leqslant \mathfrak{d}_H \leqslant \mathfrak{d}_{\ell^m}$ and $\mathfrak{s}_{\ell^m} \leqslant \mathfrak{s}_H \leqslant C'_m \cdot \mathfrak{s}_{\ell^m}$.*

Since for each $p \in (1, +\infty)$, it is well-known that the ℓ^p-space is uniformly convex and uniformly smooth, one obtains the following.

Corollary 2.3. Let H be a connected seminorming graph. Then the Banach space $(\mathcal{W}_H, \|\cdot\|_H)$ is uniformly convex and uniformly smooth.

The connectivity of H in Theorem 2.2 was in fact neglected in the original statement in [5], but it is certainly necessary. For example, by taking a disjoint union of two isomorphic norming graphs with $m/2$ edges (assume m is even), one obtains another norming graph with m edges that gives exactly the same norm, whose correct parameters in Theorem 2.2 are $\mathfrak{d}_H = \Theta(\mathfrak{d}_{\ell^{m/2}})$ and $\mathfrak{s}_H = \Theta(\mathfrak{s}_{\ell^{m/2}})$. Indeed, in Theorem 4.1 below we obtain a general statement without assuming connectivity by using Theorem 1.2. But first, let us point out the negligence in [5] which causes that the proof of Theorem 2.2 does not work for disconnected graphs. This subtle error lies in proving $\mathfrak{d}_H \leqslant \mathfrak{d}_{\ell^m}$ and $\mathfrak{s}_{\ell^m} \leqslant \mathfrak{s}_H$ by claiming that the Banach space $(\mathcal{W}_H, \|\cdot\|_H)$ contains a subspace isomorphic to $(\ell^m, \|\cdot\|_m)$. Here we give a full proof of the claim, which in turn reveals where the connectivity of H is used. To this end, we introduce the following notation, which will also be useful in Section 3.

Definition 2.4. Let Ω be partitioned as $\Omega = \Omega_1 \sqcup \Omega_2 \sqcup \dots$ with countably many parts such that $\nu(\Omega_i) = 2^{-i}$ for every $i \in \mathbb{N}$. For each $m \in \mathbb{N}$, $\gamma > 0$, and $\mathbf{a} = (a_1, a_2, \dots) \in \ell^m$, $W_{\gamma,\mathbf{a}}$ denotes the function satisfying $W_{\gamma,\mathbf{a}}(x, y) = 2^{i\gamma} \cdot a_i$ whenever $(x, y) \in \Omega_i^2$ and $W_{\gamma,\mathbf{a}} = 0$ outside $\bigcup_i \Omega_i^2$.

Suppose that H is a norming graph with n vertices and m edges. In particular this implies that m is even (see [8, Exercise 14.8]). The map $\mathbf{a} \mapsto W_{\frac{n}{m},\mathbf{a}}$ is linear, and thus, proving that this map preserves the respective norms is enough to conclude that the subspace spanned by $W_{\frac{n}{m},\mathbf{a}}$ is isomorphic to ℓ^m. For each $\mathbf{a} = (a_1, a_2, \dots) \in \ell^m$,

$$\|\mathbf{a}\|_m^m = \sum_i a_i^m = \sum_i \frac{1}{2^{in}} \cdot (2^{in/m} \cdot a_i)^m = t\big(H, W_{\frac{n}{m},\mathbf{a}}\big) \,.$$

Indeed, if $x_1, \dots, x_n$ do not fall into any single Ω_i, connectedness of H implies that the product in (1) evaluates to 0. Otherwise, if $(x_1, \dots, x_n) \in \Omega_i^n$ for some $i \in \mathbb{N}$, then $\nu^{\otimes n}(\Omega_i^n) = \frac{1}{2^{in}}$ and the product in (1) evaluates to constant $(2^{in/m} \cdot a_i)^m$, which proves the last equality. This is exactly where the proof of the claim relies on H being connected.

Now, turning to weakly norming graphs, Theorem 1.1 is a direct consequence of the following result.

Theorem 2.5. *Let H be a weakly norming graph. Then for each $\varepsilon \in (0,1)$,*

(a) $\mathfrak{d}_{r(H)}(\varepsilon) = 0$, *and*
(b) $\mathfrak{s}_{r(H)}(\varepsilon) \geqslant \frac{1}{2}\varepsilon$.

For the proof, we introduce a random graphon model that generalises graphon representations of the Erdős–Rényi random graph. Let $\mathcal{D}$ be a probability distribution on $[0,1]$ and let $\Omega = \Omega_1 \sqcup \ldots \sqcup \Omega_n$ be an arbitrary partition of Ω into sets of measure $\frac{1}{n}$. Denote by $\mathbb{U}(n,\mathcal{D})$ the random graphon obtained by assigning a constant value generated independently at random by the distribution $\mathcal{D}$ on each $(\Omega_i \times \Omega_j) \cup (\Omega_j \times \Omega_i)$, $1 \leqslant i \leqslant j \leqslant n$. Although $\mathbb{U}(n,\mathcal{D})$ depends on the partition $\Omega_1 \sqcup \ldots \sqcup \Omega_n$, we shall suppress the dependency parameter as different $\mathbb{U}(n,\mathcal{D})$'s are 'isomorphic' in the sense that there exists a measure-preserving bijection that maps one partition to the other. We use the term *asymptotically almost surely*, or *a.a.s.* for short, in the standard way, i.e., a property $\mathcal{P}$ of $\mathbb{U}(n,\mathcal{D})$ holds a.a.s. if the probability that $\mathcal{P}$ occurs tends to 1 as $n \to \infty$. We write $a = b \pm \epsilon$ if and only if $a \in [b-\epsilon, b+\epsilon]$.

Proposition 2.6. *Let $\mathcal{D}$ be a probability distribution on $[0,1]$ and let $d = \mathbb{E}[\mathcal{D}]$. Then for any fixed graph H, $U \sim \mathbb{U}(n,\mathcal{D})$ satisfies $t(H,U) = d^{e(H)} \pm o_n(1)$ a.a.s.*

We omit the proof, as it is a straightforward application of the standard concentration inequalities to subgraph densities in Erdős–Rényi random graphs (see, for example, [8, Corollary 10.4]).

Proof of Theorem 2.5. Throughout the proof, we briefly write $\|\cdot\|_{r(H)} = \|\cdot\|$. For $x \in [0,1]$, denote by $\mathbf{1}\{x\}$ the Dirac measure on x. Set

$$\mathcal{D}_1 := \tfrac{1}{2} \cdot \mathbf{1}\{0\} + \tfrac{1}{2} \cdot \mathbf{1}\{1\}.$$

Let U_1 and U_2 be two independent copies of $\mathbb{U}(n,\mathcal{D}_1)$. Proposition 2.6 then implies a.a.s.

$$\|U_i\| = t(H,U_i)^{1/e(H)} = \tfrac{1}{2} \pm o_n(1), \quad \text{for } i = 1,2. \tag{4}$$

For each $i = 1,2$, let $U_i^* := \frac{1}{2\|U_i\|} \cdot U_i$ be the normalisation of U_i that satisfies $\|U_i^*\| = \frac{1}{2}$. Then by substituting $U_i = 2\|U_i\| \cdot U_i^*$ and using (4) we get that

$$\|U_i^* - U_i\| = \big|1 - 2\|U_i\|\big| \cdot \|U_i^*\| = o_n(1)\,. \tag{5}$$

Since the random graphon $|U_1 - U_2|$ is also distributed like $\mathbb{U}(n,\mathcal{D}_1)$, we again have $\|U_1 - U_2\| = \frac{1}{2} \pm o_n(1)$ a.a.s. Thus, by the triangle inequality and (5), $2U_1^*$ and $2U_2^*$ are two symmetric functions with $\|2U_1^*\| = \|2U_2^*\| = 1$ whose linear combination is always close to the corresponding one of U_1 and U_2, i.e., for any fixed $\alpha, \beta \in \mathbb{R}$,

$$\big|\, \|\alpha U_1 + \beta U_2\| - \|\alpha U_1^* + \beta U_2^*\| \,\big| \leqslant |\alpha|\, \|U_1 - U_1^*\| + |\beta|\, \|U_2 - U_2^*\| = o_n(1). \tag{6}$$

In particular, $\alpha = 2$ and $\beta = -2$ give $\|2U_1^* - 2U_2^*\| \geqslant \|2U_1 - 2U_2\| - o_n(1) = 1 \pm o_n(1)$. That is, $2U_1^*$ and $2U_2^*$ are points on the unit sphere that are 'far' apart. Setting $\alpha = \beta = 1$ in (6) gives $\big|\, \|U_1 + U_2\| - \|U_1^* + U_2^*\| \,\big| = o_n(1)$, and therefore, for any $0 < \varepsilon < 1$,

$$\mathfrak{d}_{r(H)}(\varepsilon) \leqslant 1 - \left\| \frac{2U_1^* + 2U_2^*}{2} \right\| = 1 - \left\| \frac{2U_1 + 2U_2}{2} \right\| \pm o_n(1). \tag{7}$$

Now let

$$\mathcal{D}_2 := \tfrac{1}{4} \cdot \mathbf{1}\{0\} + \tfrac{1}{2} \cdot \mathbf{1}\{\tfrac{1}{2}\} + \tfrac{1}{4} \cdot \mathbf{1}\{1\}.$$

Then, since $\frac{1}{2}(U_1 + U_2)$ has distribution $\mathbb{U}(n, \mathcal{D}_2)$ and $\mathbb{E}[\mathcal{D}_2] = \frac{1}{2}$, we have by Proposition 2.6 a.a.s. $\|U_1 + U_2\| = 1 \pm o_n(1)$. Substituting this into (7) proves that the modulus of convexity of $\|\cdot\|$ is 0 for each $\varepsilon \in (0, 1)$.

For $\varepsilon \in (0, 1)$ given in (b), let

$$\mathcal{D}_3 := \tfrac{1}{4}\big(\mathbf{1}\{0\} + \mathbf{1}\{\varepsilon\} + \mathbf{1}\{1 - \varepsilon\} + \mathbf{1}\{1\}\big) \quad \text{and} \quad \mathcal{D}_4 := \tfrac{1}{4}\big(\mathbf{1}\{0\} + \mathbf{1}\{\tfrac{\varepsilon}{2}\} + \mathbf{1}\{\tfrac{1}{2}\} + \mathbf{1}\{\tfrac{1+\varepsilon}{2}\}\big).$$

The distributions of $|U_1 - \varepsilon U_2|$ and $\frac{1}{2}|U_1 + \varepsilon U_2|$ are $\mathbb{U}(n, \mathcal{D}_3)$ and $\mathbb{U}(n, \mathcal{D}_4)$, respectively. As $\mathbb{E}[\mathcal{D}_3] = \frac{1}{2}$ and $\mathbb{E}[\mathcal{D}_4] = \frac{1+\varepsilon}{4}$, Proposition 2.6 yields that, a.a.s., $\|2U_1 - 2\varepsilon U_2\| = 1 \pm o_n(1)$ and $\|2U_1 + 2\varepsilon U_2\| = 1 + \varepsilon \pm o_n(1)$. Therefore, by (6), $\|2U_1^* - 2\varepsilon U_2^*\| = 1 \pm o_n(1)$ and $\|2U_1^* + 2\varepsilon U_2^*\| = 1 + \varepsilon \pm o_n(1)$ a.a.s. Hence, substituting $2U_1^*$ and $2\varepsilon U_2^*$ into (3) gives

$$\mathfrak{s}_X(\varepsilon) \geqslant \frac{1}{2}\left(\|2U_1^* + 2\varepsilon U_2^*\| + \|2U_1^* - 2\varepsilon U_2^*\| - 2\right) = \frac{\varepsilon}{2} \pm o_n(1),$$

which proves (b). □

3. Disconnected (semi-)norming and weakly norming graphs

To be precise, we expand Theorem 1.2 to two parallel statements, also omitting any isolated vertices from H (this operation does not change $t(H, \cdot)^{1/e(H)}$).

Theorem 1.2 (Restated)**.** *For a graph H without isolated vertices, the following holds:*

- *(a) A graph H is weakly norming if and only if all connected components of H are isomorphic and weakly norming.*
- *(b) A graph H is (semi-)norming if and only if all connected components of H are isomorphic and (semi-)norming.*

To prove this theorem, we need some basic facts about weakly norming graphs. Given a graph H and a collection $\mathbf{w} = (W_e)_{e \in E(H)} \in \mathcal{W}^{E(H)}$, define the $\mathbf{w}$-*decorated homomorphism density* by

$$t(H, \mathbf{w}) := \int_{x_1 \in \Omega} \dots \int_{x_n \in \Omega} \prod_{e = ij \in E(H)} W_e(x_i, x_j)\ .$$

That is, we assign a possibly different W_e to each $e \in E(H)$ and are counting such 'multicoloured' copies of H. In particular, if $W_e = W$ for all $e \in E(H)$, we obtain $t(H, \mathbf{w}) = t(H, W)$. Hatami [5] observed that the (weakly) norming property is equivalent to a Hölder-type inequality for the decorated homomorphism density.

Lemma 3.1 (Theorem 2.8 in [5])**.** *Let H be a graph. Then*

(a) H is weakly norming if and only if, for every $w \in \mathcal{W}_{\geqslant 0}^{E(H)}$,

$$t(H, \mathbf{w})^{e(H)} \leqslant \prod_{e \in E(H)} t(H, W_e).$$

(b) H is seminorming if and only if, for every $w \in \mathcal{W}^{E(H)}$,

$$t(H, \mathbf{w})^{e(H)} \leqslant \prod_{e \in E(H)} |t(H, W_e)|.$$

As the second inequality is more general than the first, it immediately follows that every seminorming graph is weakly norming. Another easy consequence of this characterisation is that, for a weakly norming graph H, its subgraph F, and $W \in \mathcal{W}_{\geqslant 0}$, we have the inequality

$$t(F, W) \leqslant t(H, W)^{e(F)/e(H)}. \tag{8}$$

Indeed, one can easily prove this by setting $W_e = W$ for $e \in E(F)$ and $W_e \equiv 1$ otherwise. For yet another application, we use Lemma 3.1 to prove that a weakly norming graph essentially has no subgraph with larger average degree.

Lemma 3.2. *Let H be a weakly norming graph without isolated vertices and let F be its subgraph. Then $\frac{e(F)}{v(F)} \leqslant \frac{e(H)}{v(H)}$.*

Proof. We may assume F has no isolated vertices either, as adding isolated vertices only reduces the average degree. Let $X \subseteq \Omega$ be a subset with $\nu(X) = 1/2$ and let $U : \Omega^2 \to [0, 1]$ be the graphon defined by $U(x, y) = 1$ if $x, y \in X$ and 0 otherwise. Then, for any graph J without isolated vertices, $t(J, U) = 2^{-v(J)}$. Choosing $W_e = U$ for $e \in E(F)$ and $W_e \equiv 1$ otherwise for $\mathbf{w} \in \mathcal{W}_{\geqslant 0}^{E(H)}$ then gives

$$t(F, U)^{e(H)} = t(H, \mathbf{w})^{e(H)} \leqslant t(H, U)^{e(F)} t(H, 1)^{e(H)-e(F)} = t(H, U)^{e(F)}.$$

Comparing $t(F, U)^{e(H)} = 2^{-v(F)e(H)}$ and $t(H, U)^{e(F)} = 2^{-v(H)e(F)}$ concludes the proof. □

Remark 3.3. This is reminiscent of [5, Theorem 2.10(i)], which states that $\frac{e(F)}{v(F)-1} \leqslant \frac{e(H)}{v(H)-1}$ whenever H is weakly norming and F is a subgraph of H with $v(F) > 1$. However, this theorem is only true if H is connected and hence also needs to be corrected. To see this, let H be a vertex disjoint union of two copies of $K_{1,2}$, which is a norming graph. Then $\frac{e(H)}{v(H)-1} = 4/5$ but $\frac{e(F)}{v(F)-1} = 1$ for $F = K_{1,2}$.

Suppose now that a weakly norming graph H without isolated vertices consists of two vertex-disjoint subgraphs F_1 and F_2. If $e(F_1)/v(F_1) > e(F_2)/v(F_2)$, then

$$\frac{e(H)}{v(H)} = \frac{e(F_1)+e(F_2)}{v(F_1)+v(F_2)} < \frac{e(F_1)}{v(F_1)},$$

which contradicts Lemma 3.2. By iterating this, we obtain the following fact.

Corollary 3.4. Every component in a weakly norming graph without isolated vertices has the same average degree.

Before proceeding to the next step, we recall some basic facts about ℓ^p-spaces. For $0 < p < q \leqslant +\infty$ we have $\|\cdot\|_p \geqslant \|\cdot\|_q$. Furthermore, there exists $\mathbf{c} \in \ell^\infty$ such that

$$\|\mathbf{c}\|_p > \|\mathbf{c}\|_q \,. \tag{9}$$

Lemma 3.5. *In a weakly norming graph H without isolated vertices, every connected component has the same number of edges.*

Proof. Let $F_1, \ldots, F_k$ be the connected components of H and let $\gamma := \frac{v(F_1)}{e(F_1)}$. By Corollary 3.4, $\frac{2}{\gamma}$ is the average degree of all F_i, $i = 1, 2, \ldots, k$. Recall the definition of $W_{\gamma,\mathbf{a}}$ given in Definition 2.4. For each $\mathbf{a} = (a_1, a_2, \ldots) \in \ell^\infty$ and each connected graph F which also has average degree $\frac{2}{\gamma}$, and, say, m edges, we have

$$t(F, |W_{\gamma,\mathbf{a}}|) = \sum_i |a_i|^m = \|\mathbf{a}\|_m^m \,. \tag{10}$$

Suppose that not all the components have the same number of edges. Let $p = \min_j e(F_j)$. We may assume that $p = e(F_1)$. Let $q > p$ be the number of edges in a component with more edges than F_1 and let $\mathbf{c} \in \ell^\infty$ be given by (9). Define the collection $\mathbf{w} = (W_e)_{e \in E(H)}$ by $W_e = |W_{\gamma,\mathbf{c}}|$ for $e \in E(F_1)$ and $W_e \equiv 1$ otherwise. Lemma 3.1 then gives

$$t(F_1, |W_{\gamma,\mathbf{c}}|)^{e(H)} = t(H, \mathbf{w})^{e(H)} \leqslant \prod_{e \in E(H)} t(H, W_e) = t(H, |W_{\gamma,\mathbf{c}}|)^p. \tag{11}$$

Expanding the term $t(H, |W_{\gamma,\mathbf{c}}|)$ on the right-hand side of (11) using (10) yields

$$t(H, |W_{\gamma,\mathbf{c}}|) = \prod_{j=1}^{k} t(F_j, |W_{\gamma,\mathbf{c}}|) = \prod_{j=1}^{k} \|\mathbf{c}\|_{e(F_j)}^{e(F_j)} \,.$$

On the left-hand side of (11), we have by (10) that $t(F_1, |W_{\gamma,\mathbf{c}}|) = \|\mathbf{c}\|_p^p$. Substituting these back to (11) gives

$$\|\mathbf{c}\|_p^{p \cdot e(H)} \leqslant \left(\prod_{j=1}^{k} \|\mathbf{c}\|_{e(F_j)}^{e(F_j)} \right)^p ,$$

which contradicts the fact that $\|\mathbf{c}\|_p \geqslant \|\mathbf{c}\|_{e(F_j)}$ for each $j \in [k]$ with at least one of the inequalities being strict. □

Lemma 3.6. *For a weakly norming graph H without isolated vertices, all the components of H are isomorphic.*

Proof. Suppose that there are at least two non-isomorphic graphs amongst all the components $F_1, \dots, F_k$. By Lemma 3.5 we may assume that all F_i have the same number of edges, say m. In particular, $e(H) = mk$. By Theorem 5.29 in [8], there exists a graphon U so that the numbers $t(F_1, U), \dots, t(F_k, U)$ are not all equal. We may assume that $t(F_1, U)$ attains the maximum amongst $t(F_1, U), \dots, t(F_k, U)$. Then we have $t(H, U) = \prod_{i=1}^{k} t(F_i, U) < t(F_1, U)^k$, which contradicts

$$t(F_1, U) \leqslant t(H, U)^{m/e(H)} = t(H, U)^{1/k}$$

which follows from (8). □

Proof of Theorem 1.2. Suppose first that H is weakly norming. Let F be the graph given by Lemma 3.6 which is isomorphic to every component of H and let k be the number of components of H. Now enumerate the edges in H by $(e, i) \in E(F) \times [k]$, where each (e, i) denotes the edge e in the i-th copy of F. Then each $\mathbf{w} \in \mathcal{W}^{E(H)}$ can be written as $(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k)$, where $\mathbf{w}_i = (W_{e,i})_{e \in E(F)}$ such that $t(H, \mathbf{w}) = \prod_{i=1}^{k} t(F, \mathbf{w}_i)$. Let $\mathbf{u} = (U_e)_{e \in E(F)} \in \mathcal{W}_{\geqslant 0}^{E(F)}$ be arbitrary. Then Lemma 3.1 together with the choice $\mathbf{w}_1 = \mathbf{w}_2 = \ldots = \mathbf{w}_k = \mathbf{u}$, i.e., $W_{e,i} = U_e$, implies

$$\begin{aligned} t(F, \mathbf{u})^{k^2 \cdot e(F)} &= t(F, \mathbf{u})^{k \cdot e(H)} = t(H, \mathbf{w})^{e(H)} \\ &\leqslant \prod_{(e,i) \in E(H)} t(H, W_{e,i}) = \prod_{(e,i) \in E(F) \times [k]} t(F, U_e)^k = \prod_{e \in E(F)} t(F, U_e)^{k^2}. \end{aligned} \tag{12}$$

Taking the k^2-th root proves that F is weakly norming.

When H is seminorming, we can still apply Lemma 3.6 to obtain a graph F isomorphic to each component, since H is also weakly norming. Thus, the enumeration $E(F) \times [k]$ of $E(H)$ and the factorisation $t(H, \mathbf{w}) = \prod_{i=1}^{k} t(F, \mathbf{w}_i)$ for each $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k) \in \mathcal{W}^{E(H)}$ remain the same. Now let $\mathbf{u} = (U_f)_{f \in E(F)} \in \mathcal{W}^{E(F)}$ be arbitrary. Then again by taking $\mathbf{w}_1 = \mathbf{w}_2 = \ldots = \mathbf{w}_k = \mathbf{u}$ in Lemma 3.1, we obtain

$$t(F, \mathbf{u})^{k^2 \cdot e(F)} = t(H, \mathbf{w})^{e(H)} \leqslant \prod_{(e,i) \in E(H)} |t(H, W_{e,i})| = \prod_{e \in E(F)} |t(F, U_e)|^{k^2},$$

which proves that F is seminorming. If H is norming, then $|t(F, W)| = |t(H, W)|^{1/k}$ must be nonzero for each nonzero $W \in \mathcal{W}$. Thus, F is also norming. □

4. Concluding remarks

As mentioned in Section 2, Theorem 1.2 yields a full generalisation of Theorem 2.2.

Theorem 4.1. *For each $m \in \mathbb{N}$, there exist constants $C_m, C'_m > 0$ such that the following holds: let H be a seminorming graph with m edges in each (isomorphic) non-singleton component. Then the Banach space $(\mathcal{W}_H, \|\cdot\|_H)$ satisfies $C_m \cdot \mathfrak{d}_{\ell^m} \leqslant \mathfrak{d}_H \leqslant \mathfrak{d}_{\ell^m}$ and $\mathfrak{s}_{\ell^m} \leqslant \mathfrak{s}_H \leqslant C'_m \cdot \mathfrak{s}_{\ell^m}$.*

As a consequence, the connectivity condition in Corollary 2.3 can also be removed, i.e., $(\mathcal{W}_H, \|\cdot\|_H)$ is always uniformly convex and uniformly smooth whenever H is seminorming.

There is more literature in the area that has been imprecise when it comes to connectivity, but which can be corrected by Theorem 1.2 to hold in full generality. For instance, Exercise 14.7(b) in [8] states that every seminorming graph is either a star or Eulerian, which is true only if the seminorming graph is connected. To correct the statement, we may replace a star by a vertex disjoint union of isomorphic stars by using Theorem 1.2. Likewise, whenever studying properties of graph norms, one can invoke Theorem 1.2 and focus on connected graphs. We finally remark that the theorems used in our proofs have no errors concerning connectivity. In particular, Theorem 2.8 in [5] is still valid regardless of connectivity.

In [6], the *step Sidorenko property* is defined to prove that there exists an edge-transitive graph that is not weakly norming (for the precise definition, we refer to [6]), where the proof relies on the fact from [8] that every weakly norming graph is step Sidorenko. Moreover, it is shown in [4] that the converse is also true for connected graphs, i.e., every connected step Sidorenko graph is weakly norming. However, Theorem 1.2 proves that the converse no longer holds for disconnected graphs, as a vertex-disjoint union of non-isomorphic step Sidorenko graphs is again step Sidorenko but not weakly norming.

Acknowledgements. Part of this work was carried out while the third author visited the other authors in Prague and he is grateful for their support and hospitality. We would also like to thank the anonymous referees for their careful reviews of the manuscript and for their helpful comments.

References

[1] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[2] D. Conlon and J. Lee. Sidorenko's conjecture for blow-ups. To appear in Discrete Anal.

[3] D. Conlon and J. Lee. Finite reflection groups and graph norms. *Adv. Math.*, 315:130–165, 2017.

[4] M. Doležal, J. Grebík, J. Hladký, I. Rocha, and V. Rozhoň. Cut distance identifying graphon parameters over weak* limits. arXiv:1809.03797v3.

[5] H. Hatami. Graph norms and Sidorenko's conjecture. *Israel J. Math.*, 175:125–150, 2010.

[6] D. Král', T. Martins, P. P. Pach, and M. Wrochna. The step Sidorenko property and non-norming edge-transitive graphs. *J. Combin. Theory Ser. A*, 162:34–54, 2019.

[7] L. Lovász. Graph homomorphisms: Open problems. manuscript available at http://www. cs. elte. hu/lovasz/problems. pdf, 2008.

[8] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[9] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory, Ser. B*, 96(6):933–957, 2006.

[10] A. Thomason. Pseudorandom graphs. In *Random graphs '85 (Poznań, 1985)*, volume 144 of *North-Holland Math. Stud.*, pages 307–331. North-Holland, Amsterdam, 1987.

(FG,JH) Institute of Mathematics of the Czech Academy of Sciences, Žitná 25, 115 67 Prague, Czechia. Supported by GAČR project 18-01472Y. With institutional support RVO:67985840.

(JL) Department of Mathematics, University College London, Gower Street, London WC1E 6BT, UK. Supported by ERC Consolidator Grant PEPCo 724903.

Part II

Classical graph-theoretic concepts in graphons

TILINGS IN GRAPHONS

JAN HLADKÝ, PING HU, AND DIANA PIGUET

Abstract. We introduce a counterpart to the notion of tilings, that is vertex-disjoint copies of a fixed graph F, to the setting of graphons. The case $F = K_2$ gives the notion of matchings in graphons. We give a transference statement that allows us to switch between the finite and limit notion, and derive several favorable properties, including the LP-duality counterpart to the classical relation between the fractional vertex covers and fractional matchings/tilings, and discuss connections with property testing.

As an application of our theory, we determine the asymptotically almost sure F-tiling number of inhomogeneous random graphs $\mathbb{G}(n, W)$. As another application, in an accompanying paper [Hladký, Hu, Piguet: Komlós's tiling theorem via graphon covers, J. Graph Theory, 2019] we give a proof of a strengthening of a theorem of Komlós [Komlós: Tiling Turán Theorems, Combinatorica, 2000].

1. Towards limits of tilings[1]

The emergence of graph limit theories has brought numerous novel views on classical problems in graph theory. More precisely, the problems in which these theories helped concern comparing subgraph densities. This is a very broad area lying in the heart of extremal graph theory. In this explanatory section, we focus only on the applications of dense graphs limits, firstly because these have been richer, and secondly because this is the direction which we pursue in this paper. Two closely related theories have emerged. Razborov's flag algebras method [21] represents an abstract approach to graph limits. The method gives universal methods for calculations with subgraph densities (most notably the semidefinite method and differential calculus), and has led to the complete solutions of, or at least to a breakthrough progress on, several prominent problems in the area. Of these breakthroughs let us mention a result of Razborov [22] who determined an optimal function $f : [0, 1] \to [0, 1]$ such that if an n-vertex graph G contains at least $\alpha\binom{n}{2}$ edges then

$$G \text{ contains at least } (f(\alpha) + o(1))\binom{n}{3} \text{ triangles,} \tag{1.1}$$

Hladký: The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement number 628974.

Hu: This work has received funding from the Leverhulme Trust 2014 Philip Leverhulme Prize of Daniel Kral and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 648509).

Piguet was supported by the Czech Science Foundation, grant number GJ16-07822Y.

[1]An extended abstract describing also the main results of this paper appeared in the proceedings of Eurocomb 2017, [9].

thus resolving an old question of Lovász and Simonovits. On the other hand, the theory developed by Borgs, Chayes, Lovász, Szegedy, Sós and Vesztergombi [5, 18] provides explicit limit objects, the so-called graphons. This theory has been successfully applied in various parts of graph theory (and in particular provided insights into the properties of Szemerédi regularity partitions) and random graphs. In extremal graph theory, the theory of graphons has been used to prove a certain "local" version of Sidorenko's conjecture, [16].

As said, the graph limit theories have been very powerful in relating subgraph densities. Some other concepts, like the one of the minimum degree have been translated to the setting of graph limits and have been explored. Extremal graph theory, however, is a much richer field, and which other statements or features can be formulated in the language of graph limits is interesting in its own right. In the present paper we develop a theory of tiling in graphons. A *tiling* $\mathcal{T}$ by a finite graph F in another graph G ("F-tiling in G", in short) is a collection of vertex-disjoint copies of F (not necessarily induced) in G. This concept is used also under the name of F-matching, as we get the usual notion of matchings, when $F = K_2$. The *size* of the tiling $\mathcal{T}$ is simply the cardinality $|\mathcal{T}|$. We write $\mathsf{til}(F, G)$ for the size of a maximum F-tiling in G. For example, a "tiling counterpart" to (1.1) would entail finding an optimal function $g : [0, 1] \to [0, 1]$ such that any n-vertex graph G containing at least $\alpha\binom{n}{2}$ edges satisfies

$$G \text{ contains a tiling of } K_3 \text{ of size at least } (g(\alpha) + o(1))\,\tfrac{n}{3}.$$

Such a function g was indeed determined in [1]. Another example is the basic result of Erdős and Gallai [11], where they determined the size of a matching guaranteed in a graph of a given density.

There is a fractional relaxation of the notion of tilings. In that notion, we put $[0, 1]$-weights on copies of F in G, and require that the total weight at each vertex in G is at most 1. However, for us it is more convenient to choose a slightly different notion in which we replace copies by homomorphic copies. More precisely, we assume that the graph F is on the vertex set $[k]$. We denote by $\mathcal{F}_F(G)$ all the copies of F in the graph G,

$$\mathcal{F}_F(G) = \Big\{(u_1, u_2, \ldots, u_k) \in V(G)^k : u_i u_j \in E(G) \text{ for each pair } ij \in E(F)\Big\}\ . \tag{1.2}$$

More precisely, the members of $\mathcal{F}_F(G)$ correspond to ordered vertex-sets of G that represent homomorphisms of F into G. With a slight abuse of notation we shall call members of $\mathcal{F}_F(G)$ *copies of* F *in* G. Given a copy F' of F in G, the *vertex set* $V(F')$ is defined in an obvious way. Note that $|V(F')| \le k$, but equality need not hold.

A *fractional* F*-tiling* in a finite graph G is a weight function $\mathfrak{t} : \mathcal{F}_F(G) \to [0, 1]$ that satisfies that for each $v \in V(G)$,

$$\sum_{F' \in \mathcal{F}_F(G), V(F') \ni v} \mathfrak{t}(F') \le 1\ .$$

The *size* $\|\mathfrak{t}\|$ of $\mathfrak{t}$ is the total weight of $\mathcal{F}_F(G)$, i.e., $\|\mathfrak{t}\| = \sum_{F' \in \mathcal{F}_F(G)} \mathfrak{t}(F')$. A standard compactness argument shows, that there exists an F-tiling of maximum size. We denote this maximum size by $\mathsf{ftil}(F, G)$ and call it the *fractional F-tiling number*. Each F-tiling $\mathcal{T} \subset \mathcal{F}_F(G)$ can be represented as a fractional F-tiling by simply putting weight 1 on the copies of $\mathcal{T}$ and 0 on the copies of $\mathcal{F}_F(G) \setminus \mathcal{T}$. Thus we have $\mathsf{til}(F, G) \le \mathsf{ftil}(F, G)$.

Remark 1.1. *To illustrate the notion, and in particular to emphasize the difference between copies and homomorphic copies, let us compute* $\mathsf{ftil}(F, G)$ *when* F *is a five-cycle and* G *is a triangle (so,* G *is the smaller of the graphs!). Clearly,* $\mathsf{ftil}(C_5, K_3) \le \frac{3}{5}$*, which is just a particular instance of the general bound* $\mathsf{ftil}(F, G) \le \frac{v(G)}{v(F)}$*. On the other hand,* $1 \mapsto 1$*,* $2 \mapsto 2$*,* $3 \mapsto 1$*,* $4 \mapsto 2$*,* $5 \mapsto 3$ *is a homomorphism of* C_5 *to* K_3*. Now we can consider 2 further variants of this homomorphism obtained by cyclic shifts of* K_3*, and putting weight* $\frac{1}{5}$ *on each of these 3 homomorphic copies. This shows that* $\mathsf{ftil}(C_5, K_3) \ge \frac{3}{5}$*.*

The linear programming (LP) duality provides a very useful way of expressing $\mathsf{ftil}(F, G)$. To this end recall that a function $\mathfrak{c} : V(G) \to [0, 1]$ is a *fractional F-cover* of G if $\sum_{v \in F'} \mathfrak{c}(v) \ge 1$ for each copy $F' \in \mathcal{F}_F(G)$ (so, in the summation, we view F' as a multiset). The *size* of the fractional cover $\mathfrak{c}$ is the total weight of $V(G)$, and is denoted by $\|\mathfrak{c}\|$. Again, a compactness argument shows that there exists a fractional F-cover of G of minimum size, which is denoted by $\mathsf{fcov}(F, G)$ and called the *fractional F-cover number*. Then the Duality Theorem asserts that $\mathsf{fcov}(F, G) = \mathsf{ftil}(F, G)$. Let us note that showing the "$\ge$" direction is easy and the difficulty lies in proving the "$\le$" direction.

In this paper we define the notions of (fractional) F-tilings and fractional F-covers for graphons (Definition 3.1 and Definition 3.12). As we show, when the graphon in question is taken to be a representation of a finite graph then there is a correspondence between the classical finite notion and the new graphon notion (see Propositions 3.2 and 3.13). We relate these graphon notions to the limits of the corresponding finite parameters (see Theorem 3.4, Proposition 3.8, Corrolary 5.2) and treat their continuity properties on the graphon space (see Theorem 3.7, Theorem 3.7, and Theorem 3.14). We derive the LP-duality between the fractional F-tiling number and fractional F-cover number for graphons (Theorem 3.16).

This paper is organized as follows. In Section 2 we introduce common notation and provide preliminaries regarding functional analysis, graph limits and the regularity lemma. In Section 3 we introduce all the main concepts in our theory and state the main results. In Section 4 we give proofs of these results. In Section 5 we determine the F-tiling number of inhomogeneous random graphs. In an accompanying paper [13] we give another application of the theory by proving a strengthened version of a theorem of Komlós [15] regarding tilings in finite graphs.

Let us also mention that Doležal and Hladký [8] study in more detail the case $F = K_2$, that is, the case of matchings in graphons.

2. Notation and preliminaries

Suppose that X is an arbitrary set. Given $\mathbf{x} = (x_1, \dots, x_k) \in X^k$ and two numbers $1 \le i < j \le k$, we write $\pi_{ij}(\mathbf{x})$ for the projection of $\mathbf{x}$, $\pi_{ij}(\mathbf{x}) = (x_i, x_j)$. We extend this to projecting sets, i.e., given $Y \subset X^k$, we write $\pi_{ij}(Y) = \{\pi_{ij}(\mathbf{x}) : \mathbf{x} \in Y\}$. We will use this notion only with regard to the inverse map. That is, if $A \subset X^2$ then $\pi_{ij}^{-1}(A) = \big\{\mathbf{x} \in X^k : (x_i, x_j) \in A\big\}$.

Given a function f and a number a we define its *support* $\operatorname{supp} f = \{x : f(x) \neq 0\}$ and its variant $\operatorname{supp}_a f = \{x : f(x) \ge a\}$.

Our notation follows [17]. Throughout the paper we shall assume that Ω is an atomless Borel probability space equipped with a measure ν (defined on an implicit σ-algebra). The product measure on Ω^k is denoted by ν^k. Recall that a set is *null* if it has zero measure.

2.1. **Banach–Alaoglu Theorem.** As Ω is a Borel probability space, it is in particular a separable measure space. It is well-known that then the Banach space $\mathcal{L}^1(\Omega)$ is separable (see e.g. [6, Theorem 13.8]). The dual of $\mathcal{L}^1(\Omega)$ is $\mathcal{L}^\infty(\Omega)$.

Let us now recall the defining property of the weak* topology on $\mathcal{L}^\infty(\Omega)$: a sequence $f_1, f_2, \dots \in \mathcal{L}^\infty(\Omega)$ converges weak* to a function $f \in \mathcal{L}^\infty(\Omega)$ if for each $g \in \mathcal{L}^1(\Omega)$ we have that

$$\int_{x\in\Omega} f_n(x)g(x)\,\mathsf{d}\nu \to \int_{x\in\Omega} f(x)g(x)\,\mathsf{d}\nu\ .$$

Weak* convergence is denoted by $f_n \xrightarrow{\text{w}^*} f$.

These initial preparations can be used to verify that in the current context the assumptions of the sequential Banach–Alaoglu Theorem (as stated for example in [23, Theorem 1.9.14]) are fulfilled. Thus, let us state the theorem in the setting of $\mathcal{L}^\infty(\Omega)$.

Theorem 2.1. *If Ω is a Borel probability space then each sequence of functions of $\mathcal{L}^\infty(\Omega)$-norm at most 1 contains a weak* convergent subsequence.*

2.2. **Graphons.** Our graphons will be defined on Ω^2, i.e., symmetric measurable functions $W : \Omega^2 \to [0, 1]$. For a finite graph G on vertex set $\{v_1, \dots, v_n\}$ we can partition Ω into n sets $\Omega_1, \dots, \Omega_n$ of measure $\frac{1}{n}$ each and obtain a *representation* of G. This graphon, always denoted by W_G, is defined to be one or zero on each square $\Omega_i \times \Omega_j$ depending on whether the pair v_iv_j does or does not form an edge, respectively. Note that W_G is not uniquely defined; it depends on the partition $\Omega_1, \dots, \Omega_n$. Given a bounded symmetric measurable function $W : \Omega^2 \to \mathbb{R}$, define its *cut norm* $\|W\|_\square$ to be $\sup_{S,T\subseteq\Omega} \left|\int_{(x,y)\in S\times T} W(x, y)\,\mathsf{d}\nu^2\right|$, where the supremum is taken over all measurable subsets S and T. Let W be a graphon

and let $\varphi : \Omega \to \Omega$ be a measure preserving map. Define W^φ by $W^\varphi(x,y) = W(\varphi(x),\varphi(y))$. Given two graphons U and W, we define the *cut distance* of them by $\delta_\square(U,W) = \inf_\varphi \|U - W^\varphi\|_\square$, where the infimum is taken over all measure preserving map $\varphi : \Omega \to \Omega$. We denote $U \le W$ ν^2-almost everywhere by $U \le W$.

Suppose that we are given an arbitrary graphon $W : \Omega^2 \to [0,1]$ and a graph F whose vertex set is $[k]$. We write $W^{\otimes F} : \Omega^k \to [0,1]$ for a function defined by $W^{\otimes F}(x_1,\dots,x_k) = \prod_{1\le i<j\le k, ij\in E(F)} W(x_i,x_j)$, which we call the *density* of F in W. Let $\mathcal{F}_F(W) = \operatorname{supp} W^{\otimes F}$. Note that if we take W to be a graphon representation a finite graph G then there is a natural correspondence between $\mathcal{F}_F(W)$ and $\mathcal{F}_F(G)$ defined in (1.2).

We shall need the following easy statement about graphons, which is essentially given in [18, Lemma 4.1].

Lemma 2.2. *Suppose that U and W are two graphons such that $\|W-U\|_\square < \delta$, and let F be a finite graph with k vertices. Then for arbitrary measurable sets $P_1,\dots,P_k \subset \Omega$ we have*

$$\left| \int_{(x_1,\dots,x_k)\in\prod_i P_i} U^{\otimes F}(x_1,\dots,x_k)\,\mathsf{d}\nu^k - \int_{(x_1,\dots,x_k)\in\prod_i P_i} W^{\otimes F}(x_1,\dots,x_k)\,\mathsf{d}\nu^k \right| \le \binom{k}{2}\delta.$$

□

2.3. **Removal lemma.** Let us state a graphon version of the Removal lemma.

Lemma 2.3. *Suppose that F is a graph on a vertex set $[k]$. Then for every $\epsilon > 0$, there exists a constant $\delta > 0$ such that whenever $W : \Omega^2 \to [0,1]$ is a graphon with $\int_{(x_1,\dots,x_k)\in\Omega^k} W^{\otimes F}\,\mathsf{d}\nu^k < \delta$, then W can be made F-free by decreasing it in a suitable way such that it changes by at most ϵ in the $\mathcal{L}^1(\Omega^2)$-distance.*

2.4. **Partite versions of graphons.** In this section we introduce an auxiliary notion of partite graphons which we shall use in Section 4.5. If A and B are (not necessarily disjoint) sets then we write $A \amalg B$ for their disjoint union, i.e., for the set $(\{1\}\times A)\cup(\{2\}\times B)$. If A and B are measurable subsets of a measure space (Ω,ν) then we can view A as a measure space of total measure $\nu(A)$ and B as a measure space of total measure $\nu(B)$. Then, $A \amalg B$ can be viewed as a measure space of total measure $\nu(A)+\nu(B)$.

Suppose that $W : \Omega^2 \to [0,1]$ is a graphon and $A_1,\dots,A_k$ are (not necessarily disjoint) measurable subsets of Ω. Then we can construct a $(A_1,\dots,A_k)$-*partite version of* W, which is a function $U : (A_1 \amalg \dots \amalg A_k)^2 \to [0,1]$ defined by the formula

$$U_{\restriction(\{i\}\times A_i)\times(\{j\}\times A_j)} = W_{\restriction A_i\times A_j},$$

for each $i,j \in [k]$ distinct, and by setting $U_{\restriction(\{i\}\times A_i)\times(\{i\}\times A_i)}$ to 0 for each $i \in [k]$. U is almost a graphon; the only reason why it need not be is that the measure on $A_1 \amalg \dots \amalg A_k$

defined above, and henceforth denoted by $\nu_{A_1 \amalg \ldots \amalg A_k}$, need not have total measure 1. So, we shall use graphon terminology even for partite versions.

If $X \subset (A_1 \amalg \ldots \amalg A_k)^2$ then the *folded version of* X is simply the set of all $(x, y) \in \Omega$ such that there exist some $i, j \in [k]$ such that $((i, x), (i, y)) \in X$. Note that

$$\text{the } \Omega^2\text{-measure of a folded set is at most}$$
$$\text{the } (A_1 \amalg \ldots \amalg A_k)^2\text{-measure of the original set} \tag{2.1}$$

When k is fixed, the total measure of $\nu_{A_1 \amalg \ldots \amalg A_k}$ is upper-bounded (by k), and thus it is easy to translate Lemma 2.3 as follows.

Lemma 2.4. *Suppose that F is a graph on a vertex set $[k]$. Then for every $\epsilon > 0$ there exists a constant $\delta > 0$ such that the following holds. If $W : \Omega^2 \to [0, 1]$ is a graphon and $A_1, \ldots, A_k$ are measurable subsets of Ω with the property that for the $(A_1, \ldots, A_k)$-version of W, denoted by U, we have $\int_{(x_1,\ldots,x_k) \in (A_1 \amalg \ldots \amalg A_k)^k} U^{\otimes F} \, \mathsf{d}\nu^k_{A_1 \amalg \ldots \amalg A_k} < \delta$ then U can be made F-free by decreasing it in a suitable way such that it changes by at most ϵ in the $\mathcal{L}^1\left((A_1 \amalg \ldots \amalg A_k)^2\right)$-distance.* □

Let us make one easy and well-known comment about the Removal lemma (this comment applies equally to Lemma 2.3 and 2.4). When making a graphon F-free with a small change in the $\mathcal{L}^1$-distance it only makes sense to nullify the graphon at each point where its value is changed. In particular, when positive values of the graphon are uniformly lower bounded, the nullification occurs on a set of small measure. This is stated in the lemma below.

Lemma 2.5. *Suppose that F is a graph on a vertex set $[k]$. Then for every $\alpha > 0$ and $d > 0$ there exists a constant $\beta > 0$ such that the following holds. If $W : \Omega^2 \to \{0\} \cup [d, 1]$ is a graphon and $A_1, \ldots, A_k$ are measurable subsets of Ω with the property that for the $(A_1, \ldots, A_k)$-version of W, denoted by U, we have $\int_{(x_1,\ldots,x_k) \in (A_1 \amalg \ldots \amalg A_k)^k} U^{\otimes F} \, \mathsf{d}\nu^k_{A_1 \amalg \ldots \amalg A_k} < \beta$ then U can be made F-free by nullifying it on a suitable set of $\nu^2_{A_1 \amalg \ldots \amalg A_k}$-measure at most α.* □

2.5. **Regularity lemma.** Here, we introduce Szemerédi's Regularity Lemma in a form that is convenient for our later purposes. Recall that the *density* of a bipartite graph with colour classes A and B is defined as $d(A, B) = \frac{e(A,B)}{|A||B|}$. We say that (A, B) forms an *ϵ-regular pair* if $|d(A, B) - d(X, Y)| < \epsilon$ for each $X \subset A$ and $Y \subset B$ with $|X| \geq \epsilon|A|$ and $|Y| \geq \epsilon|B|$. Suppose that G is a graph on a vertex set V. Let $\mathcal{V} = \{V_0, V_1, \ldots, V_\ell\}$ be a partition of V. We say that a subgraph H of G is an *(ϵ, d)-regularization* of G if

- $V(H) = V$,
- $|V_0| < \epsilon n$,
- for each $1 \leq i < j \leq \ell$, we have that $|V_i| = |V_j|$,

- for each $1 \leq i < j \leq \ell$, we have that the density of the pair (V_i, V_j) in the graph H is either 0 or at least d, and the pair is ϵ-regular in H,
- there are no edges in H incident to any vertex of V_0,
- there are no edges in the graphs $H[V_1], H[V_2], \ldots, H[V_\ell]$, and
- $e(H) \geq e(G) - 2d|V|^2$.

In the setting above, the sets $V_1, \ldots, V_\ell$ are called *clusters*, and the number ℓ is called the *complexity* of the regularization. The Szemerédi Regularity Lemma then can be stated as follows.

Lemma 2.6. *For each $d, \epsilon > 0$ there exists a number L such that any graph admits an (ϵ, d)-regularization of complexity at most L.*

Suppose that G is a graph and H is a (ϵ, d)-regularization of G for a partition $\mathcal{V} = \{V_0, V_1, \ldots, V_\ell\}$. We can then create the *cluster graph* R corresponding to the pair $(H, \mathcal{V})$. This is an edge-weighted graph on the vertex set $[\ell]$ in which an edge ij is present if the bipartite graph $H[V_i, V_j]$ has positive density. The weight on the edge ij in R is then the density of $H[V_i, V_j]$. A standard and well-known fact states that if H is an (ϵ, d)-regularization of a graph G, and R is the corresponding cluster graph, then for the representations W_G and W_R we have that

$$\mathrm{dist}_\square(W_G, W_R) \leq 4d + 2\epsilon \,. \tag{2.2}$$

The so-called Slicing Lemma tells us that a subpair (X, Y) of an ϵ-regular pair (A, B) is (ϵ/x)-regular, if $|X| \geq x|A|$ and $|Y| \geq x|B|$. This fact, however, is useless when we take x smaller than ϵ. The following lemma shows that we can still save the situation if we take X and Y at random. The lemma is a weak version of the the main result of [12]. There, the statement ([12, Theorem 3.6]) is given even in the setting of so-called sparse regular pairs which is a generalization of the regularity concept which we do not need in the present paper.

Lemma 2.7. *Suppose that $\beta, \epsilon_1 > 0$ are given. Then there exist numbers $\epsilon_0 > 0$ and $C > 0$ such that for every ϵ_0-regular pair (A, B) of an arbitrary density $d = d(A, B)$ and for any $a, b > C/d$, the number of pairs (A', B'), $A' \subset A$, $B' \subset B$, $|A'| = a$, $|B'| = b$ which induce ϵ_1-regular pairs of density at least $d - \epsilon_1$ is at least*

$$\exp(-\beta^a - \beta^b)\binom{|A|}{a}\binom{|B|}{b} \,.$$

Last, we need a weak version of the Blow-up Lemma. In our version, we do not require perfect tiling.

Lemma 2.8. *Suppose that we are given a graph F on a vertex set $[k]$ and numbers $\gamma, d > 0$. Then there exist numbers $\epsilon > 0$ and $m_0 \in \mathbb{N}$ such that the following holds. If B is a graph*

whose vertices are partitioned into sets $V_1, \dots, V_k$ with $|V_1| = \dots = |V_k| > m_0$ and which has the property that for each $ij \in E(F)$ the pair (V_i, V_j) is ϵ-regular with density at least d, then $\mathsf{til}(F, B) \geq (1-\gamma)|V_1|$.

2.6. **Subgraphons converging to a subgraphon.** In this section, we provide a proof of the following simple lemma (which seems to be new).

Lemma 2.9. *Suppose that $(W_n)_n$ is a sequence of graphons on Ω converging to a graphon $W : \Omega^2 \to [0,1]$ in the cut-norm. Let $U \leq W$ be an arbitrary graphon. Then there exists a sequence $(U_n)_n$ with $U_n \leq W_n$ which converges to U in the cut-norm.*

Proof. Let $\epsilon > 0$ be arbitrary. There exist k and a partition of Ω into sets of measure $1/k$ each, $\Omega = \Omega_1 \sqcup \dots \sqcup \Omega_k$, such that there exist step-functions with steps $\Omega_i \times \Omega_j$ which approximate U and W up to an error at most ϵ^2 in the $\mathcal{L}^1(\Omega^2)$-norm (each one separately). Such partition and approximation exist since squares generate the sigma-algebra on Ω^2. It is also well known that the said step-functions can be taken to be constant $U^{ij} := k^2 \int_{(x,y)\in\Omega_i\times\Omega_j} U(x,y)\,\mathsf{d}\nu^2$ and $W^{ij} := k^2 \int_{(x,y)\in\Omega_i\times\Omega_j} W(x,y)\,\mathsf{d}\nu^2$ on each square $\Omega_i \times \Omega_j$. A routine calculation gives that for all but at most $2\epsilon k^2$ pairs (i,j) we have

(2.3)
$$\int_{(x,y)\in\Omega_i\times\Omega_j} |U(x,y) - U^{ij}|\,\mathsf{d}\nu^2 \leq \epsilon/k^2 \quad \text{and} \quad \int_{(x,y)\in\Omega_i\times\Omega_j} |W(x,y) - W^{ij}|\,\mathsf{d}\nu^2 \leq \epsilon/k^2\,.$$

Let $\mathcal{P} \subset [k]^2$ be the set of the pairs that fail (2.3).

Since $U^{ij} \leq W^{ij}$ for each i and j, when we define U_n by

$$U_n(x,y) = W_n(x,y) \cdot \frac{U^{ij}}{W^{ij}} \text{ for } (x,y) \in \Omega_i \times \Omega_j,$$

we indeed get a graphon with $U_n \leq W_n$. Let us now bound $\|U - U_n\|_\square$. Let $A, B \subset \Omega$ be arbitrary. We write $R^{ij} = (A \times B) \cap (\Omega_i \times \Omega_j)$. Then by the triangle inequality we have

$$\begin{aligned}
&\left| \int_{(x,y)\in A\times B} U(x,y)\,\mathsf{d}\nu^2 - \int_{(x,y)\in A\times B} U_n\,\mathsf{d}\nu^2 \right| \\
&\leq \sum_{i,j} \left| \int_{(x,y)\in R^{ij}} U(x,y)\,\mathsf{d}\nu^2 - \int_{(x,y)\in R^{ij}} U_n(x,y)\,\mathsf{d}\nu^2 \right| \\
&= \sum_{i,j} \left| \int_{(x,y)\in R^{ij}} U(x,y)\,\mathsf{d}\nu^2 - \int_{(x,y)\in R^{ij}} W_n(x,y)\cdot\frac{U^{ij}}{W^{ij}}\,\mathsf{d}\nu^2 \right| . \qquad (2.4)
\end{aligned}$$

The total contribution of the pairs $(i,j) \in \mathcal{P}$ to the right-hand side of (2.4) is at most 2ϵ. Suppose now that $(i,j) \notin \mathcal{P}$, and let us find an upper bound for the corresponding term

in (2.4).

$$\begin{aligned}
&\left|\int_{(x,y)\in R^{ij}} U(x,y)\,\mathsf{d}\nu^2 - \int_{(x,y)\in R^{ij}} W_n(x,y)\cdot\frac{U^{ij}}{W^{ij}}\,\mathsf{d}\nu^2\right| \\
\overset{(2.3)}{\leq}\ &\left|\int_{(x,y)\in R^{ij}} U^{ij}(x,y)\,\mathsf{d}\nu^2 - \int_{(x,y)\in R^{ij}} W_n(x,y)\cdot\frac{U^{ij}}{W^{ij}}\,\mathsf{d}\nu^2\right| + \epsilon/k^2 \\
\leq\ & U^{ij}\left|\int_{(x,y)\in R^{ij}} 1\,\mathsf{d}\nu^2 - \int_{(x,y)\in R^{ij}} \frac{W_n(x,y)}{W^{ij}}\,\mathsf{d}\nu^2\right| + \epsilon/k^2 \\
\leq\ & U^{ij}\left|\int_{(x,y)\in R^{ij}} 1 - \frac{1}{W^{ij}}\left(\|W_n - W\|_\square + \int_{(x,y)\in R^{ij}} W(x,y)\,\mathsf{d}\nu^2\right)\right| + \epsilon/k^2 \\
=\ & U^{ij}\left|\frac{1}{W^{ij}}\left(\|W_n - W\|_\square + \int_{(x,y)\in R^{ij}} \big(W(x,y) - W^{ij}\big)\,\mathsf{d}\nu^2\right)\right| + \epsilon/k^2 \\
\overset{(2.3)}{\leq}\ & U^{ij}\left(\frac{1}{W^{ij}}\|W_n - W\|_\square + \epsilon/k^2\right) + \epsilon/k^2 \\
\leq\ & \|W_n - W\|_\square + 2\epsilon/k^2\ .
\end{aligned}$$

Thus, the total contribution of the terms $(i,j)\notin\mathcal{P}$ is at most $k^2\|W_n - W\|_\square + 2\epsilon$ which is less than 3ϵ for large enough n. Consequently, $\left|\int_{(x,y)\in A\times B} U\,\mathsf{d}\nu^2 - \int_{(x,y)\in A\times B} U_n\,\mathsf{d}\nu^2\right| \leq 5\epsilon$.

Thus, the proof proceeds by letting ϵ go to zero and diagonalizing. □

3. The theory of tilings in graphons

3.1. **A naive approach.** We want to gain some understanding what an F-tiling in a graphon should be. Let us take our first motivation from the world of finite graphs. Suppose that F and G are finite graphs, F is on the vertex set $[k]$. Then each F-tiling can be viewed as a set $\mathbf{X}\subset V(G)^k$ such that

- (a) for each $T\in\mathbf{X}$ we have that T_iT_j forms an edge of G for each $ij\in E(F)$, $i<j$, and
- (b) for each $v\in V(G)$ there is at most one pair $(T,i)\in\mathbf{X}\times[k]$ such that $v=T_i$.

The size of the F-tiling $\mathbf{X}$ is obviously $|\mathbf{X}|$, which can be rewritten using the projection $\pi_1 : V(G)^k\to V(G)$ on the first coordinate as

$$|\mathbf{X}| = |\pi_1(\mathbf{X})|\ . \tag{3.1}$$

Let A be the adjacency matrix of G. Then Condition (a) can be rewritten as

- (a') for each $T\in\mathbf{X}$ we have that $\prod_{ij\in E(F),i<j} A_{T_iT_j} > 0$.

These conditions can be translated directly to graphons. That is, we might want to say that a set $\mathbf{Y}\subset\Omega^k$ is an F-tiling in a graphon $W:\Omega^2\to[0,1]$ if

- for each $1 \le i < j \le k$ the map $\pi_j \circ \pi_i^{-1}$ is a ν-measure preserving bijection between $\pi_i(\mathbf{Y})$ and $\pi_j(\mathbf{Y})$,
- for each $T \in \mathbf{Y}$ we have that $\prod_{ij \in E(F), i<j} W(T_i, T_j) > 0$,
- and for each $x \in \Omega$ there is at most one pair $(T, i) \in \mathbf{Y} \times [k]$ such that $x = T_i$.

Using (3.1), we would then say that $\nu(\pi_1(\mathbf{Y}))$ is the size of the F-tiling $\mathbf{Y}$.

There is however a substantial problem with this approach in that whether the condition $\prod_{ij \in E(F), i<j} W(T_i, T_j) > 0$ holds depends on values of W of ν^2-measure zero. The theory of graphons cannot in principle achieve this level on sensitivity. Thus, a different approach is needed.

3.2. **Introducing tilings in graphons.** As we saw in the previous section the main problem with the straightforward graphon counterpart to F-tilings of finite graphs was that the limiting object was too centralized; so centralized that all the information was needed to be carried on a set of measure zero. Observe that in a setting of a finite graph G, fractional F-tilings can be more "spread out" than integer F-tilings in the sense that the weight of the latter is always supported on at most $v(G)/v(F)$ many copies of F while it can be supported on up to $\binom{v(G)}{v(F)}$ many copies in the former. Thus, the notion of fractional F-tilings seems better admissible to a limit counterpart. Indeed, as we shall see in Section 3.3, in the world of graphons there is no difference between F-tilings and fractional F-tilings. Thus, we want to approach the notion of F-tilings in graphons by looking at fractional F-tilings in finite graphs.

Suppose that F and G are finite graphs, F is on the vertex set $[k]$, the order of G is n, and let $\mathsf{t}_G : \mathcal{F}_F(G) \to [0,1]$ be a fractional F-tiling in G. Let $W_G : \Omega^2 \to [0,1]$ be a graphon representation of G. Let $(\Omega_u)_{u \in V(G)}$ be the partition of Ω corresponding to this representation. We can then associate a measure μ on Ω^k corresponding to t_G by the defining formula

$$\mu(A) = \sum_{u_1 u_2 \cdots u_k \in \mathcal{F}_F(G)} \frac{\mathsf{t}_G(u_1 u_2 \cdots u_k)}{n} \cdot \frac{\nu^k\left((\Omega_{u_1} \times \Omega_{u_2} \times \cdots \times \Omega_{u_k}) \cap A\right)}{\nu^k\left(\Omega_{u_1} \times \Omega_{u_2} \times \cdots \times \Omega_{u_k}\right)}, \tag{3.2}$$

for each measurable $A \subset \Omega^k$. In words, we introduce a linear renormalization on t_G. Then the measure μ spreads the total of $\frac{\mathsf{t}_G(u_1 u_2 \cdots u_k)}{n}$ uniformly over each rectangle $\Omega_{u_1} \times \Omega_{u_2} \times \cdots \times \Omega_{u_k}$. An example is given in Figure 3.1. The following properties of μ are obvious:

(1) μ is absolutely continuous with respect to ν^k,
(2) μ is supported on a subset of the set $\mathcal{F}_F(W_G)$,
(3) for each $X \subset \Omega$ we have $\sum_{\ell=1}^{k} \mu\left(\prod_{i=1}^{\ell-1} \Omega \times X \times \prod_{i=\ell+1}^{k} \Omega\right) \le \nu(X)$.

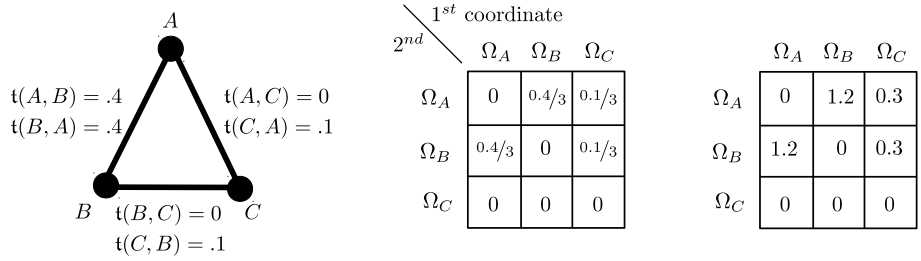


FIGURE 3.1. The left-hand picture shows a sample graph G and a fractional K_2-tiling in it. The middle picture shows the total masses of the nine rectangles with respect to the measure μ defined by (3.2). The right-hand picture then shows the point-wise values of the Radon–Nikodym derivative $\frac{\mathrm{d}\mu}{\mathrm{d}\nu^2}$.

Property (1) allows us to take the Radon–Nikodym derivative $\frac{\mathrm{d}\mu}{\mathrm{d}\nu^k}$ of μ. We define functions that arise in this way as F-tilings.[2]

Definition 3.1. Suppose that $W : \Omega^2 \to [0,1]$ is a graphon, and that F is a graph on the vertex set $[k]$. An $\mathcal{L}^1(\Omega^k)$-function $\mathsf{t} : \Omega^k \to [0,+\infty)$ is called an *F-tiling* in W if

$$\operatorname{supp} \mathsf{t} \subset \mathcal{F}_F(W) \,, \tag{3.3}$$

and we have for each $x \in \Omega$ that

$$\sum_{\ell=1}^{k} \int_{(x_1,\dots x_{\ell-1},x_{\ell+1},\dots,x_k)\in\Omega^{k-1}} \mathsf{t}(x_1,\dots,x_{\ell-1},x,x_{\ell+1},\dots,x_k)\,\mathsf{d}\nu^{k-1} \le 1 \,. \tag{3.4}$$

The *size* of an F-tiling t is $\|\mathsf{t}\| = \int_{(x_1,\dots,x_k)\in\Omega^k} \mathsf{t}(x_1,\dots,x_k)\,\mathsf{d}\nu^k$. The *$F$-tiling number* of W, denoted by $\mathsf{til}(F,W)$, is the supremum of sizes over all F-tilings in W.

A couple of remarks is in place. First, if F and F' are copies of the same graph on the vertex set $[k]$ with the vertex-labels permuted we can get an F'-tiling from an F-tiling by permuting the corresponding coordinates. Second, observe that the definition does not depend on the values of W, only on $\operatorname{supp} W$. Third, we have $\mathsf{til}(F,W) \le \frac{1}{k}$. This is a counterpart to the finite statement $\mathsf{til}(F,G) \le \frac{v(G)}{k}$. Fourth, the supremum in the definition of $\mathsf{til}(F,W)$ need not be attained. An example is given in Proposition 3.3. Fifth, when W_G is a representation of a finite graph G, we have $\mathsf{ftil}(F,G) = v(G)\mathsf{til}(F,W_G)$ (we emphasize that the relevant graph parameter is $\mathsf{ftil}(F,G)$, not $\mathsf{til}(F,G)$). While this would easily follow from tools we develop later (specifically, from Proposition 3.13 and Theorem 3.16), here we give a self-contained proof.

[2]As said, later we shall see that there is no distinction between integral and fractional tilings for graphons. Thus, even though the motivation for the current notion comes from *fractional* tilings in finite graphs, we call the resulting graphon concept simply *tilings*.

Proposition 3.2. *Suppose that F and G are finite graphs and that W_G is a graphon representation of G. Let n be the number of vertices of G. Then we have* $\mathsf{ftil}(F,G) = n \cdot \mathsf{til}(F,W_G)$.

Proof. Let us assume that the vertex set of G is $V(G) = [n]$. Let us consider the partition $\Omega = \Omega_1 \sqcup \ldots \sqcup \Omega_n$ of the space that hosts W_G into sets representing the individual vertices of G.

Suppose that $\mathsf{t}_G : \mathcal{F}_F(G) \to [0,1]$ is an arbitrary fractional F-tiling in G. Define $\mathsf{t} : \Omega^k \to [0,+\infty)$ to be constant $\mathsf{t}_G(i_1, i_2, \cdots, i_k)/n$ on each set $\Omega_{i_1} \times \Omega_{i_2} \times \cdots\cdots \times \Omega_{i_k}$. It is straightforward to check that t is an F-tiling in W of size $\|\mathsf{t}\|/n$. We conclude that $\mathsf{ftil}(F,G) \leq n \cdot \mathsf{til}(F,W_G)$.

Suppose that $\mathsf{t} : \Omega^k \to [0,+\infty)$ is an arbitrary F-tiling in W. Define $\mathsf{t}_G : \mathcal{F}_F(G) \to [0,1]$ to be constant $n \cdot \int_{(x_1,\cdots,x_k)\in\Omega_{i_1}\times\Omega_{i_2}\times\cdots\times\Omega_{i_k}} \mathsf{t}(x_1,\cdots,x_k)\,\mathsf{d}\nu^k$ for each $(i_1,i_2,\ldots,i_k) \in \mathcal{F}_F(G)$. It is straightforward to check that t_G is an F-tiling in W of size $n\|\mathsf{t}\|$. We conclude that $\mathsf{ftil}(F,G) \geq n \cdot \mathsf{til}(F,W_G)$. □

Proposition 3.3. *Consider the graphon $W : [0,1]^2 \to [0,1]$ defined as*

$$W(x,y) = \begin{cases} 0 & \text{if } x+y > 1/2 \\ 1 & \text{if } x+y \leq 1/2 \end{cases}. \tag{3.5}$$

Then $\mathsf{til}(K_2, W) = 1/2$*, but there exists no K_2-tiling of size $1/2$.*

Proof. For an arbitrary $C \geq 1$ we can take a function $\mathsf{t}_C : [0,1]^2 \to [0,+\infty)$ defined by

$$\mathsf{t}_C(x,y) = \begin{cases} 0 & \text{if } x+y > 1/2 \text{ or } x+y < 1/2 - 1/C \\ C & \text{if } x+y \in [1/2 - 1/C, 1/2] \end{cases}.$$

It is easy to see that t_C is a K_2-tiling and that $\|\mathsf{t}_C\| \geq 1/2 - 1/C$. Thus $\mathsf{til}(F,W) \geq 1/2$. Yet, there is no K_2-tiling of size $1/2$. To see this, assume for a contradiction that t is such a K_2-tiling. By replacing $\mathsf{t}(x,y)$ by $\frac{1}{2}(\mathsf{t}(x,y) + \mathsf{t}(y,x))$, we can assume that $\mathsf{t}(x,y)$ is symmetric. To get the contradiction, it is enough to show that t is constant zero almost everywhere on each rectangle $[0,\alpha]\times[0,1-\alpha]$ (see Figure 3.2). Firstly, observe that because t is symmetric, (3.4) tells us that $\int_{y\in[0,1]} \mathsf{t}(x_0,y)\,\mathsf{d}y \leq 1/2$ for every $x_0 \in [0,1]$. Since $1/2 = \|\mathsf{t}\| = \int_{x_0\in[0,1]} \left(\int_{y\in[0,1]} \mathsf{t}(x_0,y)\,\mathsf{d}y\right) \mathsf{d}x_0$, we conclude that $\int_{y\in[0,1]} \mathsf{t}(x_0,y)\,\mathsf{d}y = 1/2$ for almost every $x_0 \in [0,1]$. Applying this for $x_0 \geq 1-\alpha$, we get that the integral $\int_{x\in[1-\alpha,1]} \left(\int_{y\in[0,x]} \mathsf{t}(x,y)\,\mathsf{d}y\right) \mathsf{d}x$ (over the green triangle in Figure 3.2) is exactly $\alpha/2$. The value $\int_{x\in[0,1]} \left(\int_{y\in[0,\alpha]} \mathsf{t}(x,y)\,\mathsf{d}y\right) \mathsf{d}x$ is at most $\alpha/2$ by (3.4). We conclude that t is zero almost everywhere on $[0,\alpha]\times[0,1-\alpha]$. □

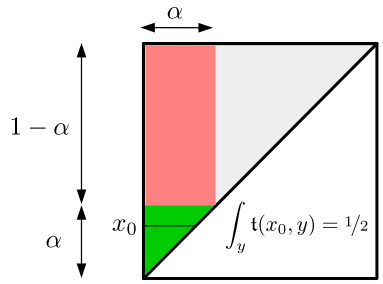


FIGURE 3.2. For the graphon defined by (3.5), any symmetric K_2-tiling t of weight $\|\mathsf{t}\| = 1/2$ satisfies that t is zero almost everywhere on the red rectangle $[0, \alpha] \times [0, 1 - \alpha]$.

3.3. **Graphon tilings versus fractional graphon tilings.** Let us now explain that there should be no difference between F-tilings and fractional F-tilings for graphons. In the world of finite graphs, the former is a proper subset of the latter. So, let us show that in the graphon world we have the opposite inclusion as well. That is, we want to show that a fractional F-tiling t of certain size in W yields an F-tiling of approximately the same size (after rescaling) in finite graphs G_n of a sequence that converges to W. To this end, fix a sequence $(R_n)_n$ of finer and finer Szemerédi regularizations of the graphs G_n. That is, R_n is a cluster graph whose edges carry weights in the interval $[0, 1]$ and which comes from regularizing G_n with an error ϵ_n, with $\lim_n \epsilon_n = 0$. Since the sequence $(R_n)_n$ converges to W in the cut-distance we can take t and turn into a fractional F-tiling t_n in R_n of size $\approx \|\mathsf{t}\| v(R_n)$ (for n large). Of course, the fact that we can transfer a fractional F-tiling on a graphon to a fractional F-tiling on a graph in a sequence converging to this graphon needs a formal statement, which we give in Theorem 3.4. In the last step, we recall that Lemma 2.8 provides a standard tool for finding an F-tiling of size $\approx \|\mathsf{t}\| n$ in G_n based on an fractional F-tiling of size $\approx \|\mathsf{t}\| v(R_n)$ in the cluster graph R_n. A description by picture how this step is done is given in Figure 3.3.

Note that the above also explains our remark in Section 3.2 that the notion of tilings depends only on the support of the graphon, and not on its values themselves. Indeed, Lemma 2.8 was applied to subpairs of regular pairs. And Lemma 2.8 works for regular pairs of an arbitrary positive density.[3]

3.4. **Tilings and convergence.** Suppose F is a finite graph and that we have a sequence (G_n) of graphs, $v(G_n) = n$, that converges to a graphon $W : \Omega^2 \to [0, 1]$ in the cut-distance. Observe that the numbers $\mathsf{til}(F, G_n)$ grow (up to) linearly in n. Hence, we could hope that $\frac{\mathsf{til}(F,G_n)}{n} \to \mathsf{til}(F, W)$ holds. Unfortunately, this is not always true. Indeed,

[3]Lower density will require a finer error parameter of regularity. But in the limit case, we have "infinitely fine regular pairs".

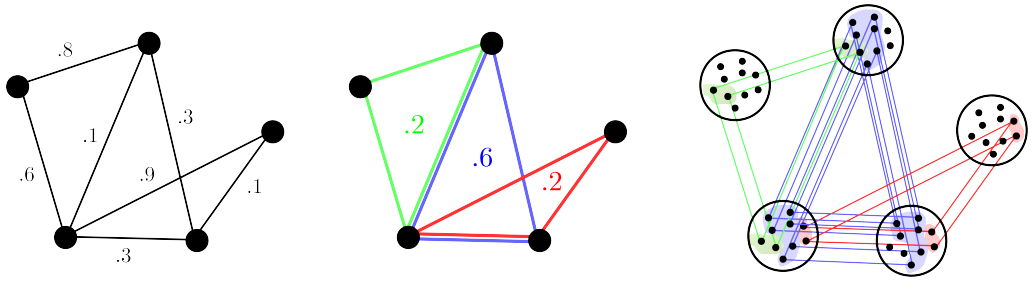


FIGURE 3.3. The left-hand picture shows the graph R_n with the edge-weights. The middle picture shows a fractional triangle-tiling t in R_n (the different copies of a triangle are depicted with different colors). On the right-hand picture, we split the clusters of G_n according to t, thus obtaining subpairs of the original regular pairs. Using Lemma 2.8, we get an F-tiling in G_n of size proportional to the size of the fractional F-tiling in R_n. (This example is oversimplified: If the triangle-tiling is "spread out", meaning that $\mathsf{t}(H)$ is of order ϵ_n or less for many copies H of the triangle, then the regularity of the pairs is not inherited to the subpairs and Lemma 2.8 cannot be applied. As we shall see, this issue can be resolved.)

consider $F = K_2$, and the graphs G_n are defined as a perfect matching on n vertices for n even, and as an edgeless n-vertex graph for n odd. Then $(G_n)_n$ converges to the zero graphon but $\frac{\mathsf{til}(F,G_n)}{n}$ oscillates between 0 and $\frac{1}{2}$. So, while we see that the quantity $\mathsf{til}(F,\cdot)$ cannot be continuous on the space of graphons, here we establish lower semicontinuity for graphs converging to graphon.

Theorem 3.4. *Suppose that F is a finite graph and let (G_n) be a sequence of graphs of growing orders converging to a graphon $W : \Omega^2 \to [0,1]$ in the cut-distance. Then we have that $\liminf_n \frac{\mathsf{til}(F,G_n)}{v(G_n)} \geq \mathsf{til}(F,W)$.*

The proof of Theorem 3.4 is given in Section 4.4.

Remark 3.5. *Lower semicontinuity is the more applicable half of continuity for the purposes of extremal graph theory. Indeed, in extremal graph theory one typically wants to bound the F-tiling number of graphs from a graph class of interest from below (as opposed to bounding from above). Thanks to Theorem 3.4 this can be achieved (in the asymptotic sense) by finding a lower bound $\mathsf{til}(F,W)$ for each limiting graphon W. This proof scheme is used in* [13].

Remark 3.6. *Let us compare the situation with the sparse case. That is, suppose that we have a Benjamini–Schramm convergent sequence $(G_n)_n$ of bounded degree graphs, and*

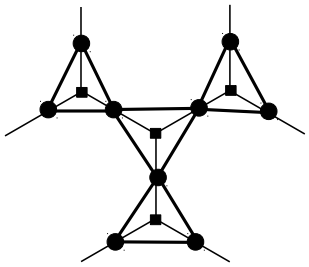


FIGURE 3.4. The transformation described in Remark 3.6. A part of the random (bipartite) graph H_n shown with thin lines and box vertices. The corresponding line graph G_n shown with thick lines and circle vertices.

we are concerned with the sequence $\left(\mathsf{til}(F,G_n)/v(G_n)\right)_n$ *for some fixed graph* F*. A theorem of Nguyen and Onak* [20] *(reproved later by Elek and Lippner* [10] *and by Bordanave, Lelarge, and Salez* [4]*) tells us that indeed the normalized matching ratios (i.e., the most important case* $F = K_2$*) converge. In contrast, Endre Csóka has communicated to us that the sequence* $\left(\mathsf{til}(K_3,G_n)/v(G_n)\right)_n$ *needs not to converge. The construction he uses to this end is as follows. For* n *even, consider a random cubic graph* H_n *of order* $2n$*, for* n *odd consider a random bipartite cubic graph* H_n *of order* $2n$*. It is well-known that the resulting sequence is Benjamini–Schramm convergent almost surely, and that the limit is the rooted infinite 3-regular tree. Let* G_n *be the line graphs of* H_n*. This is shown in Figure 3.4. There is a one-to-one correspondence between independent sets in* H_n*'s and triangle tilings in* G_n*'s. Thus, when* H_n *is bipartite (*n *even), we have* $\mathsf{til}(K_3, G_n) = v(G_n)/3$*. On the other hand, a result of Bollobás about independence number of random cubic graphs* [3] *translates as* $\mathsf{til}(K_3, G_n) \le 0.92 \cdot v(G_n)/3$ *asymptotically almost surely for* n *odd.*

A similar construction shows that the parameter $\mathsf{til}(F,G)/v(G)$ *is discontinuous in the Benjamini–Schramm topology for each 2-connected graph* F*. In that construction, random cubic graphs and random bipartite cubic graphs are replaced by random* $v(F)$*-regular graphs and random bipartite* $v(F)$*-regular graphs, respectively. It would be interesting to fully characterize the graphs* F *for which the quantity* $\mathsf{til}(F,\cdot)/v(\cdot)$ *is continuous. The case when* F *is a path or a star seems particularly interesting.*

In Theorem 3.7 below we state a version of Theorem 3.4 for a sequence (W_n) of graphons converging to a graphon W. Then we have $\liminf_n \mathsf{til}(F, W_n) \ge \mathsf{til}(F, W)$. Note that this is not a strengthening of Theorem 3.4 since in general we do not have (even approximately) that $\frac{\mathsf{til}(F,G)}{v(G)} = \mathsf{til}(F, U_G)$ for a graphon representation U_G of the graph G.

Theorem 3.7. *Suppose that* (W_n) *is a sequence of graphons* $W_n : \Omega^2 \to [0,1]$ *converging to a graphon* $W : \Omega^2 \to [0,1]$ *in the cut-distance and let* F *be an arbitrary graph. Then we have that* $\liminf_n \mathsf{til}(F, W_n) \ge \mathsf{til}(F, W)$*.*

The proof of Theorem 3.7 is given in Section 4.3.

3.5. **Robust tiling number.** While — as we have explained — the function $\mathsf{til}(F,\cdot)$ is not upper-semicontinuous, it is "upper-semicontinuous in the cut-distance after small $\mathcal{L}^1$-perturbations". This is stated below.

Proposition 3.8. *Suppose that F is an arbitrary graph and that $W:\Omega^2\to[0,1]$ is a graphon. Then for an arbitrary $\eta>0$ there exists a number $\delta>0$ such that each graphon U with $\|W-U\|_\square<\delta$ can be decreased in the $\mathcal{L}^1(\Omega^2)$-distance by at most η (in a suitable way) so that we obtain a graphon U^* for which $\mathsf{til}(F,U^*)\le\mathsf{til}(F,W)+\eta$.*

The proof of Proposition 3.8 is given in Section 4.5. Let us state a version of Proposition 3.8 which deals with the situation when the the graphon W is approximated by a finite graph and not a graphon.

Proposition 3.9. *Suppose that F is an arbitrary graph and that $W:\Omega^2\to[0,1]$ is a graphon. Then for an arbitrary $\eta>0$ there exists a number $\delta>0$ such that in each finite graph G with $\mathrm{dist}_\square(W,G)<\delta$ we can erase at most $\eta v(G)^2$ edges (in a suitable way) so that for the resulting graph G^* we have $\mathsf{til}(F,G^*)\le n\cdot(\mathsf{til}(F,W)+\eta)$.*

Proposition 3.9 is not stronger than Proposition 3.8 since it deals with graphs only. On the other hand, Proposition 3.9 does not follow directly from Proposition 3.8 since it contains the extra assertion that each edge is either entirely deleted, or kept entirely untouched. We do not include a proof of Proposition 3.9 as it is analogous to that of Proposition 3.8.

Propositions 3.8 motivates the following definition. Given a finite graph F, a number $\epsilon>0$ and a graphon $W:\Omega^2\to[0,1]$ we define

$$\mathsf{til}_\epsilon(F,W)=\inf_{W^-}\mathsf{til}\left(F,W^-\right)\ , \tag{3.6}$$

where W^- ranges over all graphons with $W^-\le W$ and with $\|W-W^-\|_1\le\epsilon$. Similarly, for an n-vertex graph G, we define

$$\mathsf{til}_\epsilon(F,G)=\inf_{G^-}\mathsf{til}\left(F,G^-\right)\ ,$$

where G^- ranges over all subgraphs of G with at most ϵn^2 edges deleted from G. These "robust versions of the tiling number" are continuous.

Theorem 3.10. *Suppose that F is a finite graph and $\epsilon>0$. Then the quantity $\mathsf{til}_\epsilon(F,\cdot)$ is continuous on the space of graphons equipped with the cut-norm.*

Theorem 3.11. *Suppose that F is a finite graph and $\epsilon>0$. Suppose that $(G_n)_n$ is a sequence of graphs of growing orders that converges in the cut-distance to a graphon W. Then the sequence $\left(\mathsf{til}_\epsilon(F,G_n)/v(G_n)\right)_n$ converges to $\mathsf{til}_\epsilon(F,W)$.*

We shall give a proof of Theorem 3.10 in Section 4.6. We omit a proof of Theorem 3.11 but it follows easily by combining the proofs of Theorem 3.10 and Theorem 3.4.

Let us give an interpretation of Theorem 3.11 in property testing in the so-called dense model. Formally, let $\mathcal{G}$ be the class of all isomorphism classes of finite graphs. We say that a function (called often a *parameter*) $f : \mathcal{G} \to \mathbb{R}$ is *testable* if for each $\epsilon > 0$ there exists a number $r = r(\epsilon)$ and a function (called often a *tester*) $g : \mathcal{G} \to \mathbb{R}$ such that for each $G \in \mathcal{G}$, we have

$$\mathbf{P}\left[|f(G) - g(H)| > \epsilon\right] < \epsilon \,,$$

where $H = G[A]$ is the subgraph of G induced by selecting a set A of r vertices at random. A very convenient and concise characterization of testable parameters was provided in [5] in the language of graph limits: *A graph parameter f is testable if and only if it is continuous in the cut-distance.*[4] Also, in the positive case, we can take the tester to be $g = f$.

Thus, Theorem 3.11 tells us that $\mathsf{til}_\epsilon(F, \cdot)$ is testable. For example, suppose that we have a large computer network G and we want to estimate the size of the largest possible matching in it. But a small number of links between the computers may become broken, and we do not know in advance which links these will be, so we actually want to estimate $\mathsf{til}_\epsilon(K_2, G)$. Then a good estimate to this quantity can be provided just by computing $\mathsf{til}_\epsilon(K_2, H)$ for a large, randomly selected induced subgraph H of G.

3.6. **Fractional covers and LP duality.** We define fractional covers in graphons in a complete analogy to the finite notion.

Definition 3.12. Suppose that $W : \Omega^2 \to [0,1]$ is a graphon, and F is a graph on the vertex set $[k]$. A measurable function $\mathfrak{c} : \Omega \to [0,1]$ is called a *fractional F-cover* in W if

$$\nu^k\left(\mathcal{F}_F(W) \cap \left\{(x_1, x_2, \ldots, x_k) \in \Omega^k : \sum_{i=1}^{k} \mathfrak{c}(x_i) < 1\right\}\right) = 0 \,.$$

The *size* of $\mathfrak{c}$, denoted by $\|\mathfrak{c}\|$, is defined by $\|\mathfrak{c}\| = \int_{x\in\Omega} \mathfrak{c}(x)\, \mathsf{d}\nu$. The *fractional F-cover number* $\mathsf{fcov}(F, W)$ of W is the infimum of the sizes of fractional F-covers in W.

As with tilings, let us note that the notion of fractional covers does not depend on the values of the graphon but only on its support.

The following proposition gives us a simple but useful relation between the fractional F-cover number between a graph and its graphon representation.

Proposition 3.13. *Suppose that F and G are finite graphs and that W_G is a graphon representation of G. Let n be the number of vertices of G. Then we have* $\mathsf{fcov}(F, G) = n \cdot \mathsf{fcov}(F, W_G)$.

[4] See also [19] for more advanced connections between property testing and graph limits.

Proof. Suppose that $\Omega_1 \sqcup \Omega_2 \sqcup \ldots \sqcup \Omega_n$ is the partition of the underlying probability space Ω given by the graphon representation W. Any fractional F-cover of G can be represented as a function with steps $\Omega_1, \Omega_2, \ldots, \Omega_n$. This step-function is a fractional F-cover of W_G. This shows that $\mathsf{fcov}(F, G) \geq n \cdot \mathsf{fcov}(F, W_G)$. On the other hand, let $\mathfrak{c} : \Omega \to [0, 1]$ be an arbitrary fractional F-cover of W_G. Let us modify $\mathfrak{c}$ on each set Ω_i by replacing it by its essential infimum on that set. Since W_G is constant on the rectangles $\Omega_i \times \Omega_j$ we get that the modified step-function $\mathfrak{c}'$ is still a fractional F-cover of W_G, obviously with $\|\mathfrak{c}'\| \leq \|\mathfrak{c}\|$. But $\mathfrak{c}'$ can be viewed as a fractional F-cover of G. This shows that $\mathsf{fcov}(F, G) \leq n \cdot \mathsf{fcov}(F, W_G)$. □

The sequence of graphons representing a growing number of isolated copies of F converges to the constant zero graphon. This shows that the fractional F-cover number is not continuous. We can, however, establish lower semicontinuity using the following theorem.

Theorem 3.14. *Suppose that F is a graph on the vertex set $[k]$. Let (W_n) be a sequence of graphons $W_n : \Omega^2 \to [0, 1]$ converging to a graphon W in the cut-norm. Suppose that $\mathfrak{c}_n$ is a fractional F-cover of W_n. Then an arbitrary weak* accumulation point $\mathfrak{c}$ of $(\mathfrak{c}_n)$ is a fractional F-cover of W.*

We give the proof of Theorem 3.14 in Section 4.1.

Recall that if a sequence of functions (f_n) weak* converges to a function f, we have $\lim \int f_n = \int f$. Also, recall that by Theorem 2.1 the set of all functions of $\mathcal{L}^\infty$-norm at most 1 is sequentially compact with respect to the weak* topology. Thus, we get the following.

Corollary 3.15. *Suppose that F is a finite graph. Suppose that (W_n) is a sequence of graphons converging to W. Then $\liminf_n \mathsf{fcov}(F, W_n) \geq \mathsf{fcov}(F, W)$.*

Corollary 3.15 has two important consequences. Firstly, it tells us that $\mathsf{fcov}(F, \cdot)$ is lower-semicontinuous with respect to the cut-distance. Secondly, if we take $W_1 = W_2 = \cdots = W$, and $\mathfrak{c}_n$ a sequence of fractional F-covers whose sizes tend to $\mathsf{fcov}(F, W)$, we get a fractional F-tiling which attains the value $\mathsf{fcov}(F, W)$. Thus, unlike with $\mathsf{til}(F, W)$,

(3.7) the value of $\mathsf{fcov}(F, W)$ is attained by some fractional F-cover.

We are now ready to state the LP duality theorem for tilings in graphons.

Theorem 3.16. *Suppose that $W : \Omega^2 \to [0, 1]$ is a graphon and F is an arbitrary finite graph. Then we have $\mathsf{til}(F, W) = \mathsf{fcov}(F, W)$.*

We give the proof of Theorem 3.16 in Section 4.2.

Theorem 3.16 is a very convenient tool for extremal graph theory. More specifically, suppose that we want to prove that for a fixed graph F and for a family of graphs $\mathcal{G}$ we

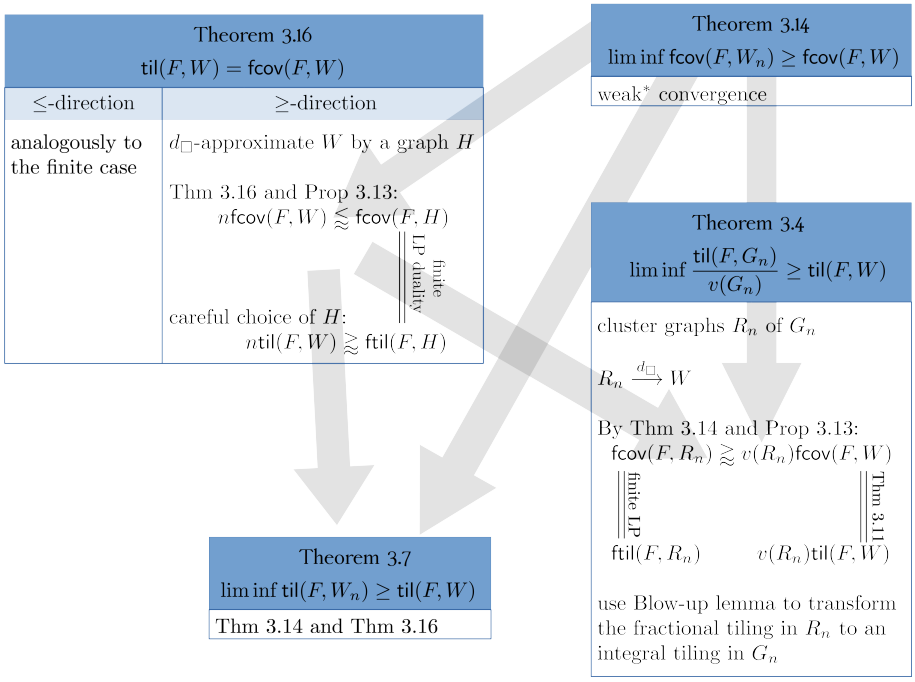


FIGURE 4.1. Main proof steps and dependencies of main results from Section 3.

have that $\mathsf{til}(F,G) \geq (\gamma + o(1))v(G)$ for each $G \in \mathcal{G}$. (Here, $o(1)$ tends to 0 as $v(G)$ goes to infinity.) By combining Theorem 3.4 and Theorem 3.16 it suffices to show that no graphon arising as a limit of graphs from $\mathcal{G}$ has a fractional F-cover of size less than γ. We shall see one particular application of this scheme in Section 5 and another one is used in the accompanying paper [13] on Komlós's Theorem.

4. Proofs

Figure 4.1 shows the dependencies and the key steps in the proofs of our main results.

4.1. **Proof of Theorem 3.14.** Without loss of generality let us assume that $\mathfrak{c}_n \xrightarrow{w^*} \mathfrak{c}$. In order to show that $\mathfrak{c}$ is a fractional F-cover of W, we need to show that for any measurable subsets $A_1, \dots, A_k \subset \Omega$ for which there are numbers $\alpha_1, \dots, \alpha_k$, $\sum \alpha_i < 1$ with $\mathfrak{c}_{\restriction A_i} < \alpha_i$, we have $\int_{(x_1,\dots,x_k)\in\prod A_i} W^{\otimes F}(x_1,\dots,x_k)\,\mathsf{d}\nu^k = 0$. Let $a = 1 - \sum \alpha_i$. Suppose for contradiction that $\int_{(x_1,\dots,x_k)\in\prod A_i} W^{\otimes F}(x_1,\dots,x_k)\,\mathsf{d}\nu^k > 0$. Let $\xi > 0$ be such that the set

$$\mathbf{X} = \left\{ \mathbf{x} \in \prod_i A_i : W^{\otimes F}(\mathbf{x}) > \xi \right\}$$

has positive ν^k-measure. Since k-dimensional boxes generate the sigma-algebra on Ω^k, for each $d > 0$ we can find sets $B_1 \subset A_1, \ldots, B_k \subset A_k$ of positive measure such that

$$\nu^k\left(\mathbf{X} \cap \prod_i B_i\right) \geq (1-d)\prod_i \nu(B_i)\ . \tag{4.1}$$

Let us fix such sets for $d := (a/12k)^k$.

Let us take N_1 such that for each $n > N_1$ we have $\|W - W_n\|_\square < (a/13k)^{k+2} \cdot \prod_i \nu(B_i) \cdot \xi$. Lemma 2.2 tells us that for each collection $D_1, \ldots, D_k$ of measurable subsets of Ω we have

$$\begin{aligned}&\left|\int_{(x_1,\ldots,x_k)\in\prod D_i} W_n^{\otimes F}(x_1,\ldots,x_k)\,\mathsf{d}\nu^k - \int_{(x_1,\ldots,x_k)\in\prod D_i} W^{\otimes F}(x_1,\ldots,x_k)\,\mathsf{d}\nu^k\right| \\ &< (a/13k)^k \cdot \prod_i \nu(B_i)\cdot\xi\ .\end{aligned} \tag{4.2}$$

Since $\mathfrak{c}_n \xrightarrow{\text{w}^*} \mathfrak{c}$, there exists a number N_2 such that for each $n > N_2$ we have

$$\int_{x\in B_i} \mathfrak{c}_n(x)\,\mathsf{d}\nu \leq \int_{x\in B_i} \mathfrak{c}(x) + \frac{a}{3k}\,\mathsf{d}\nu \leq \nu(B_i)\left(\alpha_i + \frac{a}{3k}\right) \quad \text{for each } i = 1,\ldots,k. \tag{4.3}$$

Now, fix $n > \max(N_1, N_2)$. For $i = 1,\ldots,k$, define $D_i = B_i \cap \mathfrak{c}_n^{-1}([0, \alpha_i + 2a/3k])$. Then we have

$$\begin{aligned}\int_{x\in B_i} \mathfrak{c}_n(x)\,\mathsf{d}\nu &= \int_{x\in D_i} \mathfrak{c}_n(x)\,\mathsf{d}\nu + \int_{x\in B_i\setminus D_i} \mathfrak{c}_n(x)\,\mathsf{d}\nu \\ &\geq \nu(D_i)\cdot 0 + \nu(B_i\setminus D_i)\cdot(\alpha_i + 2a/3k) \\ &= \nu(B_i)\left(\alpha_i + 2a/3k - \frac{\nu(D_i)}{\nu(B_i)}\cdot(\alpha_i + 2a/3k)\right)\ .\end{aligned}$$

Plugging this into (4.3) we get

$$\nu(B_i)\left(\alpha_i + 2a/3k - 2\frac{\nu(D_i)}{\nu(B_i)}\right) \leq \nu(B_i)\,(\alpha_i + a/3k)\ ,$$

and consequently,

$$\nu(D_i) \geq \frac{a}{6k}\cdot\nu(B_i)\ . \tag{4.4}$$

Observe also that for each $\mathbf{x} \in \prod D_i$ we have $\sum_i \mathfrak{c}_n(\mathbf{x}_i) \leq \sum_i (\alpha_i + 2a/3k) = 1 - a/3 < 1$. Since $\mathfrak{c}_n$ is a fractional F-cover of W_n, we conclude that

$$\int_{(x_1,\cdots,x_k)\in\prod D_i} W_n^{\otimes F}(x_1,\ldots,x_k)\,\mathsf{d}\nu^k = 0\ . \tag{4.5}$$

On the other hand,

$$\begin{aligned}\int_{(x_1,\cdots,x_k)\in\prod D_i} W_n^{\otimes F}(x_1,\dots,x_k)\,\mathsf{d}\nu^k &\geq \int_{(x_1,\cdots,x_k)\in\mathbf{X}\cap\prod D_i} \xi\,\mathsf{d}\nu^k\\ \boxed{\text{by (4.1)}} &\geq \left(\prod_i \nu(D_i) - d\prod_i \nu(B_i)\right)\xi\\ \boxed{\text{by (4.4)}} &\geq (a/12k)^k \prod_i \nu(B_i)\cdot\xi\ .\end{aligned}$$

This, together with (4.5) contradicts (4.2).

4.2. **Proof of Theorem 3.16.** We split the proof into the easy "$\leq$"-part and the difficult "$\geq$"-part. The former is given in Proposition 4.1 and its proof is a straightforward modification of the finite version. The latter is given in Proposition 4.2. In the proof of Proposition 4.2 we first approximate a graphon by a finite graph on which we use finite LP-duality as a black-box.

We remark that this approach seems ad-hoc and it is reasonable to ask for a more conceptual proof. Indeed, as was pointed out to us by an anonymous referee, there is fairly rich literature on "algebraic infinite-dimensional duality" (see [2] for a survey). This theory would certainly not be able to give the results we need. The referee also pointed out to us "topological infinite-dimensional duality" developed in [7] which they suggested might provide a blackbox proof of Theorem 3.16. We were, however, unable to apply this theory in our setting.

Proposition 4.1. *Suppose that $W:\Omega^2\to[0,1]$ is a graphon defined on Ω and F is a graph on the vertex set $[k]$. Suppose that $\mathfrak{c}:\Omega\to[0,1]$ is a fractional F-cover of W and that $\mathfrak{t}:\Omega^k\to[0,1]$ is an F-tiling in W. Then $\|\mathfrak{t}\|\leq\|\mathfrak{c}\|$.*

Proof. We have

$$\begin{aligned}\|\mathfrak{t}\| &= \\ \boxed{\text{by (3.3)}} &= \int_{(x_1,\dots,x_k)\in\mathcal{F}_F(W)} \mathfrak{t}\,(x_1,\dots,x_k)\ \mathsf{d}\nu^k\\ \boxed{\text{Definition 3.12}} &\leq \int_{(x_1,\dots,x_k)\in\mathcal{F}_F(W)} \mathfrak{t}\,(x_1,\dots,x_k)\left(\sum_{i=1}^k \mathfrak{c}\,(x_i)\right)\mathsf{d}\nu^k\\ &= \int_{x\in\Omega}\left(\mathfrak{c}(x)\sum_{i=1}^k\int_{\substack{(x_1,\dots,x_{i-1},x_{i+1},\dots,x_k)\in\Omega^{k-1}\\ (x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_k)\in\mathcal{F}_F(W)}} \mathfrak{t}\,(x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_k)\right)\mathsf{d}\nu\\ \boxed{\text{by (3.4)}} &\leq \int_{x\in\Omega}\mathfrak{c}(x)\,\mathsf{d}\nu\ ,\end{aligned}$$

as was needed. □

Proposition 4.2. *Suppose that $W : \Omega^2 \to [0,1]$ is a graphon defined on Ω and F is a graph on the vertex set $[k]$. Then for an arbitrary $\epsilon > 0$ there exists an F-tiling $\mathfrak{t} : \Omega^k \to [0,1]$ in W with $\|\mathfrak{t}\| \geq \mathsf{fcov}(F,W) - \epsilon$.*

Proof. For each $n = 1, 2, \ldots$, we find a suitable $k_n \in \mathbb{N}$, a partition $\Omega = \Omega_1^{(n)} \sqcup \Omega_2^{(n)} \sqcup \ldots \sqcup \Omega_{k_n}^{(n)}$ into k_n sets of measure $\frac{1}{k_n}$ each, and a graphon $W_n : \Omega^2 \to [0,1]$ such that W_n is constant on each rectangle $\Omega_i^{(n)} \times \Omega_j^{(n)}$ $(i,j \in [k_n])$, and $\|W - W_n\|_1 < 1/n$. This is possible as step functions with square steps are dense in $\mathcal{L}^1(\Omega^2)$. We shall now modify the graphons W_n in two steps.

First, let $W_n' : \Omega^2 \to [0,1]$ be defined as

$$W_n'(x,y) = \begin{cases} W_n(x,y) & \text{if } W_n(x,y) \geq \sqrt{1/n} \\ 0 & \text{otherwise} \end{cases} .$$

This way, we have that

$$\|W_n - W_n'\|_1 < \sqrt{1/n} \,. \tag{4.6}$$

For a fixed n, we say that a rectangle $\Omega_i^{(n)} \times \Omega_j^{(n)}$ is *shoddy* if the measure of the set

$$\left(\operatorname{supp} W_n' \setminus \operatorname{supp} W\right) \cap \left(\Omega_i^{(n)} \times \Omega_j^{(n)}\right)$$

is at least $\sqrt[4]{1/n} \cdot \frac{1}{k_n^2}$. Note that at each point of $\operatorname{supp} W_n' \setminus \operatorname{supp} W$ the difference between W_n and W is at least $\sqrt{1/n}$. This gives that $\nu^2\left(\operatorname{supp} W_n' \setminus \operatorname{supp} W\right) \leq \frac{\|W_n - W\|_1}{\sqrt{1/n}} \leq \sqrt{1/n}$, which in turn implies that at most $\sqrt[4]{1/n} \cdot k_n^2$ rectangles $\Omega_i^{(n)} \times \Omega_j^{(n)}$ are shoddy. Let us define

$$W_n''(x,y) = \begin{cases} 0 & (x,y) \text{ lies in a shoddy rectangle} \\ W_n'(x,y) & \text{otherwise} \end{cases} .$$

Observe that shoddy rectangles are symmetric, so W_n'' is indeed a graphon. We changed W_n'' on a set of measure at most $\sqrt[4]{1/n}$ which gives that

$$\|W_n' - W_n''\|_1 \leq \sqrt[4]{1/n} \,. \tag{4.7}$$

To summarize, we ended up with a graphon W_n'' which is a step-function on the rectangles $\Omega_i^{(n)} \times \Omega_j^{(n)}$ and for which we have

$$\|W_n'' - W\|_1 \leq \|W_n - W\|_1 + \|W_n' - W_n\|_1 + \|W_n'' - W_n'\|_1 \overset{(4.6),(4.7)}{<} \frac{1}{n} + \sqrt{1/n} + \sqrt[4]{1/n} \overset{n\to\infty}{\longrightarrow} 0 \,.$$

Since the $\mathcal{L}^1(\Omega^2)$-norm is stronger than the cut-norm, we have that the graphons (W_n'') converge to W. In particular, Theorem 3.14 tells us that $\liminf_n \mathsf{fcov}(F, W_n'') \geq \mathsf{fcov}(F,W)$. Let N_1 be such that for each $n \geq N_1$ we have $\mathsf{fcov}(F, W_n'') \geq \mathsf{fcov}(F,W) - \epsilon/2$. Let

$N_2 = \lceil k/\epsilon \rceil^8$. Now, fix an arbitrary number $N \geq \max(N_1, N_2)$, and consider the graphon W''_N. This graphon represents a finite graph on the vertex set $V = \{1, \dots, k_N\}$. We call this graph H. We forget the edge-weights, that is we put an edge ij in H whenever W''_N is positive on $\Omega_i^{(N)} \times \Omega_j^{(N)}$. By Proposition 3.13, $\mathsf{fcov}(F, H) = k_N \cdot \mathsf{fcov}(F, W''_N)$.

The LP-duality of finite graphs tells us that we can find a fractional F-tiling $\mathfrak{t}_H : V^k \to [0,1]$ on H with $\sum \mathfrak{t}_H(v_1, \dots, v_k) = \mathsf{fcov}(F, H)$. It is easy to transform $\mathfrak{t}_H$ to a fractional F-tiling $\mathfrak{t} : \Omega^k \to [0, \infty)$ on W''_N by defining $\mathfrak{t}$ to be the constant $k_N \cdot \mathfrak{t}_H(v_1, \dots, v_k)$ on the entire rectangle $\prod_i \Omega_{v_i}^{(N)}$ (for each $v_1, v_2, \dots, v_k \in [k_N]$). Then $\mathfrak{t}$ has weight $\frac{1}{k_N} \sum \mathfrak{t}_H(v_1, \dots, v_k) = \mathsf{fcov}(F, W''_N)$. The function $\mathfrak{t}$ is not necessarily a fractional F-tiling on W, as condition (3.3) may be violated due to values at points $P = \operatorname{supp}\left((W''_N)^{\otimes F}\right) \setminus \operatorname{supp}\left(W^{\otimes F}\right)$. We have

$$P \subset \bigcup_{i<j, ij \in E(F)} \pi_{ij}^{-1}\left(\operatorname{supp} W''_n \setminus \operatorname{supp} W\right) . \tag{4.8}$$

Let $\mathfrak{t}'$ be zero on P, and equal to $\mathfrak{t}$ elsewhere. Then $\mathfrak{t}'$ is a fractional F-tiling on W.

Since $\operatorname{supp} W''_n \setminus \operatorname{supp} W$ is disjoint from points of shoddy rectangles, we have that for each choice of $v_1, \dots, v_k$, the product set $\prod_{i=1}^k \Omega_{v_i}^{(N)}$ satisfies

$$\begin{aligned} \nu^k\left(\prod_{i=1}^k \Omega_{v_i}^{(N)} \cap P\right) &\overset{(4.8)}{\leq} \sum_{\ell<j, ij \in E(F)} \nu^k\left(\prod_{i=1}^k \Omega_{v_i}^{(N)} \cap \pi_{\ell j}^{-1}\left(\operatorname{supp} W''_N \setminus \operatorname{supp} W\right)\right) \\ &\leq \nu^k\left(\prod_{i=1}^k \Omega_{v_i}^{(N)}\right) \binom{k}{2} \sqrt[4]{1/N} . \end{aligned}$$

Using that $\mathfrak{t}$ is constant on $\prod_{i=1}^k \Omega_{v_i}^{(N)}$, we get that

$$\int_{(x_1,\dots,x_k) \in \prod_{i=1}^k \Omega_{v_i}^{(N)}} \mathfrak{t}'(x_1, \dots, x_k) \mathsf{d}\nu^k \geq \left(1 - \tbinom{k}{2} \sqrt[4]{1/N}\right) \int_{(x_1,\dots,x_k) \in \prod_{i=1}^k \Omega_{v_i}} \mathfrak{t}(x_1, \dots, x_k) \mathsf{d}\nu^k .$$

Summing over all product sets $\prod_{i=1}^k \Omega_{v_i}^{(N)}$ we get that $\mathfrak{t}'$ is a fractional F-tiling on W with

$$\begin{aligned} \int_{(x_1,\dots,x_k) \in \Omega^k} \mathfrak{t}'(x_1, \dots, x_k) \, \mathsf{d}\nu^k &\geq \left(1 - \tbinom{k}{2} \sqrt[4]{1/N}\right) \int_{(x_1,\dots,x_k) \in \Omega^k} \mathfrak{t}(x_1, \dots, x_k) \, \mathsf{d}\nu^k \\ &\geq \int_{(x_1,\dots,x_k) \in \Omega^k} \mathfrak{t}(x_1, \dots, x_k) \, \mathsf{d}\nu^k - \frac{\epsilon}{2} \\ &= \mathsf{fcov}(F, W''_N) - \frac{\epsilon}{2} \geq \mathsf{fcov}(F, W) - \epsilon , \end{aligned}$$

as was needed. □

4.3. **Proof of Theorem 3.7.** By Theorem 3.16 it suffices to prove that $\liminf_n \mathsf{fcov}(F, W_n) \geq \mathsf{fcov}(F, W)$. This, however, is the subject of Corollary 3.15.

4.4. **Proof of Theorem 3.4.** For simplicity, let us assume that $v(G_n) = n$. Suppose that $\epsilon > 0$ is arbitrary. We want to show that for n sufficiently large, $\mathsf{til}(F, G_n) \geq (\mathsf{til}(F, W) - \epsilon)n = (\mathsf{fcov}(F, W) - \epsilon)n$ (where the last equality uses Theorem 3.16). Let $\delta > 0$ be such that whenever U is a graphon of cut-distance at most δ from W we have that

$$\mathsf{fcov}(F, U) \geq \mathsf{fcov}(F, W) - \epsilon/3 \,. \tag{4.9}$$

Such a number δ exists by Corollary 3.15. Let $d = \delta/10$. Let ϵ_1 be given by Lemma 2.8 for the error parameter $\gamma = \epsilon/10$ and density $d/2$. Let ϵ_0 be given by Lemma 2.7 for the input parameter $\beta = 0.1$ and error parameter ϵ_1. Set $\epsilon_2 = \min(\epsilon_0, \delta/10)$. Lemma 2.6 with error parameter ϵ_2 gives us a bound L on the complexity of partitions.

Suppose n is large enough, so that we have

$$\mathrm{dist}_\square(W_{G_n}, W) < \delta/2 \,. \tag{4.10}$$

Consider a cluster graph R_n corresponding to a (ϵ_2, d)-regularization of the graph G_n. Let $\mathcal{V} = (V_1, \ldots, V_\ell)$ be the corresponding partition of $V(G_n)$ into clusters, with $\ell \leq L$. Let $m = |V_1|$. By (2.2) we have $\mathrm{dist}_\square(W_{G_n}, W_{R_n}) < \delta/2$. Combining with (4.10), triangle inequality gives $\mathrm{dist}_\square(W, W_{R_n}) < \delta$. In particular, (4.9) can be applied to W_{R_n}. Proposition 3.13 then gives that $\mathsf{fcov}(F, R_n) = \ell \cdot \mathsf{fcov}(F, W_{R_n}) \geq \ell\,(\mathsf{fcov}(F, W) - \epsilon/3)$. We ignore the edge weights in R_n. We apply the (standard) LP duality on R_n. We get a fractional F-tiling $\mathsf{t}_{R_n} : [\ell]^k \to [0, 1]$. We now delete from t_{R_n} very small weights. Formally, for a k-tuple $\mathbf{v} \in [\ell]^k$ we define $\mathsf{t}'_{R_n}(\mathbf{v})$ to be 0 if $\mathsf{t}_{R_n}(\mathbf{v}) < \epsilon/(10\ell^{k-1})$ and to be $\mathsf{t}_{R_n}(\mathbf{v})$ otherwise. Observe that $\|\mathsf{t}'_{R_n}\| \geq \|\mathsf{t}_{R_n}\| - \epsilon\ell/10$. For each copy $\tilde{F}$ of the graph F inside R_n which is in the support of t'_{R_n}, consider an arbitrary integer $m_{\tilde{F}}$ in the range between $\left[(1 - \epsilon/10) \cdot \mathsf{t}'_{R_n}(\tilde{F}) \cdot m, \mathsf{t}'_{R_n}(\tilde{F}) \cdot m\right]$. Since we can assume that n is sufficiently large, for the gap between the endpoints of this range we have

$$\frac{\epsilon}{10} \cdot \mathsf{t}'_{R_n}(\tilde{F}) \cdot m \geq \frac{\epsilon}{10} \cdot \frac{\epsilon}{10\ell^{k-1}} \cdot \frac{(1 - \epsilon_2)n}{\ell} \geq 1 \,.$$

In particular, we can indeed find an integer $m_{\tilde{F}}$ in the specified range.

Now, for each cluster V_i, we construct a partition of V_i into sets $V_{i,0} \sqcup \bigsqcup V_{i,\tilde{F}}$ where $\tilde{F}$ runs over all copies of F in R_n that touch the vertex i. In this partition, the size of each set $V_{i,\tilde{F}}$ is $m_{\tilde{F}}$, and $V_{i,0}$ is the remainder. Since t'_{R_n} is an F-tiling $\sum_{\tilde{F}:\tilde{F}\ni i} m_{\tilde{F}} \leq \sum_{\tilde{F}:\tilde{F}\ni i} \mathsf{t}'_{R_n}(\tilde{F}) \cdot m \leq m$. That means that at least one such partition exists. Among all such partitions $V_{i,0} \sqcup \bigsqcup V_{i,\tilde{F}}$, consider a random one.

Suppose that this way we randomly partitioned all the clusters $V_1, \ldots, V_\ell$. For a given copy $\tilde{F}$ of F in R_n in the support of t'_{R_n}, we set up a random variable $X_{\tilde{F}}$ as follows. If for each edge $ij \in \tilde{F}$ the pair $(V_{i,\tilde{F}}, V_{j,\tilde{F}})$ is ϵ_1-regular, we set $X_{\tilde{F}} = (1 - \epsilon/10)m_{\tilde{F}}$. If there

exists an edge $ij \in \tilde{F}$ for which the pair $(V_{i,\tilde{F}}, V_{j,\tilde{F}})$ fails to be ϵ_1-regular, we set $X_{\tilde{F}} = 0$. Lemma 2.8 gives that one can always find an F-tiling of size $X_{\tilde{F}}$ in $G_n\left[\bigcup_{i\in V(\tilde{F})} V_{i,\tilde{F}}\right]$.

Lemma 2.7 implies that $\mathbf{E}\left[X_{\tilde{F}}\right] \geq \exp\left(-2e(F)0.1^{m_{\tilde{F}}}\right)(1-\epsilon/10)m_{\tilde{F}} \geq (1-\epsilon/5)m_{\tilde{F}}$. By linearity of expectation,

$$\begin{aligned}\mathbf{E}\left[\sum_{\tilde{F}} X_{\tilde{F}}\right] &\geq (1-\epsilon/5)\sum_{\tilde{F}}(1-\epsilon/10)\cdot \mathfrak{t}'_{R_n}(\tilde{F})\cdot m \geq (1-\epsilon/3)\|\mathfrak{t}'_{R_n}\|m \\ &\geq (1-\epsilon/3)\|\mathfrak{t}_{R_n}\|m - \epsilon n/10\ .\end{aligned}$$

So, let us fix one collection of partitions such that $\sum_{\tilde{F}} X_{\tilde{F}} \geq (1-\epsilon/3)\|\mathfrak{t}_{R_n}\|m - \epsilon n/10$. Then by the above there is an F-tiling in G_n of size $\sum_{\tilde{F}} X_{\tilde{F}}$. Using the fact that $\mathsf{fcov}(F, R_n) \geq \ell\left(\mathsf{fcov}(F,W) - \epsilon/3\right)$, we obtain that $\mathsf{til}(F, G_n) \geq (\mathsf{fcov}(F,W) - \epsilon)n$, as needed.

4.5. **Proof of Proposition 3.8.** Write $k = v(F)$. Given $\eta > 0$, let the number β be given by the Removal lemma 2.5 for the graph F and the parameters $\alpha = (\eta/2)^{k+2}$ and $d = \eta/2$. Set $\delta = \beta(\eta/2)^{4k^2}/k^2$.

Suppose now that U is δ-close to W in the cut-norm. Let us fix an optimal fractional F-cover $\mathfrak{c}$ of W (recall Theorem 3.16 and (3.7)). Let $\widehat{U}$ and $\mathfrak{c}'$ be defined by

$$\begin{aligned}\widehat{U}(x,y) &= \begin{cases} U(x,y) & \text{if } U(x,y) > \eta/2 \\ 0 & \text{if } U(x,y) \leq \eta/2 \end{cases}, \\ \mathfrak{c}'(x) &= \max\left(\mathfrak{c}(x) + \eta, 1\right)\ .\end{aligned}$$

Obviously, $\|U - \widehat{U}\|_1 \leq \eta/2$ and $\|\mathfrak{c}'\| \leq \|\mathfrak{c}\| + \eta$. Let $\ell = \lceil \eta^{-1} \rceil$. Let us consider the sets A_i, $i = 1, \ldots, \ell$ defined by $A_i = \mathfrak{c}^{-1}[0, i/\ell)$. Let $X \subset \Omega^k$ be the set of points $(x_1, \ldots, x_k)$ where $\widehat{U}^{\otimes F}(x_1, \ldots, x_k) > 0$ but $\sum_i \mathfrak{c}'(x_i) < 1$. Let $\mathcal{I} \subset \mathbb{N}^k$ be the set of all k-tuples of integers whose sum is ℓ. We have $|\mathcal{I}| \leq \left(\frac{2}{\eta}\right)^k$. Observe that

$$X \subset \bigcup_{\mathbf{i}\in\mathcal{I}} A_{\mathbf{i}_1} \times A_{\mathbf{i}_2} \times \cdots \times A_{\mathbf{i}_k}\ . \tag{4.11}$$

The next claim is crucial.

Claim. For each k-tuple $\mathbf{i} \in \mathcal{I}$, we have for $\mathbf{A} = A_{\mathbf{i}_1} \times A_{\mathbf{i}_2} \times \cdots \times A_{\mathbf{i}_k}$ that $\nu^k(\mathbf{A} \cap X) \leq \beta$.

Proof. Suppose for a contradiction that $\nu^k(\mathbf{A} \cap X) > \beta$. Therefore, we have

$$\int_{(x_1,\ldots,x_k)\in\mathbf{A}} U^{\otimes F}(x_1, \ldots, x_k)\, \mathsf{d}\nu^k \geq \int_{(x_1,\ldots,x_k)\in\mathbf{A}} \widehat{U}^{\otimes F}(x_1, \ldots, x_k)\, \mathsf{d}\nu^k \geq (\eta/2)^{e(F)}\cdot\beta > k^2\delta\ . \tag{4.12}$$

On the other hand, for each k-tuple $\mathbf{x} \in \mathbf{A}$ we have $\sum_j \mathfrak{c}(\mathbf{x}_j) < 1$. Since $\mathfrak{c}$ is a fractional F-cover of W, we conclude that $W^F(\mathbf{x}) = 0$. Therefore,

$$\int_{(x_1,\dots,x_k)\in\mathbf{A}} W^{\otimes F}(x_1,\dots,x_k)\,\mathsf{d}\nu^k = 0\,. \tag{4.13}$$

Combining 4.12 and 4.13 with Lemma 2.2 we get that $\|U - W\|_\square > \delta$, which is a contradiction. □

We shall use tools introduced in Section 2.4. For each $\mathbf{i} \in \mathcal{I}$, let $U_\mathbf{i}$ be the $(A_{\mathbf{i}_1},\dots,A_k)$-partite version of $\widehat{U}$ defined on an auxiliary measure space $\Omega_\mathbf{i} = A_{\mathbf{i}_1} \amalg \dots \amalg A_{\mathbf{i}_k}$ with a measure $\nu_{\Omega_\mathbf{i}}$. The above Claim tells us that $\int_{(x_1,\dots,x_k)\in(\Omega_\mathbf{i})^k} \widehat{U}_\mathbf{i}^{\otimes F}\,\mathsf{d}\,(\nu_{\Omega_\mathbf{i}}) \le \beta$. The Removal Lemma (Lemma 2.5) tells us that there exists a set $S_\mathbf{i} \subset \Omega_\mathbf{i}^2$ of $\Omega_\mathbf{i}^2$-measure at most α such that nullifying $\widehat{U}_\mathbf{i}$ on $S_\mathbf{i}$ yields an F-free graphon. Let $B_\mathbf{i} \subset \Omega^2$ be the folded version of $S_\mathbf{i}$. We have that $\nu^2(B_\mathbf{i}) \le \alpha$ by (2.1). Let us now nullify $\widehat{U}$ on $B = \bigcup_{\mathbf{i}\in\mathcal{I}} B_\mathbf{i}$. The set B has ν^2-measure at most $|\mathcal{I}| \cdot \alpha \le \frac{2}{\eta^k} \cdot \alpha \le \eta/2$, and thus the resulting graphon U^* satisfies $\|U - U^*\|_1 \le \eta$. The nullification together with (4.11) tells us that $\mathfrak{c}'$ is a fractional F-cover of U^*. This finishes the proof.

4.6. **Proof of Theorem 3.10 .** Suppose that a sequence of graphons $(W_n)_n$ converges to a graphon W in the cut-norm. We shall prove the statement in two steps:

$$\limsup_n \mathsf{til}_\epsilon(F, W_n) \;\ge\; \mathsf{til}_\epsilon(F, W)\,, \text{and} \tag{4.14}$$

$$\limsup_n \mathsf{til}_\epsilon(F, W_n) \;\le\; \mathsf{til}_\epsilon(F, W)\,. \tag{4.15}$$

(Note that by passing to a subsequence, we could turn the limit superior into a limit.)

First, let us prove (4.14). For each n, suppose that $U_n \le W_n$ is an arbitrary graphon of $\mathcal{L}^1(\Omega^2)$-norm at most ϵ. Since the space of graphons is sequentially compact, let us consider the limit U of a suitable subsequence $(U_{n_i})_i$. Since $U_{n_i} \le W_{n_i}$ for each i, we also have $U \le W$. Furthermore, since the $\mathcal{L}^1(\Omega^2)$-norm is continuous with respect to the cut-distance, we have that the $\mathcal{L}^1(\Omega^2)$-norm of U is at most ϵ. In particular, U appears in the infimum in (3.6) for the graphon W. We then have

$$\limsup_n \mathsf{til}(F, W_n - U_n) \ge \liminf_i \mathsf{til}(F, W_{n_i} - U_{n_i}) \overset{T3.7}{\ge} \mathsf{til}(F, W - U) \ge \mathsf{til}_\epsilon(F, W)\,,$$

as was needed for (4.14).

For (4.15) we shall need that the function $\mathsf{til}_\delta(F, W)$, considered as a function in $\delta \in (0,1)$ is left-continuous.

Claim. Suppose that $\delta \in (0,1)$. Then in the setting above, for each $\eta > 0$ and each graphon $U \le W$ with $\|U\|_1 = \delta$ there exists a graphon $U' \le U$ with $\|U'\|_1 < \delta$ such that $\mathsf{til}(F, W - U) \le \mathsf{til}(F, W - U') + \eta$.

Proof of Claim. Clearly, there exists a set $A \subset \Omega$ of measure at most η such that $U_{\restriction A\times\Omega}$ is positive on a set of positive measure. Let us nullify U on $(A \times \Omega) \cup (\Omega \times A)$. For the resulting graphon U' we have $\|U'\|_1 < \delta$.

Suppose now that $\mathfrak{t}$ is an arbitrary F-tiling in $W - U'$. By nullifying $\mathfrak{t}$ on those $v(F)$-tuples whose at least one coordinate lies in A, we obtain an F-tiling $\mathfrak{t}^*$ in $W-U$. By (3.4), we have $\|\mathfrak{t}\| \le \|\mathfrak{t}^*\| + \eta$. We conclude that $\mathsf{til}(F, W - U) \le \mathsf{til}(F, W - U') + \eta$. □

By the above claim, it suffices to prove the following weaker form of (4.15):

$$\limsup_n \mathsf{til}_\epsilon(F, W_n) \le \mathsf{til}_{\epsilon-1/\ell}(F, W) + 1/\ell\,, \text{for each } \ell \in \mathbb{N}. \tag{4.16}$$

So, suppose that $\ell \in \mathbb{N}$ is arbitrary. Let $U \le W$ be an arbitrary graphon of $\mathcal{L}^1(\Omega^2)$-norm at most $\epsilon - 1/\ell$. Let $\mathfrak{c} : \Omega \to [0, 1]$ be an arbitrary fractional F-cover of $W - U$ of size $\mathsf{til}(F, W - U)$. Such a cover exists by Theorem 3.16. Let us consider the sets A_i, $i = 1, \dots, \ell - 1$ defined by $A_i = \mathfrak{c}^{-1}[(i-1)/\ell, i/\ell)$ and $A_\ell = \mathfrak{c}^{-1}[(\ell-1)/\ell, 1]$. We therefore have another fractional F-cover $\mathfrak{c}^* = \sum_i \frac{i}{\ell}\mathbf{1}_{A_i}$ of size at most $\|\mathfrak{c}\| + 1/\ell$.

For each n, let us take a graphon $U_n \le W_n$ so that the sequence $(U_n)_n$ converges to U in the cut-norm. Such a sequence $(U_n)_n$ exists by Lemma 2.9. Since the $\mathcal{L}^1(\Omega^2)$-norm of U is at most $\epsilon - 1/\ell$ and since the $\mathcal{L}^1(\Omega^2)$ topology is stronger than the cut-norm topology, we can additionally assume that the $\mathcal{L}^1(\Omega^2)$-norm of each U_n is at most $\epsilon - 1/\ell$. By the same argument as in the proof of Proposition 3.8, for large enough n, the density of copies of F in $W_n - U_n$ not covered by the $\mathfrak{c}^*$ is $o_n(1)$. Using the Removal lemma, for large enough n, we can find graphons $U'_n \le W_n - U_n$ so that U'_n has $\mathcal{L}^1(\Omega^2)$-norm at most $1/2\ell$ and that $W_n - (U_n + U'_n)$ has zero density of copies of F in not covered by the $\mathfrak{c}^*$. Since $U_n + U'_n$ has $\mathcal{L}^1(\Omega^2)$-norm at most ϵ, we get that

$$\mathsf{til}_\epsilon(F, W_n) \le \|\mathfrak{c}^*\| \le \|\mathfrak{c}\| + 1/\ell \le \mathsf{til}(F, W - U) + 1/\ell\,,$$

as was needed for (4.16).

5. Tilings in inhomogeneous random graphs

In this section we give a simple application of our theory. It concerns the random graph model $\mathbb{G}(n, W)$ which was introduced by Lovász and Szegedy in [18]. Let us briefly recall the model. If $W : \Omega^2 \to [0, 1]$ is a graphon, then to sample a graph from the distribution $\mathbb{G}(n, W)$, $G \sim \mathbb{G}(n, W)$, we take deterministically $V(G) = [n]$. Further, we sample points $x_1, \dots, x_n \in \Omega$ independently at random according to the law ν. To define the edges of G, we include each pair ij as an edge in G with probability $W(x_i, x_j)$, independently of the other choices. When W is constant p, $\mathbb{G}(n, W)$ is the usual Erdős–Rényi random graph $\mathbb{G}(n, p)$. See [17, Chapter 10] for more properties of the model $\mathbb{G}(n, W)$.

We prove that the ratios $\frac{\mathsf{til}(F,\mathbb{G}(n,W))}{n}$ and $\frac{\mathsf{ftil}(F,\mathbb{G}(n,W))}{n}$ converge to $\mathsf{til}(F,W)$ asymptotically almost surely. The proof of this statement is a short application of Theorem 3.16.

Theorem 5.1. *Suppose that $W : \Omega^2 \to [0,1]$ is a graphon. Then the values $\frac{\mathsf{til}(F,\mathbb{G}(n,W))}{n}$ and $\frac{\mathsf{ftil}(F,\mathbb{G}(n,W))}{n}$ converge in probability to the constant $\mathsf{til}(F,W)$.*

Proof. It is well-known (see e.g. [17, Lemma 10.16]) that the sequence of graphs $(\mathbb{G}(n,W))_n$ converges to W in the cut-distance almost surely. Thus by Theorem 3.4, $\frac{\mathsf{til}(F,\mathbb{G}(n,W))}{n}$ is asymptotically almost surely at least $\mathsf{til}(F,W) - o(1)$. The analogous statement for $\frac{\mathsf{ftil}(F,\mathbb{G}(n,W))}{n}$ holds, as for any graphs F and G we have $\mathsf{til}(F,G) \leq \mathsf{ftil}\,(F,G)$.

Now, we pick an arbitrary $\ell \in \mathbb{N}$ and we show that asymptotically almost surely, $\frac{\mathsf{ftil}(F,\mathbb{G}(n,W))}{n}$ is at most $\mathsf{til}(F,W) + 3/\ell$. Let us apply Theorem 3.16 and fix a fractional F-cover $\mathfrak{c} : \Omega \to [0,1]$ of size $\mathsf{til}(F,W)$. Let us round $\mathfrak{c}$ up to the closest multiple of $1/2\ell$. This way, the size of the modified fractional F-cover $\mathfrak{c}'$ increased by at most $1/2\ell$. For $i = 0, \ldots, 2\ell$, define Ω_i to be the preimage of $i/2\ell$ under $\mathfrak{c}'$. Since $\mathfrak{c}'$ is a fractional F-cover, we have that for each k-tuple $i_1, i_2, \ldots, i_k$ with $\sum_j i_j < 2\ell$ that

$$\int_{(x_1,\ldots,x_k)\in\prod_{\ell=1}^{k}\Omega_{i_\ell}} W^{\otimes F}(x_1,\ldots,x_k)\,\mathsf{d}\nu^k = 0\,.$$

Since the integrand is non-negative, we get $W^{\otimes F}(\mathbf{x}) = 0$ for almost every k-tuple $\mathbf{x} = (x_1,\ldots,x_k)$ as above. Then, for such a tuple $\mathbf{x}$, there exist two indices $p_{\mathbf{x}}, q_{\mathbf{x}}$, $1 \leq p_{\mathbf{x}} < q_{\mathbf{x}} \leq k$, $p_{\mathbf{x}}q_{\mathbf{x}} \in E(F)$ such that

$$W(x_{p_{\mathbf{x}}}, x_{q_{\mathbf{x}}}) = 0. \tag{5.1}$$

Let us now sample the random graph $G \sim \mathbb{G}(n,W)$. Let $y_1, y_2, \ldots, y_n \in \Omega$ be the random points that represent the n vertices $1, 2, \ldots, n$ of $\mathbb{G}(n,W)$. By the Law of Large Numbers, asymptotically almost surely we have for each $j = 0, 1, \ldots, 2\ell$,

$$|\{y_1,\ldots,y_n\} \cap \Omega_j| \leq (\nu(\Omega_j) + 1/\ell^2)\,n\,. \tag{5.2}$$

Define a function $\mathfrak{b} : V(G) \to [0,1]$ by mapping a vertex i to $\mathfrak{c}'(y_i)$. We then have

$$\begin{aligned} \|\mathfrak{b}\| &= \sum_{i=1}^{n} \mathfrak{c}'(y_i) = \sum_{j=0}^{2\ell} |\{y_1,\ldots,y_n\} \cap \Omega_j| \cdot \frac{j}{2\ell} \\ &\overset{(5.2)}{\leq} \sum_{j=0}^{2\ell} (\nu(\Omega_j) + 1/\ell^2)\,n \cdot \frac{j}{2\ell} \leq (\|\mathfrak{c}'\| + 2/\ell)n \leq (\|\mathfrak{c}\| + 3/\ell)n\,. \end{aligned}$$

We claim that $\mathfrak{b}$ is a fractional F-cover of G with probability 1. Indeed, let $m_1, m_2, \dots, m_k \in V(G)$ be arbitrary with

$$\sum_{j=1}^{k} \mathfrak{b}(m_j) < 1 \ . \tag{5.3}$$

It is our task to show that there exist $pq \in E(F)$ such that $m_p m_q \notin E(G)$. To this end, we observe that (5.3) translates as $y_{m_1} \in \Omega_{i_1}, y_{m_2} \in \Omega_{i_2}, \dots, y_{m_k} \in \Omega_{i_k}$ for some k-tuple $i_1, i_2, \dots, i_k$ with $\sum_j i_j < 2\ell$. Thus, (5.1) applies for some numbers $p, q \in [k]$. But then indeed the edge $m_p m_q$ was included with probability $W(y_{m_p}, y_{m_q}) = 0$ in $\mathbb{G}(n, W)$, as was needed.

Since ℓ was arbitrary, we obtain that $\frac{\mathsf{ftil}(F, \mathbb{G}(n,W))}{n}$ is asymptotically almost surely at most $\mathsf{til}(F, W) + o(1)$. The analogous statement for $\frac{\mathsf{til}(F, \mathbb{G}(n,W))}{n}$ follows from the fact that for any graphs F and G we have $\mathsf{til}(F, G) \geq \mathsf{ftil}\,(F, G)$. □

It is plausible that Theorem 5.1 can be be extended even to sparse inhomogeneous random graphs $\mathbb{G}(n, p_n W)$, where $(p_n)_n$ is a sequence of positive reals tending to zero.

As an immediate corollary of Theorem 5.1 and Theorem 3.4 we obtain the following corollary.

Corollary 5.2. *Suppose that F is an arbitrary graph and W is an arbitrary graphon. Then*

$$\liminf \frac{\mathsf{til}(F, G_n)}{v(G_n)} = \mathsf{til}(F, W) \ ,$$

where the limit inferior ranges over all graph sequences (G_n) that converge to W.

6. A concluding remark: Analytic approach to linear programming in the setting of graphons

Our main results are proven by first approximating a graphon by a step-function and then applying discrete tools, such as linear programming. Arguments that would work directly in the setting of functional analysis would certainly be much shorter and more elegant. As was pointed out by the referee, some such tools are already available in [2] and [7], but we were unable to make use of them in the current context. In a recent paper [14], Hladký and Rocha investigate concepts of the chromatic number and the clique number of graphons. Unlike in finite graphs, the LP duality fails for the fractional versions of these parameters.

7. Acknowledgments

JH would like thank Dan Král and András Máthé for useful discussions that preceded this project. He would also like to thank Martin Doležal for the discussions they have had regarding functional analysis. We thank two anonymous referees for their comments.

Part of this paper was written while JH was participating in the program *Measured group theory* at The Erwin Schrödinger International Institute for Mathematics and Physics.

The contents of this publication reflects only the authors' views and not necessarily the views of the European Commission of the European Union. This publication reflects only its authors' view; the European Research Council Executive Agency is not responsible for any use that may be made of the information it contains.

References

[1] P. Allen, J. Böttcher, J. Hladký, and D. Piguet. A density Corrádi-Hajnal theorem. *Canad. J. Math.*, 67(4):721–758, 2015.

[2] E. J. Anderson and P. Nash. *Linear programming in infinite-dimensional spaces*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Ltd., Chichester, 1987. Theory and applications, A Wiley-Interscience Publication.

[3] B. Bollobás. The independence ratio of regular graphs. *Proc. Amer. Math. Soc.*, 83(2):433–436, 1981.

[4] Ch. Bordanave, M. Lelarge, and J. Salez. Matchings on infinite graphs. *Probab. Theory Related Fields*, 157(1-2):183–208, 2013.

[5] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[6] A. Bruckner, J. Bruckner, and B. Thomson. *Real Analysis*. ClassicalRealAnalysis.com, second edition, 2008.

[7] S. A. Clark. An infinite-dimensional LP duality theorem. *Math. Oper. Res.*, 28(2):233–245, 2003.

[8] M. Doležal and J. Hladký. Matching polytons. *Electron. J. Combin.*, 26(4):Paper No. 4.38, 33, 2019.

[9] M. Doležal, J. Hladký, P. Hu, and D. Piguet. First steps in combinatorial optimization on graphons: Matchings. In *European Conference on Combinatorics, Graph Theory and Applications (EuroComb 2017)*, volume 61 of *Electron. Notes Discrete Math.*, pages 359–365. Elsevier Sci. B. V., Amsterdam, 2017.

[10] G. Elek and G. Lippner. Borel oracles. An analytical approach to constant-time algorithms. *Proc. Amer. Math. Soc.*, 138(8):2939–2947, 2010.

[11] P. Erdős and T. Gallai. On maximal paths and circuits of graphs. *Acta Math. Acad. Sci. Hungar*, 10:337–356 (unbound insert), 1959.

[12] S. Gerke, Y. Kohayakawa, V. Rödl, and A. Steger. Small subsets inherit sparse ϵ-regularity. *J. Combin. Theory Ser. B*, 97(1):34–56, 2007.

[13] J. Hladký, P. Hu, and D. Piguet. Komlós's tiling theorem via graphon covers. *J. Graph Theory*, 90(1):24–45, 2019.

[14] J. Hladký and I. Rocha. Independent sets, cliques, and colorings in graphons. *European J. Combin.*, 88:103108, 18, 2020.

[15] J. Komlós. Tiling Turán theorems. *Combinatorica*, 20(2):203–218, 2000.

[16] L. Lovász. Subgraph densities in signed graphons and the local Simonovits-Sidorenko conjecture. *Electron. J. Combin.*, 18(1):Paper 127, 21, 2011.

[17] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[18] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[19] L. Lovász and B. Szegedy. Testing properties of graphs and functions. *Israel J. Math.*, 178:113–156, 2010.

[20] H. N. Nguyen and K. Onak. Constant-Time Approximation Algorithms via Local Improvements. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 327–336, 2008.

[21] A. A. Razborov. Flag algebras. *J. Symbolic Logic*, 72(4):1239–1282, 2007.

[22] A. A. Razborov. On the minimal density of triangles in graphs. *Combin. Probab. Comput.*, 17(4):603–618, 2008.

[23] T. Tao. *An epsilon of room, I: real analysis*, volume 117 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010. Pages from year three of a mathematical blog.

Institute of Mathematics, Czech Academy of Science, Žitná 25, 110 00, Praha, Czechia.
The Institute of Mathematics of the Czech Academy of Sciences is supported by RVO:67985840.

School of Mathematics, Sun Yat-sen University, Guangzhou, 510275, China. This work has been carried out while at: Department of Computer Science, University of Warwick, Coventry, CV4 7AL, United Kingdom.

Institute of Computer Science, Czech Academy of Sciences, Pod Vodárenskou věží 2, 182 07 Prague, Czech Republic. With institutional support RVO:67985807.

KOMLÓS'S TILING THEOREM VIA GRAPHON COVERS

JAN HLADKÝ, PING HU, AND DIANA PIGUET

Abstract. Komlós [Komlós: Tiling Turán Theorems, Combinatorica, 2000] determined the asymptotically optimal minimum-degree condition for covering a given proportion of vertices of a host graph by vertex-disjoint copies of a fixed graph H, thus essentially extending the Hajnal–Szemerédi theorem which deals with the case when H is a clique. We give a proof of a graphon version of Komlós's theorem. To prove this graphon version, and also to deduce from it the original statement about finite graphs, we use the machinery introduced in [Hladký, Hu, Piguet: Tilings in graphons, arXiv:1606.03113]. We further prove a stability version of Komlós's theorem.

1. Introduction

Questions regarding the number of vertex-disjoint copies of a fixed graph H that can be found in a given graph G are an important part in extremal graph theory. The corresponding quantity, i.e., the maximum number of vertex-disjoint copies of H in G, is denoted $\mathsf{til}(H,G)$, and called the *tiling number of H in G*. The by far most important case is when $H = K_2$ because then $\mathsf{til}(H,G)$ is the matching number of G. For example, a classical theorem of Erdős–Gallai [5] gives an optimal lower bound on the matching ratio of a graph in terms of its edge density.

Recall that the theory of dense graph limits (initiated in [13, 2]) and the related theory of flag algebras (introduced in [16]) have led to breakthroughs on a number of long-standing problems that concern relating subgraph densities. It is natural to attempt to broaden the toolbox available in the graph limits world to be able to address extremal problems that involve other parameters than subgraph densities. In [9] we worked out such a set of tools for working with tiling numbers. In this paper we use this theory to prove a strengthened version of a tiling theorem of Komlós, [10].

1.1. **Komlós's Theorem.** Suppose that H is a fixed graph with chromatic number r. We want to find a minimum degree threshold that guarantees a prescribed lower bound on $\mathsf{til}(H,G)$ for a given (large) n-vertex graph G. Consider first the special case $H = K_r$. Then one end of the range for the problem is covered by Turán's Theorem: if $\delta(G) > {}^{(r-2)n}/_{r-1}$ then $\mathsf{til}(H,G) \geq 1$. The other end is covered by the Hajnal–Szemerédi Theorem, [8]:

Hladký: The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement number 628974.

Hu: This work has received funding from the Leverhulme Trust 2014 Philip Leverhulme Prize of Daniel Kral and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 648509).

Piguet was supported by the Czech Science Foundation, grant number GJ16-07822Y.

if $\delta(G) \geq \lfloor (r-1)n/r \rfloor$ then $\mathsf{til}(H,G) = \lfloor n/r \rfloor$ (which is the maximum possible value for $\mathsf{til}(H,G)$). If $\delta(G) = m < \lfloor (r-1)n/r \rfloor$, then we apply Hajnal-Szemerédi Theorem to the complement of G to get an equitable coloring with $n-m+1$ colors, such that the size of color classes are r or $r-1$. And therefore we get $\mathsf{til}(H,G) = n-(r-1)(n-m+1)$.

When H is a general r-chromatic graph, the asymptotically optimal minimum degree condition $\delta(G) \geq (1+o_n(1))(r-1)n/r$ for the property $\mathsf{til}(H,G) \geq 1$ is given by the Erdős–Stone Theorem (see Section 2.5). Komlós's Theorem then determines the optimal threshold for greater values of $\mathsf{til}(H,G)$. To this end we need to introduce the critical chromatic number.

Definition 1.1. Suppose that H is a graph of order h whose chromatic number is r. We write ℓ for the order of the smallest possible color class in any r-coloring of H. The *critical chromatic number* of H is then defined as

$$\chi_{\mathrm{cr}}(H) = \frac{(r-1)h}{h-\ell}\ . \tag{1.1}$$

Observe that

$$\chi_{\mathrm{cr}}(H) \in (\chi(H)-1, \chi(H)]\ . \tag{1.2}$$

We can now state Komlós's Theorem.

Theorem 1.2 ([10])**.** *Let H be an arbitrary graph, and $x \in [0,1]$. Then for every $\epsilon > 0$ there exists a number n_0 such that the following holds. Suppose that G is a graph of order $n > n_0$ with minimum degree at least*

$$\left(x\left(1-\frac{1}{\chi_{\mathrm{cr}}(H)}\right)+(1-x)\left(1-\frac{1}{\chi(H)-1}\right)\right)n\ . \tag{1.3}$$

Then $\mathsf{til}(H,G) \geq \frac{(x-\epsilon)n}{v(H)}$.

This result is tight (up to the error term $\frac{\epsilon n}{v(H)}$) as shown by an $\chi(H)$-partite n-vertex graph whose $\chi(H)-1$ colour classes are of size $n\cdot(\chi(H)-x(\chi_{\mathrm{cr}}(H)+1-\chi(H)))/\chi(H)(\chi(H)-1)$ each, and the $\chi(H)$-th colour class is of size $n\cdot x(\chi_{\mathrm{cr}}(H)+1-\chi(H))/\chi(H)$.[a] Additional edges can be inserted into the last colour class arbitrarily. Komlós calls these graphs *bottleneck graphs with parameters x and $\chi_{\mathrm{cr}}(H)$*.[b]

Note also that Theorem 1.2 does not cover the case of perfect tilings, i.e., when $\mathsf{til}(H,G) = \left\lfloor \frac{n}{v(H)} \right\rfloor$. Indeed, the answer to this "exact problem" (as opposed to approximate) is more complicated as was shown by Kühn and Osthus [11].

Here, we reprove Komlós's Theorem. Actually, our proof also gives a stability version of Theorem 1.2. This stability version seems to be new.

[a] Again, we neglect rounding issues.

[b] Note that the parameter $\chi(H)$ need not be an input as it can be reconstructed from $\chi_{\mathrm{cr}}(H)$ using (1.2).

Theorem 1.3. *Let H be an arbitrary graph, and $x \in [0,1]$. Then for every $\epsilon > 0$ there exists a number $n_0 \in \mathbb{N}$ such that the following holds. Suppose that G is a graph of order $n > n_0$ with minimum degree at least as in (1.3). Then*

$$\mathsf{til}(H,G) \geq \frac{(x-\epsilon)\,n}{v(H)}\,.$$

Furthermore, if $x \in [0,1)$ then for every $\epsilon > 0$ there exist numbers $n_0 \in \mathbb{N}$ and $\delta > 0$ such that the following holds. Suppose that G is a graph of order $n > n_0$ with minimum degree at least as in (1.3). Then we have

$$\mathsf{til}(H,G) \geq \frac{(x+\delta)\,n}{v(H)}\,,$$

unless G is ϵ-close in the edit distance[c] *to a bottleneck graph with parameters x and $\chi_{\mathrm{cr}}(H)$.*

The original proof of Theorem 1.2 is not lengthy but uses an ingenious recursive regularization of the graph G.[d] Our proof offers an alternative point of view on the problem. In fact we believe it follows the most natural strategy: *If G had only a small tiling number then, by the LP duality,*[e] *it would have a small fractional F-cover. This would lead to a contradiction to the minimum degree assumption.* The actual execution of this proof strategy, using the graphon formalism, is quite technical, in particular in the stability part. Tools that we need to use to this end involve the Banach–Alaoglu Theorem, and arguments about separability of function spaces. While the amount of analytic tools needed may be viewed as a disincentive we actually believe that working out these techniques will be useful in bringing more tools from graph limit theories to extremal combinatorics.

1.2. **Organization of the paper.** In Section 2 we introduce the notation and recall background regarding measure theory, graphons and extremal graph theory. In Section 3 we give a digest of those parts of the theory of tilings in graphons developed in [9] that are needed in the present paper. Thus, any reader familiar with the general theory of graphons should be able to read this paper without having to study [9]. In Section 4 we state the graphon version of Komlós's Theorem, and use it to deduce Theorem 1.3. This graphon version of Komlós's Theorem is then proved in Section 5. Sections 6 and 7 contain some concluding comments.

[c] see Section 2.4 for a definition

[d] See Section 6.

[e] Normally, the LP duality would require the *fractional* version of the tiling number to be considered. However, we are able to overcome this matter.

2. Preliminaries

2.1. **Basic measure theory and weak* convergence.** Throughout, we shall work with an atomless Borel probability space Ω equipped with a measure ν (defined on an implicit σ-algebra).

Given a function f and a number a we define its support $\operatorname{supp} f = \{x : f(x) \neq 0\}$ and its variant $\operatorname{supp}_a f = \{x : f(x) \geq a\}$. Recall that a set is *null* if it has zero measure. "*Almost everywhere*" is a synonym to "up to a null-set". If f is a measurable function, we write $\operatorname{essinf} f := \sup\{a : \{f(x) \leq a\} \text{ is null}\}$ for the *essential infimum* of f and $\operatorname{esssup} f := \inf\{a : \{f(x) \geq a\} \text{ is null}\}$ for the *essential supremum* of f.

The product measure on Ω^k is denoted by ν^k. Recall, that this measure can be constructed by Caratheodory's construction from the k-th power of the σ-algebra underlying Ω. In particular, we have the following basic fact (which we state only for the case $k = 2$, which will be needed later).

Fact 2.1. *Suppose that $P \subset \Omega^2$ is a set of positive measure. Then for every $\epsilon > 0$ there exist sets $X, Y \subset \Omega$ of positive measure so that*

$$\nu^2 (X \times Y \cap P) \geq (1 - \epsilon)\nu(X)\nu(Y) .$$

If Ω is a Borel probability space, then it is a separable measure space. The Banach space $\mathcal{L}^1(\Omega)$ is separable (see e.g. [3, Theorem 13.8]). The dual of $\mathcal{L}^1(\Omega)$ is $\mathcal{L}^\infty(\Omega)$. Recall that a sequence $f_1, f_2, \ldots \in \mathcal{L}^\infty(\Omega)$ *converges weak** to a function $f \in \mathcal{L}^\infty(\Omega)$ if for each $g \in \mathcal{L}^1(\Omega)$ we have that $\int_\Omega f_n g \to \int_\Omega fg$. This convergence notion defines the so-called *weak* topology* on $\mathcal{L}^\infty(\Omega)$. The sequential Banach–Alaoglu Theorem (as stated for example in [18, Theorem 1.9.14]) in this setting reads as follows.

Theorem 2.2. *If Ω is a Borel probability space then each sequence of functions of $\mathcal{L}^\infty(\Omega)$-norm at most 1 contains a weak* convergent subsequence.*

2.2. **Graphons.** Our notation follows mostly [12]. Our graphons will be defined on Ω^2. Recall that Ω is an atomless Borel probability space with probability measure ν.

We refer the reader to [12] to the key notions of *cut-norm* $\|\cdot\|_\square$ and *cut-distance* $\operatorname{dist}_\square(\cdot,\cdot)$. We just emphasize that to derive the latter from the former, one has to involve certain measure-preserving bijections. This step causes that the cut-distance is coarser (in the sense of topologies) than then cut-norm. When we say that a sequence of graphs *converges* to a graphon we refer to the cut-distance.

Suppose that we are given an arbitrary graphon $W : \Omega^2 \to [0, 1]$ and a graph F whose vertex set is $[k]$. We write $W^{\otimes F} : \Omega^k \to [0, 1]$ for a function defined by

$$W^{\otimes F}(x_1, \ldots, x_k) = \prod_{\substack{1 \leq i < j \leq k \\ ij \in E(F)}} W(x_i, x_j) .$$

Last, let us recall the notion of neighborhood and degree in a graphon $W : \Omega^2 \to [0,1]$. If $x_1, \ldots, x_\ell \in \Omega$, then the *common neighborhood* $N(x_1, \ldots, x_\ell)$ is the set $\bigcap_{i=1}^{\ell} (\operatorname{supp} W(x_i, \cdot))$. The *degree* of a vertex $x \in \Omega$ is $\deg_W(x) = \int_{y\in\Omega} W(x,y)$. The *minimum degree of* W is $\delta(W) = \operatorname{essinf} \deg_W(x)$. It is well-known (see for example [16, Theorem 3.15]) that any limit graphon of sequence of graphs with large minimum degrees has a large minimum degree.

Lemma 2.3. *Suppose $\alpha > 0$ and that $G_1, G_2, \ldots$ are finite graphs converging to a graphon W, and that their minimum degrees satisfy $\delta(G_i) \geq \alpha v(G_i)$. Then $\delta(W) \geq \alpha$.* □

2.3. **Independent sets in graphons.** If $W : \Omega^2 \to [0,1]$ is a graphon then we say that a measurable set $A \subset \Omega$ is an *independent set* in W if W is 0 almost everywhere on $A \times A$. The next (standard) lemma asserts that a weak* limit of independent sets is again an independent set.

Lemma 2.4. *Let $W : \Omega^2 \to [0,1]$ be a graphon. Suppose that $(A_n)_{n=1}^{\infty}$ is a sequence of independent sets in W. Suppose that the indicator functions of the sets A_n converge weak* to a function $f : \Omega \to [0,1]$. Then* $\operatorname{supp} f$ *is an independent set in W.*

Proof. It is enough to prove that for each $\epsilon > 0$, the set $P = \operatorname{supp}_\epsilon f$ is independent. There is nothing to prove if P is null, so assume that P has positive measure. Suppose that the statement is false. Then by by Fact 2.1 there exist sets $X, Y \subset P$ of positive measure such that

$$\nu^2\left(X \times Y \cap \left\{(x,y) \in \Omega^2 : W(x,y) = 0\right\}\right) < \frac{\epsilon^2}{5}\nu(X)\nu(Y)\,. \tag{2.1}$$

Recall that $\int_X f \geq \epsilon\nu(X)$ and $\int_Y f \geq \epsilon\nu(Y)$. By weak* convergence, for n sufficiently large, $\nu(X \cap A_n) \geq \frac{\epsilon}{2}\nu(X)$ and $\nu(Y \cap A_n) \geq \frac{\epsilon}{2}\nu(Y)$. Since A_n is an independent set, we have that W is 0 almost everywhere on $(X \cap A_n) \times (Y \cap A_n)$. This contradicts (2.1). □

2.4. **Edit distance.** Given two n-vertex graphs G and H, the *edit distance from G to H* is the number of edges of G that need to be edited (i.e., added or deleted) to get H from G. Here, we minimize over all possible identifications of $V(G)$ and $V(H)$. So, for example if G and H are isomorphic then their edit distance is 0. We say that H is *ϵ-close* to G in the edit distance if its distance from H is at most $\epsilon\binom{n}{2}$.

2.5. **Erdős–Stone–Simonovits Stability Theorem.** Suppose that H is a graph of chromatic number r. The Erdős–Stone–Simonovits Stability Theorem [6, 17] asserts that if G is an H-free graph on n vertices then $e(G) \leq \left(1 - \frac{1}{r-1} + o_n(1)\right)\binom{n}{2}$. This is accompanied by a stability statement: for each $\epsilon > 0$ there exists numbers $\delta > 0$ and n_0 such that if G is an H-free graph on n vertices, $n > n_0$ and $e(G) > \left(1 - \frac{1}{r-1} - \delta\right)\binom{n}{2}$, then G must be ϵ-close to the $(r-1)$-partite Turán graph in the edit distance. We shall need the min-degree version of this (which is actually weaker and easier to prove): if the minimum degree of G

is at least $\left(1-\frac{1}{r-1}-\delta\right)n$ and G is H-free, then G must be ϵ-close to the $(r-1)$-partite Turán graph in the edit distance.

We say that $W:\Omega^2\to[0,1]$ is a $(r-1)$-*partite Turán graphon* if there exists a partition $\Omega=\Omega_1\sqcup\ldots\sqcup\Omega_{r-1}$ into sets of measure $1/r-1$ each, such that $W_{\restriction\Omega_i\times\Omega_j}$ equals 1 almost everywhere for $i\neq j$ and equals 0 almost everywhere for $i=j$. The stability part of the min-degree version of the Erdős–Stone–Simonovits Theorem yields the following:

Theorem 2.5. *Suppose that H is a graph of chromatic number r. If W is a graphon with $\int_{\Omega^{V(H)}}W^{\otimes H}=0$ and minimum degree at least $1-\frac{1}{r-1}$, then W is a $(r-1)$-partite Turán graphon.*

3. Tilings in graphons

In this section, we recall the main concepts and results from [9]. Let us first recall the most important definitions of an *F-tiling* and a *fractional F-cover* in a graphon. The definition of F-tilings in graphons is inspired by the definition of *fractional* F-tilings in finite graphs (we explained in [9, Section 3.2] that there should be no difference between integral and fractional F-tilings in graphons).

Definition 3.1. Suppose that $W:\Omega^2\to[0,1]$ is a graphon, and that F is a graph on the vertex set $[k]$. A function $\mathfrak{t}:\Omega^k\to[0,+\infty)$ is called an *F-tiling* in W if

$$\operatorname{supp}\mathfrak{t}\subset\operatorname{supp}W^{\otimes F}\ ,$$

and we have for each $x\in\Omega$ that

$$\sum_{\ell=1}^{k}\int_{(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k)\in\Omega^{k-1}}\mathfrak{t}(x_1,\ldots,x_{\ell-1},x,x_{\ell+1},\ldots,x_k)\leq 1\ .$$

The *size* of an F-tiling $\mathfrak{t}$ is $\|\mathfrak{t}\|=\int_{\Omega^k}\mathfrak{t}$. The *$F$-tiling number* of W, denoted by $\mathsf{til}(F,W)$, is the supremum of sizes over all F-tilings in W.

For the definition of fractional F-covers in graphons one just rewrites *mutatis mutandis* the usual axioms of fractional F-covers in finite graphs.

Definition 3.2. Suppose that $W:\Omega^2\to[0,1]$ is a graphon, and F is a graph on the vertex set $[k]$. A measurable function $\mathfrak{c}:\Omega\to[0,1]$ is called a *fractional F-cover* in W if

$$\nu^k\left(\left(\operatorname{supp}W^{\otimes F}\right)\cap\left\{(x_1,x_2,\ldots,x_k)\in\Omega^k:\sum_{i=1}^{k}\mathfrak{c}(x_i)<1\right\}\right)=0\ .$$

The *size* of $\mathfrak{c}$, denoted by $\|\mathfrak{c}\|$, is defined by $\|\mathfrak{c}\|=\int_\Omega\mathfrak{c}$. The *fractional F-cover number* $\mathsf{fcov}(F,W)$ of W is the infimum of the sizes of fractional F-covers in W.

Let us note that in [9, (3.7)], we established that

(3.1) the value of $\mathsf{fcov}(F,W)$ is attained by some fractional F-cover.

With these notions at hand, we can state two key results from [9]: the lower-semicontinuity of the F-tiling number, and the graphon LP-duality.

Theorem 3.3 ([9, Theorem 3.4])**.** *Suppose that F is a finite graph and suppose that (G_n) is a sequence of graphs of growing orders converging to a graphon $W:\Omega^2\to[0,1]$ in the cut-distance. Then we have that $\liminf_n \frac{\mathsf{til}(F,G_n)}{v(G_n)} \geq \mathsf{til}(F,W)$.*

Theorem 3.4 ([9, Theorem 3.16])**.** *Suppose that $W:\Omega^2\to[0,1]$ is a graphon and F is an arbitrary finite graph. Then we have $\mathsf{til}(F,W)=\mathsf{fcov}(F,W)$.*

The following useful proposition relates qualitatively the F-tiling number and the F-homomorphism density.

Proposition 3.5. *Suppose that F is a finite graph on a vertex set $[k]$. Then for an arbitrary graphon W we have that $\mathsf{til}(F,W)=0$ if and only if*

$$\int_{\Omega^k} W^{\otimes F} = 0\,. \tag{3.2}$$

Proof. By Theorem 3.4 and (3.1) we know, that $\mathsf{til}(F,W)=0$ if and only if the constant zero function (up to a null set) is a fractional F -cover of W . The latter property is equivalent to (3.2). □

4. Komlós's Theorem

We state our result as a graphon counterpart of Theorem 1.2. First, in analogy to bottleneck graphs we define the class of bottleneck graphons.

Definition 4.1. Suppose that numbers $x\in[0,1)$ *and* $\chi_{\mathrm{cr}}\in(1,+\infty)$ are given. Let us write $r=\lceil\chi_{\mathrm{cr}}\rceil$. We say that a graphon $W:\Omega^2\to[0,1]$ is a *bottleneck graphon with parameters x and χ_{cr}* if there exists a partition $\Omega=\Omega_1\sqcup\Omega_2\sqcup\ldots\sqcup\Omega_r$ such that $\nu(\Omega_r)={}^{x(\chi_{\mathrm{cr}}+1-r)}\!/_{r}$, $\nu(\Omega_1)=\nu(\Omega_2)=\ldots=\nu(\Omega_{r-1})={}^{(r-x(\chi_{\mathrm{cr}}+1-r))}\!/_{r(r-1)}$, and such that

- for each $1\leq i<j\leq r$, W is 1 almost everywhere on $\Omega_i\times\Omega_j$,
- for each $1\leq i\leq r-1$, W is 0 almost everywhere on $\Omega_i\times\Omega_i$.

A set of graphons on a given probability space Ω is called *a graphon class* if with each graphon it contains all graphons isomorphic to it. Given a graphon W and a graphon class $\mathcal{C}$, we define $\mathrm{dist}_\square(W,\mathcal{C})=\inf_{U\in\mathcal{C}}\|W-U\|_\square$. We also define $\mathrm{dist}_1(W,\mathcal{C})=\inf_{U\in\mathcal{C}}\|W-U\|_1$.

For a given $x\in[0,1]$ and $\chi_{\mathrm{cr}}\in(1,\infty)$, we write $\mathcal{C}_{x,\chi_{\mathrm{cr}}}$ for the set of all bottleneck graphons with parameters x and χ_{cr}. This is obviously a graphon class. The next standard

lemma asserts that convergence to $\mathcal{C}_{x,\chi_{\mathrm{cr}}}$in the cut-norm implies convergence in the $\mathcal{L}^1$-norm.

Lemma 4.2. *Suppose that $x \in [0,1]$ and $\chi_{\mathrm{cr}} \in (1,\infty)$. If (W_n) is a sequence of graphons with $\mathrm{dist}_\square(W_n, \mathcal{C}_{x,\chi_{\mathrm{cr}}}) \to 0$ then $\mathrm{dist}_1(W_n, \mathcal{C}_{x,\chi_{\mathrm{cr}}}) \to 0$.*

Proof. Let $B_{x,\chi_{\mathrm{cr}}}$ be (any representative of the isomorphism class of) the bottleneck graphons with parameters x and χ_{cr} in which $B_{x,\chi_{\mathrm{cr}}}$ restricted to $\Omega_r \times \Omega_r$ is zero. The fact that $\mathrm{dist}_\square(W_n, \mathcal{C}_{x,\chi_{\mathrm{cr}}}) \to 0$ allows us to find partitions $\Omega^{(n)} = \Omega_1^{(n)} \sqcup \ldots \sqcup \Omega_r^{(n)}$ where the sets $\Omega_i^{(n)}$ have measures as in Definition 4.1 and approximately satisfy the other properties. Let us modify each graphon W_n by making it zero on $\Omega_r^{(n)} \times \Omega_r^{(n)}$. For the modified graphons W_n', we have $\mathrm{dist}_\square(W_n', B_{x,\chi_{\mathrm{cr}}}) \to 0$. The graphon $B_{x,\chi_{\mathrm{cr}}}$ is 0-1-valued. Thus, [12, Proposition 8.24] tells us that $\mathrm{dist}_1(W_n', B_{x,\chi_{\mathrm{cr}}}) \to 0$. Consequently, $\mathrm{dist}_1(W_n, \mathcal{C}_{x,\chi_{\mathrm{cr}}}) \to 0$. □

Theorem 4.3. *Let H be an arbitrary graph with chromatic number at least two, and $x \in [0,1]$. Suppose that W is a graphon with minimum degree at least*

$$x\left(1 - \frac{1}{\chi_{\mathrm{cr}}(H)}\right) + (1-x)\left(1 - \frac{1}{\chi(H)-1}\right). \tag{4.1}$$

Then $\mathsf{fcov}(H,W) \geq \frac{x}{v(H)}$. Furthermore, if $x < 1$ and $\mathsf{fcov}(H,W) = \frac{x}{v(H)}$ then W is a bottleneck graphon with parameters x and $\chi_{\mathrm{cr}} := \chi_{\mathrm{cr}}(H)$.[f]

The proof of Theorem 4.3 occupies Section 5. Let us now employ the transference results from Section 3 to see that Theorem 4.3 indeed implies Theorem 1.3.

Proof of Theorem 1.3. We first prove the main assertion, and leave the "furthermore" part for later. Suppose that $(G_n)_n$ is a sequence of graphs with

$$\delta(G_n) \geq \left(x\left(1 - \frac{1}{\chi_{\mathrm{cr}}(H)}\right) + (1-x)\left(1 - \frac{1}{\chi(H)-1}\right)\right) v(G_n) \tag{4.2}$$

whose orders tend to infinity for some fixed $x > 0$ and a finite graph H. Let W be a graphon that is an accumulation point of this sequence with respect to the cut-distance. Then the minimum degree of W is at least $x\left(1 - \frac{1}{\chi_{\mathrm{cr}}(H)}\right) + (1-x)\left(1 - \frac{1}{\chi(H)-1}\right)$ by Lemma 2.3. Thus Theorem 4.3 tells us that $\mathsf{fcov}(H,W) \geq \frac{x}{v(H)}$. Then Theorems 3.3 and 3.4 imply that $\liminf_n \frac{\mathsf{til}(H,G_n)}{v(G_n)} \geq \mathsf{til}(H,W) = \mathsf{fcov}(H,W)$, as needed.

Let us now move to the "furthermore" part of the statement. Suppose that $(G_n)_n$ is a sequence of graphs whose orders tend to infinity which satisfies (4.2) for some fixed $x > 0$ and a finite graph H. Suppose that for each $\delta > 0$, when n is sufficiently large, we have that $\mathsf{til}(H,G_n) \leq \frac{x+\delta}{v(H)} \cdot n$. Let us now pass to any limit graphon W. We have $\delta(W) \geq x\left(1 - \frac{1}{\chi_{\mathrm{cr}}(H)}\right) + (1-x)\left(1 - \frac{1}{\chi(H)-1}\right)$ and, by Theorems 3.3 and 3.4, we have

[f] Clearly, there is no uniqueness for $x = 1$.

that $\mathsf{til}(H,W) \le \frac{x}{v(H)}$. Theorem 4.3 tells us that W must be a bottleneck graphon with parameters x and $\chi_{\mathrm{cr}}(H)$. We conclude, that for large enough n, the graph G_n is ϵ-close in the cut-distance to a bottleneck graph with parameters x and $\chi_{\mathrm{cr}}(H)$. Furthermore, by Lemma 4.2, we can actually infer ϵ-closeness in the edit distance, as was needed. □

5. Proof of Theorem 4.3

In Section 5.2 we prove the main part of the statement, and in Section 5.4 we refine our arguments to get the stability asserted in the "furthermore" part. Prior to each of these two section, an overview of the proof is given.

Throughout the section, we shall work with "slices of W", i.e., one-variable functions $W(x,\cdot)$ for some fixed $x \in \Omega$. Recall that measurability of $W(\cdot,\cdot)$ gives that $W(x,\cdot)$ is measurable for almost every $x \in \Omega$. We shall assume that $W(x,\cdot)$ is measurable for every $x \in \Omega$. This is only for the sake of notational simplicity; in the formal proofs we would first take away the exceptional set of x's.

Let us write $\delta = \delta(W)$.

Let us first deal with the case $x = 0$. Then the only non-trivial assertion in Theorem 4.3 is the stability. So, suppose that the conditions of the theorem are fulfilled with $x = 0$, and we have $\mathsf{fcov}(H,W) = 0$. Then Theorem 3.4 and Proposition 3.5 tell us that $\int_{\Omega^{V(H)}} W^{\otimes H} = 0$. Recall that $\delta \ge 1 - \frac{1}{\chi(H)-1}$ by (4.1). The Erdős–Stone–Simonovits Stability Theorem 2.5 tells us that W must be a $\chi(H)$-partite Turán graphon. By Definition 4.1, this is equivalent to being a bottleneck graphon with parameters 0 and $\chi_{\mathrm{cr}}(H)$, which was to be proven.

Thus, throughout the remainder of the proof, we shall assume that x is positive.

5.1. **Overview of the proof of the main part of the statement.** Here, we provide an overview of the proof of the main part of Theorem 4.3. The proof itself, as written in Section 5.2 requires to deal with several technicalities stemming from our infinitesimal approach to the problem (e.g., infima need not be attained). To separate these technicalities from the key ideas, in this overview we shall assume that Ω is a *finite* probability space, $\Omega = \{\omega_1, \dots, \omega_z\}$. (We shall assume that each ω_j has positive measure.) The reader can then view W as a finite cluster graph with "clusters" $\omega_1, \dots, \omega_z$. (The clusters are not required to have the same size.) In this overview, we try to make use of this analogy and explain the ideas behind our proof from the Regularity lemma perspective. We essentially use the same notation as in Section 5.2; the only difference is that our objects are simpler due to the discrete setting. That is, in the actual execution of the proof in Section 5.2, we will have to incorporate small additional error parameters to the setting. We comment on the differences at the end of this overview.

Among all proper colourings of H with $r = \chi(H)$ colours consider one that minimizes the size of the smallest colour class and let $V(H) = V_1 \sqcup V_2 \sqcup \ldots \sqcup V_r$ be the partition of the vertex set into the colour classes of this colouring such that $\ell_1 \ge \ell_2 \ge \ldots \ge \ell_r > 0$, for $\ell_i = |V_i|$. Let $h = \sum_i \ell_i$ be the order of H. Let $\mathfrak{c} : \Omega \to [0,1]$ be an arbitrary fractional

H-cover of W. Notice that Definition 3.2 is consistent with the usual graph-theoretic definition of a fractional cover when the target W is viewed as a finite graph ("cluster graph"). However, we emphasize that this corresponding graph-theoretic definition of a fractional cover is about homomorphisms rather than copies. That is, the requirement is that

$$\sum_{k=1}^{h} \mathfrak{c}(x_k) \geq 1 \,, \tag{5.1}$$

whenever $\{x_v \in \{\omega_1, \dots, \omega_z\}\}_{v \in V(H)}$ is an h-tuple of not necessarily different clusters with the property that $W(x_u, x_v) > 0$ for each $uv \in E(H)$. This definition makes sense even if not all the clusters x_v are distinct as regularity embedding techniques allow us to embed H into the corresponding collection of clusters even in this setting.

We need to show that $\int_\Omega \mathfrak{c} \geq \frac{x}{v(H)}$. To get such a lower-bound, we start focusing on those parts of Ω where the value of $\mathfrak{c}$ is small. More precisely, our idea is to take a cluster $B_1 \in \{\omega_1, \dots, \omega_z\}$ with the smallest value of $\mathfrak{c}$. Then, having defined the clusters $B_1, \dots, B_i$ (for some $i < r$), we take $B_{i+1} \in \{\omega_1, \dots, \omega_z\}$ to be the cluster that has the smallest value of $\mathfrak{c}$ in the common neighborhood of $B_1, \dots, B_i$. Notice that since our minimum-degree is bigger than $1 - \frac{1}{r-1}$, these common neighborhood are indeed nonempty. In particular, the clusters $B_1, B_2, \dots, B_r$ form a copy of K_r. Since by mapping the colour class V_i of H into B_i for each $i \in [r]$ we obtain a graph homomorphism, (5.1) implies that

$$\sum_{i=1}^{r} \ell_i \mathfrak{c}(B_i) \geq 1 \,. \tag{5.2}$$

It can then be calculated that $\int_\Omega \mathfrak{c} \geq \frac{x}{v(H)}$, as was needed.

In the actual proof, the counterparts to common neighborhoods are denoted A_i and the counterparts to the smallest values of $\mathfrak{c}$ are denoted by α_i. The extra difficulty coming from the infinitesimal setting is that

(a) the infimum α_i of $\mathfrak{c}$ on A_i need not be attained, and
(b) there is no notion of a "cluster", neighborhood of which could be taken.

A lower bound that implies that the actual sets A_i are nonempty is given in Claim 5.2. In Claim 5.3 we then show that the actual sets B_i are indeed "pairwise adjacent", thus providing a counterpart to (b). In Claim 5.4 we prove a counterpart of (5.2). These facts can be used to deduce that $\int_\Omega \mathfrak{c} \geq \frac{x}{v(H)}$ in a relatively straightforward way.

5.2. **The main part of the statement.** We start the proof with a simple auxiliary claim.

Claim 5.1. Suppose that $t > 0$, $f \in \mathcal{L}^\infty(\Omega)$, $0 \leq f \leq 1$ is such that

$$\nu\{w \in \Omega : \|W(w, \cdot) - f\|_1 < t\} > 0 \,.$$

Then $\|f\|_1 \geq \delta - t$.

Proof. Recall that for almost every $w \in \Omega$, we have $\|W(w,\cdot)\|_1 \geq \delta$. Let us fix one such w which additionally satisfies $\|W(w,\cdot) - f\|_1 < t$. By the triangle inequality,

$$\|f\|_1 \geq \|W(w,\cdot)\|_1 - \|W(w,\cdot) - f\|_1 \geq \delta - t \,.$$

□

Among all proper colourings of H with $r = \chi(H)$ colours consider one that minimizes the size of the smallest colour class and let $V(H) = V_1 \sqcup V_2 \sqcup \ldots \sqcup V_r$ be the partition of the vertex set into the colour classes of this colouring such that $\ell_1 \geq \ell_2 \geq \ldots \geq \ell_r > 0$, for $\ell_i = |V_i|$. Let $h = \sum_i \ell_i$ be the order of H. Fix an arbitrarily small $\gamma \in (0,1)$.

Let $\mathfrak{c} : \Omega \to [0,1]$ be an arbitrary fractional H-cover of W. It is enough to show that $\int_\Omega \mathfrak{c} \geq \frac{x}{v(H)} - \gamma$. Set

$$\epsilon = \gamma \cdot \left(\frac{\delta - \left(1 - \frac{1}{r-1}\right)}{3r^2} \right)^4 \,. \tag{5.3}$$

The fact that $x > 0$ together with (4.1) tells us that $\delta > 1 - \frac{1}{r-1}$ and $\epsilon > 0$.

Let $A_1 = \Omega$. Sequentially, for $i = 1, \ldots, r$, given sets

$$A_1, \ldots, A_i, B_1, \ldots, B_{i-1}, F_1, \ldots, F_{i-1} \subset \Omega$$

of positive measure and numbers $\alpha_1, \ldots, \alpha_{i-1}$, define number α_i and sets B_i, F_i, A_{i+1} as follows. Set $\alpha_i = \operatorname{essinf} \mathfrak{c}_{\restriction A_i}$, $B_i = \left\{w \in A_i : \mathfrak{c}(w) \leq \alpha_i + \frac{\gamma}{h}\right\}$. It follows that $\nu(B_i) > 0$. By the separability of the space $\mathcal{L}^\infty(\Omega)$ there exists a function $f_i \in \mathcal{L}^\infty(\Omega)$, $0 \leq f_i \leq 1$ such that the set $F_i := \{w \in B_i : \|W(w,\cdot) - f_i(\cdot)\|_1 < \epsilon\}$ has positive measure. Finally, define

$$A_{i+1} := \left\{w \in A_i : \nu\left\{y \in F_i : W(w,y) > 0\right\} \geq \left(1 - \sqrt[4]{\epsilon}\right) \nu(F_i)\right\} \,. \tag{5.4}$$

In order to be able to proceed with the construction for step $i+1$, we need to show that A_{i+1} has positive measure. The following claim gives an optimal quantitative lower-bound.

Claim 5.2. We have $\nu(A_i) \geq \delta - (1 - \nu(A_{i-1})) - 3 \cdot \sqrt[4]{\epsilon} = \nu(A_{i-1}) + \delta - 1 - 3 \cdot \sqrt[4]{\epsilon}$.

Before proving Claim 5.2, we note that as an immediate consequence of Claim 5.2, we have that

$$\nu\left(A_{i+1}\right) \geq 1 - i \cdot (1 - \delta) - 3i \cdot \sqrt[4]{\epsilon} \tag{5.5}$$

for each $i+1 \leq r$. Recall that $\delta > 1 - \frac{1}{r-1}$ by (4.1), then together with (5.3) we know that for $i+1 \leq r$, the set A_{i+1} has positive measure.

Proof of Claim 5.2. We want to prove that A_{i+1} contains almost all of $A_i \cap \left(\operatorname{supp}_{\sqrt[4]{\epsilon}} f_i\right)$. To this end, we consider the quantity

$$\begin{aligned} &\int_{w\in A_i\cap\left(\operatorname{supp}_{\sqrt[4]{\epsilon}} f_i\right)\setminus A_{i+1}} \int_{y\in F_i} |W(w,y)-f_i(w)| \\ &= \int_{y\in F_i} \int_{w\in A_i\cap\left(\operatorname{supp}_{\sqrt[4]{\epsilon}} f_i\right)\setminus A_{i+1}} |W(w,y)-f_i(w)| \ . \end{aligned} \tag{5.6}$$

First, we consider the left-hand side of (5.6). Fix $w \in A_i \cap \left(\operatorname{supp}_{\sqrt[4]{\epsilon}} f_i\right) \setminus A_{i+1}$. Since $w \in \operatorname{supp}_{\sqrt[4]{\epsilon}} f_i$, we have $f_i(w) \geq \sqrt[4]{\epsilon}$. Since $w \notin A_{i+1}$, we have that the sets of $y \in F_i$, for which $W(w,y) = 0$ has measure at least $\sqrt[4]{\epsilon}\nu(F_i)$. Therefore, $\int_{y\in F_i} |W(w,y) - f_i(w)| \geq \sqrt[4]{\epsilon} \cdot \sqrt[4]{\epsilon}\nu(F_i)$. Integrating over w, we get

$$\int_{w\in A_i\cap\left(\operatorname{supp}_{\sqrt[4]{\epsilon}} f_i\right)\setminus A_{i+1}} \int_{y\in F_i} |W(w,y)-f_i(w)| \geq \sqrt{\epsilon}\nu\left(A_i \cap \left(\operatorname{supp}_{\sqrt[4]{\epsilon}} f_i\right) \setminus A_{i+1}\right)\nu(F_i) \ . \tag{5.7}$$

Next, consider the right-hand side of (5.6). Fix $y \in F_i$. Then

$$\int_{w\in A_i\cap\left(\operatorname{supp}_{\sqrt[4]{\epsilon}} f_i\right)\setminus A_{i+1}} |W(w,y)-f_i(w)| \leq \int_{w\in\Omega} |W(w,y)-f_i(w)| = \|W(y,\cdot)-f_i(\cdot)\|_1 \leq \epsilon \ ,$$

where the last inequality uses the definition of F_i. Integrating over y, we get

$$\int_{y\in F_i} \int_{w\in A_i\cap\left(\operatorname{supp}_{\sqrt[4]{\epsilon}} f_i\right)\setminus A_{i+1}} |W(w,y)-f_i(w)| \leq \epsilon\nu(F_i) \ . \tag{5.8}$$

Putting (5.7) and (5.8) together, we get that

$$\nu\left(A_i \cap \left(\operatorname{supp}_{\sqrt[4]{\epsilon}} f_i\right) \setminus A_{i+1}\right) \leq \sqrt{\epsilon} \ .$$

By Claim 5.1 and the definition of f_i, we have $\|f_i\|_1 \geq \delta - \epsilon$, therefore the set $\operatorname{supp}_{\sqrt[4]{\epsilon}} f_i$ has measure at least $\delta - \epsilon - \sqrt[4]{\epsilon} \geq \delta - 2\sqrt[4]{\epsilon}$. Plugging these estimates into

$$\nu(A_{i+1}) \geq \nu(A_i) - \left(1 - \nu\left(\operatorname{supp}_{\sqrt[4]{\epsilon}} f_i\right)\right) - \nu\left(A_i \cap \left(\operatorname{supp}_{\sqrt[4]{\epsilon}} f_i\right) \setminus A_{i+1}\right) \ ,$$

we get the desired result. □

Having defined the sets $A_1, \ldots, A_r, B_1, \ldots, B_r$ and $F_1, \ldots, F_r$, we want to proceed with getting control on the numbers $\alpha_1, \ldots, \alpha_r$. The following claim is crucial to this end.

Claim 5.3. We have that

$$\int_{F_1\times\ldots\times F_r} W^{\otimes K_r} > 0 \ .$$

Proof. Note that

$$\int_{x_r\in F_r}\int_{x_{r-1}\in F_{r-1}}\cdots\int_{x_1\in F_1} W^{\otimes K_r}(x_1,\dots x_r) =$$
$$\int_{x_r\in F_r}\int_{x_{r-1}\in N(x_r)\cap F_{r-1}}\cdots\int_{x_1\in N(x_r,x_{r-1},\dots,x_2)\cap F_1} W^{\otimes K_r}(x_1,\dots x_r)\ .$$

The advantage of rewriting the integral in this way is that the integrand on the right-hand side is positive for every choice of $x_r,\dots,x_1$. So, we only need to show that we are integrating over a set of positive measure. Indeed, suppose that numbers $x_r\in F_r$, $x_{r-1}\in N(x_r)\cap F_{r-1}$, …, $x_{r-i}\in N(x_r,\dots,x_{r-i+1})\cap F_{r-i}$ were given. It is our task to show that the measure of $N(x_r,\dots,x_{r-i})\cap F_{r-i-1}$ is positive. To this end, we use that $x_r,\dots,x_{r-i}\in A_{r-i}$. Then (5.4) tells us that

$$\min\left\{\nu\left(N(x_r)\cap F_{r-i-1}\right)\ ,\ \nu\left(N(x_{r-1})\cap F_{r-i-1}\right)\ ,\ \dots\ ,\ \nu\left(N(x_{r-i})\cap F_{r-i-1}\right)\right\}$$
$$\geq\ (1-\sqrt[4]{\epsilon})\nu(F_{r-i-1})\ .$$

We conclude that

$$\nu(N(x_r,\dots,x_{r-i})\cap F_{r-i-1})\geq(1-(i+1)\sqrt[4]{\epsilon})\nu(F_{r-i-1})>0\ ,$$

as was needed. □

The advertised gain of control on the numbers $\alpha_1,\dots,\alpha_r$ now follows easily.

Claim 5.4. We have

$$\ell_1\alpha_1+\ell_2\alpha_2+\ldots+\ell_r\alpha_r\geq 1-\gamma. \tag{5.9}$$

Proof. Claim 5.3 gives that $\int_{F_1\times\ldots\times F_r} W^{\otimes K_r}>0$. Since H is r-colorable, and since $F_i\subset B_i$, we also have that

$$\int_{(B_1)^{\ell_1}}\int_{(B_2)^{\ell_2}}\cdots\int_{(B_{r-1})^{\ell_{r-1}}}\int_{(B_r)^{\ell_r}} W^{\otimes H}>0\ . \tag{5.10}$$

Recall that for each $w\in B_i$, $\mathfrak{c}(w)\leq\alpha_i+\frac{\gamma}{h}$. Thus, for each $\mathbf{w}\in\prod_j(B_j)^{\ell_j}$, we have

$$\sum_{i=1}^{h}\mathfrak{c}(\mathbf{w}_i)\leq\sum_{j=1}^{r}\left(\alpha_j+\frac{\gamma}{h}\right)\ell_j=\gamma+\sum_{j=1}^{r}\ell_j\alpha_j\ .$$

Combining (5.10) with the fact that $\mathfrak{c}$ is a fractional H-cover, we get (5.9). □

Observe that

$$\begin{aligned}\int_\Omega\mathfrak{c}\ &\geq\ \nu(A_r)\alpha_r+\left(\nu(A_{r-1})-\nu(A_r)\right)\alpha_{r-1}+\ldots+\left(\nu(A_1)-\nu(A_2)\right)\alpha_1\\ &=\ \sum_{i=2}^{r}\nu(A_i)(\alpha_i-\alpha_{i-1})+\alpha_1\ .\end{aligned} \tag{5.11}$$

Using (5.5) and (5.11) we obtain

$$\int_\Omega \mathfrak{c} \geq \sum_{i=2}^{r} \nu(A_i)\,(\alpha_i - \alpha_{i-1}) + \alpha_1 \geq \alpha_1 + \sum_{i=2}^{r} \left(1 - (i-1)(1-\delta) - 3(i-1)\sqrt[4]{\epsilon}\right)(\alpha_i - \alpha_{i-1})\,.$$

Combined with the observation that $\sum_{i=2}^{r}(\alpha_i - \alpha_{i-1}) = \alpha_r - \alpha_1$, we get

$$\begin{aligned}\int_\Omega \mathfrak{c} &\geq \alpha_r + (\delta - 1 - 3\sqrt[4]{\epsilon})\left(\sum_{i=2}^{r}(i-1)\,(\alpha_i - \alpha_{i-1})\right).\\ &= \alpha_r + (\delta - 1 - 3\sqrt[4]{\epsilon})\left((r-1)\alpha_r - \sum_{i=1}^{r-1}\alpha_i\right). \qquad (5.12)\end{aligned}$$

Recall that $\delta = 1 + x\left(\frac{1}{r-1} - \frac{1}{\chi_{\mathrm{cr}}(H)}\right) - \frac{1}{r-1}$. Plugging this equality in (5.12) we obtain

$$\begin{aligned}\int_\Omega \mathfrak{c} \quad &\geq \quad \alpha_r + \left(\frac{x}{r-1} - \frac{x}{\chi_{\mathrm{cr}}} - \frac{1}{r-1} - 3\sqrt[4]{\epsilon}\right)\left((r-1)\alpha_r - \sum_{i=1}^{r-1}\alpha_i\right)\\ &\overset{(5.9)}{\geq} \quad \underbrace{\sum_{i=1}^{r-1}\frac{\alpha_i}{r-1} - 3\sqrt[4]{\epsilon}(r-1)}_{\mathsf{(R1)}}\\ &\quad + \underbrace{\left(\frac{x}{r-1} - \frac{x}{\chi_{\mathrm{cr}}}\right)\left[\frac{r-1}{\ell_r}\left(1 - \sum_{i=1}^{r-1}\ell_i\alpha_i - \gamma\right) - \sum_{i=1}^{r-1}\alpha_i\right]}_{\mathsf{(R2)}}, \qquad (5.13)\end{aligned}$$

where we use the fact $\alpha_r \leq 1$ to get (R1) and use (5.9) to get (R2). Using Definition 1.1, we infer that

$$\frac{x}{r-1} - \frac{x}{\chi_{\mathrm{cr}}} = x\left(\frac{1}{r-1} - \frac{h-\ell_r}{(r-1)h}\right) = \frac{x\ell_r}{(r-1)h}\,. \qquad (5.14)$$

This allows us to express the term (R2) in (5.13) as

$$\mathsf{(R2)} = \frac{x}{h}(1-\gamma) - \frac{x}{(r-1)h}\sum_{i=1}^{r-1}\alpha_i\left((r-1)\ell_i + \ell_r\right). \qquad (5.15)$$

The term (R1) from (5.13) can be decomposed as follows:

$$\mathsf{(R1)} = \frac{x}{(r-1)h}\sum_{i=1}^{r-1}\alpha_i h + \frac{1-x}{r-1}\sum_{i=1}^{r-1}\alpha_i - 3\sqrt[4]{\epsilon}(r-1)\,. \qquad (5.16)$$

Plugging the equalities (5.3), (5.15) and (5.16) in (5.13) and using the fact that $h = \sum_i \ell_i$ we get

$$\begin{aligned}\int_\Omega \mathfrak{c} &= \frac{x}{h}(1-\gamma) + \frac{x}{(r-1)h}\sum_{i=1}^{r-1}\alpha_i\left(h-\ell_r-(r-1)\ell_i\right) + \frac{1-x}{r-1}\sum_{i=1}^{r-1}\alpha_i - \sqrt[4]{\gamma} \\ &= \frac{x}{h}(1-\gamma) + \frac{x}{(r-1)h}\underbrace{\sum_{i=1}^{r-1}\left(\alpha_i\sum_{j=1}^{r-1}(\ell_j-\ell_i)\right)}_{(\mathsf{T1})} + \underbrace{\frac{1-x}{r-1}\sum_{i=1}^{r-1}\alpha_i}_{(\mathsf{T2})} - \sqrt[4]{\gamma}\,. \qquad (5.17)\end{aligned}$$

Let us expand the term (T1).

$$\begin{aligned}\sum_{i=1}^{r-1}\alpha_i\left(\sum_{j=1}^{r-1}(\ell_j-\ell_i)\right) &= \sum_{i=1}^{r-1}\alpha_i\left[\sum_{1\le j<i}(\ell_j-\ell_i) + \sum_{i<j\le r-1}(\ell_j-\ell_i)\right] \\ &= \sum_{i=1}^{r-1}\sum_{j<i}(\ell_j-\ell_i)(\alpha_i-\alpha_j)\,.\end{aligned}$$

Recall that for $j<i$, we have $\ell_j \ge \ell_i$ and $\alpha_j \le \alpha_i$. So, (T1) is non-negative. As $x \le 1$, we have that (T2) is non-negative as well. As $\gamma > 0$ is arbitrarily small, we obtain that $\int_\Omega \mathfrak{c} \ge \frac{x}{h}$ for any fractional H-cover $\mathfrak{c}$.

5.3. **Overview of the proof of the furthermore part of the statement.** Before describing the proof, let us make some observations about the bottleneck graphon (structure of which we want to force). The only fractional H-cover $\mathfrak{c}$ which satisfies $\int_\Omega \mathfrak{c} \le \frac{x}{v(H)}$ is constant 0 almost everywhere on $\Omega_1 \cup \ldots \cup \Omega_{r-1}$ (using notation as described in Definition 4.1) and constant $1/\ell_r$ almost everywhere on Ω_r. Also, in the idealized/discretized setting of Section 5.1, the sets $A_1, A_2, \ldots, A_r$ would start with $A_1 = \Omega$ and then each A_{i+1} would be obtained from A_i by subtracting one set $\Omega_{\pi(i)}$ for one (but arbitrary) permutation $\pi(1), \pi(2), \ldots, \pi(r-1)$ of $1, 2, \ldots, r-1$. In the infinitesimal setting of Section 5.2, we cannot make such a precise statement: Recall that Section 5.2 starts with fixing an error parameter $\gamma > 0$, and then defining objects based on this error parameter. Below, for a given choice of γ, we shall denote these objects with superscript.

So, the goal is clear on an intuitive level: if $\mathfrak{c}$ is a fractional H-cover that satisfies $\int_\Omega \mathfrak{c} = \frac{x}{v(H)}$, we want to describe properties of the "limits sets" $A_i^{(\gamma)}$ as $\gamma \to 0$, and assert that they indeed have the same structure as in the bottleneck graph.

The first step towards this is complementing Claim 5.2. Indeed, in Claim 5.5 below we prove that $\nu(A_j^{(\gamma)} \setminus A_{j+1}^{(\gamma)}) \ge 1-\delta-\phi$, where $\phi \to 0$ as $\gamma \to 0$. Then, in Claim 5.6 we prove that the essential range of $\mathfrak{c}$ is indeed $\{0, 1/\ell_r\}$. Now, we proceed to the key construction of the "limits sets" advertised above. Namely, we define sets O_j to be the supports of weak* accumulation points the indicator functions of the sets $A_j^{(\gamma)} \setminus A_{j+1}^{(\gamma)}$ as $\gamma \to 0$. By the

discussion above, we are hoping that the sets O_j are the individual blocks of a bottleneck graphon. In Claims 5.7, 5.8, 5.9 we prove some basic properties of these sets: namely that $\nu(O_j) \geq 1-\delta$, the sets O_j are disjoint, and that $\mathfrak{c}$ is zero on each O_j. In the remaining claim, the structure of W is completely forced.

5.4. **The furthermore part of the statement.** Suppose that $\mathsf{fcov}(H,W) = \frac{x}{h}$ and let $\mathfrak{c}$ be a fractional H-cover attaining this value (see (3.1)). For any given $\gamma > 0$, we have numbers $\epsilon^{(\gamma)}$, $\alpha_1^{(\gamma)}, \dots, \alpha_r^{(\gamma)}$, sets $A_1^{(\gamma)}, \dots, A_r^{(\gamma)}$, $B_1^{(\gamma)}, \dots, B_r^{(\gamma)}$ and $F_1^{(\gamma)}, \dots, F_r^{(\gamma)}$, and functions $f_1^{(\gamma)}, \dots, f_r^{(\gamma)}$ defined in the previous part (the superscript denotes the dependence on γ).

Since the term $(\mathsf{T1})$ in (5.17) is non-negative, we get from (5.17) that

$$\frac{x}{h} = \int_\Omega \mathfrak{c} \geq \frac{x}{h}(1-\gamma) - \sqrt[4]{\gamma} + \frac{1-x}{r-1}\sum_{i=1}^{r-1} \alpha_i^{(\gamma)} .$$

This implies that

$$\sum_{i=1}^{r-1} \alpha_i^{(\gamma)} \leq \frac{2(r-1)\sqrt[4]{\gamma}}{(1-x)} , \tag{5.18}$$

and consequently

$$\alpha_r^{(\gamma)} \overset{(5.9)}{\geq} \frac{1-\gamma-\frac{2h(r-1)\sqrt[4]{\gamma}}{(1-x)}}{\ell_r} . \tag{5.19}$$

Claim 5.5. For any $\gamma > 0$ and any $j \in [r-1]$, we have $\nu(A_j^{(\gamma)} \setminus A_{j+1}^{(\gamma)}) \geq 1-\delta-\phi$, where $\phi = \frac{16hr\sqrt[4]{\gamma}}{1-x}$.

Proof. Let us first show that

$$\nu\left(A_{j+1}^{(\gamma)}\right) \leq 1 - j(1-\delta) + \frac{\phi}{2} . \tag{5.20}$$

Indeed, suppose not. Then applying Claim 5.2 repeatedly for $i = j+2, \dots, r-1$, we get that

$$\nu\left(A_r^{(\gamma)}\right) \geq 1-(r-1)(1-\delta) + \frac{\phi}{4} \overset{(5.14)}{\geq} \frac{x\ell_r}{h} + \frac{\phi}{4} .$$

We then have

$$\int_\Omega \mathfrak{c} \geq \alpha_r^{(\gamma)} \cdot \nu\left(A_r^{(\gamma)}\right) \overset{(5.19)}{\geq} \frac{x}{h} + \frac{\phi}{4\ell_r} - \frac{4r\sqrt[4]{\gamma}}{1-x} > \frac{x}{h} ,$$

which is a contradiction to the choice of $\mathfrak{c}$. This establishes (5.20).

We have $\nu(A_j^{(\gamma)} \setminus A_{j+1}^{(\gamma)}) = \nu(A_j^{(\gamma)}) - \nu(A_{j+1}^{(\gamma)})$. The measure of the former set is bounded from below by $1-(j-1)(1-\delta)-3(j-1)\cdot\sqrt[4]{\epsilon}$ by (5.5), and the measure of the latter set is bounded from above by $1-j(1-\delta)+\frac{\phi}{2}$ by (5.20). The claim follows. □

Claim 5.6. The essential range of $\mathfrak{c}$ is $\{0, 1/\ell_r\}$.

Proof. First assume that for some $\phi > 0$ there is a set S of measure at least ϕ such that $\mathfrak{c}(S) \subseteq (\phi, \frac{1}{\ell_r} - \phi)$. Fix $\gamma = \left(\frac{(1-x)\phi^2}{2(r+1)}\right)^4$. Then $\alpha_r^{(\gamma)} > \frac{1}{\ell_r} - \phi$ by (5.19). In particular, S is disjoint from $A_r^{(\gamma)}$. We get

$$\int_\Omega \mathfrak{c} \geq \nu(S)\,\phi + \nu\left(A_r^{(\gamma)}\right)\alpha_r^{(\gamma)} \geq \phi^2 + \left(\frac{x}{h}\cdot\ell_r - \sqrt[4]{\gamma}\right)\frac{1-\gamma-\frac{2h(r-1)\sqrt[4]{\gamma}}{(1-x)}}{\ell_r} > \frac{x}{h}\,,$$

a contradiction. Now assume that for some $\phi > 0$ there is a set S of measure at least ϕ such that $\mathfrak{c}(S) \subseteq (\frac{1}{\ell_r} + \phi, 1]$. Fix $\gamma = \left(\frac{(1-x)\phi}{4hr}\right)^4$. Then

$$\int_\Omega \mathfrak{c} \geq \nu\left(A_r^{(\gamma)} \setminus S\right)\alpha_r^{(\gamma)} + \nu(S)\left(\frac{1}{\ell_r} + \phi\right) > \frac{x}{h}\,,$$

again a contradiction, proving the claim. □

Let $\left(\gamma_n^{(r)}\right)_{n=1}^\infty$ be a sequence of numbers, with $\gamma_n^{(r)} \overset{n\to\infty}{\longrightarrow} 0$. Now, for a fixed $i = r-1, r-2, \ldots, 1$, we inductively derive $\left(\gamma_n^{(i)}\right)_{n=1}^\infty$ from $\left(\gamma_n^{(i+1)}\right)_{n=1}^\infty$ in the following way. Consider the sequence of sets

$$\left(A_i^{\left(\gamma_n^{(i+1)}\right)} \setminus A_{i+i}^{\left(\gamma_n^{(i+1)}\right)}\right)_{n=1}^\infty$$

viewed as indicator functions. These functions have an accumulation point $\chi_i : \Omega \to [0,1]$ in the weak* topology by by Theorem 2.2. Let $O_i = \operatorname{supp}\chi_i$. Let $\left(\gamma_n^{(i)}\right)_{n=1}^\infty \subset \left(\gamma_n^{(i+1)}\right)_{n=1}^\infty$ be a subsequence along which these indicator functions converge to χ_i. Since O_i arises from the weak* limit of the sets $A_i^{\left(\gamma_n^{(i)}\right)} \setminus A_{i+1}^{\left(\gamma_n^{(i)}\right)}$, we have that

$$\nu\left(O_i \setminus A_i^{\left(\gamma_n^{(i)}\right)}\right) = o_n(1)\,. \tag{5.21}$$

Claim 5.7. We have $\nu(O_i) \geq 1 - \delta$.

Proof of Claim 5.7. By Claim 5.5, we have that $\nu\left(A_i^{\left(\gamma_n^{(i)}\right)} \setminus A_{i+1}^{\left(\gamma_n^{(i)}\right)}\right) \geq 1 - \delta - o_n(1)$. Since χ_i is the weak* limit of the indicator functions of the sets $A_i^{\left(\gamma_n^{(i)}\right)} \setminus A_{i+1}^{\left(\gamma_n^{(i)}\right)}$, we have that

$$\int_\Omega \chi_i \geq 1 - \delta\,. \tag{5.22}$$

Since esssup $\chi_i \leq 1$, we get that $\nu(O_i) \geq 1 - \delta$. □

Claim 5.8. The sets O_1, O_2, …, O_{r-1} are pairwise disjoint.

Proof of Claim 5.8. Let $i \in [r-2]$ be arbitrary. We want to show that the set O_i is disjoint from $O_{i+1} \cup O_{i+2} \cup \ldots \cup O_{r-1}$. We have that

$$\left(A_i^{\left(\gamma_n^{(i)}\right)} \setminus A_{i+1}^{\left(\gamma_n^{(i)}\right)}\right) \cap (O_{i+1} \cup O_{i+2} \cup \ldots \cup O_{r-1}) \subset (O_{i+1} \cup O_{i+2} \cup \ldots \cup O_{r-1}) \setminus A_{i+i}^{\left(\gamma_n^{(i)}\right)} .$$

Recall that the support of the weak* limit of the indicator functions of the sets $A_{i+1}^{\left(\gamma_n^{(i)}\right)}$ contains the set $O_{i+1} \cup O_{i+2} \cup \ldots \cup O_{r-1}$. This proves the claim. □

Claim 5.9. The function $\mathfrak{c}_{\restriction O_i}$ is zero almost everywhere.

Proof of Claim 5.9. Suppose that this is not the case, i.e., $\mathfrak{c}$ is at least some $\theta > 0$ on a subset $P \subset O_i$ of measure θ. Recall that O_i arises as the weak* limit of the sets $A_i^{\left(\gamma_n^{(i+1)}\right)} \setminus A_{i+1}^{\left(\gamma_n^{(i+1)}\right)}$. Therefore, for each n sufficiently large, $\mathfrak{c}$ is at least θ on a subset $P' \subset O_i \cap \left(A_i^{\left(\gamma_n^{(i+1)}\right)} \setminus A_{i+1}^{\left(\gamma_n^{(i+1)}\right)}\right)$ of measure $\theta/2$. By Claim 5.6, $\mathfrak{c}_{\restriction P'} = 1/\ell_r$. Also, combining Claim 5.6 and (5.19) we get that

$$\mathfrak{c}_{\restriction A_r^{\left(\gamma_n^{(i+1)}\right)}} = 1/\ell_r \, .$$

Assume further that n is such that $\gamma_n^{(i+1)} < \left(\frac{r^2\theta}{2\ell_r}\right)^4$. Then

$$\begin{aligned}
\int_\Omega \mathfrak{c} &\geq \nu\left(P' \sqcup A_r^{\left(\gamma_n^{(i+1)}\right)}\right) \cdot \frac{1}{\ell_r} \\
\boxed{\text{by (5.3) and (5.5)}} \quad &\geq \left(\frac{\theta}{2} + 1 - (r-1)\cdot(1-\delta) - 3(r-1)\cdot \sqrt[4]{\gamma_n^{(i+1)}} \cdot \frac{\delta - (1-\frac{1}{r-1})}{3r^2}\right) \cdot \frac{1}{\ell_r} \\
\boxed{\text{by (4.1)}} \quad &= \left(\frac{\theta}{2} + \frac{x\ell_r}{h} - \sqrt[4]{\gamma_n^{(i+1)}} \cdot \frac{x\ell_r}{r^2}\right) \cdot \frac{1}{\ell_r} > \frac{x}{h} \, ,
\end{aligned}$$

which is a contradiction to the fact that $\int_\Omega \mathfrak{c} = \frac{x}{h}$. □

We can now proceed with the inductive step for $i-1$ in the same manner.

Having defined the functions χ_i, the sets O_i and the sequences $\left(\gamma_n^{(i)}\right)_{n=1}^\infty$ for $i = r-1, \ldots, 1$, we now derive some further properties of these.

Claim 5.10. For $\ell = r-1, r-2, \ldots, 1$ and each j, $\ell < j \leq r-1$, if $F_\ell^{\left(\gamma_n^{(j)}\right)} \cap O_j$ is not null then $\nu\left(O_j \setminus F_\ell^{\left(\gamma_n^{(j)}\right)}\right) = o_n(1)$.

Claim 5.11. For $\ell = r-1, r-2, \ldots, 1$ and each j, $\ell < j \leq r-1$, and each $n \in \mathbb{N}$ sufficiently large, we have that $F_\ell^{\left(\gamma_n^{(j)}\right)} \cap O_j$ is a null-set.

Claim 5.12. For $\ell = r-1, r-2, \dots, 1$ and for each sufficiently large $n \in \mathbb{N}$ the set

$$\left(A_\ell^{\left(\gamma_n^{(\ell)}\right)} \setminus A_{\ell+1}^{\left(\gamma_n^{(\ell)}\right)} \right) \setminus (O_{\ell+1} \cup O_{\ell+2} \cup \ldots \cup O_{r-1} \cup \operatorname{supp} \mathfrak{c})$$

is independent in W.

Claim 5.13. For $\ell = r-1, r-2, \dots, 1$ the set O_ℓ is independent in W.

Claim 5.14. For $\ell = r-1, r-2, \dots, 1$ we have that χ_ℓ is constant 1 almost everywhere on O_ℓ and constant 0 almost everywhere on $\Omega \setminus O_\ell$.

Claim 5.15. For $\ell = r-1, r-2, \dots, 1$, W is 1 almost everywhere on $O_\ell \times (\Omega \setminus O_\ell)$.

We shall now prove Claims 5.10–5.15 by induction. That is, first we prove Claim 5.10, Claim 5.11, Claim 5.12, Claim 5.13, Claim 5.15 (in this order) for $\ell = r-1$, and then continue proving the same batch of claims for $\ell = r-2, \dots, 1$. Note that Claims 5.10 and 5.11 are vacuous for $\ell = r-1$.

Proof of Claim 5.10. Suppose that $F_\ell^{\left(\gamma_n^{(j)}\right)} \cap O_j$ is not null. Claim 5.13 and 5.15 (applied to $\ell_{\text{Cl5.13}} = \ell_{\text{Cl5.15}} = j$) assert that the one-variable functions $W(w, \cdot)$ are the same for almost all $w \in O_j$. Consequently,

$$\left\| W(w,\cdot) - f_\ell^{\left(\gamma_n^{(j)}\right)}(\cdot) \right\|_1 < \epsilon^{\left(\gamma_n^{(j)}\right)} , \tag{5.23}$$

for almost all $w \in O_j$.

Combining (5.21) with $A_j^{\left(\gamma_n^{(j)}\right)} \subset A_\ell^{\left(\gamma_n^{(j)}\right)}$, we get

$$\nu\left(O_j \setminus A_\ell^{\left(\gamma_n^{(j)}\right)} \right) = o_n(1) . \tag{5.24}$$

By Claim 5.9, $\mathfrak{c}$ is zero almost everywhere on O_j. Therefore, (5.24) can be rewritten as $\nu\left(O_j \setminus B_\ell^{\left(\gamma_n^{(j)}\right)} \right) = o_n(1)$. The claim follows by plugging (5.23) into the definition of $F_\ell^{\left(\gamma_n^{(j)}\right)}$. □

Proof of Claim 5.11. Suppose that the statement of the claim does not hold. Then there exists an infinite sequence of numbers n for which $F_\ell^{\left(\gamma_n^{(j)}\right)} \cap O_j$ is not null. Let n be such that $F_\ell^{\left(\gamma_n^{(j)}\right)} \cap O_j$ is not null, and suppose that it is sufficiently large. We then have that

$$\nu\left(O_j \cap F_\ell^{\left(\gamma_n^{(j)}\right)} \cap A_j^{\left(\gamma_n^{(j)}\right)} \right) \geq \nu(O_j) - \nu\left(O_j \setminus F_\ell^{\left(\gamma_n^{(j)}\right)} \right) - \nu\left(O_j \setminus A_j^{\left(\gamma_n^{(j)}\right)} \right) .$$

The first term is at least $1-\delta$ by Claim 5.7. The second term is $o_n(1)$ by Claim 5.10. The third term is $o_n(1)$ by (5.21). We conclude that

$$\nu\left(O_j \cap F_\ell^{\left(\gamma_n^{(j)}\right)} \cap A_j^{\left(\gamma_n^{(j)}\right)}\right) > \frac{1}{2}(1-\delta)\ . \tag{5.25}$$

Consider an arbitrary $w \in O_j \cap F_\ell^{\left(\gamma_n^{(j)}\right)} \cap A_j^{\left(\gamma_n^{(j)}\right)}$. As $w \in A_j^{(\gamma_n^{(j)})}$, the definition from (5.4) gives,

$$\nu\left(N(w) \cap F_\ell^{\left(\gamma_n^{(j)}\right)}\right) \geq \left(1 - \sqrt[4]{\epsilon^{\left(\gamma_n^{(j)}\right)}}\right) \nu\left(F_\ell^{\left(\gamma_n^{(j)}\right)}\right)\ .$$

In particular,

$$\nu\left(N(w) \cap O_j \cap F_\ell^{\left(\gamma_n^{(j)}\right)}\right) \geq \nu\left(O_j \cap F_\ell^{\left(\gamma_n^{(j)}\right)}\right) - \sqrt[4]{\epsilon^{\left(\gamma_n^{(j)}\right)}}\nu\left(F_\ell^{\left(\gamma_n^{(j)}\right)}\right) \overset{(5.25)}{\geq} \frac{1}{4}(1-\delta)\ .$$

Integrating w over the set $O_j \cap F_\ell^{\left(\gamma_n^{(j)}\right)} \cap A_j^{\left(\gamma_n^{(j)}\right)}$ of positive measure (by (5.25)), and get that

$$\int_{w \in O_j \cap F_\ell^{\left(\gamma_n^{(j)}\right)} \cap A_j^{\left(\gamma_n^{(j)}\right)}} \int_{y \in O_j \cap F_\ell^{\left(\gamma_n^{(j)}\right)}} W(w,y) > 0\ .$$

Hence $O_j \cap F_\ell^{\left(\gamma_n^{(j)}\right)}$ is not an independent set, a contradiction to Claim 5.13. □

Proof of Claim 5.12. Suppose that the statement of the claim fails for ℓ. Then, we can find two sets $P, Q \subset \left(A_\ell^{\left(\gamma_n^{(\ell)}\right)} \setminus A_{\ell+1}^{\left(\gamma_n^{(\ell)}\right)}\right) \setminus (O_{\ell+1} \cup O_{\ell+2} \cup \ldots \cup O_{r-1} \cup \operatorname{supp} \mathfrak{c})$ such that $\int_{P\times Q} W > 0$.

Consider an r-tuple $\mathbf{w} \in F_1^{\left(\gamma_n^{(\ell)}\right)} \times F_2^{\left(\gamma_n^{(\ell)}\right)} \times \ldots \times F_{\ell-1}^{\left(\gamma_n^{(\ell)}\right)} \times P \times Q \times O_{\ell+1} \times \ldots \times O_{r-1}$. For $j = 1, 2, \ldots, \ell-1$, $\mathbf{w}_j \in F_j^{\left(\gamma_n^{(\ell)}\right)} \subset B_j^{\left(\gamma_n^{(\ell)}\right)} \subset \mathfrak{c}^{-1}(0)$, where the last inclusion uses (in addition to the definition of the set $B_j^{\left(\gamma_n^{(\ell)}\right)}$) Claim 5.6. For $j = \ell, \ell+1$, we have $\mathfrak{c}(\mathbf{w}_j) = 0$ since P and Q are disjoint from $\operatorname{supp} \mathfrak{c}$. For $j = \ell+2, \ldots, r$, we have $\mathfrak{c}(\mathbf{w}_j) = 0$ by Claim 5.9, except possibly a null set of exceptional values of $\mathbf{w}_j$. We conclude that $\sum_j \mathfrak{c}(\mathbf{w}_j) = 0$, except possibly a null set exceptional vectors $\mathbf{w}$. In particular, for almost every $\mathbf{w} \in F_1^{\left(\gamma_n^{(\ell)}\right)} \times F_2^{\left(\gamma_n^{(\ell)}\right)} \times \ldots \times F_{\ell-1}^{\left(\gamma_n^{(\ell)}\right)} \times P \times Q \times O_{\ell+1} \times \ldots \times O_{r-1}$,

$$v(H) \cdot \sum_j \mathfrak{c}(\mathbf{w}_j) = 0\ . \tag{5.26}$$

As the chromatic number of H is r and each color-class of H has size at most $v(H)$, we get that the function $v(H) \cdot \mathfrak{c}$ is a fractional K_r-cover. Combined with (5.26), we get that

$W^{\otimes K_r}(\mathbf{w}) = 0$ (for almost every $\mathbf{w}$). Therefore,

$$\int_{F_1^{\left(\gamma_n^{(\ell)}\right)}} \int_{F_2^{\left(\gamma_n^{(\ell)}\right)}} \dots \int_{F_{\ell-1}^{\left(\gamma_n^{(\ell)}\right)}} \int_P \int_Q \int_{O_{\ell+1}} \dots \int_{O_{r-1}} W^{\otimes K_r} = 0 \,. \tag{5.27}$$

We abbreviate $\mathcal{O} = O_{\ell+1} \cup \ldots \cup O_{r-1}$. Let us now take an arbitrary $w \in A_\ell^{\left(\gamma_n^{(\ell)}\right)}$. Recall that $A_\ell^{\left(\gamma_n^{(\ell)}\right)} \subset A_{\ell-1}^{\left(\gamma_n^{(\ell)}\right)} \subset \ldots \subset A_2^{\left(\gamma_n^{(\ell)}\right)}$. Therefore, (5.4) tells us that

$$\nu\left(F_j^{\left(\gamma_n^{(\ell)}\right)} \cap N(w)\right) \geq \left(1 - \sqrt[4]{\epsilon^{\left(\gamma_n^{(\ell)}\right)}}\right) \nu\left(F_j^{\left(\gamma_n^{(\ell)}\right)}\right)$$

for each $j = \ell-1, \ell-2, \dots, 1$. Similarly, given an arbitrary $u_t \in F_t^{\left(\gamma_n^{(\ell)}\right)}$ $(t = 2, \dots, \ell-1)$, we make use of the fact that $F_t^{\left(\gamma_n^{(\ell)}\right)} \subset A_t^{\left(\gamma_n^{(\ell)}\right)}$ and deduce that

$$\nu\left(F_j^{\left(\gamma_n^{(\ell)}\right)} \cap N(u_t)\right) \geq \left(1 - \sqrt[4]{\epsilon^{\left(\gamma_n^{(\ell)}\right)}}\right) \nu\left(F_j^{\left(\gamma_n^{(\ell)}\right)}\right)$$

for each $j = t-1, \ell-2, \dots, 1$. Claim 5.11 tells us that

$$\nu\left(F_j^{\left(\gamma_n^{(\ell)}\right)} \cap N(u_t) \setminus \mathcal{O}\right) \geq \left(1 - \sqrt[4]{\epsilon^{\left(\gamma_n^{(\ell)}\right)}}\right) \nu\left(F_j^{\left(\gamma_n^{(\ell)}\right)}\right) > (1 - \frac{1}{2r}) \nu\left(F_j^{\left(\gamma_n^{(\ell)}\right)}\right) \,.$$

That is, starting from any $w \in A_\ell^{\left(\gamma_n^{(\ell)}\right)}$, we can plant a positive ν^ℓ-measure of K_ℓ-cliques $w u_{\ell-1} u_{\ell-2} \dots u_1$ as above. The situation is illustrated on Figure 5.1. We can refine this construction to find a positive ν^r-measure of K_r-cliques as follows. First we take $w_P \in P$ and $w_Q \in Q$ such that $W(w_P, w_Q) > 0$ (we have a ν^2-positive measure of such choices). Then we sequentially find vertices

$$u_{\ell-1} \in F_{\ell-1}^{\left(\gamma_n^{(\ell)}\right)} \setminus \mathcal{O}, \dots, u_1 \in F_1^{\left(\gamma_n^{(\ell)}\right)} \setminus \mathcal{O}$$

that are neighbors of w_P, w_Q and the vertices fixed in the previous rounds. Having chosen the $K_{\ell+1}$-clique $w_P w_Q u_{\ell-1} u_{\ell-2} \dots u_1$, Claim 5.8 tells us that $O_{\ell+1}, O_{\ell+2}, \dots, O_{r-1}$ are disjoint, then together with Claim 5.15 we know that padding arbitrary elements from $O_{\ell+1}, O_{\ell+2}, \dots, O_{r-1}$ yields a copy of K_r. Since all these sets have positive measure, we get a contradiction to (5.27). □

Proof of Claim 5.13. Recall that O_ℓ arises from the weak* limit of the sets $A_\ell^{\left(\gamma_n^{(\ell)}\right)} \setminus A_{\ell+1}^{\left(\gamma_n^{(\ell)}\right)}$. Claims 5.8 and 5.9 tell us that O_ℓ can also be seen as the weak* limit of the sets

$$\left(A_\ell^{\left(\gamma_n^{(\ell)}\right)} \setminus A_{\ell+1}^{\left(\gamma_n^{(\ell)}\right)}\right) \setminus (O_{\ell+1} \cup O_{\ell+2} \cup \ldots \cup O_{r-1} \cup \operatorname{supp} \mathfrak{c}) \,.$$

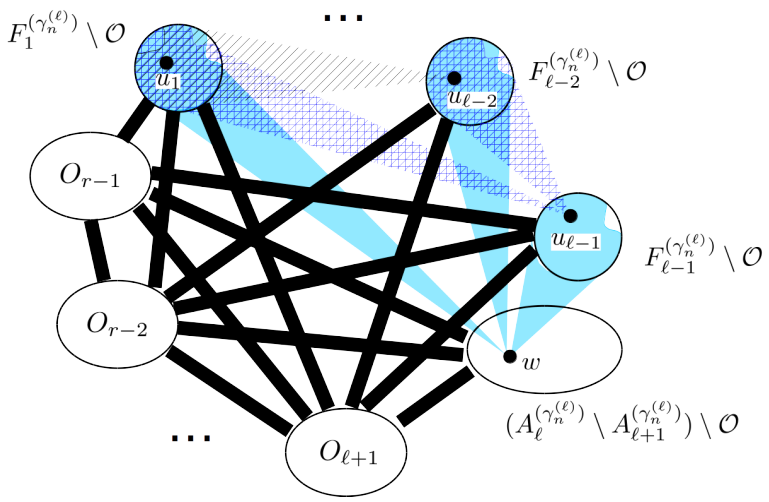


FIGURE 5.1. The black complete bipartite graphs are forced by Claim 5.15. The almost complete connections depicted with colours follow from the fact that the respective vertices lie in the sets $A_j^{\left(\gamma_n^{(\ell)}\right)}$ $(j = \ell, \ell-1, \dots, 2)$, and thus are well-connected to the sets $F_t^{\left(\gamma_n^{(\ell)}\right)}$ (for each $t \in [j-1]$).

Thus the claim follows by combining Claim 5.12 and Lemma 2.4. □

Proof of Claim 5.14. The fact that $(\chi_\ell)_{\Omega \setminus O_\ell} = 0$ follows simply because O_ℓ is the indicator of supp χ_ℓ. Suppose now for contradiction that $(\chi_\ell)_{O_\ell}$ is less than 1 on a set of positive measure. Combining this with (5.22) gives that $\nu(O_\ell) > 1 - \delta$.[g] This, however cannot be the case since $\delta(W) \geq \delta$ and O_ℓ is an independent set by Claim 5.13. □

Proof of Claim 5.15. This follows by combining Claim 5.7, Claim 5.13, and the fact that the minimum degree of W is at least δ. □

6. Comparing the proofs

If not counting preparations related to the Regularity method, then the heart of Komlós's proof of Theorem 1.2 in [10] is a less than three pages long calculation. In comparison, the corresponding part of our proof in Section 5.2 has circa four pages. So, our proof is not shorter, but it is conceptually much simpler. Indeed, Komlós's proof proceeds by an ingenious iterative regularization of the host graph, a technique which was novel at that time and which is rare even today (apart from proofs of variants of Komlós's Theorem, such as [19, 7]).

[g]Note this is stronger than Claim 5.7 because the inequality is strict.

Our graphon formalism, on the other hand, allows us to proceed with the most pedestrian thinkable proof strategy. That is, to show using relatively straightforward calculations that no small fractional H-covers exist.

Let us note that our proof can be de-graphonized as follows. Consider a graph G satisfying the minimum-degree condition as in (1.3). Apply the min-degree form of the Regularity lemma, thus arriving to a cluster graph R. Now, the calculations from Section 5.2 can be used *mutatis mutandis* to prove that R contains no small fractional H-cover. Thus, by LP duality, the cluster graph R contains a large fractional H-tiling. This fractional H-tiling in R can be pulled back to a proportionally sized integral H-tiling in G by Blow-up lemma type techniques. The advantage of this approach is that it allows the above mentioned argument "take a vertex which has the smallest value of $\mathfrak{c}$ and consider its neighborhood" (on the level of the cluster graph), but this is compensated by the usual technical difficulties like irregular or low density pairs.

7. Further possible applications

While Komlós's Theorem provides a complete answer (at least asymptotically) for lower-bounding $\mathsf{til}(H, G)$ in terms of the minimum degree of G, the average degree version of the problem is much less understood. Apart from the Erdős–Gallai Theorem ($H = K_2$) mentioned in Section 1, the only other known graphs for which the asymptotic F-tiling thresholds have been determined are all bipartite graphs, [7] and K_3, [1]. The current graphon formalism may be of help in finding further density thresholds.

After this paper was made public, Piguet and Saumell [15, 14] used a similar approach (with the de-graphonized formalism, as described in Section 6) to obtain a strengthening of Komlós's Theorem. In that strengthening, the lower-bound (1.3) is not required for all vertices but rather only for a certain (and optimal) proportion (which depends on x and the graph H) of them.

Let us remark that in [4], the authors provide a graphon proof of the Erdős–Gallai Theorem. The key tool to this end is to establish the half-integrality property of the fractional vertex cover "polyt*on*". These objects are defined in analogy to fractional vertex cover potypes of graphs, but for graphons (hence the "-*on*" ending). This half-integrality property is a direct counterpart to the well-known statement about fractional vertex cover polytopes of finite graphs.

8. Acknowledgments

JH would like thank Dan Král and András Máthé for useful discussions that preceded this project. He would also like to thank Martin Doležal for the discussions they have had regarding functional analysis. We thank the referees for their helpful comments.

Part of this paper was written while JH was participating in the program *Measured group theory* at The Erwin Schrödinger International Institute for Mathematics and Physics.

The contents of this publication reflects only the authors' views and not necessarily the views of the European Commission of the European Union. This publication reflects only its authors' view; the European Research Council Executive Agency is not responsible for any use that may be made of the information it contains.

References

[1] P. Allen, J. Böttcher, J. Hladký, and D. Piguet. A density Corrádi-Hajnal theorem. *Canad. J. Math.*, 67(4):721–758, 2015.

[2] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[3] A. Bruckner, J. Bruckner, and B. Thomson. *Real Analysis*. ClassicalRealAnalysis.com, second edition, 2008.

[4] M. Doležal and J. Hladký. Matching polytons. arXiv:1606.06958.

[5] P. Erdős and T. Gallai. On maximal paths and circuits of graphs. *Acta Math. Acad. Sci. Hungar*, 10:337–356 (unbound insert), 1959.

[6] P. Erdős and A. H. Stone. On the structure of linear graphs. *Bulletin of the American Mathematical Society*, 52:1087–1091, 1946.

[7] C. Grosu and J. Hladký. The extremal function for partial bipartite tilings. *European J. Combin.*, 33(5):807–815, 2012.

[8] A. Hajnal and E. Szemerédi. Proof of a conjecture of P. Erdős. In *Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969)*, pages 601–623. North-Holland, Amsterdam, 1970.

[9] J. Hladký, P. Hu, and D. Piguet. Tilings in graphons. arXiv:1606.03113v1.

[10] J. Komlós. Tiling Turán theorems. *Combinatorica*, 20(2):203–218, 2000.

[11] D. Kühn and D. Osthus. The minimum degree threshold for perfect graph packings. *Combinatorica*, 29(1):65–107, 2009.

[12] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[13] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[14] D. Piguet and M. Saumell. A median-type condition for graph tiling.

[15] D. Piguet and M. Saumell. A median-type condition for graph tiling. In *European Conference on Combinatorics, Graph Theory and Applications (EuroComb 2017)*, volume 61 of *Electron. Notes Discrete Math.*, pages 979–985. Elsevier Sci. B. V., Amsterdam, 2017.

[16] A. A. Razborov. Flag algebras. *J. Symbolic Logic*, 72(4):1239–1282, 2007.

[17] M. Simonovits. A method for solving extremal problems in graph theory, stability problems. In *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pages 279–319. Academic Press, New York, 1968.

[18] T. Tao. *An epsilon of room, I: real analysis*, volume 117 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010. Pages from year three of a mathematical blog.

[19] A. Treglown. A degree sequence Hajnal–Szemerédi theorem. *J. Combin. Theory Ser. B*, 118:13–43, 2016.

Current address: Institute for geometry, TU Dresden, 01062 Dresden, Germany. *This work has been carried out while at:* Institute of Mathematics, Czech Academy of Science. Žitná 25, 110 00, Praha, Czech Republic. The Institute of Mathematics of the Czech Academy of Sciences is supported by RVO:67985840.

Current address: School of Mathematics, Sun Yat-sen University, Guangzhou, 510275, China. *This work has been carried out while at:* Department of Computer Science, University of Warwick, Coventry, CV4 7AL, United Kingdom.

The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, 182 07 Prague, Czech Republic. With institutional support RVO:67985807.

MATCHING POLYTONS

MARTIN DOLEŽAL AND JAN HLADKÝ

Abstract. Hladký, Hu, and Piguet [Tilings in graphons, preprint] introduced the notions of matching and fractional vertex covers in graphons. These are counterparts to the corresponding notions in finite graphs.

Combinatorial optimization studies the structure of the matching polytope and the fractional vertex cover polytope of a graph. Here, in analogy, we initiate the study of the structure of the set of all matchings and of all fractional vertex covers in a graphon. We call these sets the matching polyton and the fractional vertex cover polyton.

We also study properties of matching polytons and fractional vertex cover polytons along convergent sequences of graphons.

As an auxiliary tool of independent interest, we prove that a graphon is r-partite if and only if it contains no graph of chromatic number $r+1$. This in turn gives a characterization of bipartite graphons as those having a symmetric spectrum.

1. Introduction

Theories of graph limits are arguably one of the most important directions in discrete mathematics in the last decade. They link graph theory to analytic parts of mathematics, and through this connection introduce new tools to graph theory. With regard to such applications, the most fruitful theory has been that of flag algebras, [16]. Here, we deal with the theory of graphons developed by Borgs, Chayes, Lovász, Szegedy, Sós, and Vesztergombi, [1, 13]. This theory, too, has found numerous applications in extremal graph theory (e.g., [11]), theory of random graphs [2], and in our understanding of properties of Szemerédi regularity partitions, see e.g. [14]. Much of the theory is built on counterparts of concepts well-known from the world of finite graphs (such as subgraph densities or cuts). Ideally, such counterparts are continuous with respect to the cut-distance, and are equal to the original concept when the graphon in question is a representation of a finite graph.

Hladký, Hu, and Piguet in [6] translated the concept of vertex-disjoint copies of a fixed finite graph F in a (large) host graph to graphons. Following preceding literature on this topic, they use the name *F-tiling* (in a graph or in a graphon). This allows them to introduce the *F-tiling ratio* of a graphon. They also translate the closely related concept of (fractional) F-covers in finite graphs to graphons which is a dual concept (in the sense

Martin Doležal was supported by GA ČR Grant 16-07378S.

Jan Hladký: The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement number 628974.

An extended abstract describing these results has appeared in the proceedings of the EuroComb2017 conference, [4].

of linear programming) to F-tilings. The case when F is an edge, $F = K_2$, is the most important. Then F-tilings are exactly matchings, the F-tiling ratio is just the matching ratio[1], and (fractional) F-covers are exactly (fractional) vertex covers.

In this paper we deal exclusively with the case $F = K_2$ and from now on we specialize our description to this case. We discuss the possible generalizations to other F-tilings of some of the results presented in this paper in Section 5.4.

Hladký, Hu, and Piguet mostly study the numerical quantities provided by the theory they develop, that is, the matching ratio and the fractional vertex cover ratio. Given a graphon W, we denote these two quantities (which we define in Section 2.5) by $\mathsf{match}(W)$ and $\mathsf{fcov}(W)$. One of their main results is a transference result about the matching ratio between finite graphs and graphons, as follows.[2]

Theorem 1 (Theorem 3.4 in [6]). *Suppose that $(G_n)_n$ is a sequence of graphs of growing orders that converge to a graphon W in the cut-distance. Then for every $\varepsilon > 0$ there exists n_0 so that for each $n > n_0$ the graph G_n contains a matching of size at least $(\frac{\mathsf{match}(W)}{2} - \varepsilon) \cdot v(G_n)$.*

Another main result from [6] is the counterpart of the prominent linear programming duality between the fractional matching number of a graph and its fractional vertex cover number. Since — as Hladký, Hu, and Piguet argue — in the graphon world there is no distinction between matchings and fractional matchings, their LP duality has the form $\mathsf{match}(W) = \mathsf{fcov}(W)$. In [6] and [7] they give applications of this LP duality in extremal graph theory.[3]

In this paper, on the other hand, we study the sets of all fractional matchings, and of all fractional vertex covers. In the case of a finite graph G, these sets are known as the *fractional matching polytope* and the *fractional vertex cover polytope*. We shall denote them by $\mathrm{MATCH}(G)$ and $\mathrm{FCOV}(G)$. Study of $\mathrm{MATCH}(G)$ and $\mathrm{FCOV}(G)$ (and study of related polytopes such as the (integral) matching polytope and the perfect matching polytope) is central in polyhedral combinatorics and in combinatorial optimization. From the numerous results on the geometry of these polytopes, let us mention integrality of the fractional matching polytope and fractional vertex cover polytope of a bipartite graph, or the Edmonds' perfect matching polytope theorem. Here, we initiate a parallel study in the context of graphons. While in the finite case we have $\mathrm{MATCH}(G) \subseteq \mathbb{R}^{E(G)}$ and $\mathrm{FCOV}(G) \subseteq \mathbb{R}^{V(G)}$, given a graphon $W : \Omega^2 \to [0,1]$, for the corresponding objects $\mathrm{MATCH}(W)$ and $\mathrm{FCOV}(W)$ it turns out that we have $\mathrm{MATCH}(W) \subseteq L^1(\Omega^2)$ and $\mathrm{FCOV}(W) \subseteq L^\infty(\Omega)$. So, while $\mathrm{MATCH}(G)$ and $\mathrm{FCOV}(G)$ are studied using tools from

[1] the *matching ratio* is just the matching number divided by the number of vertices

[2] See Section 2 for the definition of cut-distance convergence.

[3] These applications become particularly interesting when general F-tilings are considered.

linear algebra, in order to study MATCH(W) and FCOV(W) we need to use the language of functional analysis. We employ the *-on* word ending used among others for graph*on*s and for permut*on*s and call the limit counterparts to polytopes (such as MATCH(W) and FCOV(W)) polyt*on*s.

1.1. **Overview of the paper.** In Section 2 we recall the necessary background concerning graphons and the theory of matchings/tilings in graphons developed in [6]. As an auxiliary tool for our main results, we prove in Section 2.4 that a graphon is r-partite if and only if it has zero density of every graph of chromatic number $r+1$ (see Proposition 5). This in turn gives a characterization of bipartite graphons as those having a symmetric spectrum (see Theorem 8), which is a graphon counterpart to a well-known property for finite graphs.

In Section 3 we treat (half)-integrality of the extreme points of the fractional vertex cover polyton of a graphon. The main results of this section are summarized in Theorems 13,14 and 15. As an application, we deduce a graphon version of the Erdős–Gallai theorem on matchings in dense graphs (see Theorem 19). In Section 4 we show that if a sequence of graphons $(W_n)_n$ converges to a graphon W in the cut-distance then "MATCH(W_n) asymptotically contains MATCH(W)" (see Corollary 24, or Theorem 23 for a slightly stronger statement). This result is dual in the sense of linear programming to results from [6] on the relation between FCOV(W_n) and FCOV(W), which we recall in Section 4.1.

Section 5 contains some concluding remarks.

2. Notation and preliminaries

2.1. Functional analysis.

2.1.1. *Weak* convergence and the Banach–Alaoglu theorem.* We now recall the concept of weak* convergence and the Banach–Alaoglu theorem (which we present in a form that is tailored for our purposes). Suppose that (X, λ) is a Borel probability space. Recall that a sequence of measurable functions $f_1, f_2, \ldots : X \to [0,1]$ *converges weak** to a measurable function $f : X \to [0,1]$ if for every measurable set $A \subseteq X$ we have $\lim_n \int_A f_n = \int_A f$. The Banach–Alaoglu theorem then asserts that the space of measurable functions from X to $[0,1]$ is sequentially compact.

Since the product of finitely many sequentially compact spaces is sequentially compact, the Banach–Alaoglu theorem generalizes as follows. Suppose that $k \in \mathbb{N}$ is fixed, and suppose that we are given k sequences $f_1^1, f_2^1, f_3^1, \ldots$, $f_1^2, f_2^2, f_3^2, \ldots$, $\ldots$, $f_1^k, f_2^k, f_3^k, \ldots$ of measurable functions from X to $[0,1]$. Then there exist a sequence of indices $n_1 < n_2 < \ldots$ and functions $f^1, f^2, \ldots, f^k : X \to [0,1]$ such that for each $i \in [k]$, the sequence $f_{n_1}^i, f_{n_2}^i, f_{n_3}^i, \ldots$ converges weak* to f^i.

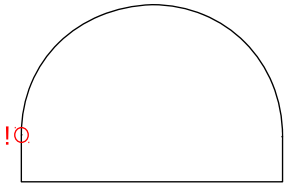


FIGURE 1. By joining a half-circle and a rectangle in $\mathbb{R}^2$ we get an extreme point that is not exposed.

2.1.2. *The Krein–Milman Theorem.* In this section, we briefly recall the Krein–Milman theorem. This theorem will allow to talk about "vertices" in our counterparts to polytopes.

Suppose that $\mathcal{L}$ is a vector space, and suppose that $X \subseteq \mathcal{L}$ is a convex set. Recall that a point $x \in X$ is called an *extreme point of* X if the only pair $x', x'' \in X$ for which $x = \frac{1}{2}(x' + x'')$ is the pair $x' = x$, $x'' = x$. We shall write $\mathcal{E}(X)$ to denote the set of all extreme points of X.

When $\mathcal{L}$ is finite-dimensional and X is a polytope in $\mathcal{L}$ then the extreme points of X are exactly its vertices. The importance of the notion of extreme points comes from the Krein–Milman theorem which states that in a locally convex topological vector space, each compact convex set equals the closed convex hull of its extreme points.

The notion of extreme points is not the only generalization of vertices of a polytope. Another basic notion from convex analysis is that of exposed points. (Note that its stronger variant, the notion of strongly exposed points, can be used for a characterization of the Radon–Nikodym property of Banach spaces which is an extensively studied topic.) A point x in a convex set X is *exposed* if there exists a continuous linear functional for which x attains its strict maximum on X.

It is easy to see that every exposed point is extreme. The converse does not hold; a well-known counterexample in $\mathbb{R}^2$ is shown in Figure 1.

2.2. **Measure theory.** Suppose that λ is a probability measure on a space Λ. Recall that for a measurable function $f : \Lambda \to \mathbb{R}$, its *essential supremum* is defined as

$$\operatorname{esssup}(f) = \inf\{a : \lambda(\{x : f(x) > a\}) = 0\} .$$

That is, the essential supremum is like the ordinary supremum, except that high values are not considered as long as they occur on a null-set. The *essential infimum* $\operatorname{essinf}(f)$ is defined analogously.

Recall also that the product measure $\lambda^{\otimes 2}$ is defined by

$$\lambda^{\otimes 2}(R) = \inf\left\{\sum_{i=1}^{\infty} \lambda(A_i)\lambda(B_i) \colon A_i, B_i \subseteq \Lambda \text{ are measurable such that } R \subseteq \bigcup_{i=1}^{\infty} A_i \times B_i\right\}$$

for every $R \subseteq \Lambda^2$ from the product σ-algebra.

Below, we state an auxiliary result, which says that for every set $D \subseteq \Lambda^2$ we can find a rectangle (i.e., a set which is a product of two sets) which is almost entirely contained in D.

Lemma 2. *Let (Λ, λ) be a probability space and let $A, B \subseteq \Lambda$ be given. Let $D \subseteq A \times B$ be a set of positive $\lambda^{\otimes 2}$ measure. Then for every $\varepsilon > 0$ there is a measurable rectangle $R \subseteq A \times B$ such that $\lambda^{\otimes 2}(R \setminus D) < \varepsilon \lambda^{\otimes 2}(R)$.*

Proof. Let us fix $\varepsilon > 0$. By the definition of the product measure, we can find measurable rectangles $R_1, R_2, \ldots \subseteq A \times B$ such that $D \subseteq \bigcup_{i=1}^{\infty} R_i$ and

$$\lambda^{\otimes 2}\left(\bigcup_{i=1}^{\infty} R_i \setminus D\right) < \varepsilon \lambda^{\otimes 2}(D) \leq \varepsilon \lambda^{\otimes 2}\left(\bigcup_{i=1}^{\infty} R_i\right) .$$

Then there is a natural number m such that

$$\lambda^{\otimes 2}\left(\bigcup_{i=1}^{m} R_i \setminus D\right) < \varepsilon \lambda^{\otimes 2}\left(\bigcup_{i=1}^{m} R_i\right) . \tag{1}$$

Now the finite union $\bigcup_{i=1}^{m} R_i$ can obviously be decomposed into finitely many *pairwise disjoint* measurable rectangles $S_1, \ldots, S_l$. Then the inequality (1) can be rewritten as

$$\sum_{i=1}^{l} \lambda^{\otimes 2}(S_i \setminus D) < \varepsilon \sum_{i=1}^{l} \lambda^{\otimes 2}(S_i) .$$

Thus there is some $i \in \{1, \ldots, l\}$ such that $\lambda^{\otimes 2}(S_i \setminus D) < \varepsilon \lambda^{\otimes 2}(S_i)$. The corresponding S_i is the wanted measurable rectangle R. □

2.3. **Graphon basics.** Our notation follows [12]. Throughout the paper we shall assume that Ω is an atomless Borel probability space equipped with a measure ν (defined on an implicit σ-algebra). We denote by $\nu^{\otimes k}$ the product measure on Ω^k.

A *graphon* is a symmetric Lebesgue measurable function $W : \Omega^2 \to [0, 1]$. We refer to [12, Part 3] for a detailed treatment of the concept. On an intuitive level, a graphon can be viewed as an adjacency matrix (scaled down to fit into Ω^2) of a large graph, where the values of the graphon represent frequencies of 1's in the adjacency matrix.

Given a finite graph G we can construct its *graphon representation* $W_G : \Omega^2 \to [0, 1]$ as follows. We partition Ω arbitrarily into sets $\{\Omega_v\}_{v \in V(G)}$ of measure $1/v(G)$ each. We then define W_G to be 1 on $\Omega_u \times \Omega_v$ if $uv \in E(G)$ and 0 otherwise. The actual graphon representation W_G depends on the choice of the partition $\{\Omega_v\}_{v \in V(G)}$, but the cut-distance $\mathrm{dist}_{\square}(W_G, \cdot)$ does not (see below for the definition of cut-distance).

Suppose that F is a graph on the vertex set $[k]$. Then the *density* of F in a graphon W is defined as

$$t(F,W) = \int_{x_1}\int_{x_2}\cdots\int_{x_k}\prod_{\substack{ij\in E(F)\\ i<j}} W(x_i,x_j)\ .$$

Recall that the *cut-norm* $\|\cdot\|_\square$ and the *cut-distance* $\mathrm{dist}_\square(\cdot,\cdot)$ are defined by

$$\|U\|_\square = \sup_{S,T\subseteq\Omega}\left|\int_{S\times T} U\right|\ ,\ U\in L^1(\Omega^2)\ ,$$

where the supremum ranges over all measurable sets $S,T\subseteq\Omega$, and

$$\mathrm{dist}_\square(U,W) = \inf_\phi \|U-W^\phi\|_\square\ ,\ U,W\in L^1(\Omega^2)\ ,$$

where the infimum ranges over all measure-preserving bijections on Ω, and W^ϕ is defined by $W^\phi(x,y)=W(\phi(x),\phi(y))$.

2.3.1. *Independent sets and partite graphons.* A set $A\subseteq\Omega$ is an *independent set* in a graphon $W:\Omega^2\to[0,1]$ if W is zero almost everywhere on $A\times A$. In [7, Lemma 2.3] it was proven that the support of any weak* accumulation point of a sequence of (indicator functions of) independent sets is an independent set. Here, we extend the statement by allowing the sets in the sequence to be only "asymptotically independent". The proof is however similar to that in [7].

Lemma 3. *Let $W:\Omega^2\to[0,1]$ be a graphon. Suppose that $(A_n)_{n=1}^\infty$ is a sequence of sets in Ω with the property that*

$$\int_{(x,y)\in A_n\times A_n} W(x,y)\overset{n\to\infty}{\longrightarrow}0\ .$$

Suppose that the indicator functions of the sets A_n converge weak to a function $f:\Omega\to[0,1]$. Then* $\mathrm{supp}\, f$ *is an independent set in W.*

Proof. It is enough to prove that for every $\delta>0$, the set $P=\{x\in\Omega : f(x)>\delta\}$ is independent. Suppose that the statement is false. Then there exist sets $X,Y\subseteq P$ of positive measure and a number $\alpha>0$ such that

$$\nu\left(X\times Y\cap\left\{(x,y)\in\Omega^2 : W(x,y)\le\alpha\right\}\right) < \frac{\delta^2}{5}\nu(X)\nu(Y)\ . \tag{2}$$

Recall that $\int_X f\ge\delta\nu(X)$ and $\int_Y f\ge\delta\nu(Y)$. By weak* convergence, for n sufficiently large, $\nu(X\cap A_n)\ge\frac12\delta\nu(X)$ and $\nu(Y\cap A_n)\ge\frac12\delta\nu(Y)$. In particular, (2) gives that

$$\int_{X\cap A_n}\int_{Y\cap A_n} W(x,y)\ge\left(\frac14\delta^2-\frac15\delta^2\right)\alpha\nu(X)\nu(Y)=\frac{1}{20}\delta^2\alpha\nu(X)\nu(Y)\ , \tag{3}$$

for each large enough n. On the other hand,

$$\int_{X\cap A_n}\int_{Y\cap A_n} W(x,y) \le \int_{A_n}\int_{A_n} W(x,y) \overset{n\to\infty}{\longrightarrow} 0\,, \tag{4}$$

by the assumptions of the lemma. Clearly, (3) contradicts (4). □

Given a cardinal r, let us recall that a graphon $W : \Omega^2 \to [0,1]$ is *r-partite* if there exists a partition $\Omega = \bigsqcup_{i\in I} \Omega_i$, where for the index set I we have $|I| = r$, into measurable sets such that for each $i \in I$, we have that Ω_i is an independent set in W. While most of the time we are interested in r finite, note that this definition allows to define even *countably-partite* graphons.

2.3.2. *Inhomogeneous random graphs.* Let us recall that given a graphon $W : \Omega^2 \to [0,1]$ and a number $n \in \mathbb{N}$, the *W-random graph* is defined on the vertex set $\{1,\dots,n\}$ as follows. First, draw independent random points $x_1,\dots,x_n$ from Ω according to the measure ν, and then for each $ij \in \binom{n}{2}$, put an edge into the graph with probability $W(x_i,x_j)$ (and independently). For more details, see [12, Section 10.1].

2.4. **Characterization of partite graphons using forbidden subgraphs.** In this section, we prove that for a given $r \in \mathbb{N}$, a graphon is r-partite if and only if it has zero density of each finite graph of chromatic number at least $r+1$. This is first proven for $r = 2$ in Lemma 4 (in which case the characterization of bipartiteness can be phrased in terms of odd cycles, just as in the world of finite graphs) and then for general r in Proposition 5. A strengthening of Lemma 4 which we give in Lemma 7 is needed for proving Theorem 15. These results may be of independent interest.[4] Further, in Section 2.4.2 we give a characterization of bipartite graphons in terms of their spectrum.

Lemma 4. *Suppose that $W : \Omega^2 \to [0,1]$ is a graphon. Then W is bipartite if and only if for every odd integer $k \ge 3$ it holds $t(C_k, W) = 0$.*

Proof. Suppose first that there is an odd integer $k \ge 3$ such that $t(C_k, W) > 0$. Let $\Omega = \Omega_0 \sqcup \Omega_1$ be an arbitrary decomposition of Ω into two disjoint measurable subsets. Then there exists $(i_j)_{j=1}^k \in \{0,1\}^k$ such that

$$\int\limits_{x_1\in\Omega_{i_1}} \int\limits_{x_2\in\Omega_{i_2}} \cdots \int\limits_{x_k\in\Omega_{i_k}} W(x_1,x_2)W(x_2,x_3)\dots W(x_k,x_1) > 0\,.$$

[4] In subsequent work [8], Hladký and Rocha used this lemma to obtain properties of the "independent set polyton".

As k is odd, there is $j \in \{1, \dots, k\}$ such that $i_j = i_{j+1}$ (here we use the cyclic indexing, i.e. $k+1 = 1$). By Fubini's theorem

$$\int\limits_{x_j \in \Omega_{i_j}} \int\limits_{x_{j+1} \in \Omega_{i_{j+1}}} W(x_j, x_{j+1}) > 0 \,,$$

in other words $\int_{\Omega_i^2} W > 0$ for some $i \in \{0, 1\}$. As the decomposition $\Omega = \Omega_0 \sqcup \Omega_1$ was chosen arbitrarily, this proves that W is not bipartite.

Now suppose that $t(C_k, W) = 0$ for every odd integer $k \geq 3$. By transfinite induction we define a transfinite sequence $\{(A_\alpha, B_\alpha) \colon \alpha \leq \gamma\}$ (for some countable ordinal γ) consisting of pairs of measurable subsets of Ω such that $\nu(A_\gamma \cup B_\gamma) = 1$, and such that for every $\alpha \leq \gamma$ the following conditions hold:

- (i)$_\alpha$ $A_\beta \subseteq A_\alpha$ and $B_\beta \subseteq B_\alpha$ for every $\beta \leq \alpha$,
- (ii)$_\alpha$ $\nu(A_\beta \cup B_\beta) < \nu(A_\alpha \cup B_\alpha)$ for every $\beta < \alpha$,
- (iii)$_\alpha$ $W \restriction_{A_\alpha^2} = 0$ a.e. and $W \restriction_{B_\alpha^2} = 0$ a.e.,
- (iv)$_\alpha$ $W \restriction_{(A_\alpha \cup B_\alpha) \times (\Omega \setminus (A_\alpha \cup B_\alpha))} = 0$ a.e.,
- (v)$_\alpha$ $\nu(A_\alpha \cap B_\alpha) = 0$.

Once we are done with the construction, the bipartiteness of W immediately follows by the equation $\nu(A_\gamma \cup B_\gamma) = 1$ together with (iii)$_\gamma$ and (v)$_\gamma$.

We start the construction by setting $A_0 = B_0 = \emptyset$, then the conditions (i)$_\emptyset$-(v)$_\emptyset$ hold trivially. Now suppose that we have already constructed $\{(A_\alpha, B_\alpha) \colon \alpha < \alpha_0\}$ for some countable ordinal α_0 such that the conditions (i)$_\alpha$-(v)$_\alpha$ hold for every $\alpha < \alpha_0$. If α_0 is a limit ordinal then we set $A_{\alpha_0} = \bigcup_{\alpha < \alpha_0} A_\alpha$ and $B_{\alpha_0} = \bigcup_{\alpha < \alpha_0} B_\alpha$, then (i)$_{\alpha_0}$-(v)$_{\alpha_0}$ clearly hold. Otherwise, $\alpha_0 = \alpha + 1$ for some ordinal $\alpha < \alpha_0$. If $\nu(A_\alpha \cup B_\alpha) = 1$ then the construction is finished (with $\gamma := \alpha$). So suppose that $\nu(A_\alpha \cup B_\alpha) < 1$ and denote $\Omega'_\alpha := \Omega \setminus (A_\alpha \cup B_\alpha)$. If $W \restriction_{\Omega'_\alpha \times \Omega'_\alpha} = 0$ a.e. then it suffices (by (iv)$_\alpha$) to set $A_{\alpha_0} = A_\alpha \cup \Omega'_\alpha$ and $B_{\alpha_0} = B_\alpha$. Otherwise there is $x_0 \in \Omega'_\alpha$ such that $\nu\left(\{y \in \Omega'_\alpha \colon W(x_0, y) > 0\}\right) > 0$. By Fubini's theorem, we may also assume that for every odd integer $k \geq 3$ we have that

$$\int\limits_{\Omega^{k-1}} W(x_0, x_1) W(x_1, x_2) \dots W(x_{k-1}, x_0) = 0 \,. \tag{5}$$

We set

$$D_0 = \{y \in \Omega'_\alpha \colon W(x_0, y) > 0\} \quad \text{, and}$$

$$D_\ell = \left\{ y \in \Omega'_\alpha \colon \int\limits_{\Omega^\ell} W(x_0, x_1) W(x_1, x_2) \dots W(x_\ell, y) > 0 \right\}, \text{ for } \ell \geq 1 \,.$$

Then we set

$$E_\alpha = \bigcup_{\ell \text{ even}} D_\ell \quad \text{and} \quad F_\alpha = \bigcup_{\ell \text{ odd}} D_\ell \,.$$

Finally, we define $A_{\alpha 0} = A_\alpha \cup E_\alpha$ and $B_{\alpha 0} = B_\alpha \cup F_\alpha$. The conditions $(\mathrm{i})_{\alpha 0}$ and $(\mathrm{ii})_{\alpha 0}$ are clearly satisfied, so let us verify only $(\mathrm{iii})_{\alpha 0}$-$(\mathrm{v})_{\alpha 0}$.

As for $(\mathrm{iii})_{\alpha 0}$, suppose for a contradiction that $W \restriction_{A^2_{\alpha 0}}$ is positive on a set of positive measure. By the induction hypothesis, namely by $(\mathrm{iii})_\alpha$ and $(\mathrm{iv})_\alpha$, we easily conclude that $W \restriction_{E^2_\alpha}$ is also positive on a set of positive measure. So there are even integers $\ell_1, \ell_2 \geq 0$ such that $W \restriction_{D_{\ell_1} \times D_{\ell_2}}$ is positive on a set of positive measure. But then a simple application of Fubini's theorem leads to a contradiction with (5) for $k = \ell_1 + \ell_2 + 3$. So we have $W \restriction_{A^2_{\alpha 0}} = 0$ a.e. Similarly, we get $W \restriction_{B^2_{\alpha 0}} = 0$ a.e. This proves $(\mathrm{iii})_{\alpha 0}$.

As for $(\mathrm{iv})_{\alpha 0}$, suppose for a contradiction that $W \restriction_{(A_{\alpha 0} \cup B_{\alpha 0}) \times (\Omega \setminus (A_{\alpha 0} \cup B_{\alpha 0}))}$ is positive on a set of positive measure. By the induction hypothesis, namely by $(\mathrm{iv})_\alpha$, we have that $W \restriction_{(E_\alpha \cup F_\alpha) \times (\Omega \setminus (A_{\alpha 0} \cup B_{\alpha 0}))} = W \restriction_{(E_\alpha \cup F_\alpha) \times (\Omega'_\alpha \setminus (E_\alpha \cup F_\alpha))}$ is also positive on a set of positive measure. By Fubini's theorem, there is $z \in \Omega'_\alpha \setminus (E_\alpha \cup F_\alpha)$ such that

$$\nu\left(\{y \in E_\alpha \cup F_\alpha \colon W(y,z) > 0\}\right) > 0 \,.$$

So there is an integer $\ell \geq 0$ such that

$$\nu\left(\{y \in D_\ell \colon W(y,z) > 0\}\right) > 0 \,.$$

By Fubini's theorem, we easily conclude that $z \in D_{\ell+1}$. But this contradicts the fact that $z \notin E_\alpha \cup F_\alpha$.

As for $(\mathrm{v})_{\alpha 0}$, suppose for a contradiction that $\nu(A_{\alpha 0} \cap B_{\alpha 0}) > 0$. By the induction hypothesis, namely by $(\mathrm{v})_\alpha$, we also have $\nu(E_\alpha \cap F_\alpha) > 0$. So there are an even integer ℓ_1 and an odd integer ℓ_2 such that $\nu(D_{\ell_1} \cap D_{\ell_2}) > 0$. But then an application of Fubini's theorem leads to a contradiction with (5) for $k = \ell_1 + \ell_2 + 2$.

To finish the proof, it suffices to observe that for some countable ordinal γ we get $\nu(A_\gamma \cup B_\gamma) = 1$ (and at that point, the construction stops by the description above). So suppose for a contradiction that this is not the case. Then, by conditions $(\mathrm{i})_\alpha$ and $(\mathrm{ii})_\alpha$, we can find uncountably many pairwise disjoint subsets of Ω of positive measure, namely $(A_{\alpha+1} \cup B_{\alpha+1}) \setminus (A_\alpha \cup B_\alpha)$ (where α is not bounded by any countable ordinal γ). But this is obviously not possible. □

We can now state Proposition 5 which generalizes Lemma 4 to higher chromatic number. Its proof stems from discussions with András Máthé.

Proposition 5. *Suppose that $k \in \mathbb{N}$ and $W : \Omega^2 \to [0,1]$ is a graphon for which $t(H,W) = 0$ for each graph H of chromatic number $k+1$. Then W is k-partite.*

Proof. In the proof, we first approximate W a sequence of samples of W-random graphs $G_n \sim \mathbb{G}(n, W)$. As a second step, we show that with probability one, each G_n is k-colorable. Last, we can transfer the k-colorings of G_n to a k-coloring of W. As for the second step, observe that indeed we have

$$\mathbf{P}_{G_n \sim \mathbb{G}(n,W)}[G_n \text{ is not } k\text{-colorable}] \leq \sum_H t(H, W) \ ,$$

where the sum runs over all not k-colorable graphs H on n vertices. Since all the terms are 0 by the assumptions of the proposition, we conclude that G_n is indeed k-colorable almost surely.

Let $(G_n)_n$ be samples of the inhomogeneous random graphs $\mathbb{G}(n, W)$. It is well-known that the graphs G_n converge to W in the cut-distance almost surely, see e.g., [12, Lemma 10.16]. We can therefore map the vertices $\{1, 2, \ldots, n\}$ of G_n to sets $\Omega_n^{(1)}, \Omega_n^{(2)}, \ldots, \Omega_n^{(n)}$ which partition Ω into sets of measure $\frac{1}{n}$ each, in a way that this partition witnesses ε_n-closeness of G_n to W in the cut-distance (where $\varepsilon_n \to 0_+$). In particular, whenever $O \subseteq V(G_n)$ is an independent set, we have

$$\int_{x \in \bigcup_{v \in O} \Omega_n^{(v)}} \int_{y \in \bigcup_{v \in O} \Omega_n^{(v)}} W(x, y) < \varepsilon_n \ .$$

Recall G_n is k-colorable (with probability one). Let us fix a partition

$$V_n^{(1)} \sqcup V_n^{(2)} \sqcup \ldots \sqcup V_n^{(k)} = \Omega$$

according to one fixed k-coloring of G_n.

Consider now a weak* accumulation point $(f_1, f_2, \ldots, f_k)$ of the sequence of k-tuples of functions

$$\left(\mathbf{1}_{V_n^{(1)}} \ , \ \mathbf{1}_{V_n^{(2)}} \ , \ldots \ , \ \mathbf{1}_{V_n^{(k)}}\right)_n \ .$$

Such an accumulation point exists by the sequential Banach–Alaoglu theorem, see Section 2.1.1.

We have $f_1 + f_2 + \ldots + f_k = 1$ almost everywhere on Ω. Consequently we can find measurable sets $A_i \subseteq \operatorname{supp} f_i$ so that $\Omega = A_1 \sqcup A_2 \sqcup \ldots \sqcup A_k$. Lemma 3 tells us that $A_1 \sqcup A_2 \sqcup \ldots \sqcup A_k$ is a k-coloring of W. □

Let us note that in [3] a result in a similar direction was proven:

Theorem 6. *Suppose that W is a graphon, and H is a finite graph with the property that $t(H, W) = 0$. Then W is countably partite.*

2.4.1. *A technical lemma.* Lemma 4 tells us that if a graphon is not bipartite, then it has a positive density of some odd cycle C_k. The next lemma allows us to zoom in into some location, where some of these C_k's are very densely located. That is, we will find sets $A_1, \dots, A_k \subseteq \Omega$ such that W is positive on most of $A_h \times A_{h+1}$ for each $h \in [k]$. The additional properties, which we will need for our proof of Theorem 15 later, are that the sets A_h have the same measure, and are disjoint.

Lemma 7. *Suppose that $W : \Omega^2 \to [0,1]$ is a graphon. If W is not bipartite then there exists an odd integer $k \geq 3$ with the following property. For each $\varepsilon > 0$ there exist pairwise disjoint sets $A_1, \dots, A_k \subseteq \Omega$ of the same positive measure α, such that for each $h \in [k]$, W is positive everywhere on $A_h \times A_{h+1}$ except a set of measure at most $\varepsilon\alpha^2$. Here, we use cyclic indexing, $A_{k+1} = A_1$.*

Proof. Suppose that W is not bipartite. By Lemma 4 there is an odd integer $k \geq 3$ such that

$$t := \int_{\Omega^k} W(x_1, x_2)W(x_2, x_3)\dots W(x_k, x_1) > 0 \,. \tag{6}$$

We find a natural number n such that

$$n > \frac{k^2}{t} \,. \tag{7}$$

We fix a decomposition $\Omega = \bigcup_{i=1}^{n} \Omega_i$ of Ω into pairwise disjoint sets of the same measure $\frac{1}{n}$. We also set

$$\begin{aligned} D = \big\{(x_1, \dots, x_k) \in \Omega^k \colon &\text{ there are } i, j \in \{1, \dots, k\} \text{ and } \ell \in \{1, \dots, n\} \\ &\text{such that } i \neq j \text{ and } x_i, x_j \in \Omega_\ell\big\} \,. \end{aligned} \tag{8}$$

Then we have

$$\nu^{\otimes k}(D) \leq \sum_{\substack{i,j=1,\dots,k \\ i \neq j}} \frac{1}{n} \leq \frac{k^2}{n} \overset{(7)}{<} t \,,$$

and so

$$\int_D W(x_1, x_2)W(x_2, x_3)\dots W(x_k, x_1) \leq \nu^{\otimes k}(D) < t \,. \tag{9}$$

By (6) and (9), we get

$$\int_{\Omega^k \setminus D} W(x_1, x_2)W(x_2, x_3)\dots W(x_k, x_1) > 0 \,.$$

By this and (8) there are pairwise distinct integers $\ell_1, \dots, \ell_k \in \{1, \dots, n\}$ such that

$$\int_{\Omega_{\ell_1}} \int_{\Omega_{\ell_2}} \cdots \int_{\Omega_{\ell_k}} W(x_1, x_2) W(x_2, x_3) \dots W(x_k, x_1) > 0 \,, \tag{10}$$

and so the set

$$E := \{(x_1, x_2, \dots, x_k) \in \Omega_{\ell_1} \times \Omega_{\ell_2} \times \dots \times \Omega_{\ell_k} : W(x_1, x_2) W(x_2, x_3) \dots W(x_k, x_1) > 0\} \tag{11}$$

is of positive measure.

Now let us fix $\varepsilon > 0$, and let $\delta > 0$ be such that

$$\frac{\nu^{\otimes k}(E) - \delta}{\nu^{\otimes k}(E) + \delta} \geq 1 - \frac{\varepsilon}{2} \,. \tag{12}$$

Recall that the σ-algebra of all measurable subsets of $\Omega_{\ell_1} \times \dots \times \Omega_{\ell_k}$ is generated by the algebra consisting of all finite unions of measurable rectangles. Thus there is a finite union $S = \bigcup_{i=1}^{m} R_i$ of measurable rectangles $R_1, \dots, R_m$ in $\Omega_{\ell_1} \times \dots \times \Omega_{\ell_k}$ such that

$$\nu^{\otimes k}(E \triangle S) = \nu^{\otimes k}(E \setminus S) + \nu^{\otimes k}(S \setminus E) \leq \delta \,.$$

Without loss of generality, we may assume that the measurable rectangles $R_1, \dots, R_m$ are pairwise disjoint. Then we have

$$\frac{\nu^{\otimes k}(E \cap S)}{\nu^{\otimes k}(S)} \geq \frac{\nu^{\otimes k}(E) - \nu^{\otimes k}(E \setminus S)}{\nu^{\otimes k}(E) + \nu^{\otimes k}(S \setminus E)} \geq \frac{\nu^{\otimes k}(E) - \delta}{\nu^{\otimes k}(E) + \delta} \overset{(12)}{\geq} 1 - \frac{\varepsilon}{2} \,. \tag{13}$$

Now the left-hand side of (13) can be expressed as

$$\frac{\nu^{\otimes k}(E \cap S)}{\nu^{\otimes k}(S)} = \sum_{i=1}^{m} \frac{\nu^{\otimes k}(R_i)}{\nu^{\otimes k}(S)} \cdot \frac{\nu^{\otimes k}(E \cap R_i)}{\nu^{\otimes k}(R_i)} \,,$$

i.e. as a convex combination of $\frac{\nu^{\otimes k}(E \cap R_i)}{\nu^{\otimes k}(R_i)}$, $i = 1, \dots, m$. Therefore by (13), there is an index $i_0 \in \{1, \dots, m\}$ such that

$$\frac{\nu^{\otimes k}(E \cap R_{i_0})}{\nu^{\otimes k}(R_{i_0})} \geq 1 - \frac{\varepsilon}{2} \,. \tag{14}$$

Let R_{i_0} be of the form $R_{i_0} = B_1 \times \dots \times B_k$. Find a natural number p such that

$$p \geq \frac{2k}{\varepsilon \nu^{\otimes k}(R_{i_0})} \,. \tag{15}$$

For every $i = 1, \dots, k$, we fix a finite decomposition $B_i = B_i^0 \cup \bigcup_{j=1}^{q_i} B_i^j$ of B_i into $1 + q_i$ many pairwise disjoint sets, such that $q_i := \lfloor p\nu(B_i) \rfloor$, $\nu(B_i^0) \leq \frac{1}{p}$ and $\nu(B_i^j) = \frac{1}{p}$ for $j = 1, \dots, q_i$.

Then we clearly have

$$\nu^{\otimes k}\left(R_{i_0}\setminus\prod_{i=1}^{k}\bigcup_{j=1}^{q_i}B_i^j\right)\le\frac{k}{p}\,, \tag{16}$$

and so

$$\frac{\nu^{\otimes k}\left(E\cap\prod_{i=1}^{k}\bigcup_{j=1}^{q_i}B_i^j\right)}{\nu^{\otimes k}\left(\prod_{i=1}^{k}\bigcup_{j=1}^{q_i}B_i^j\right)}\overset{(16)}{\ge}\frac{\nu^{\otimes k}(E\cap R_{i_0})-\frac{k}{p}}{\nu^{\otimes k}(R_{i_0})}\overset{(14)}{\ge}1-\frac{\varepsilon}{2}-\frac{k}{p\nu^{\otimes k}(R_{i_0})}\overset{(15)}{\ge}1-\varepsilon\,. \tag{17}$$

The left-hand side of (17) can be expressed as the following convex combination:

$$\frac{\nu^{\otimes k}\left(E\cap\prod_{i=1}^{k}\bigcup_{j=1}^{q_i}B_i^j\right)}{\nu^{\otimes k}\left(\prod_{i=1}^{k}\bigcup_{j=1}^{q_i}B_i^j\right)}=\sum_{j_1=1}^{q_1}\cdots\sum_{j_k=1}^{q_k}\frac{\nu^{\otimes k}\left(\prod_{i=1}^{k}B_i^{j_i}\right)}{\nu^{\otimes k}\left(\prod_{i=1}^{k}\bigcup_{j=1}^{q_i}B_i^j\right)}\cdot\frac{\nu^{\otimes k}\left(E\cap\prod_{i=1}^{k}B_i^{j_i}\right)}{\nu^{\otimes k}\left(\prod_{i=1}^{k}B_i^{j_i}\right)}\,.$$

Therefore by (17), there are $j_i\in\{1,\dots,q_i\}$, $i=1,\dots,k$, such that

$$\frac{\nu^{\otimes k}\left(E\cap\prod_{i=1}^{k}B_i^{j_i}\right)}{\nu^{\otimes k}\left(\prod_{i=1}^{k}B_i^{j_i}\right)}\ge1-\varepsilon\,,$$

or equivalently

$$\frac{\nu^{\otimes k}\left(\prod_{i=1}^{k}B_i^{j_i}\setminus E\right)}{\nu^{\otimes k}\left(\prod_{i=1}^{k}B_i^{j_i}\right)}\le\varepsilon\,. \tag{18}$$

We set $A_i=B_i^{j_i}$ for $i=1,\dots,k$. Then $A_1,\dots,A_k$ are pairwise disjoint (as $A_i\subseteq B_i\subseteq\Omega_{\ell_i}$ for every i), and each of these sets has the same measure $\alpha=\frac{1}{p}$.

By (11) we have

$$\begin{aligned}&\{(x_1,x_2,\dots,x_k)\in A_1\times A_2\times\ldots\times A_k\colon W(x_1,x_2)W(x_2,x_3)\dots W(x_k,x_1)=0\}\\&=\prod_{i=1}^{k}B_i^{j_i}\setminus E\,.\end{aligned} \tag{19}$$

By (18),

$$\nu^{\otimes k}\left(\prod_{i=1}^{k}B_i^{j_i}\setminus E\right)\le\varepsilon\nu^{\otimes k}\left(\prod_{i=1}^{k}B_i^{j_i}\right)=\frac{\varepsilon}{p^k}\,.$$

For each $h \in [k]$, we have

$$\prod_{i=1}^{k} B_i^{j_i} \setminus E \supseteq A_1 \times \ldots \times A_{h-1} \times \{(x_h, x_{h+1}) \in A_h \times A_{h+1} \colon W(x_h, x_{h+1}) = 0\} \times A_{h+2} \times \ldots A_k \ .$$

Plugging this into (19), we get

$$\nu^{\otimes 2}\left(\{(x_h, x_{h+1}) \in A_h \times A_{h+1} \colon W(x_h, x_{h+1}) = 0\}\right) \leq \frac{\varepsilon}{p^2} \ ,$$

as required. □

2.4.2. *Application: spectra of bipartite graphons.* It is a well-known fact that a finite graph is bipartite if and only if the spectrum of its adjacency matrix is symmetric. In Theorem 8, we prove a counterpart of this fact for graphons. This result seems to be new. Indeed, just as in the finite case, to obtain this result one needs the characterization of bipartite graphs using odd cycles, which we gave in Lemma 4. This section is not needed for our main results concerning matching polytons.

Let us briefly recall the notion of eigenvalues of graphons, following [12, Section 7.5], where more details can be found. To a given graphon $W : \Omega^2 \to [0, 1]$ we can associate a kernel operator $T_W : L^2(\Omega) \to L^2(\Omega)$,

$$\left(T_W(f)\right)(x) := \int_{y \in \Omega} W(x, y) f(y) \ .$$

Then T_W is a Hilbert–Schmidt operator, and thus it has a real spectrum of finitely or countably many nonzero eigenvalues. The spectrum is said to be *symmetric* if each $\lambda \in \mathbb{R}$ is an eigenvalue if and only if $-\lambda$ is, and their multiplicities are the same.

Theorem 8. *A graphon is bipartite if and only if the associated kernel operator has a symmetric spectrum.*

Proof. Suppose first that $W : \Omega^2 \to [0, 1]$ is a bipartite graphon, and the colour classes are $A \sqcup B = \Omega$. Let $\lambda \in \mathbb{R}$ be an arbitrary eigenvalue and let $f \in L^2(\Omega)$ be the corresponding eigenfunction. Define a function $g \in L^2(\Omega)$ by setting $g(x) = f(x)$ for $x \in A$ and $g(x) = -f(x)$ for $x \in B$. Now, for almost every $x \in A$ we have

$$\begin{aligned} T_W(g)(x) &= \int_{y \in \Omega} W(x, y) g(y) = \int_{y \in B} W(x, y) \cdot (-f(y)) = -\int_{y \in B} W(x, y) f(y) \\ &= -\int_{y \in \Omega} W(x, y) f(y) = -T_W(f)(x) = -\lambda f(x) = -\lambda g(x) \ , \end{aligned}$$

where the fact that A is an independent set justifies the second and the fourth equality, and the fact that f is an eigenvalue of T_W is used for the sixth equality. Similar calculations give that for almost every $x \in B$ we have that $T_W(g)(x) = -\lambda g(x)$. We conclude that g is an eigenfunction for eigenvalue $-\lambda$. Hence, T_W has a symmetric spectrum.

Let us now prove the converse direction. Suppose that W is a graphon such that T_W has a symmetric spectrum. In particular, for every $k \in \{3, 5, 7, \ldots\}$ we have that $\sum_{\lambda \in \mathrm{Spec}(T_W)} \lambda^k = 0$. Just as in the graph case, the sum of the k-th powers of the eigenvalues corresponds to the k-cycle density (see [12, Equation (7.22)]), i.e., $t(C_k, W) = 0$. Lemma 4 then tells us that W is bipartite. □

2.5. **Introducing matchings and vertex covers in graphons.** We introduce the notion of matchings in a graphon. Our definitions follow [6], where they were given in the more general context of F-tilings; for completeness, we recall this more general notation in Section 2.5.1.

Definition 9. *Suppose that $W : \Omega^2 \to [0,1]$ is a graphon. We say that a function $\mathfrak{m} \in L^1(\Omega^2)$ is a* matching *in W if*

(M1) $\mathfrak{m} \geq 0$ almost everywhere,
(M2) $\operatorname{supp} \mathfrak{m} \subseteq \operatorname{supp} W$ up to a null-set, and
(M3) for almost every $x \in \Omega$, we have $\int_y \mathfrak{m}(x,y) + \int_y \mathfrak{m}(y,x) \leq 1$.

In [6] it is argued in detail why this is "the right" notion of matchings. We do not repeat this discussion here and only make two comments. Firstly, the requirements in Definition 9 are counterparts to fractional matchings in finite graphs. Namely, a fractional matching in a graph G can be represented as a function $f : V(G)^2 \to \mathbb{R}$ such that

(F1) $f \geq 0$,
(F2) if $f(x,y) > 0$ then $xy \in E(G)$, and
(F3) for every $x \in V(G)$, we have $\sum_y f(x,y) + \sum_y f(y,x) \leq 1$.

Note that usually fractional matchings are represented using symmetric functions. This is however only a notational matter. The current choice for these functions being not-necessarily symmetric is adopted from [6]. (There, this choice was dictated not by matching, but rather by creating a general concept of F-tilings even for graphs which are not vertex-transitive.)

Secondly, note that the actual values of W do not play any role in Defition 9, only the support of W matters. To explain this to a curious reader, we need to assume some familiarity with the regularity method, which is a finite counterpart of the theory of graphons. (Readers who are not familiar with the regularity method or not curious may skip this text.) The fact that the values of W are not relevant (as long as they are positive) reflects the situation in the blow-up lemma [10], which says that one can find a perfect matching in any super-regular pair of any density (as long as it is positive).

Given a matching $\mathfrak{m}$ in a graphon W we define its *size*, $\|\mathfrak{m}\| = \int_x \int_y \mathfrak{m}(x,y)$. The *matching ratio* of W, denoted by $\mathsf{match}(W)$, is defined as the supremum of the sizes of all matchings in W.

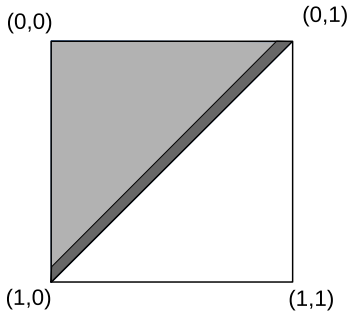


FIGURE 2. The graphon U discussed in the text. A matching $\mathfrak{m}_\varepsilon$ shown in dark gray. The support of U is shown in light gray.

Remark 10. *As said already in the Introduction, even though Definition 9 is inspired by fractional matchings in finite graphs, the resulting graphon concept is referred to as "matchings". This is because in the graphon world every function* $\mathfrak{m}$ *from Definition 9 behaves in many ways as an integral matching. One demonstration of this is Theorem 1 which relates sizes of matchings in finite graphs to the matching ratio of the limit graphon.*

On the other hand, for a graphon representation W_G *of one finite graph* G *there is correspondence between matchings in* W_G *and integral matchings in* G*; see Section 2.5.2. Thus, the previous paragraph applies only to the limit setting.*

We write $\mathrm{MATCH}(W) \subseteq L^1(\Omega^2)$ for the set of all matchings in W. It is straightforward to check that this set is convex (like the set of fractional matchings in a finite graph) and closed (if we consider the norm topology on $L^1(\Omega^2)$). But — unlike the finite case — it need not be compact. To see this consider the graphon $U : [0,1]^2 \to [0,1]$ defined as $U(x,y) = 1$ for $x + y \leq 1$ and $U(x,y) = 0$ for $x + y > 1$. This example was first given in [6] in a somewhat different context. For ε positive, consider a matching $\mathfrak{m}_\varepsilon$ defined to be $1/(2\varepsilon)$ on a stripe of width ε along the diagonal $x+y = 1$ and zero otherwise. This is shown on Figure 2. It is clear that the matchings $\mathfrak{m}_\varepsilon$ do not contain any convergent subsequence, as we let $\varepsilon \to 0_+$. Considering the weak topology on the space $L^1(\Omega^2)$ (that is the topology generated by the dual space $L^\infty(\Omega^2)$) does not help as the same counterexample easily shows. Therefore considering the set $\mathrm{MATCH}(W)$ as a subset of the second dual of $L^1(\Omega^2)$ equipped with its weak* topology seems to be the only reasonable way to have a natural compactification of $\mathrm{MATCH}(W)$. However, we did not need go that far.

We can now proceed with the definition of fractional vertex covers of a graphon. First, recall that a function $c : V(G) \to [0,1]$ is a *fractional vertex cover* of a finite graph G if we have $c(x) + c(y) \geq 1$ for each $xy \in E(G)$. Thus, the graphon counterpart is as follows.

Definition 11. *Suppose that* $W : \Omega^2 \to [0,1]$ *is a graphon. We say that a function* $\mathfrak{c} \in L^\infty(\Omega)$ *is a* fractional vertex cover *of* W *if* $0 \le \mathfrak{c} \le 1$ *almost everywhere and the set*

$$\operatorname{supp} W \setminus \{(x,y) : \mathfrak{c}(x) + \mathfrak{c}(y) \ge 1\}$$

has measure 0.

A fractional vertex cover is called an *integral vertex cover* if its values are from the set $\{0,1\}$ almost everywhere.

Given a fractional vertex cover $\mathfrak{c}$ of a graphon W we define its *size*, $\|\mathfrak{c}\| = \int_x \mathfrak{c}(x)$. The fractional cover number, $\mathsf{fcov}(W)$ is the infimum of sizes of all fractional vertex covers of W.

We write $\mathrm{FCOV}(W) \subseteq L^\infty(\Omega)$ for the set of all fractional vertex covers of W. It is straightforward to check that this set is convex. Further, as was first shown in [6, Theorem 3.14], it is also compact in the space $L^\infty(\Omega)$ equipped with the weak* topology.

2.5.1. *F-tilings.* In this brief section we show, how the concept of matchings is generalized in [6] to general F-tilings. This section is just for a comparison and these more general definitions are not needed in the rest of the paper. Let us recall that given a graph F, an *F-tiling* in a graph G is a collection of (not necessarily induced) vertex-disjoint copies of F in G. Let us now include a corresponding definition for graphons.

Definition 12. *Suppose that* $W : \Omega^2 \to [0,1]$ *is a graphon, and that* F *is a graph on the vertex set* $[k]$*. A function* $\mathfrak{t} \in L^1(\Omega^k)$ *is called an* F-tiling *in* W *if*
(1) $\mathfrak{t} \ge 0$,
(2) $\operatorname{supp} \mathfrak{t} \subseteq \Big\{(x_1,\dots,x_k) \in \Omega^k : \prod_{i,j:ij\in E(F)} W(x_i,x_j) > 0\Big\}$, *and*
(3) for each $\ell \in [k]$ *and for almost every* $x_\ell \in \Omega$*, we have that*

$$\sum_{\ell=1}^{k} \int_{x_1,\dots x_{\ell-1},x_{\ell+1},\dots,x_k} \mathfrak{t}(x_1,\dots,x_k)\, \mathrm{d}\nu^{\otimes k-1} \le 1\,.$$

The size *of an F-tiling* $\mathfrak{t}$ *is* $\|\mathfrak{t}\| = \int \mathfrak{t}(x_1,\dots,x_k)\, \mathrm{d}\nu^k$*. The* F-tiling number *of* W *is the supremum of sizes over all F-tilings in* W*.*

2.5.2. *Matchings and fractional vertex covers of representations of finite graphs.* Suppose that G is a graph, and $W_G : \Omega^2 \to [0,1]$ is a graphon representation of G corresponding to a partition $\bigsqcup_{v\in V(G)} \Omega_v = \Omega$. Given a function $\mathfrak{m} \in L^1(\Omega^2)$, define a function $f_\mathfrak{m} : V(G)^2 \to \mathbb{R}$ by

$$f_\mathfrak{m}(u,v) := v(G) \cdot \int_{\Omega_u}\int_{\Omega_v} \mathfrak{m}\,.$$

It is straightforward to check that if $\mathfrak{m}$ is a matching in W_G in the sense of Definition 9 then $f_\mathfrak{m}$ is a fractional matching in G in the sense of (F1)–(F3). On the other hand, if

a function $f : V(G)^2 \to \mathbb{R}$ is a fractional matching in G, then the function $\mathfrak{m}_f \in L^1(\Omega^2)$ defined by

$$\mathfrak{m}_f(x,y) = v(G) \cdot f(u,v) \quad \text{for } x \in \Omega_u, y \in \Omega_v$$

is a matching in W_G. In fact, it is easy to verify that we have

(20)

$$\mathrm{MATCH}(W_G) = \{\mathfrak{m} \in L^1(\Omega^2) \colon \mathfrak{m} \geq 0 \text{ a.e.}, \int_y \mathfrak{m}(x,y) + \int_y \mathfrak{m}(y,x) \leq 1 \text{ for a.e. } x \in \Omega, \\ f_{\mathfrak{m}} \text{ is a fractional matching in } G\} \,.$$

Given a function $\mathfrak{c} : \Omega \to [0,1]$, define a function $g_{\mathfrak{c}} : V(G) \to [0,1]$ by

$$g_{\mathfrak{c}}(u) := \operatorname{essinf}(\mathfrak{c}_{\restriction \Omega_u}) \,. \tag{21}$$

It is straightforward to check that $\mathfrak{c}$ is a fractional vertex cover in W_G in the sense of Definition 11 if and only if $g_{\mathfrak{c}}$ is a fractional vertex cover in G in the usual sense.

Let us try to reverse this: for a function $g : V(G) \to [0,1]$, let $\mathfrak{c}_g : \Omega \to [0,1]$ be defined by

$$\mathfrak{c}_g(x) := g(u) \quad \text{for each } u \in V(G) \text{ and each } x \in \Omega_u \,. \tag{22}$$

Then one can easily verify that

$$\mathrm{FCOV}(W_G) = \{\mathfrak{c} \in L^\infty(\Omega) \colon \text{ there is a fractional vertex cover } g \text{ of } G \\ \text{such that } \mathfrak{c}_g(x) \leq \mathfrak{c}(x) \leq 1 \text{ for a.e. } x \in \Omega\} \,. \tag{23}$$

3. Extreme points of fractional vertex cover polytons

3.1. **Digest of properties of fractional vertex cover polytopes.** Let G be a finite graph. Then the set $\mathrm{FCOV}(G) \subseteq [0,1]^{V(G)}$ is a polytope. It is a fundamental fact in combinatorial optimization that all the vertices of this polytope are half-integral (i.e., in the form $\{0, \frac{1}{2}, 1\}^{V(G)}$), [17, Theorem 30.2]. Furthermore, all its vertices are integral if and only if G is bipartite, [17, Theorem 18.3].

3.2. **Graphon counterparts.** In view of the Krein–Milman Theorem (see Section 2.1.2) the graphon counterparts to the results described in Section 3.1 will be expressed in terms of $\mathcal{E}(\mathrm{FCOV}(W))$. Let us now state these counterparts. We say that a function $f \in L^\infty(\Omega)$ is *integral* if for almost all $x \in \Omega$ we have $f(x) \in \{0,1\}$. We say that a function $f \in L^\infty(\Omega)$ is *half-integral* if for almost all $x \in \Omega$ we have $f(x) \in \{0, \frac{1}{2}, 1\}$.

The following three theorems (proven later) are the main results of this section.

Theorem 13. *Suppose that $W : \Omega^2 \to [0,1]$ is a graphon. Then all the extreme points of* $\mathrm{FCOV}(W)$ *are half-integral.*

Theorem 14. *Suppose that $W : \Omega^2 \to [0,1]$ is a bipartite graphon. Then all the extreme points of* $\mathrm{FCOV}(W)$ *are integral.*

Theorem 15. *Suppose that $W : \Omega^2 \to [0,1]$ is a graphon. If all the extreme points of* $\mathrm{FCOV}(W)$ *are integral then W is bipartite.*

Observe that every extreme point of the fractional vertex cover polyton of a bipartite graphon is in fact exposed. Indeed, let $\phi \in \mathcal{E}(\mathrm{FCOV}(W))$ for some bipartite graphon $W : \Omega^2 \to [0,1]$. Then Theorem 14 tells us that the sets $A = \phi^{-1}(0)$ and $B = \phi^{-1}(1)$ partition Ω up to a 0-measure set. It is now clear that the linear functional f,

$$f(\psi) := \int_B \psi(x)\mathrm{d}x - \int_A \psi(x)\mathrm{d}x$$

is strictly maximized at ϕ on $\mathrm{FCOV}(W)$. We do not know whether the same can be proved for non-bipartite graphons as well (see Section 5.5).

3.3. **Proof of Theorem 14.** During the proof of Theorem 14, we employ the notation $\partial(z) := \min(z, 1-z)$ for $z \in [0,1]$. We shall need the following easy fact.

Fact 16. *Suppose that $x, y \in [0,1]$ are two reals which satisfy $x+y \geq 1$. Then the numbers $x^+ = x + \partial(x)$ and $y^- = y - \partial(y)$ satisfy $x^+, y^- \in [0,1]$ and $x^+ + y^- \geq 1$.*

Proof. The fact that $x^+, y^- \in [0,1]$ is obvious. To prove that $x^+ + y^- \geq 1$, we distinguish three cases. First, suppose that $0 \leq x \leq \frac{1}{2} \leq y \leq 1$. Then $\partial(x) = x$, $\partial(y) = 1-y$, and consequently, $x^+ + y^- = 2x + 2y - 1 \geq 1$. Second, suppose that $0 \leq y \leq \frac{1}{2} \leq x \leq 1$. Then $\partial(x) = 1-x$, $\partial(y) = y$, and consequently, $x^+ + y^- = 1$. Third, suppose that $\frac{1}{2} \leq x, y \leq 1$. Then $\partial(x) = 1-x$, $\partial(y) = 1-y$, and consequently, $x^+ + y^- = 2y \geq 1$. □

Proof of Theorem 14. Let $\Omega = \Omega_A \sqcup \Omega_B$ be a partition into two sets of positive measure such that W is zero almost everywhere on $(\Omega_A \times \Omega_A) \cup (\Omega_B \times \Omega_B)$. Suppose that $\mathfrak{c} \in \mathrm{FCOV}(W)$ is not integral. Using the notation from Fact 16, define two functions $\mathfrak{c}', \mathfrak{c}'' : \Omega_A \sqcup \Omega_B \to \mathbb{R}$ by $\mathfrak{c}'(a) = \mathfrak{c}(a)^+$, $\mathfrak{c}'(b) = \mathfrak{c}(b)^-$, $\mathfrak{c}''(a) = \mathfrak{c}(a)^-$, $\mathfrak{c}''(b) = \mathfrak{c}(b)^+$, for each $a \in \Omega_A$ and $b \in \Omega_B$. By Fact 16, we have that $\mathfrak{c}', \mathfrak{c}'' \in \mathrm{FCOV}(W)$. Further $\mathfrak{c} = \frac{1}{2}(\mathfrak{c}' + \mathfrak{c}'')$. As $\mathfrak{c}$ is not integral, we have that $\mathfrak{c}$ is distinct from $\mathfrak{c}'$ and $\mathfrak{c}''$. We conclude that $\mathfrak{c}$ is not an extreme point of $\mathrm{FCOV}(W)$. □

3.4. **Proof of Theorem 13.** The proof of Theorem 13 is very similar to that of Theorem 14. We first state the counterpart of Fact 16 we need to this end. We omit the proof as it is almost the same as that of Fact 16. Here, we employ the notation $\partial^\bullet(z) := \min(z, 1-z, |\frac{1}{2} - z|)$ for $z \in [0,1]$.

Fact 17. *Suppose that $x, y \in [0,1]$ are two reals which satisfy $x+y \geq 1$. Then the numbers $x^{+\bullet} = x + \partial^\bullet(x)$ and $y^{-\bullet} = y - \partial^\bullet(y)$ satisfy $x^{+\bullet}, y^{-\bullet} \in [0,1]$ and $x^{+\bullet} + y^{-\bullet} \geq 1$.*

Proof of Theorem 13. Suppose that $\mathfrak{c} \in \mathrm{FCOV}(W)$ is not half-integral. Consider the sets $\Omega_A = \{x \in \Omega : 0 \leq \mathfrak{c}(x) \leq \frac{1}{2}\}$ and $\Omega_B = \{x \in \Omega : \frac{1}{2} < \mathfrak{c}(x) \leq 1\}$. Using the notation from Fact 17, define two functions $\mathfrak{c}', \mathfrak{c}'' : \Omega_A \sqcup \Omega_B \to \mathbb{R}$ by $\mathfrak{c}'(a) = \mathfrak{c}(a)^{+\bullet}$, $\mathfrak{c}'(b) = \mathfrak{c}(b)^{-\bullet}$, $\mathfrak{c}''(a) = \mathfrak{c}(a)^{-\bullet}$, $\mathfrak{c}''(b) = \mathfrak{c}(b)^{+\bullet}$, for each $a \in \Omega_A$ and $b \in \Omega_B$. By Fact 17, we have that $\mathfrak{c}', \mathfrak{c}'' \in \mathrm{FCOV}(W)$. Further $\mathfrak{c} = \frac{1}{2}(\mathfrak{c}' + \mathfrak{c}'')$. As $\mathfrak{c}$ is not half-integral, we have that $\mathfrak{c}$ is distinct from $\mathfrak{c}'$ and $\mathfrak{c}''$. We conclude that $\mathfrak{c}$ is not an extreme point of $\mathrm{FCOV}(W)$. □

3.5. **Proof of Theorem 15.**

Proof of Theorem 15. We shall prove the counterpositive. Suppose that W is not bipartite. Let $\mathrm{COV}(W)$ be the closure (in the weak* topology) of the convex hull of all integral vertex covers of W. Clearly, we have $\mathrm{COV}(W) \subseteq \mathrm{FCOV}(W)$, and each integral vertex cover of $\mathrm{FCOV}(W)$ is contained in $\mathrm{COV}(W)$. Below, we shall show that

$$\mathrm{FCOV}(W) \setminus \mathrm{COV}(W) \neq \emptyset \,. \tag{24}$$

The Krein–Milman Theorem (see Section 2.1.2) then tells us that $\mathcal{E}\big(\mathrm{FCOV}(W)\big) \setminus \mathrm{COV}(W) \neq \emptyset$. It will thus follow that there exists a non-integral fractional vertex cover in $\mathcal{E}\big(\mathrm{FCOV}(W)\big)$, as was needed to show.

Take $\mathfrak{c} : \Omega \to [0,1]$ to be constant $\frac{1}{2}$. Clearly, $\mathfrak{c} \in \mathrm{FCOV}(W)$. In order to show (24), it suffices to prove that $\mathfrak{c} \notin \mathrm{COV}(W)$. Let k be the odd integer given by Lemma 7. Let $\varepsilon = \frac{1}{32k^2}$, and let the sets $A_1, \ldots, A_k$ of measure $\alpha > 0$ be given by Lemma 7.

In order to prove that $\mathfrak{c}$ is not in the weak* closure of the convex hull of integral vertex covers, consider an arbitrary ℓ-tuple of integral vertex covers $\mathfrak{c}_1, \ldots, \mathfrak{c}_\ell$ of W, and numbers $\gamma_1, \ldots, \gamma_\ell \geq 0$ with $\sum \gamma_i = 1$.

Consider an arbitrary $i \in [\ell]$. We say that $\mathfrak{c}_i$ *marks the set* A_h (where $h \in [k]$) if $\mathfrak{c}_i$ restricted to A_h attains the value 0 on a set of measure at most $\frac{\alpha}{4k}$. Put equivalently, $\mathfrak{c}_i$ marks the set A_h if $\mathfrak{c}_i$ restricted to A_h attains the value 1 on a set of measure at least $\frac{\alpha(4k-1)}{4k}$.

Claim 1. *For each $i \in [\ell]$, the vertex cover $\mathfrak{c}_i$ marks at least $\frac{k+1}{2}$ many of the sets $A_1, \ldots, A_k$.*

Proof. Suppose that this is not the case. Recall that k is odd. We can find an index $h \in [k]$ that A_h and A_{h+1} are not marked (again, using the cyclic notation $A_{k+1} = A_k$). Therefore, the $\mathfrak{c}_i$-preimages $B_h \subseteq A_h$ and $B_{h+1} \subseteq A_{h+1}$ of 0 have both measures more than $\frac{\alpha}{4k}$. It follows from Lemma 7 and the way we set ε that W is positive on a set of positive measure on $B_h \times B_{h+1}$. This contradicts the fact that $\mathfrak{c}_i$ is a vertex cover. □

Let us write $A = \bigcup_h A_h$. It follows from Claim 1 that

$$\int_A \mathfrak{c}_i \geq \frac{k+1}{2} \cdot \frac{\alpha(4k-1)}{4k} \geq \alpha \left(\frac{k}{2} + \frac{1}{6} \right) \,. \tag{25}$$

By convexity, we can replace $\int_A \mathfrak{c}_i$ by $\int_A (\sum_i \gamma_i \mathfrak{c}_i)$ in (25).

We now have

$$\int_A \left(\sum_i \gamma_i \mathfrak{c}_i(x) - \mathfrak{c}(x) \right) \mathrm{d}x = \int_A \left(\sum_i \gamma_i \mathfrak{c}_i(x) \right) \mathrm{d}x - \int_A \mathfrak{c}(x)\,\mathrm{d}x \tag{26}$$
$$\overset{(25)}{\geq} \alpha \left(\frac{k}{2} + \frac{1}{6} \right) - k\alpha \cdot \frac{1}{2} = \frac{\alpha}{6}\,.$$

Since neither the set A nor the bound on the right-hand side of (26) depend on the choice of the number ℓ, the vertex covers $\mathfrak{c}_i$, and the constants γ_i, we get that $\mathfrak{c}$ is not in the weak* closure of convex combinations of integral vertex covers, as was needed. □

3.6. **An application: the Erdős–Gallai Theorem.** In this section, we prove a graphon counterpart to the following classical result of Erdős and Gallai, [5].

Theorem 18 (Erdős–Gallai, 1959)**.** *Suppose that n and ℓ are positive integers that satisfy $\ell \leq n/2$. Then any n-vertex graph with more than $\max\{(\ell-1)(n-\ell+1) + \binom{\ell-1}{2}, \binom{2\ell-1}{2}\}$ edges contains a matching with at least ℓ edges.*

The bound in Theorem 18 is optimal. Indeed, when $\ell \leq 0.4(n+1)$, the *extremal graph* (denoted by $\mathrm{ExG}(n,\ell)$) is the complete graph $K_{\ell-1,n-\ell+1}$ together with the complete graph inserted into the $(\ell-1)$-part. When $\ell \geq 0.4(n+1)$, the extremal graph is the complete graph of order $2\ell-1$ with $n+1-2\ell$ isolated vertices padded. That is, the extremal graph undergoes a transition at edge density (asymptotically) 0.64.

Motivated by this, for $e \in [0,1]$ we define graphons Ψ_e and Φ_e as follows. We partition $\Omega = B_1 \sqcup B_2$ so that $\nu(B_1) = 1 - \sqrt{1-e}$ and $\nu(B_2) = \sqrt{1-e}$. We define Ψ_e to be constant 0 on $B_2 \times B_2$ and 1 elsewhere. We partition $\Omega = C_1 \sqcup C_2$ so that $\nu(C_1) = e^2$ and $\nu(C_2) = 1 - e^2$. We define Φ_e to be constant 1 on $C_1 \times C_1$ and 0 elsewhere. These definition uniquely determine Ψ_e and Φ_e, up to isomorphism. Thus, our graphon version of the Erdős–Gallai theorem reads as follows.

Theorem 19. *Suppose that $W : \Omega^2 \to [0,1]$ is a graphon. Let $e = \int_{x,y} W(x,y)$. Then $\mathsf{fcov}(W) \geq \min\left\{\sqrt{\frac{1}{4}e}, 1 - \sqrt{1-e}\right\}$. (The maximum is attained by the former term for $e \leq 0.64$ and by the latter term for $e \geq 0.64$).*

Furthermore, we have an equality if and only if W is isomorphic to Ψ_e (if $e \leq 0.64$) or to Φ_e (if $e \geq 0.64$).

This version of the Erdős–Gallai Theorem implies an asymptotic version of the finite statement. Furthermore, it provides a corresponding stability statement.

Theorem 20. *For every $\varepsilon > 0$ there exists numbers $n_0 \in \mathbb{N}$ and $\delta > 0$ such that the following holds. If G is a graph on $n > n_0$ vertices with more than $\max\{(\ell-1)(n-\ell+1)+$*

$\binom{\ell-1}{2}, \binom{2\ell-1}{2}\}$ *edges, then* G *contains a matching with at least* $\ell-\varepsilon n$ *edges. Furthermore,* G *contains a matching with at least* $\ell+\delta n$ *edges, unless* G *is* εn^2*-close to the graph* $\mathrm{ExG}(n,\ell)$ *as above in the edit distance.*

The way of deriving Theorem 20 from Theorem 19 is standard, and we refer the reader to [7] where this was done in detail in the context of a tiling theorem of Komlós, [9], which is a statement of a similar flavor.

Let us emphasize that the original proof of Theorem 18 is simple and elementary (and the corresponding stability statement would not be difficult to prove with the same approach either). While our proof is not long, it makes use of the heavy machinery of graph limits, and in particular the results from [6] and from Section 3.2. We think that our proof offers an interesting alternative point of view on the problem.

For the proof of Theorem 19 we shall need the following fact.

Fact 21. *Suppose that* $D \in [0, \frac{1}{2}]$ *is fixed. Then the maximum of the function* $g(a,b) = a^2 + 2b - b^2$ *on the set* $\{(a,b) : a \geq 0, b \geq 0, \frac{a}{2} + b = D\}$ *is attained for* $a = 0, b = D$ *if* $D \leq 0.4$ *and* $a = 2D, b = 0$ *if* $D \geq 0.4$.

Proof. We transform this into an optimization problem in one variable by considering the function $h(b) = g(2(D-b), b)$. The function h is quadratic with limit plus infinity at $-\infty$ and at $+\infty$. Thus, the maximum of h on the interval $[0, D]$ will be either at $b = 0$ or at $b = D$. We have $h(0) = 4D^2$ and $h(D) = 2D - D^2$. A quick calculation gives that the latter is bigger for $D < 0.4$ while the the former is bigger for $D > 0.4$. □

Proof of Theorem 19. By Theorem 13 we can fix a half-integral vertex cover of size $\mathsf{fcov}(W)$. Then we get a partition $\Omega = A_0 \sqcup A_{1/2} \sqcup A_1$ given by the preimages of 0, 1/2 and 1. Let us write $\alpha_0 = \nu(A_0)$, $\alpha_{1/2} = \nu(A_{1/2})$, and $\alpha_1 = \nu(A_1)$. We have $\mathsf{fcov}(W) = \frac{1}{2}\nu(A_{1/2}) + \nu(A_1)$. Note that $W_{\restriction A_0 \times (A_0 \cup A_{1/2})} = 0$. Therefore,

$$\begin{aligned} e \leq \alpha_{1/2}^2 + \alpha_1(\alpha_1 + 2\alpha_{1/2} + 2\alpha_0) &= \alpha_{1/2}^2 + 2\alpha_1 - \alpha_1^2 \\ &\leq \max\left\{4\mathsf{fcov}(W)^2, 2\mathsf{fcov}(W) - \mathsf{fcov}(W)^2\right\} , \end{aligned} \tag{27}$$

where the last part follows from Fact 21. Pedestrian calculations show that this is equivalent to the assertion of the theorem.

We now look at the furthermore part. If $e = \max\{4\mathsf{fcov}(W)^2, 2\mathsf{fcov}(W) - \mathsf{fcov}(W)^2\}$, then the first inequality in (27) is at equality. This means that W must be 1 on $A_{1/2} \times A_{1/2}$ and on $A_1 \times \Omega$. Furthermore, Fact 21 tells us that for the second inequality in (27) to be at equality, we must have $\alpha_{1/2} = 0$ or $\alpha_1 = 1$. We conclude that W is isomorphic to Ψ_e or to Φ_e. □

4. Convergence of polytons

4.1. **Fractional vertex cover polytons of a convergent graphon sequence.** Suppose that a sequence of graphons $(W_n)_n$ converges to a graphon W in the cut-norm. We want to relate the polytons $\mathrm{FCOV}(W_n)$ to the polyton $\mathrm{FCOV}(W)$. First observe that in general, the polytons $\mathrm{FCOV}(W_n)$ do not converge to $\mathrm{FCOV}(W)$ in any reasonable sense. Indeed, for example, take W_n to be a representation of a sample of the Erdős–Rényi random graph $\mathbb{G}(2n, 1/\sqrt{n})$. It is well-known that almost surely almost all these graphs contain a perfect matching.[5] Thus, $\mathrm{FCOV}(W_n)$ contain only fractional vertex covers of size $\frac{1}{2}$ and more for almost all n. On the other hand, almost surely, the zero graphon $W = 0$ is the cut-distance limit of $(W_n)_n$, and so $\mathrm{FCOV}(W)$ consists of all $[0,1]$-valued measurable functions on Ω.

However, Theorem 22 below shows that "$\mathrm{FCOV}(W)$ asymptotically contains the polytons $\mathrm{FCOV}(W_n)$". This theorem is a special case of [6, Theorem 3.14].

Theorem 22. *Suppose that $(W_n)_n$ is a sequence of graphons on Ω that converges to a graphon $W : \Omega^2 \to [0,1]$ in the cut-norm. Suppose that $\mathfrak{c}_n \in \mathrm{FCOV}(W_n)$. Then any accumulation point of the sequence $(\mathfrak{c}_n)_n$ in the weak* topology lies in $\mathrm{FCOV}(W)$.*

4.2. **Matching polytons of a convergent graphon sequence.** The main new result of this section concerns convergence properties of the matching polytons. This result is dual to Theorem 22: if W_n converges to W in the cut-norm then "$\mathrm{MATCH}(W_n)$ asymptotically contain $\mathrm{MATCH}(W)$".

Theorem 23. *Suppose that $W : \Omega^2 \to [0,1]$ is a graphon on a probability space Ω, and let $\mathfrak{m} \in \mathrm{MATCH}(W)$ be fixed. Then for every $\varepsilon > 0$ there is $\delta > 0$ such that whenever $U \colon \Omega^2 \to [0,1]$ is a graphon with $\|U - W\|_\square < \delta$ then there is $\mathfrak{m}_U \in \mathrm{MATCH}(U)$ such that $\|\mathfrak{m}_U - \mathfrak{m}\|_\square < \varepsilon$.*

Since the cut-norm topology is stronger than the weak* topology, we get the following corollary.

Corollary 24. *Suppose that $(W_n)_n$ is a sequence of graphons on a probability space Ω that converges to a graphon $W : \Omega^2 \to [0,1]$ in the cut-norm. Suppose that $\mathfrak{m} \in \mathrm{MATCH}(W)$. Then there exists a sequence $\mathfrak{m}_n \in \mathrm{MATCH}(W_n)$ such that $(\mathfrak{m}_n)_n$ converges to $\mathfrak{m}$ in the cut-norm. In particular, the sequence $(\mathfrak{m}_n)_n$ converges to $\mathfrak{m}$ in the weak* topology.*

[5] By saying that "almost all these graphs contain a perfect matching" we mean that with probability 1 (in the product probability space $\prod_n \mathbb{G}(2n, 1/\sqrt{n})$), after sampling the graphs $G_{2n} \sim \mathbb{G}(2n, 1/\sqrt{n})$, we can find a finite set $S \subseteq \mathbb{N}$ such that for each $n \in \mathbb{N} \setminus S$, the graph G_{2n} contains a perfect matching. We provide a proof of this in Appendix B.

Recall that each function $F \in L^1(\Omega^2)$ can be approximated by a step-function up to an arbitrarily small error in the L^1-norm. Further, the approximating function can be chosen to equal the average on each step. In the proof of Theorem 23, we will need the following extension of this fact: firstly, we want to approximate two functions $F_1, F_2 \in L^1(\Omega^2)$ using the same steps, and secondly, we want the steps to be squares with sides of the same measure.

Lemma 25. *Let (Ω, ν) be a probability space. Then for every pair $(F_1, F_2) \in L^1(\Omega^2) \times L^1(\Omega^2)$ and every $\varepsilon > 0$ there exists a number $k \in \mathbb{N}$ and a partition of Ω into k pairwise disjoint subsets $\Omega_1, \dots, \Omega_k$, each of measure $\frac{1}{k}$, such that*

$$\sum_{i,j=1}^{k} \int\limits_{\Omega_i \times \Omega_j} |F_1 - F_1^{ij}| < \varepsilon \qquad \text{and} \qquad \sum_{i,j=1}^{k} \int\limits_{\Omega_i \times \Omega_j} |F_2 - F_2^{ij}| < \varepsilon \,, \tag{28}$$

where

$$F_1^{ij} = \frac{1}{\nu(\Omega_i)\nu(\Omega_j)} \int\limits_{\Omega_i \times \Omega_j} F_1 \qquad \text{and} \qquad F_2^{ij} = \frac{1}{\nu(\Omega_i)\nu(\Omega_j)} \int\limits_{\Omega_i \times \Omega_j} F_2 \,, \quad i, j = 1, \dots, n \,.$$

The proof of Lemma 25 is given in Appendix.

4.3. **Proof of Theorem 23.** Let $\varepsilon > 0$ be fixed. By basic properties of L^1-functions there is $M > 0$ (which we fix now) such that if we define

$$\tilde{\mathfrak{m}}(x, y) = \min\{\mathfrak{m}(x, y), M\}$$

then we have

$$\|\tilde{\mathfrak{m}} - \mathfrak{m}\|_{L^1(\Omega^2)} < \frac{1}{2}\varepsilon \,. \tag{29}$$

Moreover, it is obvious that such defined function $\tilde{\mathfrak{m}}$ is still a matching in the graphon W. We fix $\tilde{\varepsilon} > 0$ such that

$$3\tilde{\varepsilon} + 6\sqrt{\tilde{\varepsilon}}M + 2\tilde{\varepsilon}^{\frac{3}{2}} < \frac{1}{2}\varepsilon \,. \tag{30}$$

Claim 2. *There is $r > 0$ such that whenever $\Theta \subseteq \Omega^2$ is of positive measure such that $\frac{1}{\nu^{\otimes 2}(\Theta)} \int\limits_{\Theta} W < r$ then*

$$\nu^{\otimes 2}\left(\operatorname{supp}(W) \cap \Theta\right) < \frac{1}{2M}\tilde{\varepsilon} \,.$$

Proof. By basic properties of measurable functions, there is $s > 0$ such that

$$\nu^{\otimes 2}\{(x, y) \in \Omega^2 \colon 0 < W(x, y) < s\} < \frac{1}{4M}\tilde{\varepsilon} \,. \tag{31}$$

We will prove that $r = \frac{1}{4M}\tilde{\varepsilon}s$ works. Suppose for a contradiction that there is $\Theta \subseteq \Omega^2$ of positive measure with $\frac{1}{\nu^{\otimes 2}(\Theta)} \int\limits_{\Theta} W < r$ such that

$$\nu^{\otimes 2}\left(\operatorname{supp}(W) \cap \Theta\right) \geq \frac{1}{2M}\tilde{\varepsilon}\,. \tag{32}$$

Then we have

$$\begin{aligned} r &> \frac{1}{\nu^{\otimes 2}(\Theta)} \int\limits_{\Theta} W \geq \frac{1}{\nu^{\otimes 2}(\Theta)} \nu^{\otimes 2}\left(\{(x,y) \in \Theta \colon W(x,y) \geq s\}\right) \cdot s \\ &\overset{(31)}{>} \frac{1}{\nu^{\otimes 2}(\Theta)} \left(\nu^{\otimes 2}\left(\operatorname{supp}(W) \cap \Theta\right) - \frac{1}{4M}\tilde{\varepsilon}\right) s \overset{(32)}{\geq} \frac{1}{\nu^{\otimes 2}(\Theta)} \cdot \frac{1}{4M}\tilde{\varepsilon}s \geq \frac{1}{4M}\tilde{\varepsilon}s\,, \end{aligned}$$

which is the desired contradiction with the definition of r. □

Now we fix $r > 0$ from Claim 2, and we set

$$\eta = \frac{1}{2}\tilde{\varepsilon} \cdot \frac{1}{1 + 2M + 2\frac{M}{r}}\,. \tag{33}$$

By Lemma 25 there is a natural number k and a partition of Ω into pairwise disjoint subsets $\Omega_1, \ldots, \Omega_k$, each of measure $\frac{1}{k}$, such that

$$\sum_{i,j=1}^{k} \int\limits_{\Omega_i \times \Omega_j} |W - W^{ij}| < \eta^2 \qquad \text{and} \qquad \sum_{i,j=1}^{k} \int\limits_{\Omega_i \times \Omega_j} |\tilde{\mathfrak{m}} - m^{ij}| < \eta^2\,, \tag{34}$$

where

$$W^{ij} = \frac{1}{\nu(\Omega_i)\nu(\Omega_j)} \int\limits_{\Omega_i \times \Omega_j} W \qquad \text{and} \qquad m^{ij} = \frac{1}{\nu(\Omega_i)\nu(\Omega_j)} \int\limits_{\Omega_i \times \Omega_j} \tilde{\mathfrak{m}}\,, \qquad i,j = 1 \ldots, n\,.$$

The first inequality from (34) easily implies that for all but at most ηk^2 pairs (i,j) we have

$$\int\limits_{\Omega_i \times \Omega_j} \left|W - W^{ij}\right| \leq \frac{1}{k^2}\eta\,. \tag{35}$$

Similarly, the second inequality from (34) implies that for all but at most ηk^2 pairs (i,j) we have

$$\int\limits_{\Omega_i \times \Omega_j} \left|\mathfrak{m} - m^{ij}\right| \leq \frac{1}{k^2}\eta\,. \tag{36}$$

We are now in a position, when we can find $\delta > 0$ with the properties required by Theorem 23. Set

$$\delta = \frac{1}{k^2}\eta\,. \tag{37}$$

So, let $U\colon \Omega^2 \to [0,1]$ be a graphon such that $\|U - W\|_\square < \delta$. Let $A \subseteq [k] \times [k]$ denote the set of all those pairs (i,j) for which either (35) or (36) fails. We have that $|A| \le 2\eta k^2$. We define

$$t(x,y) = \begin{cases} \dfrac{m^{ij}}{W^{ij}} \cdot U(x,y) & \text{if } (i,j) \notin A \text{ and } W^{ij} \ge r, \\ 0 & \text{otherwise,} \end{cases} \qquad (x,y) \in \Omega_i \times \Omega_j, \qquad i,j = 1,\dots,k\,.$$

Claim 3. *We have that* $\|t - \tilde{\mathfrak{m}}\|_\square \le \tilde{\varepsilon}$.

Proof. We need to show that for every measurable sets $S, T \subseteq \Omega$ it holds $\left| \int\limits_{S\times T} (t - \tilde{\mathfrak{m}}) \right| \le \tilde{\varepsilon}$. So let us fix the sets $S, T \subseteq \Omega$. Let Θ_A denote the union of all the sets $\Omega_i \times \Omega_j$ for which $(i,j) \in A$. Similarly, let Θ_B denote the union of all the sets $\Omega_i \times \Omega_j$ for which $(i,j) \notin A$ and $W^{ij} < r$, and let Θ_C denote the union of all the sets $\Omega_i \times \Omega_j$ for which $(i,j) \notin A$ and $W^{ij} \ge r$.

The bulk of the work is in proving the following three subclaims.

Subclaim 1. We have

$$\left| \int\limits_{(S\times T)\cap\Theta_A} (t - \tilde{\mathfrak{m}}) \right| \le 2\eta M\,.$$

Subclaim 2. We have

$$\left| \int\limits_{(S\times T)\cap\Theta_B} (t - \tilde{\mathfrak{m}}) \right| < \frac{1}{2}\tilde{\varepsilon}\,.$$

Subclaim 3. We have

$$\left| \int\limits_{(S\times T)\cap\Theta_C} (t - \tilde{\mathfrak{m}}) \right| < \eta\left(1 + 2\frac{M}{r}\right)\,.$$

Indeed, Subclaims 1–3 complete the proof of Claim 3 as then

$$\left| \int\limits_{S\times T} (t - \tilde{\mathfrak{m}}) \right| < \eta\left(1 + 2M + 2\frac{M}{r}\right) + \frac{1}{2}\tilde{\varepsilon} \overset{(33)}{\le} \tilde{\varepsilon}\,.$$

Proof of Subclaim 1. Recall that by the definition of the set A, it holds $\nu^{\otimes 2}(\Theta_A) \le 2\eta$, and so we have

$$\left| \int\limits_{(S\times T)\cap\Theta_A} (t - \tilde{\mathfrak{m}}) \right| = \int\limits_{(S\times T)\cap\Theta_A} \tilde{\mathfrak{m}} \le \int\limits_{\Theta_A} \tilde{\mathfrak{m}} \le \nu^{\otimes 2}(\Theta_A) M \le 2\eta M\,.$$

□

Proof of Subclaim 2. If $\nu^{\otimes 2}(\Theta_B) = 0$ then trivially

$$\left| \int\limits_{(S\times T)\cap\Theta_B} (t - \tilde{\mathfrak{m}}) \right| = 0 \, .$$

So suppose that Θ_B is of positive measure. Note that then it clearly holds $\frac{1}{\nu^{\otimes 2}(\Theta_B)} \int\limits_{\Theta_B} W < r$, and so we have

$$\begin{aligned} \left| \int\limits_{(S\times T)\cap\Theta_B} (t - \tilde{\mathfrak{m}}) \right| &= \int\limits_{(S\times T)\cap\Theta_B} \tilde{\mathfrak{m}} \leq \nu^{\otimes 2}(\operatorname{supp}(\tilde{\mathfrak{m}}) \cap \Theta_B) \cdot M \\ &\leq \nu^{\otimes 2}(\operatorname{supp}(W) \cap \Theta_B) \cdot M \overset{\text{Claim 2}}{<} \frac{1}{2}\tilde{\varepsilon} \, . \end{aligned}$$

□

Proof of Sublclaim 3. It is enough to show that whenever a pair $(i, j) \notin A$ is such that $W^{ij} \geq r$ then

$$\left| \int\limits_{(S\times T)\cap(\Omega_i\times\Omega_j)} (t - \tilde{\mathfrak{m}}) \right| < \frac{1}{k^2}\eta\left(1 + 2\frac{M}{r}\right) \, .$$

So let us fix such a pair (i,j). Then we have

$$\begin{aligned}
&\left|\int\limits_{(S\times T)\cap(\Omega_i\times\Omega_j)} (t-\tilde{\mathfrak{m}})\right| \\
\leq&\left|\int\limits_{(S\times T)\cap(\Omega_i\times\Omega_j)} \left(\frac{m^{ij}}{W^{ij}}U - m^{ij}\right)\right| + \left|\int\limits_{(S\times T)\cap(\Omega_i\times\Omega_j)} \left(m^{ij}-\tilde{\mathfrak{m}}\right)\right| \\
\leq&\frac{m^{ij}}{W^{ij}}\left|\int\limits_{(S\times T)\cap(\Omega_i\times\Omega_j)} \left(U-W^{ij}\right)\right| + \int\limits_{(S\times T)\cap(\Omega_i\times\Omega_j)} |m^{ij}-\tilde{\mathfrak{m}}| \\
\leq&\frac{M}{r}\left|\int\limits_{(S\times T)\cap(\Omega_i\times\Omega_j)} \left((U-W)+(W-W^{ij})\right)\right| + \int\limits_{(S\times T)\cap(\Omega_i\times\Omega_j)} |m^{ij}-\tilde{\mathfrak{m}}| \\
\leq&\frac{M}{r}\left|\int\limits_{(S\times T)\cap(\Omega_i\times\Omega_j)} (U-W)\right| + \frac{M}{r}\int\limits_{(S\times T)\cap(\Omega_i\times\Omega_j)} |W-W^{ij}| \\
&+ \int\limits_{(S\times T)\cap(\Omega_i\times\Omega_j)} |m^{ij}-\tilde{\mathfrak{m}}| \\
\boxed{(i,j)\notin A}\quad \leq&\frac{M}{r}\left|\int\limits_{(S\cap\Omega_i)\times(T\cap\Omega_j)} (U-W)\right| + \frac{M}{r}\cdot\frac{1}{k^2}\eta + \frac{1}{k^2}\eta \\
\boxed{\text{by (37) and hypothesis}}\quad <&\frac{M}{r}\cdot\frac{1}{k^2}\eta + \frac{M}{r}\cdot\frac{1}{k^2}\eta + \frac{1}{k^2}\eta \\
=&\frac{1}{k^2}\eta\left(1+2\frac{M}{r}\right)\ .
\end{aligned}$$

□

□

Let B_1 denote the set of all those $x\in\Omega$ for which

$$\int\limits_{y\in\Omega} t(x,y) > \int\limits_{y\in\Omega} \tilde{\mathfrak{m}}(x,y) + \sqrt{\tilde{\varepsilon}}\ .$$

Similarly, let B_2 denote the set of all those $x \in \Omega$ for which

$$\int_{y\in\Omega} t(y,x) > \int_{y\in\Omega} \tilde{\mathfrak{m}}(y,x) + \sqrt{\tilde{\varepsilon}}\,.$$

Then we have

$$\tilde{\varepsilon} \overset{\text{Claim 3}}{\geq} \|t - \tilde{\mathfrak{m}}\|_{\square} \geq \int_{B_1\times\Omega} (t - \tilde{\mathfrak{m}}) > \nu(B_1)\sqrt{\tilde{\varepsilon}}\,,$$

and consequently $\nu(B_1) < \sqrt{\tilde{\varepsilon}}$. In the same way, we conclude that $\nu(B_2) < \sqrt{\tilde{\varepsilon}}$. Now we are ready to define $\mathfrak{m}_U$ by setting

$$\mathfrak{m}_U(x,y) = \begin{cases} 0 & \text{if } x \in B_1 \cup B_2 \text{ or } y \in B_1 \cup B_2, \\ \frac{1}{1+2\sqrt{\tilde{\varepsilon}}} t(x,y) & \text{otherwise,} \end{cases} \qquad (x,y) \in \Omega^2\,. \tag{38}$$

Claim 4. *We have that* $\|\mathfrak{m}_U - t\|_{L^1(\Omega^2)} < 6\sqrt{\tilde{\varepsilon}}M + 2\tilde{\varepsilon} + 2\tilde{\varepsilon}^{\frac{3}{2}}$.

Proof. We set $B = ((B_1 \cup B_2)\times\Omega) \cup (\Omega \times (B_1 \cup B_2))$. Then we have

$$\begin{aligned}
&\|\mathfrak{m}_U - t\|_{L^1(\Omega^2)} = \int_B |\mathfrak{m}_U - t| + \int_{\Omega^2\setminus B} |\mathfrak{m}_U - t| \\
=&\int_B t + \left(1 - \frac{1}{1+2\sqrt{\tilde{\varepsilon}}}\right) \int_{\Omega^2\setminus B} t \\
\leq& \int_{(B_1\cup B_2)\times\Omega} t + \int_{\Omega\times(B_1\cup B_2)} t + \left(1 - \frac{1}{1+2\sqrt{\tilde{\varepsilon}}}\right) \int_{(\Omega\setminus(B_1\cup B_2))^2} t \\
\boxed{\text{Claim 3}} \quad \leq& \int_{(B_1\cup B_2)\times\Omega} \tilde{\mathfrak{m}} + \tilde{\varepsilon} + \int_{\Omega\times(B_1\cup B_2)} \tilde{\mathfrak{m}} + \tilde{\varepsilon} + \left(1 - \frac{1}{1+2\sqrt{\tilde{\varepsilon}}}\right) \left(\int_{(\Omega\setminus(B_1\cup B_2))^2} \tilde{\mathfrak{m}} + \tilde{\varepsilon} \right) \\
\leq& \nu^{\otimes 2}((B_1\cup B_2)\times\Omega)\cdot M + \nu^{\otimes 2}(\Omega\times(B_1\cup B_2))\cdot M + 2\tilde{\varepsilon} + \frac{2\sqrt{\tilde{\varepsilon}}}{1+2\sqrt{\tilde{\varepsilon}}}(M+\tilde{\varepsilon}) \\
<& 4\sqrt{\tilde{\varepsilon}}M + 2\tilde{\varepsilon} + 2\sqrt{\tilde{\varepsilon}}(M+\tilde{\varepsilon}) = 6\sqrt{\tilde{\varepsilon}}M + 2\tilde{\varepsilon} + 2\tilde{\varepsilon}^{\frac{3}{2}}\,.
\end{aligned}$$

□

Now we have

$$\begin{aligned}
\|\mathfrak{m}_U - \mathfrak{m}\|_\square &\le \|\mathfrak{m}_U - t\|_\square + \|t - \tilde{\mathfrak{m}}\|_\square + \|\tilde{\mathfrak{m}} - \mathfrak{m}\|_\square \\
&\le \|\mathfrak{m}_U - t\|_{L^1(\Omega^2)} + \|t - \tilde{\mathfrak{m}}\|_\square + \|\tilde{\mathfrak{m}} - \mathfrak{m}\|_{L^1(\Omega^2)} \\
\boxed{\text{by (29), Claims 3 and 4}} \quad &< \frac{1}{2}\varepsilon + 3\tilde{\varepsilon} + 6\sqrt{\tilde{\varepsilon}}M + 2\tilde{\varepsilon}^{\frac{3}{2}} \\
\boxed{(30)} \quad &< \varepsilon \ .
\end{aligned}$$

So it remains to show that $\mathfrak{m}_U$ is a matching in the graphon U.

The fact that $\mathfrak{m}_U$ is a nonnegative function from $L^1(\Omega^2)$ is obvious, and we also have $\operatorname{supp}(\mathfrak{m}_U) \subseteq \operatorname{supp}(t) \subseteq \operatorname{supp}(U)$. So we only need to show that for almost every $x \in \Omega$ it holds

$$\int_{y\in\Omega} \mathfrak{m}_U(x,y) + \int_{y\in\Omega} \mathfrak{m}_U(y,x) \le 1 \ . \tag{39}$$

This is trivially satisfied for every $x \in B_1 \cup B_2$ as then the left-hand side of (39) equals 0. So let us fix $x \in \Omega \setminus (B_1 \cup B_2)$. We may assume that

$$\int_{y\in\Omega} \tilde{\mathfrak{m}}(x,y) + \int_{y\in\Omega} \tilde{\mathfrak{m}}(y,x) \le 1 \ , \tag{40}$$

as $\tilde{\mathfrak{m}}$ is a matching (in the graphon W). Then it holds

$$\begin{aligned}
\int_{y\in\Omega} \mathfrak{m}_U(x,y) + \int_{y\in\Omega} \mathfrak{m}_U(y,x) &\overset{(38)}{\le} \frac{1}{1+2\sqrt{\tilde{\varepsilon}}} \left(\int_{y\in\Omega} t(x,y) + \int_{y\in\Omega} t(y,x) \right) \\
\boxed{x \notin B_1 \cup B_2} \quad &\le \frac{1}{1+2\sqrt{\tilde{\varepsilon}}} \left(\int_{y\in\Omega} \tilde{\mathfrak{m}}(x,y) + \int_{y\in\Omega} \tilde{\mathfrak{m}}(y,x) + 2\sqrt{\tilde{\varepsilon}} \right) \overset{(40)}{\le} 1 \ ,
\end{aligned}$$

which completes the proof of Theorem 23.

5. Concluding remarks

5.1. **Approximating** MATCH(W) **and** FCOV(W) **using** W**-random graphs.** In Section 4, we showed that if $(W_n)_n$ converge to W in the cut-norm then, in a certain sense, MATCH(W_n) asymptotically contain MATCH(W), and FCOV(W_n) are asymptotically contained in FCOV(W). We also showed that in general, these inclusions may be proper. However, we believe that if we take W_n as a representation of $\mathbb{G}(n, W)$ then with probability one both these inclusions are asymptotically at equality.

5.2. **Bipartiteness from the matching polyton.** Theorems 14 and 15 characterize bipartiteness of a graphon in terms of its fractional vertex cover polyton. For finite graphs there is another characterization in terms of the matching polytope: a graph is bipartite if and only if MATCH(G) is integral. Recall that there seems to be no counterpart to the concept of integrality of a graphon matching (c.f. Remark 10). So, we leave as an important question to provide a characterization of bipartiteness in terms of MATCH(W).

5.3. **Perfect matching polyton.** Many variants of the above polytopes are considered in combinatorial optimization. As an example, let us mention the *perfect matching polytope* PerfMATCH(G) and the *fractional perfect matching polytope* FPerfMATCH(G) of a graph G. The corresponding graphon polyton is

$$\mathrm{PerfMATCH}(W) = \{\mathfrak{m} \in \mathrm{MATCH}(W) : \|\mathfrak{m}\| = 1\} \ .$$

(Again, we cannot distinguish between the integral and fractional version, cf. the discussion in Section 2.5.) It might be interesting to study this, and similar polytons. That said, let us emphasize that many basic results, like Edmonds' perfect matching polytope theorem, seem not to have a graphon counterpart as they concern integrality-related properties of the polytope.

5.4. **Generalizing the results to F-tilings.** Results in Section 3 are specific to matchings — even in the finite setting. Even though we have not worked out the details, we believe that our second main result, Theorem 23, extends to general F-tilings as introduced in [6] (and so does its proof).

5.5. **Extreme points of fractional vertex cover polytons.** As we explained in Section 3.2, every extreme point of the fractional vertex cover polyton of a bipartite graphon is exposed. We leave it as an open question whether every extreme point of the fractional vertex cover polyton is exposed even for non-bipartite graphons.

Acknowledgment

We thank an anonymous referee who provided very detailed and helpful comments.

The contents of this publication reflects only the authors' views and not necessarily the views of the European Commission of the European Union.

Appendix A. Proof of Lemma 25

Let $\mathcal{L}$ be the family of all pairs $(F_1, F_2) \in L^1(\Omega^2) \times L^1(\Omega^2)$ such that for every $\varepsilon > 0$ there exist a partition of Ω into finitely many pairwise disjoint subsets $B_1, \dots, B_m$ (for a

suitable natural number m), and real numbers D_1^{pq}, D_2^{pq}, $p,q=1,\dots,m$, such that

$$\sum_{p,q=1}^{m} \int\limits_{B_p\times B_q} |F_1 - D_1^{pq}| < \varepsilon \qquad \text{and} \qquad \sum_{p,q=1}^{m} \int\limits_{B_p\times B_q} |F_2 - D_2^{pq}| < \varepsilon\ .$$

It is easy to verify that $\mathcal{L}$ is a closed subspace of $L^1(\Omega^2)\times L^1(\Omega^2)$ containing all pairs consisting of characteristic functions of measurable rectangles. Therefore it holds $\mathcal{L} = L^1(\Omega^2)\times L^1(\Omega^2)$.

Now let us fix $(F_1,F_2)\in L^1(\Omega^2)\times L^1(\Omega^2)$ and $\varepsilon>0$. As $(F_1,F_2)\in\mathcal{L}$, we can find a partition of Ω into finitely many pairwise disjoint subsets $B_1,\dots,B_m$, and real numbers D_1^{pq}, D_2^{pq}, $p,q=1,\dots,m$, such that

$$\sum_{p,q=1}^{m} \int\limits_{B_p\times B_q} |F_1 - D_1^{pq}| < \frac{1}{4}\varepsilon \qquad \text{and} \qquad \sum_{p,q=1}^{m} \int\limits_{B_p\times B_q} |F_1 - D_2^{pq}| < \frac{1}{4}\varepsilon\ . \tag{41}$$

By the absolute continuity of the Lebesgue integral there is $\delta>0$ such that $\int\limits_E |F_1| < \frac{1}{4}\varepsilon$ and $\int\limits_E |F_2| < \frac{1}{4}\varepsilon$, whenever $E\subseteq\Omega\times\Omega$ is such that $\nu^{\otimes 2}(E)<\delta$. We fix a natural number k such that $\frac{m}{k} < \frac{1}{2}\delta$. For every $p=1,\dots,m$, we find a decomposition $B_p = B_p^0 \sqcup B_p^1 \sqcup \ldots \sqcup B_p^{n_p}$ of B_p into finitely many pairwise disjoint subsets such that $\nu(B_p^r)=\frac{1}{k}$ for $r=1,\dots,n_p$, and $\nu(B_p^0)\le\frac{1}{k}$. We set $F = \bigcup\limits_{p=1}^{m} B_p^0$. Then $\nu(F)$ is smaller that $\frac{1}{2}\delta$, and so

$$\int\limits_{(F\times\Omega)\cup(\Omega\times F)} |F_1| < \frac{1}{4}\varepsilon \qquad \text{and} \qquad \int\limits_{(F\times\Omega)\cup(\Omega\times F)} |F_2| < \frac{1}{4}\varepsilon\ . \tag{42}$$

Moreover, $\nu(F)$ is a multiple of $\frac{1}{k}$, and so it can be decomposed into finitely many disjoint subsets $\tilde B_1,\dots,\tilde B_l$ (for a suitable natural number l), each of measure $\frac{1}{k}$. Let $\Omega_1,\dots\Omega_k$ be some enumeration of the sets $\tilde B_1\dots,\tilde B_l$ and B_p^r, $r=1\dots,n_p$, $p=1,\dots,m$. For every $i,j=1,\dots,k$, we define real numbers C_1^{ij} and C_2^{ij} as follows. If $\Omega_i\times\Omega_j = B_p^r\times B_q^s$ for some $p,q\in\{1,\dots,m\}$, $r\in\{1,\dots,n_p\}$ and $s\in\{1,\dots,n_q\}$ then we set $C_1^{ij}=D_1^{pq}$ and $C_2^{ij}=D_2^{pq}$. Otherwise, we set $C_1^{ij}=C_2^{ij}=0$. Let us fix $t\in\{0,1\}$. Then we have

$$\begin{aligned}
&\sum_{i,j=1}^{k} \int\limits_{\Omega_i\times\Omega_j} |F_t - C_t^{ij}| \\
&= \sum_{p,q=1}^{m} \sum_{\substack{r=1,\dots,n_p\\ s=1,\dots,n_q}} \int\limits_{B_p^r\times B_q^s} |F_t - D_t^{pq}| + \int\limits_{(F\times\Omega)\cup(\Omega\times F)} |F_t| \\
&\overset{(41),(42)}{<} \frac{1}{4}\varepsilon + \frac{1}{4}\varepsilon = \frac{1}{2}\varepsilon\ .
\end{aligned} \tag{43}$$

Claim 5. *For every $i, j = 1, \dots, k$, we have that*

$$\int\limits_{\Omega_i \times \Omega_j} |F_t - F_t^{ij}| \le 2 \int\limits_{\Omega_i \times \Omega_j} |F_t - C_t^{ij}| \,.$$

Proof. Let us fix $i, j \in \{1, \dots, k\}$. It holds

$$\int\limits_{\Omega_i \times \Omega_j} |F_t - F_t^{ij}| = \int\limits_{\substack{(x,y) \in \Omega_i \times \Omega_j \\ F_t(x,y) < F_t^{ij}}} (F_t^{ij} - F_t) + \int\limits_{\substack{(x,y) \in \Omega_i \times \Omega_j \\ F_t(x,y) > F_t^{ij}}} (F_t - F_t^{ij}) \,, \tag{44}$$

and the two integrals on the right hand side of (44) equals each other (by the definition of F_t^{ij}). Therefore it is enough to show that one of these integrals is less or equal to $\int\limits_{\Omega_i \times \Omega_j} |F_t - C_t^{ij}|$. Assume for example that $C_t^{ij} \ge F_t^{ij}$ (the complementary case is similar). Then we have

$$\int\limits_{\substack{(x,y) \in \Omega_i \times \Omega_j \\ F_t(x,y) < F_t^{ij}}} (F_t^{ij} - F_t) \le \int\limits_{\substack{(x,y) \in \Omega_i \times \Omega_j \\ F_t(x,y) < C_t^{ij}}} (C_t^{ij} - F_t) \le \int\limits_{\Omega_i \times \Omega_j} |F_t - C_t^{ij}| \,,$$

as we wanted. □

Inequality (43) combined with Claim 5 gives us (28).

Appendix B. Details concerning Footnote 5

Recall a celebrated theorem of Pósa [15] that there exists a constant $C > 0$ so that the random graph $\mathbb{G}(2n, \frac{C \log n}{n})$ contains a Hamilton cycle with probability $1 - o_n(1) > \frac{1}{2}$. Note that Hamiltonicity implies the existence of a perfect matching. Let E_{2n} be the union of edges obtained in $\sqrt[3]{n}$ independent copies of $\mathbb{G}(2n, C \log n/n)$. Since we $\sqrt[3]{n} \cdot \frac{C \log n}{n} < 1/\sqrt{n}$, get that the edge set E_{2n} is stochastically dominated (with respect to inclusion) by $\mathbb{G}(2n, 1/\sqrt{n})$. Hence the probability that $\mathbb{G}(2n, 1/\sqrt{n})$ does not contain a perfect matching is at most

$$p_{2n} := \left(\frac{1}{2}\right)^{\sqrt[3]{n}} \,.$$

The sequence $(p_{2n})_n$ is summable, and hence the Borel–Cantelli lemma tells us that at most finitely many "bad events" occur.

References

[1] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[2] S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011.

[3] M. Doležal, J. Hladký, and A. Máthé. Cliques in dense inhomogeneous random graphs. *Random Structures Algorithms*, 51(2):275–314, 2017.

[4] M. Doležal, J. Hladký, P. Hu, and D. Piguet. First steps in combinatorial optimization on graphons: Matchings. In *European Conference on Combinatorics, Graph Theory and Applications (EuroComb 2017)*, volume 61 of *Electron. Notes Discrete Math.*, pages 359–365. Elsevier Sci. B. V., Amsterdam, 2017.

[5] P. Erdős and T. Gallai. On maximal paths and circuits of graphs. *Acta Math. Acad. Sci. Hungar*, 10:337–356 (unbound insert), 1959.

[6] J. Hladký, P. Hu, and D. Piguet. Tilings in graphons. arXiv:1606.03113.

[7] J. Hladký, P. Hu, and D. Piguet. Komlós's tiling theorem via graphon covers. *J. Graph Theory*, 90(1):24–45, 2019.

[8] J. Hladký and I. Rocha. Independent sets, cliques, and colorings in graphons. arXiv:1712.07367.

[9] J. Komlós. Tiling Turán theorems. *Combinatorica*, 20(2):203–218, 2000.

[10] J. Komlós, G. N. Sárközy, and E. Szemerédi. Blow-up lemma. *Combinatorica*, 17(1):109–123, 1997.

[11] L. Lovász. Subgraph densities in signed graphons and the local Simonovits-Sidorenko conjecture. *Electron. J. Combin.*, 18(1):Paper 127, 21, 2011.

[12] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[13] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[14] L. Lovász and B. Szegedy. Regularity partitions and the topology of graphons. In *An irregular mind*, volume 21 of *Bolyai Soc. Math. Stud.*, pages 415–446. János Bolyai Math. Soc., Budapest, 2010.

[15] L. Pósa. Hamiltonian circuits in random graphs. *Discrete Math.*, 14(4):359–364, 1976.

[16] A. A. Razborov. Flag algebras. *J. Symbolic Logic*, 72(4):1239–1282, 2007.

[17] A. Schrijver. *Combinatorial optimization. Polyhedra and efficiency. Vol. A*, volume 24 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2003. Paths, flows, matchings, Chapters 1–38.

Institute of Mathematics of the Czech Academy of Sciences, Žitná 25, Praha. The Institute of Mathematics is supported by RVO:67985840.

INDEPENDENT SETS, CLIQUES, AND COLORINGS IN GRAPHONS

JAN HLADKÝ AND ISRAEL ROCHA

Abstract. We study graphon counterparts of the chromatic and the clique number, the fractional chromatic number, the b-chromatic number, and the fractional clique number. We establish some basic properties of the independence set polytope in the graphon setting, and duality properties between the fractional chromatic number and the fractional clique number. We present a notion of perfect graphons and characterize them in terms of induced densities of odd cycles and its complements.

1. Introduction

The concepts of independent sets, cliques, and colorings are among the most studied in graph theory. Before stating its graphon counterparts, let us recall some fundamental concepts.

1.1. **Review of basic properties for graphs.** Suppose that $G = (V, E)$ is a graph. We say that a function $f : V \to [k]$ is a *proper coloring* of G with k colors if for all pairs $xy \in \binom{V}{2}$ we have $xy \notin E$ or $f(x) \neq f(y)$. The *chromatic number* $\chi(G)$ is defined as the minimal number of colors for which a proper coloring exists. Thus, in a proper coloring we color the vertices avoiding that neighbors share the same color.

One can generalize the concept of proper coloring by adding multiple colors to each vertex of a graph in the following manner. Let $b \in \mathbb{N}$. A map $p : V \to \binom{[k]}{b}$ is a *b-fold coloring* of G with k colors if for all pairs $xy \in \binom{V}{2}$ we have $xy \notin E$ or $p(x) \cap p(y) = \emptyset$. The *$b$-fold chromatic number* $\chi_b(G)$ is the smallest number of colors necessary to construct a proper b-fold coloring of G.

A set $A \subseteq V$ is an *independent set* of G if for all pairs $xy \in \binom{A}{2}$ we have $xy \notin E$. The size of the largest independent set of G is the *independence number* $\alpha(G)$. Dual to the concept of an independent set is that of a *clique* of a graph G, a set $A \subseteq V$ where for all pairs xy inside A we have $xy \in E$. The *clique number* $\omega(G)$ is the size of the largest clique of G.

Now denote by $\mathcal{I}(G)$ the collection of independent sets in G. A function $c : \mathcal{I}(G) \to [0, 1]$ is a *fractional coloring* of G if for every $v \in V(G)$ we have

$$\sum_{I \in \mathcal{I}(G), v \in I} c(I) \geq 1 \,. \tag{1.1}$$

The *fractional chromatic number* is then defined as the infimum of $\sum_{I \in \mathcal{I}(G)} c(I)$ taken over all fractional colorings c. Fractional colorings are indeed a fractional relaxation of ordinary

Hladký was supported by the Alexander von Humboldt Foundation.

Rocha was supported by the Czech Science Foundation, grant number GJ16-07822Y.

colorings. Indeed, when $c : \mathcal{I}(G) \to \{0,1\}$, then (1.1) can be interpreted as the condition that in a coloring every vertex has to be covered by at least one independent set.

We say that a function $f : V \to [0,+\infty)$ is a *fractional clique* if for every $I \in \mathcal{I}(G)$ we have

$$\sum_{v\in I} f(v) \le 1. \tag{1.2}$$

The *fractional clique number* of G is defined as

$$\omega_{\mathrm{frac}}(G) = \max \sum_{v\in V} f(v),$$

where the maximum is taken over all fractional cliques of G. In the same fashion as in fractional colorings, the fractional clique number is the relaxation of the problem of finding a maximum clique in a graph. Notice that when $f : V \to \{0,1\}$, then (1.2) can be interpreted as the condition that every independent set has at most one vertex of any clique.

The study of these parameter, and in particular the study of the interplay of the integral and fractional versions of these parameters is central in graph theory.

1.2. **Our contribution.** We translate the basics of the theory regarding independent sets, cliques, and colorings to the setting of graphons (see Section 2 for basics). While turning a graph definition into a graphon one is typically straightforward (and in many cases had been done previously), counterparts to many natural facts from finite graphs turned out to be quite challenging. By optimizing over all objects of this type (independent sets, fractional cliques, ...), we can define the corresponding numerical graphon parameter (independence number, fractional clique number, ...). We study relations between these parameters. We also study continuity of these graphon parameters with respect to the cut-norm. It turns out that none of the quantities we introduce is continuous, as is shown in all cases by the sequence of constant graphons $\left(Y_n \equiv \frac{1}{n}\right)_n$ converging to $Y \equiv 0$. However, most of these parameters are semicontinuous with respect to the cut-norm. These results are summarized in Table 1.

Further, in Section 3.1 we introduce a graphon counterpart to the independence set polytope, in Section 6 we treat the LP duality between fractional cliques and fractional colorings, and in Section 7 we introduce two notions of perfect graphons. Several fairly basic problems remain open.

Remark 1. *All our notions only depend on the support of a graphon. That is, replacing a graphon for example by the indicator function of its support, the notions of independent sets, colorings, etc., the corresponding numerical parameters do no change. The only exception to this is a notion of perfect graphons introduced in Section 7.*

	(A) representation	(B) semicontinuity	(C) supremum over graphs
independence	*trivial*	Corollary 7	see Section 3.2
chromatic	Proposition 13	Theorem 14(a)	Proposition 15
fractional chromatic	Proposition 17	*not true* (Example 19)	*not true* (Example 19)
clique	*trivial*	Proposition 25	*trivial*
fractional clique	Proposition 27	Theorem 28	Corollary 32

TABLE 1. Summary of results regarding graphon parameters we study. Column (A) refers to results that show equality of the graphon parameter to the corresponding graph parameter in case of graphon representation of a finite graph. Column (B) refers to results that show that these graphon parameters are lower semicontinuous (for chromatic number, fractional chromatic number, clique number, and fractional clique number) or upper semicontinuous (for independence number) in the cut-distance. In all cases the sequence $Y_n \equiv \frac{1}{n} \to Y \equiv 0$ shows that we do not have the complementary semicontinuity. Column (C) refers to results that show that for a general graphon, the value can be computed as the supremum over all finite graphs that appear in that graphon of the graph version of that parameter.

The paper is organized as follows: in Section 2 we provide some notation and prove a preliminary lemma; In Section 3 we study independent sets; Section 4 is dedicated to develop different concepts of chromatic number; In Section 5 we introduce notions of clique numbers; In Section 6, we prove certain duality properties between chromatic and clique parameters; and finally, Section 7 is devoted to the study of perfect graphons.

2. Notation and preliminaries

Throughout, we fix an atomless Borel probability space Ω equipped with a measure ν (defined on an implicit sigma-algebra). For $k \in \mathbb{N}$, we denote by $\nu^{\otimes k}$ the product measure on Ω^k.

Graphons, introduced in [11, 2], are analytic objects that capture limit properties of dense graphs. We assume the reader's familiarity with the basics of the theory. Our notation mostly follows Lovász' treatise [10]. Our graphons will be mostly defined on Ω^2 In particular, we shall work with the *density* and the *induced density* of a finite graph H

in a graphon W, defined by

$$t(H,W) = \int_{\{x_v\}_{v\in V(H)}} \prod_{uv\in E(H)} W(x_u,x_v) \quad \text{and}$$

$$t_{\mathrm{ind}}(H,W) = \int_{\{x_v\}_{v\in V(H)}} \prod_{uv\in E(H)} W(x_u,x_v) \cdot \prod_{uv\notin E(H)} (1-W(x_u,x_v)) \ .$$

Also, we shall make use of *inhomogeneous random graphs* $\mathbb{G}(n,W)$ described for example in Section 10.1 of [10]. By a *subgraphon of W obtained by restricting to a set $A\subset\Omega$* of positive measure we mean a function $W[A]: A^2\to[0,1]$ which is simply the restriction $W\restriction_{A\times A}$. When working with this notion, we need to turn A into a probability space. That is, we view $W[A]$ as a graphon on the probability space A endowed with measure $\nu_A(B) := \frac{\nu(B)}{\nu(A)}$ for every measurable set $B\subset A$.

All subsets of Ω or of Ω^2 considered will be measurable; whenever a new set is constructed it follows immediately from the construction that the set is measurable. For sets $A,B\subset\Omega$, $C,D\subset\Omega^2$ we write $A=B \mod 0$, $C=D \mod 0$ for equality up to null sets, i.e., if $\nu(A\triangle B)=0$ and $\nu^{\otimes 2}(A\triangle B)=0$. Given a function $f: X\to\mathbb{R}$, we denote by $\mathrm{supp}(f)$ the support of f, $\mathrm{supp}(f) := \{x\in X : f(x)\neq 0\}$.

A graphon $W:\Omega\times\Omega\to[0,1]$ is a *graphon representation* of a finite graph G if there exists a partition $\Omega=\bigsqcup_{v\in V(G)}\Omega_v$ of sets of measure $\frac{1}{|V(G)|}$ such that W restricted to $\Omega_u\times\Omega_v$ is either constant 0 or constant 1 (modulo a nullset), depending on whether $uv\notin E(G)$ or $uv\in E(G)$.

2.1. **Structural properties of graphons with a given subgraph.** In this section we state Lemma 2 which says that if $t(H,W)>0$ for some finite graph H and some graphon W then we can find a subgraphon similar to the adjacency matrix of H in W. The case $H=C_{2\ell+1}$ of Lemma 2 appears in [6]. Our proof of Lemma 2 closely follows [6, Lemma 11].

Lemma 2. *Suppose that $W:\Omega\times\Omega\to[0,1]$ is a graphon. Suppose that H is a graph on vertex set $[k]$ with the property that $t(H,W)>0$. For each $\epsilon>0$ there exist $\alpha>0$ and pairwise disjoint sets $A_1,\dots,A_k\subset\Omega$, each of measure α, such that for each $ij\in E(H)$, the graphon W is positive everywhere on $A_i\times A_j$ except a set of measure at most $\epsilon\alpha^2$.*

Proof. Let $n\in\mathbb{N}$ be such that

$$n > \frac{k^2}{t(H,W)} \ . \tag{2.1}$$

Take an arbitrary partition $\Omega=\bigsqcup_{i=1}^n\Omega_i$ of Ω into pairwise disjoint sets of measure $\frac{1}{n}$. Set

$$D := \left\{\mathbf{x}\in\Omega^k \ : \text{ there are } i,j\in[k] \text{ and } \ell\in[n] \text{ such that } i\neq j \text{ and } x_i,x_j\in\Omega_\ell\right\} \ . \tag{2.2}$$

Then we have

$$\int_{\mathbf{x}\in D} \prod_{1\le i<j\le k, ij\in E(H)} W(x_i,x_j) \le \nu^{\otimes k}(D) \le \sum_{\substack{i,j=1,\dots,k\\ i\neq j}} \frac{1}{n} \le \frac{k^2}{n} \overset{(2.1)}{<} t(H,W)\,, \tag{2.3}$$

From (2.3), we get

$$\int_{\mathbf{x}\in\Omega^k\setminus D} \prod_{1\le i<j\le k, ij\in E(H)} W(x_i,x_j) > 0\,.$$

Using the definition of D, we get that there are pairwise distinct integers $\ell_1,\dots,\ell_k\in[n]$ such that

$$\int_{\Omega_{\ell_1}}\int_{\Omega_{\ell_2}}\cdots\int_{\Omega_{\ell_k}} \prod_{1\le i<j\le k, ij\in E(H)} W(x_i,x_j) > 0\,.$$

We conclude that the set

$$E := \left\{ \mathbf{x}\in\Omega_{\ell_1}\times\Omega_{\ell_2}\times\ldots\times\Omega_{\ell_k} \colon \prod_{1\le i<j\le k, ij\in E(H)} W(x_i,x_j) > 0 \right\} \tag{2.4}$$

has positive measure.

Given $\epsilon>0$, let $\delta>0$ be such that

$$\frac{\nu^{\otimes k}(E)-\delta}{\nu^{\otimes k}(E)+\delta} \ge 1-\frac{\epsilon}{2}\,. \tag{2.5}$$

Recall that the σ-algebra of all measurable subsets of $\Omega_{\ell_1}\times\ldots\times\Omega_{\ell_k}$ is generated by the algebra consisting of all finite unions of boxes. Thus there is a finite union $S=\bigcup_{i=1}^m R_i$ of boxes $R_1,\dots,R_m$ in $\Omega_{\ell_1}\times\ldots\times\Omega_{\ell_k}$ such that $\nu^{\otimes k}(E\setminus S)+\nu^{\otimes k}(S\setminus E)\le\delta$. Without loss of generality, we may assume that the boxes $R_1,\dots,R_m$ are pairwise disjoint. Then we have

$$\frac{\nu^{\otimes k}(E\cap S)}{\nu^{\otimes k}(S)} \ge \frac{\nu^{\otimes k}(E)-\nu^{\otimes k}(E\setminus S)}{\nu^{\otimes k}(E)+\nu^{\otimes k}(S\setminus E)} \ge \frac{\nu^{\otimes k}(E)-\delta}{\nu^{\otimes k}(E)+\delta} \overset{(2.5)}{\ge} 1-\frac{\epsilon}{2}\,. \tag{2.6}$$

The left-hand side of (2.6) can be expressed as

$$\frac{\nu^{\otimes k}(E\cap S)}{\nu^{\otimes k}(S)} = \sum_{i=1}^m \frac{\nu^{\otimes k}(R_i)}{\nu^{\otimes k}(S)}\cdot\frac{\nu^{\otimes k}(E\cap R_i)}{\nu^{\otimes k}(R_i)}\,,$$

i.e. as a convex combination of $\frac{\nu^{\otimes k}(E\cap R_i)}{\nu^{\otimes k}(R_i)}$, $i=1,\dots,m$. Therefore, there is an index $i_0\in[m]$ such that

$$\frac{\nu^{\otimes k}(E\cap R_{i_0})}{\nu^{\otimes k}(R_{i_0})} \ge 1-\frac{\epsilon}{2}\,. \tag{2.7}$$

Let R_{i_0} be of the form $R_{i_0} = B_1 \times \ldots \times B_k$. Find a natural number p such that

$$p \geq \frac{2k}{\epsilon \nu^{\otimes k}(R_{i_0})} \,. \tag{2.8}$$

For every $i = 1, \ldots, k$, we consider a finite decomposition $B_i = B_i^0 \cup \bigcup_{j=1}^{q_i} B_i^j$ of B_i into pairwise disjoint sets, such that $\nu(B_i^0) \leq \frac{1}{p}$ and $\nu(B_i^j) = \frac{1}{p}$ for $j = 1, \ldots, q_i$. Then we clearly have

$$\nu^{\otimes k}\left(R_{i_0} \setminus \prod_{i=1}^{k} \bigcup_{j=1}^{q_i} B_i^j\right) \leq \frac{k}{p} \,, \tag{2.9}$$

and so

$$\frac{\nu^{\otimes k}\left(E \cap \prod_{i=1}^{k} \bigcup_{j=1}^{q_i} B_i^j\right)}{\nu^{\otimes k}\left(\prod_{i=1}^{k} \bigcup_{j=1}^{q_i} B_i^j\right)} \overset{(2.9)}{\geq} \frac{\nu^{\otimes k}(E \cap R_{i_0}) - \frac{k}{p}}{\nu^{\otimes k}(R_{i_0})} \overset{(2.7)}{\geq} 1 - \frac{\epsilon}{2} - \frac{k}{p\nu^{\otimes k}(R_{i_0})} \overset{(2.8)}{\geq} 1 - \epsilon \,. \tag{2.10}$$

The left-hand side of (2.10) can be expressed as the following convex combination:

$$\frac{\nu^{\otimes k}\left(E \cap \prod_{i=1}^{k} \bigcup_{j=1}^{q_i} B_i^j\right)}{\nu^{\otimes k}\left(\prod_{i=1}^{k} \bigcup_{j=1}^{q_i} B_i^j\right)} = \sum_{j_1=1}^{q_1} \cdots \sum_{j_k=1}^{q_k} \frac{\nu^{\otimes k}\left(\prod_{i=1}^{k} B_i^{j_i}\right)}{\nu^{\otimes k}\left(\prod_{i=1}^{k} \bigcup_{j=1}^{q_i} B_i^j\right)} \cdot \frac{\nu^{\otimes k}\left(E \cap \prod_{i=1}^{k} B_i^{j_i}\right)}{\nu^{\otimes k}\left(\prod_{i=1}^{k} B_i^{j_i}\right)} \,.$$

Therefore by (2.10), there are indices $j_i \in [q_i]$, $i = 1, \ldots, k$, such that

$$\frac{\nu^{\otimes k}\left(\prod_{i=1}^{k} B_i^{j_i} \setminus E\right)}{\nu^{\otimes k}\left(\prod_{i=1}^{k} B_i^{j_i}\right)} \leq \epsilon \,. \tag{2.11}$$

We set $A_i = B_i^{j_i}$ for $i = 1, \ldots, k$. Then $A_1, \ldots, A_k$ are pairwise disjoint (as $A_i \subseteq \Omega_{\ell_i}$ for every i), and each of these sets has the same measure $\alpha = \frac{1}{p}$.

By (2.4) we have

$$\left\{(x_1, x_2, \ldots, x_k) \in A_1 \times A_2 \times \ldots \times A_k \colon \prod_{1 \leq i < j \leq k, ij \in E(H)} W(x_i, x_j) = 0\right\} = \prod_{i=1}^{k} B_i^{j_i} \setminus E \,. \tag{2.12}$$

By (2.11),

$$\nu^{\otimes k}\left(\prod_{i=1}^{k} B_i^{j_i} \setminus E\right) \leq \epsilon\nu^{\otimes k}\left(\prod_{i=1}^{k} B_i^{j_i}\right) = \frac{\epsilon}{p^k} .$$

Consider an arbitrary edge $ij \in E(H)$. Observe that the set $\prod_{i=1}^{k} B_i^{j_i} \setminus E$ contains all k-tuples $\mathbf{x} \in \prod_{h=1}^{k} A_h$ that have the property that $W(x_i, x_j) = 0$. Therefore, we get from (2.12) that

$$\nu^{\otimes 2}\left(\{(x_i, x_j) \in A_i \times A_j \colon W(x_i, x_j) = 0\}\right) \leq \epsilon\alpha^2 ,$$

as required. □

3. Independent sets

While classically, the notion of independent sets and cliques are in one-to-one correspondence by taking complements, our definitions for graphons look at each of these concepts at a different scale. The independence number of a graphon is a number in $[0, 1]$ that should be interpreted as the fraction of vertices in a maximum independent set of a graph that corresponds to that graphon. That is, we scale down the independence number linearly. On the other hand, in Section 5 we define the clique number (and its fractional variant) which are all absolute with no additional rescaling introduced (see Section 5 for a discussion of subtleties). If a graphon contains a copy of K_{17} then its clique number will be at least 17. Of course, which of the two parameters is scaled and which one is not is just a matter of convention. Let us remark that in [5] a different scale of $\log n$ is put on these parameters and studied in the context of inhomogeneous random graphs $\mathbb{G}(n, W)$.

The following is then the obvious graphon counterpart to independent sets.

Definition 3. Let $W : \Omega \times \Omega \to [0, 1]$ be a graphon. A set $A \subseteq \Omega$ is an *independent set* if W is zero almost everywhere on $A \times A$. Denote by $\mathcal{I}(W)$ the set of independent sets of a graphon W and for each $x \in \Omega$ denote by $\mathcal{I}_x(W) \subseteq \mathcal{I}(W)$ the set of independent sets of W containing the point x.

The definition of $\mathcal{I}_x(W)$ may look suspicious as it involves a measure-zero condition (a point belonging to a set). This will not cause a problem since we shall always work with collections $\mathcal{I}_x(W)$ for a set of x's of positive measure.

Let us mention that the paper [1] is closely related to ours, though concerned with different questions. In that paper, the authors investigate the independence ratio (which we call independence number in our setting) and chromatic number for bounded self-adjoint operators on an L^2 space. In particular, the graphon operator as defined in [10, Section 7.5] is an example of such operators. In particular, our notion of independent set of a graphon is equivalent to the notion of independent set as introduced in [1] of the graphon operator

associated to that graphon. In [1], the authors provide bounds for these parameters in terms of the eigenvalues of the operator — an analog to the Hoffman bound — and use harmonic analysis and convex optimization to study packing and coloring problems for finite and infinite graphs.

Here, we need to describe particular properties of independent sets in order to understand the behavior of the chromatic number and the clique number for graphons.

The next lemma which appears in [8, Lemma 20] asserts that the weak* limit of independent sets is again an independent set.

Lemma 4. *Let W be a graphon. Suppose $(A_n)_n$ is a sequence of sets in Ω with the property that*

$$\lim_{n\to\infty} \int_{A_n \times A_n} W = 0.$$

Suppose that the indicator functions of the sets A_n converge weak to a function f. Then* $\operatorname{supp}(f)$ *is an independent set in W.*

It follows that for a convergent sequence of graphons, the weak* limit of independent sets in the sequence form an independent set in the limit graphon.

Corollary 5. *Let $W_n : \Omega \times \Omega \to [0,1]$ be a sequence of graphons converging in the cut-norm to $W : \Omega \times \Omega \to [0,1]$. Let $I_n \subset \Omega$ be an independent set in W_n. Suppose that the indicator functions of the sets I_n converge weak* to a function f. Then* $\operatorname{supp}(f)$ *is an independent set in W.*

Proof. Notice that

$$\int_{I_n \times I_n} W = \left| \int_{I_n \times I_n} (W_n - W) \right| \leq \|W_n - W\|_{\square}.$$

Thus, $\lim_{n\to\infty} \int_{I_n \times I_n} W = 0$ and the claim follows from Lemma 4. □

The defining property of weak* convergence gives us that in the setting of the corollary above, we have

$$\lim_n \nu(I_n) = \lim_n \int_\Omega \mathbf{1}_{I_n} = \int_\Omega f \leq \nu(\operatorname{supp}(f)) \ , \tag{3.1}$$

where the last inequality uses that f is bounded above by 1, since it is a weak* limit of functions bounded above by 1. Thus, as a consequence of Corollary 5 we get that the supremum of the measures of independent sets in a graphon is attained. This leads us to the following definition.

Definition 6. The measure of the largest independent set of a graphon $W : \Omega \times \Omega \to [0,1]$ is called the *independence number* of W and denoted by $\alpha(W)$.

Corollary 5 yields upper semicontinuity of the independence number. As mentioned in Section 1.2, the sequence $Y_n \equiv \frac{1}{n} \to Y \equiv 0$ shows that we do not have lower semicontinuity in general.

Corollary 7. *Suppose that $(W_n)_n$ is a sequence of graphons that converges to W in the cut-distance. Then $\alpha(W) \geq \limsup_n \alpha(W_n)$.*

Proof. We may as well assume that $(W_n)_n$ converges to W in the cut-norm, and that the limit $\lim_n \alpha(W_n)$ exists. Now, for each n, consider an independent set I_n in W_n of size $\alpha(W_n)$. Let f be a weak* accumulation point of the sequence of indicator functions of the sets $(I_n)_n$; at least one such accumulation point exists by the sequential Banach–Alaoglu Theorem. By Corollary 5, $\operatorname{supp}(f)$ is an independent set, and by the same calculation as in (3.1), the measure of $\operatorname{supp}(f)$ is at least $\lim_n \alpha(W_n)$. □

3.1. **Structure of independent sets.** In this section, we make some observations about the structure of independent sets in a graphon. Often, rather than dealing with all independent sets, it is convenient to restrict attention just to maximal ones, from which all the remaining ones can be easily recovered. We denote the set of maximal independent sets (modulo nullsets) in a graphon W by $\mathcal{I}_{\max}(W) \subset \mathcal{I}(W)$. Even the set $\mathcal{I}_{\max}(W)$ can be quite complicated, at least with respect to its cardinality. Indeed, let W be a graphon representing a disjoint union of countably many complete bipartite graphs $(H_n = A_n \sqcup B_n)_{n=1}^{\infty}$, which occupy measure 2^{-n} each. Then the maximal independent sets in W are all unions of maximal independent sets in all graphs H_n, (for example $A_1, A_2, A_3, B_4, B_5, \ldots$), of which there are uncountably many.

Another perspective on the structure of $\mathcal{I}(W)$ comes from polyhedral combinatorics. To motivate this, let us first recall an approach common for finite graphs. Given a finite graph G, each independent set I in G can viewed as a vector in $\{0,1\}^{V(G)} \subset \mathbb{R}^{V(G)}$. Taking the convex hull of all such vectors, one gets what is called an *independent set* (or *stable set*) *polytope* $\mathrm{IND}(G) \subset \mathbb{R}^{V(G)}$. Among many classical results about $\mathrm{IND}(G)$, let us mention perhaps the most basic one.

Proposition 8 (e.g. [7, (9.1.3)])**.** *For any graph G, each point $\mathbf{x} \in \mathrm{IND}(G)$ satisfies the following two families of conditions:*

$$0 \leq \mathbf{x}_v \leq 1 \quad \text{for each } v \in V(G) \text{ and} \tag{3.2}$$

$$\mathbf{x}_u + \mathbf{x}_v \leq 1 \quad \text{for each } uv \in E(G)\,. \tag{3.3}$$

Further, (3.2) and (3.3) characterize $\mathrm{IND}(G)$ if and only if G is bipartite.

Now, in the graphon setting we proceed as follows. We represent each set $I \in \mathcal{I}(W)$ of a graphon $W : \Omega \times \Omega \to [0,1]$ by its characteristic function, which we view as an element

in $\{0,1\}^\Omega \subset \mathbb{R}^\Omega$. We can now take the closure (in the weak* topology) of the convex hull of such functions and get what we call *independent set polyton* $\mathrm{IND}(W) \subset \mathbb{R}^\Omega$. Such a graphon approach to polyhedral combinatorics has been introduced [6], namely for the so-called matching polytope/polyton. To illustrate the potential of this area, let us prove a part of a counterpart to Proposition 8.

Proposition 9. *For any graph $W : \Omega \times \Omega \to [0,1]$, each point $\mathbf{x} \in \mathrm{IND}(W)$ satisfies the following two families of conditions:*

$$0 \le \mathbf{x}_v \le 1 \quad \textit{for almost all } v \in \Omega \textit{ and} \tag{3.4}$$

$$\mathbf{x}_u + \mathbf{x}_v \le 1 \quad \textit{for almost all } uv \in \mathrm{supp}\,(W)\,. \tag{3.5}$$

Further, if W is not bipartite, then (3.4) and (3.5) do not characterize $\mathrm{IND}(W)$.

Problem 10. In analogy with Proposition 8, establish the other direction in Proposition 9.

Proof of Proposition 9. Obviously, every indicator function of an independent set $I \in \mathcal{I}(W)$ satisfies (3.4) and (3.5). These inequalities are then inherited to convex combinations and closure, thus proving the first part.

Suppose now that W is not bipartite. We shall prove that the point $\mathbf{y} \equiv \frac{1}{2}$ is not in $\mathrm{IND}(W)$, even though it apparently satisfies (3.4) and (3.5). By [6, Proposition 21] (stated also as Proposition 15 below), we know that $t(C_{2\ell+1}, W) > 0$ for some $\ell \in \mathbb{N}$. By Lemma 2 we know that there exist disjoint sets $A_1, A_2, \ldots, A_{2\ell+1} \subset \Omega$ of the same measure $\alpha > 0$ such that that W is positive everywhere on $A_i \times A_j$ except a set of measure at most $\frac{\alpha^2}{10\ell}$. Now, suppose for contradiction that $\mathbf{y}$ is in the weak* closure of convex combinations of characteristic functions of independent sets. In particular, there exists a point $\mathbf{y}^*$ such that we have

$$\int_{\omega \in \bigcup_{i=1}^{2\ell+1} A_i} \mathbf{y}^*(\omega) > \int_{\omega \in \bigcup_{i=1}^{2\ell+1} A_i} \mathbf{y}(\omega) - \frac{\alpha}{10} = \frac{1}{2}(2\ell+1)\alpha - \frac{\alpha}{10} \tag{3.6}$$

and such that $\mathbf{y}^*$ is a convex combination of characteristic functions of independent sets, $\mathbf{y}^* = \sum_{j=1}^{t} \alpha_j \mathbf{1}_{I_j}$, where $\alpha_j \ge 0$ are the convex coefficients and $I_j \in \mathcal{I}(W)$. We get (3.6) must hold also for one of the terms appearing in the convex combination, i.e., there exists $j \in [t]$ such that

$$\nu\left(I_j \cap \bigcup_{i=1}^{2\ell+1} A_i\right) \ge \frac{1}{2}(2\ell+1)\alpha - \frac{\alpha}{10}\,. \tag{3.7}$$

Call an index $i \in [2\ell+1]$ *marked* if $\nu(I_j \cap A_i) \ge \frac{\alpha}{5\ell}$. Observe that we cannot have two consecutive marked indices i and $i+1$ (with labeling considered is modulo $2\ell+1$), since $W_{\restriction A_i \times A_{i+1}}$ is positive on most of the domain and I_j is an independent set. Hence, there

are at most 2ℓ marked indices. For a marked index i we have $\nu(I_j \cap A_i) \le \alpha$ and for an unmarked index i we have $\nu(I_j \cap A_i) \le \frac{\alpha}{5\ell}$. Hence,

$$\nu\left(I_j \cap \bigcup_{i=1}^{2\ell+1} A_i\right) \le \ell \cdot \alpha + (\ell+1) \cdot \frac{\alpha}{5\ell}\,.$$

This contradicts (3.7). □

3.2. **A recent result of Cooley, Kang and Pikhurko.** We originally thought that there is no way how to relate subgraph densities of a graphon to its independence number. However, recently Cooley, Kang and Pikhurko [4] found a natural way to fill this gap, which we state below.

Theorem 11. *Let $W : \Omega \times \Omega \to [0,1]$ be a graphon. For each $k \in \mathbb{N}$, let a_k be the probability that $\mathbb{G}(k,W)$ does not contain any edges, that is $a_k = t_{\mathrm{ind}}(I_k, W)$, where I_k the edgeless graph on k vertices. Then the limit $\lim_{k\to\infty} \sqrt[k]{a_k}$ exists, and is equal to the independence number of W.*

4. Coloring concepts

4.1. **Ordinary colorings.** First, we define a counterpart to the usual concept of coloring.

Definition 12. Let $W : \Omega \times \Omega \to [0,1]$ be a graphon. We say that a measurable function $f : \Omega \to [k]$ is a *proper coloring* of W with k colors if, for each $i \in [k]$, the set $f^{-1}(i)$ is an independent set in W. The *chromatic number* $\chi(W)$ is defined as the minimal number of colors for which a proper coloring exists. Here, as will be with other versions of the chromatic number, when no proper coloring of W exists, $\chi(W)$ is taken to be infinity. In particular, this notion splits the space of graphons into graphons of *finite* and *infinite chromatic number*.

The next easy proposition shows that Definition 12 is consistent with the graph definition of the chromatic number.

Proposition 13. *Suppose G is a finite graph and let $W : \Omega \times \Omega \to [0,1]$ be its graphon representation. Then $\chi(G) = \chi(W)$.*

Proof. Let $\Omega = \Omega_1 \sqcup \cdots \sqcup \Omega_n$ be the partition for W corresponding to the vertices of G. If $U_1 \sqcup U_2 \sqcup \ldots \sqcup U_k = V(G)$ is a proper k-coloring of G, then the independent sets $\left(\bigcup_{v\in U_i} \Omega_v\right)_{i\in[k]}$ give a proper k-coloring of W. Thus, $\chi(W) \le \chi(G)$. Conversely, suppose that $I_1 \sqcup I_2 \sqcup \ldots \sqcup I_k = \Omega$. Now, for every $j \in [n]$, there exists at least one index, say $\ell_j \in [k]$, such that $\Omega_j \cap I_{\ell_j}$ has positive measure. Obviously, the map $j \mapsto \ell_j$ is a proper coloring of G. □

The graphons $Y_n \to Y$ from Section 1.2 show that the chromatic number is not upper continuous. Indeed, we have $\chi(Y_n) = \infty$ and $\chi(Y) = 1$. On the other hand, the following theorem shows lower semicontinuity.

Theorem 14. *Let $W : \Omega \times \Omega \to [0,1]$ be a graphon.*

(a) For every sequence of graphs (H_n) that converges to W in the cut-distance we have $\chi(W) \le \liminf_{n\to\infty} \chi(H_n)$.
(b) For the sequence $(G_n \sim \mathbb{G}(n,W))$ we have $\chi(W) = \lim_{n\to\infty} \chi(G_n)$, almost surely.

The next proposition, which can be found in [6, Proposition 5], is the key towards proving Theorem 14. We remark that the proof of this proposition given in [6] is short but non-trivial.

Proposition 15. *Suppose $k \in \mathbb{N}$ and $W : \Omega \times \Omega \to [0,1]$ is a graphon. Then $\chi(W) \le k$ if and only if $t(G,W) = 0$ for each finite graph G with $\chi(G) \ge k+1$.*

Proof of Theorem 14(a). Suppose that $\chi(W) \ge k$. By (the more difficult part of) Proposition 15, there exists a graph F with chromatic number at least k for which $t(F,W)$ is positive, say $a := t(F,W)$. Since convergence in the cut-distance implies convergence of subgraph densities, we conclude that each graph H_n has subgraph density of F equal to $a + o_n(1)$. In particular, for n sufficiently large, this density is positive. We conclude that the chromatic number of such a graph H_n is at least as that of F. □

Proof of Theorem 14(b). Since the graphs G_n converge to W in the cut-distance almost surely, we have $\chi(W) \le \liminf_{n\to\infty} \chi(G_n)$ by (a). On the other hand, if $\chi(W) \le k$, then by (the easy part of) Proposition 15, each graph of chromatic number at least $k+1$ has zero density. That is, the probability of W-sampling any such graph is zero. □

4.2. **Fractional colorings.** The next definition is a straightforward counterpart to the graph definition of fractional colorings.

Definition 16. Suppose that $W : \Omega \times \Omega \to [0,1]$ is a graphon. We say that a function $c : \mathcal{I}(W) \to [0,1]$ is a *fractional coloring* of W if for almost every $x \in \Omega$ we have $\sum_{I \in \mathcal{I}_x(W)} c(I) \ge 1$. The *fractional chromatic number* of W is defined as

$$\chi_{\mathrm{frac}}(W) = \inf \sum_{I \in \mathcal{I}(W)} c(I), \tag{4.1}$$

where the infimum is taken over the set of all fractional colorings of W.

Notice that whenever $\sum_{I\in\mathcal{I}(W)} c(I)$ is finite, then the support of the function c is at most countable. Thus it suffices to take the infimum over the set of all fractional colorings of W that are not zero at most in a countable subset of $\mathcal{I}(W)$.

The next proposition shows that the definitions of fractional chromatic number for graphs and graphons are consistent.

Proposition 17. *Suppose G is a finite graph and let W be its graphon representation. Then $\chi_{\mathrm{frac}}(G) = \chi_{\mathrm{frac}}(W)$.*

Proof. Denote by $\mathcal{I}(W)$ and $\mathcal{I}(G)$ the set of independent sets of W and G, respectively. There is a map $\pi : \mathcal{I}(G) \to \mathcal{I}(W)$ which maps independent sets in $\mathcal{I}(G)$ to corresponding independent sets in $\mathcal{I}(W)$. If c_G is an arbitrary fractional coloring of G then we can define $c_W : \pi(\mathcal{I}(G)) \to [0,1]$, $c_W(\pi(I)) := c_G(I)$. Extending c_W by zeros to $\mathcal{I}(W)$, we get a fractional coloring of W of the same size. This implies $\chi_{\mathrm{frac}}(G) \geq \chi_{\mathrm{frac}}(W)$.

Now consider a fractional coloring $c_W : \mathcal{I}(W) \to [0,1]$ of W. Let $\Omega = \Omega_1 \sqcup \cdots \sqcup \Omega_n$ be the partition for W corresponding to the vertices of G. For each $v \in V(G)$, choose a point $x_v \in \Omega_v$. We shall assume that $\sum_{I \in \mathcal{I}_{x_v}(W)} c_W(I) \geq 1$ for each $v \in V(G)$; this is true for almost all choices of $x_v \in \Omega_v$. Now define $c_G : \mathcal{I}(G) \to [0,1]$ by

$$c_G(J) := \sum_I c_W(I) \,,$$

where I ranges over all sets in $\bigcap_{v \in J} \mathcal{I}_{x_v}(W) \setminus \bigcup_{v \in V(G) \setminus J} \mathcal{I}_{x_v}(W)$. The inclusion-exclusion formula implies that for each $v \in V(G)$ we have $\sum_{J \in \mathcal{I}_v(G)} c_G(J) = \sum_{I \in \mathcal{I}_{x_v}(W)} c_W(I)$. Since $\sum_{I \in \mathcal{I}_{x_v}(W)} c_W(I) \geq 1$, for each $v \in V(G)$ we have $\sum_{J \in \mathcal{I}_v(G)} c_G(J) \geq 1$. Therefore, c_G is a valid fractional coloring of G. Thus $\chi_{\mathrm{frac}}(G) \leq \chi_{\mathrm{frac}}(W)$, which finishes the proof. □

Remark 18. *Observe that in the proof of Proposition 17 we did not use that each vertex of G is represented by a set of Ω of measure $\frac{1}{|V(G)|}$. That is, the same result holds true if vertices of G are represented by arbitrary sets $\Omega_1 \sqcup \cdots \sqcup \Omega_n = \Omega$ of positive but not necessarily the same measure.*

Example 19. In [9] Leader constructs a graph R with a countable vertex set, say $\mathbb{N}$, whose fractional chromatic number is strictly greater than the supremum of the fractional chromatic numbers of its finite subgraphs. Considering an arbitrary partition $\Omega = \bigsqcup_{i \in \mathbb{N}} \Omega_i$ into sets of positive measure, and taking W to be a graphon representation of R with respct to the partition $\Omega = \bigsqcup_{i \in \mathbb{N}} \Omega_i$, we get by a version[1] of Proposition 17 (see also Remark 18) that $\chi_{\mathrm{frac}}(W) > \sup\{\chi_{\mathrm{frac}}(G) : t(G,W) > 0\}$. Note that this example also shows that $\chi_{\mathrm{frac}}(\cdot)$ is not lower semicontinuous. Indeed, for $\ell = 1, 2, 3, \ldots$, we consider graphon W_ℓ which is equal to W on $\left(\cup_{i=1}^{\ell} \Omega_i\right) \times \left(\cup_{i=1}^{\ell} \Omega_i\right)$ and 0 otherwise. Obviously, the graphons $(W_n)_n$ converge to W in the cut-norm (actually, even pointwise), but we have $\lim_n \chi_{\mathrm{frac}}(W_n) = \sup\{\chi_{\mathrm{frac}}(G) : t(G,W) > 0\} < \chi_{\mathrm{frac}}(W)$.

[1]For countable graphs; the proof of such a version is *mutatis mutandis*.

We also note that in [9] Leader constructs another countable graph T in which $\chi_{\text{frac}}(T)$ is not attained by any fractional coloring. Taking a graphon representation of T, we see that the fractional chromatic number of a graphon need not be attained.

4.3. ***b*-fold colorings.** Next, we define counterparts of b-fold colorings for graphons.

Definition 20. Suppose that $b \in \mathbb{N}$. For a graphon $W : \Omega \times \Omega \to [0,1]$, a measurable map $p : \Omega \to \binom{[k]}{b}$ is a *b-fold coloring* of W with k colors if for almost all pairs $(x,y) \in \Omega \times \Omega$ we have $W(x,y) = 0$ or $p(x) \cap p(y) = \emptyset$. The *$b$-fold chromatic number* $\chi_b(W)$ is the smallest number of colors necessary to construct a proper b-fold coloring of W.

As in the graph case, the relation between the fractional chromatic number, the 1-fold chromatic number and the ordinary chromatic number is as follows: $\chi_{\text{frac}}(W) \le \chi(W) = \chi_1(W)$. The next theorem shows that at least qualitatively, we can reverse the inequality.

Proposition 21. *Let $W : \Omega \times \Omega \to [0,1]$ be a graphon. If $\chi_{\text{frac}}(W) < \infty$ then $\chi(W) < \infty$.*

Proof. Suppose that c is a fractional coloring of W with $\sum_{I \in \mathcal{I}(W)} c(I) < \infty$. Thus, there is a finite number of independent sets $\{I_1, \dots, I_k\}$ such that

$$\sum_{i=1}^{k} c(I_i) > \sum_{I \in \mathcal{I}(W)} c(I) - 1. \tag{4.2}$$

In particular, for each $x \in \Omega$ we have

$$\sum_{I \in \mathcal{I}_x(W)} c(I) < \sum_{i \in [k], x \in I_i} c(I_i) + 1 \,.$$

This implies that if $x \in \Omega \setminus \bigcup_{i=1}^{k} I_i$ then $\sum_{I \in \mathcal{I}_x(W)} c(I) < 0 + 1$. We conclude that $\Omega \setminus \bigcup_{i=1}^{k} I_i$ is a nullset. Therefore, $\{I_1, \dots, I_k\}$ corresponds to a proper coloring of $W$with finitely many colors. □

A sequence $\{a_n\}_{n=1}^{\infty}$ is said to be *subadditive* if for every m and n we have $a_{m+n} \le a_m + a_n$. Fekete's Lemma is a useful result concerning subadditive sequences.

Lemma 22 (Fekete's Lemma)**.** *For every subadditive sequence $\{a_n\}_{n=1}^{\infty}$, the limit $\lim_{n\to\infty} \frac{a_n}{n}$ exists and is equal to $\inf \frac{a_n}{n}$.*

If we have an a-fold coloring $p : \Omega \to \binom{\{1,\dots,k\}}{a}$ of a graphon W which uses the palette $\{1, \dots, k\}$ and a b-fold coloring $q : \Omega \to \binom{\{k+1,\dots,k+\ell\}}{b}$ which uses the palette $\{k+1, \dots, k+\ell\}$, we observe that the map $x \mapsto p(x) \cup q(x)$ is an $(a+b)$-fold coloring of W which uses the palette $\{1, \dots, k+\ell\}$. Therefore, we have $\chi_{a+b}(W) \le \chi_a(W) + \chi_b(W)$. Fekete's lemma tells us that

$$\lim_{b\to\infty} \frac{\chi_b(W)}{b} = \inf \frac{\chi_b(W)}{b}.$$

The next theorem tells us that this limit is equal to the fractional chromatic number of W.

Theorem 23. *For every graphon W we have*

$$\chi_{\text{frac}}(W) = \lim_{b\to\infty} \frac{\chi_b(W)}{b} = \inf_{b\in\mathbb{N}} \frac{\chi_b(W)}{b}.$$

Proof. First, we prove that $\chi_{\text{frac}}(W) \leq \inf_{b\to\infty} \frac{\chi_b(W)}{b}$. Given $b \in \mathbb{N}$, fix a proper b-fold coloring c_b of Ω for the b-chromatic number χ_b, say using ℓ colors. We will construct a fractional coloring c such that

$$\sum_{I\in\mathcal{I}(W)} c(I) = \frac{\chi_b(W)}{b}.$$

To this end, for each $j \in [\ell]$, consider the independent set $I_j := \{x \in \Omega : j \in c_b(x)\}$, and for each such set, define $c(I_j) = 1/b$. We have,

$$\sum_{I\in\mathcal{I}_x(W)} c(I) = 1,$$

for almost all $x \in \Omega$, and

$$\sum_{I\in\mathcal{I}(W)} c(I) = \frac{\chi_b}{b},$$

as required.

It remains to prove that $\chi_{\text{frac}}(W) \geq \inf_{b\to\infty} \frac{\chi_b(W)}{b}$. We can assume $\chi_{\text{frac}}(W)$ is bounded. The next claim allows us to restrict ourselves to fractional colorings with weights that are a multiple of $\frac{1}{b}$.

Claim. Suppose that $\epsilon > 0$ is fixed. For any b sufficiently large, there exists a fractional coloring $c : \mathcal{I}(W) \to \{\frac{j}{b} : j = 0, 1, \ldots, b\}$ with finite support such that $\sum_{I\in\mathcal{I}(W)} c(I) \leq \chi_{\text{frac}}(W) + \epsilon$.

Proof of Claim. Take $\delta > 0$ such that

$$\frac{\chi_{\text{frac}}(W) + \delta}{1 - \delta} \leq \chi_{\text{frac}}(W) + \frac{\epsilon}{2}.$$

Take a fractional coloring c_0 of W such that $\sum_{I\in\mathcal{I}(W)} c_0(I) \leq \chi_{\text{frac}}(W)+\delta$. Since $\sum_{I\in\mathcal{I}(W)} c_0(I)$ is finite, we can find a finite subset $\mathcal{I}_0 \subseteq \mathcal{I}(W)$ such that

$$\sum_{I\in\mathcal{I}(W)\setminus\mathcal{I}_0} c_0(I) < \delta\,. \tag{4.3}$$

Define a function c_1 by setting it equal to c_0 if $I \subseteq \mathcal{I}_0$ and 0 otherwise. The function c_1 is not necessarily a valid fractional coloring, but we can turn it into a fractional coloring c_2,

$c_2(I) := \frac{c_1(I)}{1-\delta}$. The fact that c_2 is a valid fractional coloring follows from (4.3). We have

$$\sum_{I\in\mathcal{I}(W)} c_2(I) = \frac{\sum_{I\in\mathcal{I}(W)} c_1(I)}{1-\delta} \le \frac{\chi_{\mathrm{frac}}(W)+\delta}{1-\delta} \le \chi_{\mathrm{frac}}(W) + \frac{\epsilon}{2},$$

by our choice of δ.

Let $b > \frac{2\cdot|\mathcal{I}_0|}{\epsilon}$. We can define c by rounding up $c_2(I)$ to the closest multiple of $1/b$. This way, c is still a proper coloring, where we have increased the total sum of weight on independent sets by at most $\frac{|\mathcal{I}_0|}{b}$. Thus, we obtain

$$\sum_{I\in\mathcal{I}(W)} c(I) \le \sum_{I\in\mathcal{I}(W)} c_2(I) + \frac{|\mathcal{I}_0|}{b} \le \sum_{I\in\mathcal{I}(W)} c_2(I) + \frac{\epsilon}{2} \le \chi_{\mathrm{frac}}(W) + \epsilon.$$

Now, c is a rational fractional coloring with common denominator b, as required. This finishes the proof of the Claim. □

Let us fix $\epsilon > 0$. For each b sufficiently large, we describe how to transform a fractional coloring $c : \mathcal{I}(W) \to \{\frac{j}{b} : j = 0, 1, \dots, b\}$ from the Claim (for the given ϵ) into a b-fold coloring $p : \Omega \to \binom{[K]}{b}$, where $K = b\sum_{I\in\mathcal{I}(W)} c(I)$ (recall that K is an integer). Actually, it is slightly easier to allow p to map to subsets of $[K]$ of cardinality *at least* b. Any such p can be obviously corrected eventually. Let us partition $[K]$ into sets $\bigsqcup_{I\in\mathcal{I}(W)} A_I$ so that the size of each set A_I is $b \cdot c(I)$. Then for each $x \in \Omega$ we define $p(x) := \bigcup_{I\in\mathcal{I}_x(W)} A_I$. The map p is obviously a proper b-fold coloring with $\sum_{I\in\mathcal{I}(W)} c(I) + \epsilon \ge \frac{\chi_b(W)}{b}$, as was needed. □

5. Cliques and fractional cliques

Below, we define the clique number and the fractional clique number of a graphon.

5.1. **Cliques.** We first introduce the clique number of a graphon.

Definition 24. For a graphon W, the *clique number* $\omega(W)$ is defined as

$$\omega(W) = \max\{r : t(K_r, W) > 0\}.$$

Lower semicontinuity of the clique number follows immediately from the definition. (As mentioned in Section 1.2, the sequence $Y_n \equiv \frac{1}{n} \to Y \equiv 0$ shows that we do not have upper semicontinuity in general.)

Proposition 25. *For every sequence of graphons $(W_n)_n$ that converges to W in the cut-distance we have $\omega(W) \le \liminf_{n\to\infty} \omega(W_n)$.*

Proof. If $\omega(W) \ge r$, then $t(K_r, W) > 0$. Since subgraph densities are continuous in the cut-distance, we have $t(K_r, W_n) > 0$ for almost all n. In particular, $\omega(W_n) \ge r$ for almost all n. □

5.2. **Fractional cliques.** In analogy with the finite counterpart, we make the following definition.

Definition 26. Suppose that $W : \Omega \times \Omega \to [0,1]$ is a graphon. We say that a measurable function $f : \Omega \to [0,+\infty)$ is a *fractional clique* if for every $I \in \mathcal{I}(W)$ we have $\int_I f \le 1$. The *size* of a fractional clique f is $\|f\| := \int f$. We define the *fractional clique number* of W as

$$\omega_{\text{frac}}(W) = \sup \int_\Omega f,$$

where the supremum is taken over all fractional cliques in W.

In analogy with Proposition 17, we first prove that the fractional clique number was introduced in a way that is consistent with the graph version.

Proposition 27. *Suppose G is a finite graph and let W be its graphon representation. Then $\omega_{\text{frac}}(G) = \omega_{\text{frac}}(W)$.*

Proof. Let $\Omega = \Omega_1 \sqcup \cdots \sqcup \Omega_n$ be the partition for W corresponding to the vertices of G. For any fractional clique $x : V(G) \to [0,\infty)$ of the graph G define a function $f : \Omega \to [0,\infty)$ with constant value $x(v)n$ in the interval Ω_v for each $v \in V(G)$. There is a map $\pi : \mathcal{I}(G) \to \mathcal{I}(W)$ which maps independent sets in $\mathcal{I}(G)$ to corresponding independent sets in $\mathcal{I}(W)$. Let us verify that f is a fractional clique for W. Let $I \in \mathcal{I}(W)$ be arbitrary. Let $J \subset V(G)$ consist of the vertices v for which $I \cap \Omega_v$ has positive measure. Then J is an independent set in G. We then have

$$\int_I f \le \sum_{v \in J} \int_{\Omega_v} f = \sum_{v \in J} x(v) \le 1 \,,$$

where the last inequality follows uses that x in a fractional clique in G. Obviously, for the size of f we have $\int_\Omega f = \sum_{v \in V(G)} x(v)$. Therefore, $\omega_{\text{frac}}(G) \le \omega_{\text{frac}}(W)$.

Now suppose $f : \Omega \to [0,\infty)$ is any fractional clique of W. Define a function $x : V(G) \to [0,\infty)$, where $x(v) = \int_{\Omega_v} f$. For each independent set of $J \in \mathcal{I}(G)$, we have

$$\sum_{v \in J} x(v) = \sum_{v \in J} \int_{\Omega_v} f = \int_{\pi(J)} f \le 1 \,.$$

Thus, x is a valid fractional clique for G with size

$$\sum_{v \in V(G)} x(v) = \sum_{v \in V(G)} \int_{\Omega_v} f = \int_\Omega f.$$

That implies $\omega_{\text{frac}}(G) \ge \omega_{\text{frac}}(W_G)$, and finishes the proof. □

We now prove lower semicontinuity of the fractional clique number.

Theorem 28. *Let $(W_n)_n$ be a sequence of graphons converging to W in the cut-norm. Then*

$$\omega_{\text{frac}}(W) \leq \liminf_{n\to\infty} \omega_{\text{frac}}(W_n) \ .$$

Proof. Let $f : \Omega \to [0,\infty)$ be a fractional clique of W. We will show that for each graphon W_n there is a fractional clique f_n such that $\liminf_{n\to\infty} \int_\Omega f_n \geq \int_\Omega f$.

For each $n \in \mathbb{N}$, define

$$s_n = \sup_{I\in\mathcal{I}(W_n)} \int_I f \ .$$

We claim that $\limsup_{n\to\infty} s_n \leq 1$. Suppose by contradiction there exists $\delta > 0$ such that $\limsup_{n\to\infty} s_n > 1+\delta$. Fix a subsequence $(s_{n_i})_{i\in M}$ with $s_{n_i} > 1+\delta$ for each $i \in M$. Thus, for each $i \in M$ we can find $I_i \in \mathcal{I}(W_{n_i})$ such that $\int_{I_i} f > 1+\delta$. By the sequential Banach–Alaoglu Theorem, the sequence of indicator functions $\mathbf{1}_{I_i} \in \mathcal{L}^\infty(\Omega)$ has a weak* accumulation point, say $g \in \mathcal{L}^\infty(\Omega)$. Note that g is bounded from above by 1. Define $I = \operatorname{supp}(g)$. By Corollary 5, I is an independent set of W. We have

$$1+\delta \leq \liminf_{i\in M} \int_{I_i} f = \liminf_{i\in M} \int_\Omega f\cdot \mathbf{1}_{I_i} \leq \int_\Omega f\cdot g \leq \int_I f \ ,$$

which contradicts the fact that f is a fractional clique and I is and independent set for W.

Since $\limsup_{n\to\infty} s_n \leq 1$, we can define for each W_n a valid fractional clique $f_n : \Omega \to [0,\infty)$ for W_n by $f_n = \frac{f}{s_n}$. Furthermore, $(f_n)_n$ satisfy $\liminf_{n\to\infty} \int_\Omega f_n \geq \int_\Omega f$, as required. □

The next relations between independence number, fractional clique number, and chromatic number are expected from what we know in the graph world and this elementary proof is a straightforward adaptation from the discrete case.

Theorem 29. *For any graphon $W : \Omega\times\Omega \to [0,1]$ we have $1 \leq \omega_{\text{frac}}(W)\alpha(W)$.*

Proof. Define a function $f : \Omega \to [0,\infty)$ by $f(x) = 1/\alpha(W)$. Clearly, f is a valid fractional clique, since for every independent set $I \in \mathcal{I}(W)$ we have $\int_I f = \frac{\nu(I)}{\alpha(W)} \leq 1$. Thus, $\omega_{\text{frac}}(W) \geq \int f = 1/\alpha(W)$ and the claim follows. □

Theorem 30. *For any graphon $W : \Omega\times\Omega \to [0,1]$ we have $1 \leq \chi(W)\alpha(W)$.*

Proof. Consider a decomposition of Ω into independent sets $I_1 \sqcup I_2 \sqcup \ldots \sqcup I_{\chi(W)} = \Omega$ mod 0. Now, the inequality follows from

$$1 = \sum_{j=1}^{\chi(W)} \int_{I_j} 1 \leq \sum_{j=1}^{\chi(W)} \alpha(W) = \chi(W)\alpha(W).$$

□

Alternatively, the proof of the previous result would follow from Theorem 29 if we would have the inequality $\omega_{\text{frac}}(W) \leq \chi_{\text{frac}}(W)$. This inequality will only be obtained in Section 6, where we investigate duality properties between these parameters. That is why at this point we decided to get it from the same argument given for graphs.

However, certain properties can be more challenging to obtain. For instance, from what we know about fractional cliques in graphs it is not surprising that if a graph G satisfies $t(G, W) > 0$, then $\omega_{\text{frac}}(G) \leq \omega_{\text{frac}}(W)$. In the discrete world, this is a direct consequence of the definition. However, for graphons this statement requires a technical proof.

Proposition 31. *Suppose that G is a graph and W is a graphon with $t(G, W) > 0$. Then $\omega_{\text{frac}}(G) \leq \omega_{\text{frac}}(W)$.*

Proof. Let $\epsilon > 0$ and let $g : V(G) \to [0, 1]$ be a fractional clique of G. To prove the statement we shall find a fractional clique $f : \Omega \to [0, \infty)$ for W such that

$$\int_\Omega f \geq (1 - \epsilon) \sum_{v \in V(G)} g(v).$$

Let $\delta = \left(\frac{\epsilon}{|V(G)|}\right)^2$. Since $t(G, W) > 0$, by Lemma 2, we can find $\beta > 0$ disjoint sets $A_v \subseteq \Omega$, $v \in V(G)$, with positive measure $\nu(A_v) = \beta$ such that for each $uv \in E(G)$,

$$\mu^{\otimes 2}(A_v \times A_u \setminus \operatorname{supp}(W)) < \delta\beta^2. \tag{5.1}$$

Define

$$f(x) := \begin{cases} \frac{(1-\epsilon)}{\beta} g(v) & \text{for } x \in A_v \\ 0 & \text{otherwise.} \end{cases}$$

Thus, we obtain

$$\int_\Omega f = (1 - \epsilon) \sum_{v \in V(G)} g(v).$$

It remains to prove that for each $I \in \mathcal{I}(W)$ we have $\int_I f \leq 1$. It follows from (5.1) that for each $uv \in E(G)$ we have

$$\mu(I \cap A_v) \leq \sqrt{\delta}\beta \text{ or } \mu(I \cap A_u) \leq \sqrt{\delta}\beta. \tag{5.2}$$

Let $J \subset V(G)$ be the set of vertices for which $\mu(I \cap A_v) > \sqrt{\delta}\beta$. By (5.2) it follows that J is an independent set. Therefore, $\sum_{v \in J} g(v) \leq 1$. We have

$$\int_I f \leq \sum_{v \in J} \int_{A_v} f + \int_{I \setminus \cup_{v \in J} A_v} f \,. \tag{5.3}$$

The first term can be bounded by

$$\sum_{v \in J} \int_{A_v} f \leq \sum_{v \in J} \beta \frac{(1 - \epsilon)}{\beta} g(v) \leq (1 - \epsilon) \sum_{v \in J} g(v) \leq 1 - \epsilon. \tag{5.4}$$

For the second term, we have

$$\int_{I\setminus\cup_{v\in J}A_v} f = \sum_{v\in V(G)\setminus J}\int_{I\cap A_v} f.$$

For each $v \in V(G)\setminus J$, we have

$$\begin{aligned}\int_{I\cap A_v} f &\le \sqrt{\delta}\beta\frac{(1-\epsilon)}{\beta}g(v)\\ &= \sqrt{\delta}\,(1-\epsilon)\,g(v)\\ &\le \frac{\epsilon}{V(G)}\,(1-\epsilon)\,g(v)\\ &\le \frac{\epsilon}{|V(G)|}.\end{aligned}$$

Therefore,

$$\int_{I\setminus\cup_{v\in J}A_v} f \le \sum_{v\in V(G)\setminus J}\frac{\epsilon}{|V(G)|} \le \epsilon. \tag{5.5}$$

By (5.3), (5.4), and (5.5), it follows that $\int_I f \le 1$ and therefore f is a valid fractional clique for W. This finishes the proof. □

The results above allow us to express the fractional clique number of a graphon using the same parameter of finite graphs appearing in it.

Corollary 32. *Suppose that W is a graphon. Then $\omega_{\mathrm{frac}}(W) = \sup_{G:t(G,W)>0}\omega_{\mathrm{frac}}(G)$.*

Proof. Proposition 31 gives $\omega_{\mathrm{frac}}(W) \ge \sup_{G:t(G,W)>0}\omega_{\mathrm{frac}}(G)$. On the other hand, taking $G_n \sim \mathbb{G}(n,W)$, we get a sequence (G_n) of graphs which all satisfy (almost surely) that $t(G_n,W) > 0$. Considering arbitrary graphon representations W_n of these graphs G_n, we know that W_n converge to W in the cut-distance. Thus Theorem 28 tells us that $\omega_{\mathrm{frac}}(W) \le \liminf_{n\to\infty}\omega_{\mathrm{frac}}(W_n)$. As $\omega_{\mathrm{frac}}(W_n) = \omega_{\mathrm{frac}}(G_n)$ by Proposition 27, we also prove the inequality

$$\omega_{\mathrm{frac}}(W) \le \sup_{G:t(G,W)>0}\omega_{\mathrm{frac}}(G)\,.$$

□

Last, we relate the fractional clique number and the integral clique number of a graphon.

Proposition 33. *Suppose that W is a graphon. Then $\omega_{\mathrm{frac}}(W) \ge \omega(W)$.*

Proof. By Corollary 32, we have $\omega_{\mathrm{frac}}(W) = \sup_{G:t(G,W)>0}\omega_{\mathrm{frac}}(G)$. Further, Definition 24 gives $\omega(W) = \sup_{G:t(G,W)>0}\omega(G)$. We thus get the statement combining these two relations together with the fact that $\omega_{\mathrm{frac}}(G) \ge \omega(G)$ for each finite graph. □

6. Duality between fractional cliques and fractional coloring

The classical LP duality states that for a finite graph G we have that

$$\omega_{\text{frac}}(G) = \chi_{\text{frac}}(G)\ . \tag{6.1}$$

Note that such a relation cannot hold for graphons in general. Indeed, taking the graphon W constructed in Example 19, we get that

$$\omega_{\text{frac}}(W) = \sup_{G:t(G,W)>0} \omega_{\text{frac}}(G) \overset{(6.1)}{=} \sup_{G:t(G,W)>0} \chi_{\text{frac}}(G) < \chi_{\text{frac}}(W)\ .$$

So, in this section we establish the maximum that can be established in this situation: weak LP duality and complementary slackness.

Theorem 34 (Weak LP duality)**.** *For every graphon $W : \Omega \times \Omega \to [0,1]$ we have*

$$\omega_{\text{frac}}(W) \le \chi_{\text{frac}}(W).$$

Proof. Suppose $f : \Omega \to [0,1]$ is an arbitrary fractional clique of W and $c : \mathcal{I}(W) \to [0,1]$ is an arbitrary fractional coloring of W. Since $\sum_{I \in \mathcal{I}_x(W)} c(I) \ge 1$ for almost every $x \in \Omega$, we can write

$$\int_{x\in\Omega} f(x) \le \int_{x\in\Omega} \left(f(x) \sum_{I\in\mathcal{I}_x(W)} c(I) \right) = \sum_{I\in\mathcal{I}(W)} \left(c(I) \int_{x\in I} f(x) \right) \le \sum_{I\in\mathcal{I}(W)} c(I), \tag{6.2}$$

since $\int_I f \le 1$ for each $I \in \mathcal{I}(W)$. This finishes the proof. □

Theorem 35 (Complementary slackness)**.** *Let $W : \Omega \times \Omega \to [0,1]$ be a graphon with a fractional clique $f : \Omega \to [0,1]$ and a fractional coloring $c : \mathcal{I}(W) \to [0,1]$. We have*

$$\int_\Omega f = \sum_{I\in\mathcal{I}(W)} c(I) \tag{6.3}$$

if and only if

$$f(x) \left(\sum_{I\in\mathcal{I}_x(W)} c(I) - 1 \right) = 0 \tag{6.4}$$

for almost every $x \in \Omega$ and

$$c(I) \left(\int_I f - 1 \right) = 0 \tag{6.5}$$

for every $I \in \mathcal{I}(W)$. In this case, for every $A \subset \Omega$ with positive measure we have $f\,|_A = 0$ almost everywhere if and only if

$$\sum_{I\in\mathcal{I}_x(W)} c(I) > 1$$

for almost every $x \in A$.

Proof. Notice that (6.3) holds if and only if both inequalities in (6.2) are at equality. Since f and c are nonnegative, (6.4) and (6.5) follow. This proves the first part. To get second part , it is enough to recall that$\sum_{I\in\mathcal{I}_x(W)} c(I) \geq 1$. □

7. Perfect graphons

The notion of perfect graph is central in combinatorial optimization. Recall that a graph G is *perfect* if for every induced subgraph H of G, we have $\chi(H) = \omega(H)$. In the remarkable work of Chudnovsky, Robertson, Seymour, and Thomas [3], a forbidden subgraph characterization for perfect graphs is given. The result settles a problem that for four decades was known as the strong perfect graph conjecture. It can be stated as follows.

Theorem 36 (Strong perfect graph Theorem)**.** *A graph is perfect if and only if it contains no induced odd cycle of length more than 4 and no complement of an induced of cycle of length more than 4.*

It is not clear what should be the right definition of perfect graphons. We offer two definitions. The first definition is in terms of subgraph densities.

Definition 37. We call a graphon $W : \Omega \times \Omega \to [0, 1]$ *subgraph-perfect* if for every finite graph H with $t_{\rm ind}(H, W) > 0$ we have $\chi(H) = \omega(H)$. Further, we call W *properly subgraph-perfect* if it is subgraph-perfect and $\chi(W) < \infty$.

The other definition of perfect graphons mimics more closely the graph definition. Note that the property $\chi(G) \geq \omega(G)$ which holds for finite graphs and was a motivation for the notion of perfect graphons holds for graphons, too.

Definition 38. We call a graphon $W : \Omega \times \Omega \to [0, 1]$ *inheritance-perfect* if for every set $A \subset \Omega$ of positive measure we have $\chi(W[A]) = \omega(W[A])$. Further, we call W *properly inheritance-perfect* if it is induced-perfect and $\chi(W) < \infty$.

We can characterize subgraph-perfect graphons in the same fashion as perfect graphs by the density of induced odd cycles and its complements.

Proposition 39. *A graphon W is subgraph-perfect if and only if for every odd integer $\ell \geq 5$ we have $t_{\rm ind}(C_\ell, W) = 0$ and $t_{\rm ind}(\overline{C_\ell}, W) = 0$.*

Proof. For every odd integer $\ell \geq 5$, we have $\omega(C_\ell) < \chi(C_\ell)$ and $\omega(\overline{C_\ell}) < \chi(\overline{C_\ell})$. Thus, if W is subgraph-perfect, then $t_{\rm ind}(C_\ell, W) = 0$ and $t_{\rm ind}(\overline{C_\ell}, W) = 0$.

Now, assume W is not subgraph-perfect. Then, there is a finite imperfect graph H with $t_{\rm ind}(H, W) > 0$. By the Strong perfect graph theorem we have that $t_{\rm ind}(C_\ell, H) > 0$

or $t_{\mathrm{ind}}(\overline{C_\ell}, H) > 0$ for some odd integer $\ell \geq 5$. By Exercise 7.6[2] of [10] we get that $t_{\mathrm{ind}}(C_\ell, W) > 0$ or $t_{\mathrm{ind}}(\overline{C_\ell}, W) > 0$. This finishes the proof. □

Recall that the *complement* of a graphon W is defined as $\overline{W}(x, y) = 1 - W(x, y)$ for every $x, y \in \Omega$. Naturally, we obtain the weak version of the last theorem.

Corollary 40. *A graphon is subgraph-perfect if and only if its complement is subgraph-perfect.*

We now relate subgraph-perfectness and induced-perfectness.

Theorem 41. *Let $W : \Omega \times \Omega \to [0, 1]$ be a properly subgraph-perfect graphon. Then W is inheritance-perfect.*

Proof. We shall first prove that $\omega(W) = \chi(W)$. To this end, consider the sequence in the statement of Theorem 14 (b). Let G_n be the graphs in that sequence and let W_n be their graphon representation. Now, we have

$$\omega(W) = \lim_{n\to\infty} \omega(G_n). \tag{7.1}$$

Furthermore, by Theorem 14 we almost surely have

$$\chi(W) = \lim_{n\to\infty} \chi(G_n). \tag{7.2}$$

Since W is properly subgraph-perfect, the graphs G_n are all perfect (almost surely). Thus, $\omega(G_n) = \chi(G_n)$, and $\omega(W) = \chi(W)$ follows from (7.1) and (7.2).

Now, observe that induced-perfectness follows. Indeed, if W is subgraph-perfect and we are given a set $A \subset \Omega$ of positive measure, we can run the above argument for the graphon $W[A]$ (which is obviously subgraph-perfect) to conclude that $\chi(W[A]) = \omega(W[A])$. □

We cannot reverse Theorem 41 directly. For example, consider W to be a graphon representation of a triangle K_3. Now, take $U := \frac{1}{2}W$, that is, U is a graphon representation of a triangle, where the edges are represented by value $\frac{1}{2}$. Obviously, U is inheritance-perfect but we have $t_{\mathrm{ind}}(C_5, W) > 0$. However, we do not know whether we can reverse Theorem 41 for $\{0, 1\}$-valued graphons. We pose this as an open problem.

Problem 42. Suppose that W is a $\{0, 1\}$-valued inheritance-perfect graphon. Is W also subgraph-perfect?

[2]Exercise 7.6 in [10] is stated for the parameter $t(\cdot, \cdot)$ rather than $t_{\mathrm{ind}}(\cdot, \cdot)$. The statement and the proof hold *mutatis mutandis*.

Our last open problem is the "weak perfect graph theorem" for inheritance-perfect graphons. Note that a positive answer to this problem would be automatically implied by a positive answer to Problem 42.

Problem 43. Suppose that W is a $\{0,1\}$-valued inheritance-perfect graphon. Is its complement $\overline{W}$ also inheritance-perfect?

References

[1] C. Bachoc, E. DeCorte, F.M. de Oliveira Filho, and F. Vallentin. Spectral bounds for the independence ratio and the chromatic number of an operator. *Israel Journal of Mathematics*, 202(1):227–254, 2014.

[2] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[3] M. Chudnovsky, N. Robertson, P. Seymour, and R. Thomas. The strong perfect graph theorem. *Ann. of Math. (2)*, 164(1):51–229, 2006.

[4] O. Cooley, M. Kang, and O. Pikhurko. Size convergence of graphons. manuscript in preparation.

[5] M. Doležal, J. Hladký, and A. Máthé. Cliques in dense inhomogeneous random graphs. *Random Structures Algorithms*, 51(2):275–314, 2017.

[6] M. Doležal and J. Hladký. Matching polytons. arXiv:1606.06958.

[7] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics: Study and Research Texts*. Springer-Verlag, Berlin, 1988.

[8] J. Hladký, P. Hu, and D. Piguet. Komlós's tiling theorem via graphon covers. *J. Graph Theory*, 90(1):24–45, 2019.

[9] I. Leader. The fractional chromatic number of infinite graphs. *J. Graph Theory*, 20(4):411–417, 1995.

[10] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[11] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

Institute of Mathematics of the Czech Academy of Sciences, Žitná 25, Praha, Czechia. The Institute of Mathematics is supported by RVO:67985840. *This work was done while affiliated with* Institut für Geometrie, TU Dresden, 01062 Dresden, Germany.

The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, 182 07 Prague, Czech Republic. With institutional support RVO:67985807.

APPROXIMATING FRACTIONALLY ISOMORPHIC GRAPHONS

JAN HLADKÝ AND ENG KEAT HNG

Abstract. Grebík and Rocha [*Fractional Isomorphism of Graphons, Combinatorica 42, pp 365–404 (2022)*] extended the well studied notion of fractional isomorphism of graphs to graphons. We prove that fractionally isomorphic graphons can be approximated in the cut distance by fractionally isomorphic finite graphs. This answers the main question from *ibid*. As an easy but convenient corollary, we deduce that every regular graphon can be approximated by regular graphs.

1. Introduction

The theory of limits of dense graph sequences, introduced in [9, 1] and nicely surveyed in [8], provides a powerful correspondence between (finite) graphs on the one hand and analytic objects called graphons on the other. In the motivating part of this introduction, we shall refer to this theory rather casually and do not need exact definitions.

The two main results of the theory of dense graph limits say that

(R1) each sequence of graphs contains a subsequence converging[1] to a graphon, and
(R2) each graphon has a sequence of graphs converging to it.

A key question is how various graph properties and parameters relate to their graphon counterparts in the context of (R1) and (R2). One parameter which is well understood is homomorphism density: by the continuity of homomorphism densities, *every* convergent sequence of graphs has homomorphism densities with respect to each fixed graph which converge to that in the limit graphon. On the other hand, there are properties and parameters of graphons which are more subtle in that they are not reflected in *every* sequence of graphs converging to a certain graphon, but rather in *at least one* sequence. Take for example the concept of vertex-transitivity, which was treated in detail for graphons in [10]. It was shown there that the limit graphon of a convergent sequence of vertex-transitive graphs is vertex-transitive; this relates to (R1). Regarding (R2), we may ask if every vertex-transitive graphon is the limit of at least one sequence of vertex-transitive graphs; this seems to be a difficult question (see [2]). Clearly, not every sequence of graphs with a vertex-transitive limit graphon is a sequence of vertex-transitive graphs (hint: delete a single edge). Indeed, many of these more subtle properties concern symmetries.

Research supported by Czech Science Foundation Project GX21-21762X and with institutional support RVO:67985807.

An extended abstract titled 'Fractionally isomorphic graphs and graphons' summarizing this work will appear in the proceedings of Eurocomb2023.

[1]convergence is meant here and elsewhere to be in the cut distance (denoted by $\delta_\square$)

In this paper we use this point of view to study fractional isomorphism. The concept of fractional isomorphism for graphs was introduced by Tinhofer in 1986 [14], and several important equivalent but very different-looking definitions were later added by Ramana, Scheinerman, and Ullman [11] and another by Dvořák [4] and independently by Dell, Grohe, and Rattan [3]. We refer the reader to Section 6 of [13] for a treatise of these concepts and recall here the definition introduced in [11], which is the most convenient to work with in this paper. An *equitable partition* of a graph G is a partition $\mathcal{P} = \{P_i\}_{i\in[k]}$ of $V(G)$ into disjoint nonempty parts such that for all $i, j \in [k]$ and $u \in P_i$ we have

$$\deg_G(u; P_j) = \frac{1}{|P_i|} \sum_{v\in P_i} \deg_G(v; P_j) \; . \tag{1}$$

The *parameters* of $\mathcal{P}$ are a pair $((p_i)_{i\in[k]}, (D_{i,j})_{i,j\in[k]})$ where $p_i = |P_i|$ for all $i \in [k]$ and $D_{i,j} = \deg_G(u; P_j)$ for all $i, j \in [k]$ and $u \in P_i$. We say that two graphs G and H are *fractionally isomorphic* if they have equitable partitions $\mathcal{P}_G$ and $\mathcal{P}_H$ which can be indexed in a way so that they have the same parameters. As a first example, observe that taking equitable partitions with $k = 1$ in the definition yields that any two d-regular graphs on n vertices are fractionally isomorphic.

In [6], Grebík and Rocha developed a theory of fractional isomorphism for graphons. In particular, they showed that all the equivalent definitions of fractional isomorphism of graphs previously studied have graphon counterparts and that these are indeed all equivalent. It follows from their results that if $\{G_n\}_{n\in\mathbb{N}}$ and $\{H_n\}_{n\in\mathbb{N}}$ are sequences of graphs so that G_n is fractionally isomorphic to H_n (for each n), $\{G_n\}_{n\in\mathbb{N}}$ converges to a graphon W and $\{H_n\}_{n\in\mathbb{N}}$ converges to a graphon U, then U and W are fractionally isomorphic. Let us sketch a proof of this result. To this end, we first recall another characterization of fractional isomorphism from [4, 3], which says that two graphs G and H on the same number of vertices are fractionally isomorphic if and only if their homomorphism counts $\hom(T, G)$ and $\hom(T, H)$ coincide for all trees T. Hence, if $\{G_n\}_{n\in\mathbb{N}}$, $\{H_n\}_{n\in\mathbb{N}}$, W and U are as above, then by the continuity of homomorphism densities we get that the homomorphism density of every tree T in W and in U are equal. This in turn is a characterization of fractional isomorphism of graphons from [6].

As with vertex-transitivity, we cannot reverse this in the strong sense: for two fractionally isomorphic graphons W and U there are many sequences $\{G_n\}_{n\in\mathbb{N}}$ converging to W and $\{H_n\}_{n\in\mathbb{N}}$ converging to U where G_n is not fractionally isomorphic to H_n. However, a weak reversal was the main open question in [6].

Question 1.1 (Question 3.2 in [6])**.** *Let W and U be fractionally isomorphic graphons. Does there exist sequences $\{G_n\}_{n\in\mathbb{N}}$ and $\{H_n\}_{n\in\mathbb{N}}$ of graphs such that G_n is fractionally*

isomorphic to H_n for each $n \in \mathbb{N}$ and

$$G_n \to_{\delta_\square} W \text{ and } H_n \to_{\delta_\square} U \text{ ?}$$

The main result of this paper is a positive answer to Question 1.1. In fact, we prove a slightly stronger statement in which we simultaneously approximate an arbitrary (even infinite) family of fractionally isomorphic graphons.

Theorem 1.2. *Given $\varepsilon > 0$ and a family $\mathcal{U}$ of fractionally isomorphic graphons, there exists $n_0 \in \mathbb{N}$ such that for each $n \geq n_0$ there exists a family $\{H_U\}_{U\in\mathcal{U}}$ of fractionally isomorphic graphs on vertex set $[n]$ such that for each $U \in \mathcal{U}$ we have $\delta_\square(U, H_U) \leq \varepsilon$.*

The simplest instance of Theorem 1.2 is the regular case. Specifically, suppose that $d \in [0, 1]$ and $\mathcal{U}$ is a family of regular graphons with degree d on a probability space (Ω, π); we say that a graphon U is *regular* with degree d if for almost every $x \in \Omega$ we have $\deg_U(x) = \int_{y\in\Omega} U(x, y)\,\mathrm{d}\pi = d$. Such graphons are known to be fractionally isomorphic (and it is easy to show this). Hence, our theorem applies and says that there are fractionally isomorphic graphs $\{H_U\}_{U\in\mathcal{U}}$ of order n which approximate $\{U\}_{U\in\mathcal{U}}$. This itself does not guarantee that the degree-regularity of the graphs H_U. Indeed, for example in our construction we might add an isolated vertex to each graph H_U (which is a step which does not spoil fractional isomorphism). However, it follows from basic properties of the cut distance that most vertices in each graph H_U have degrees around dn. Despite this, a closer look at our proof does give that the graphs H_U are in fact D-regular for some $D = (1 \pm \varepsilon)dn$. This leads to the following theorem.

Theorem 1.3. *Suppose that $d \in [0, 1]$ and $\mathcal{U}$ is a family of graphons which are fractionally isomorphic to the constant-d graphon C_d. Then for every $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for each $n \geq n_0$ there exists $D \in \mathbb{N}$ and a family $\{H_U\}_{U\in\mathcal{U}}$ of D-regular graphs on vertex set $[n]$ such that for each $U \in \mathcal{U}$ we have $\delta_\square(U, H_U) \leq \varepsilon$.*

The remainder of this paper is organised as follows. We begin with a high level sketch of the strategy of our proof of Theorem 1.2 in Section 1.1. In Section 2 we introduce notation, introduce some technical tools and recall important concepts about graphons, fractional isomorphism for graphons and graphon sampling. We also introduce a version of Szemerédi's regularity lemma suited to our purposes. In Section 3 we provide our main lemmas and apply them to prove Theorem 1.2. We provide proofs of the main lemmas in Section 3.1. Finally, in Section 4 we show how to obtain a proof of Theorem 1.3 from the proof of Theorem 1.2 given in Section 3.

1.1. **Proof idea.** Let us describe our proof of Theorem 1.2 at a high level and give relevant pointers along the way. The key concept of the theory developed in [6] is a certain quotient object $W/\mathcal{C}(W)$ for each graphon W; a formal definition is given in Section 2.4. The

importance of this object is that two graphons U and W are fractionally isomorphic if and only if $U/\mathcal{C}(U)$ and $W/\mathcal{C}(W)$ are isomorphic. To start with, we give a simplified description mimicking the definition of equitable partitions of graphs. Let (Ω, π) be a probability space. Assume in this simplification that for a graphon $W\colon \Omega^2 \to [0,1]$ the object $W/\mathcal{C}(W)$ comes from a suitably chosen finite partition $\mathcal{Q} = \{Q_i\}_{i\in[M]}$ (for some $M \in \mathbb{N}$) of Ω and has the property that for all $i, j \in [M]$ and $x \in Q_i$ we have

$$\deg_W(x; Q_j) = \frac{1}{\pi(Q_i)} \cdot \int_{y\in Q_i} \deg_W(y; Q_j)\,\mathrm{d}\pi\ , \tag{2}$$

a property clearly inspired by (1). Then $W/\mathcal{C}(W)$ serves as a counterpart to the profile, that is, it records the measures of the sets Q_i into a vector $\mathbf{r} = (r_i)_{i\in[M]}$ and the densities of pairs (Q_i, Q_j) into a matrix $\mathcal{D} = (d_{i,j})_{i,j\in[M]}$. That means $r_i = \pi(Q_i)$ for all $i \in [M]$ and $d_{i,j} = \frac{1}{\pi(Q_i)\pi(Q_j)} \int_{(x,y)\in Q_i\times Q_j} W(x,y)$ for all $i, j \in [M]$. Now since fractional isomorphism of graphons is characterized by isomorphic quotients, all the graphons in $\mathcal{U}$ have the same profile $(\mathbf{r}, \mathcal{D})$.

The standard approach to approximating each $U \in \mathcal{U}$ is to take a U-random graph $H_U \sim \mathbb{G}(n, U)$ for n large. Details about U-random graphs are given in Section 2.5; for now it is enough to recall that such a random graph has vertex set $[n]$, which corresponds to a set of n i.i.d. uniformly random points $x_1, x_2, \ldots, x_n$ from Ω, and each pair $\{i, j\}$ is inserted as an edge independently with probability $U(x_i, x_j)$. In such a setting, basic concentration tells us that with high probability,

(C1) for each $i \in [M]$ the set $X_i := \{\ell \in V(H_U) \,:\, x_\ell \in Q_i\}$ has size $|X_i| \approx r_i n$, and
(C2) for each $i, j \in [M]$ and each $\ell \in X_i$ we have $\deg_{H_U}(\ell, X_j) \approx d_{i,j} r_j n$.

Observe that if the approximate equalities above were exact, then the graphs $\{H_U\}_{U\in\mathcal{U}}$ constructed would automatically be fractionally isomorphic with the collection $\{X_i\}_{i\in[M]}$ playing the role of an equitable partition. Hence, we shall modify each H_U slightly by adding $o(n)$ vertices and $o(n^2)$ edges to obtain exact control on the number of vertices in each part and on degrees between pairs of parts. To get exact control on the number of vertices in each part, for each $i \in [M]$ we add a set Y_i of newly created vertices with $|Y_i| = \lfloor r_i n(1+\beta)\rfloor - |X_i|$ (for some small $\beta > 0$); the vertex set of the modified graph G_U shall be $(X_1 \sqcup Y_1) \cup (X_2 \sqcup Y_2) \sqcup \ldots \sqcup (X_M \sqcup Y_M)$.

To get exact control on the degrees, we would like to add edges to obtain G_U where (for some $\alpha > 0$ small)

(G1) for each $i, j \in [M]$ with $i \neq j$ and each $\ell \in X_i \cup Y_i$ we have $\deg_{G_U}(\ell, X_j \cup Y_j) = \lfloor (1+\alpha) d_{i,j} r_j n\rfloor$, and
(G2) for each $i \in [M]$ the graph $G_U[X_i \cup Y_i]$ is regular with degree $\lfloor (1+\alpha) d_{i,i} r_i n\rfloor$.

To achieve the exact degrees in (G1), for each $i, j \in [M]$ with $i \neq j$ we shall add a set of edges in the bipartite pair $(X_i \cup Y_i, X_j \cup Y_j)$ which are incident to $Y_i \cup Y_j$; we obtain these

edges by applying a theorem of Gale and Ryser. To achieve the exact degrees in (G2), for each $i \in [M]$ we shall add a set $F_{i,i}$ of edges on $X_i \cup Y_i$ which are incident to Y_i; some of these edges are obtained in the bipartite pair (X_i, Y_i) by applying the theorem of Gale and Ryser referenced above, while the remainder of these edges are obtained on Y_i by applying a theorem of Erdős and Gallai. Finally, note that since the initial graph contained no edges of the types mentioned above, the addition of such edges does not create multiedges.

The key lemmas that describe the construction outlined in the previous paragraphs are Lemmas 3.3–3.5 (which are all already tailored to the general setting, which we address below). Lemma 3.3 encapsulates a cleaning procedure for graphons whose purpose goes beyond our simplified setting, Lemma 3.4 formalizes the outcome of the random graph sampling procedure (represented by (C1) and (C2) in our simplified setting), and Lemma 3.5 deals with the subtle procedure of adding vertices and balancing degrees.

In the general setting, $U/\mathcal{C}(U)$ is defined in terms of a suitable sub-sigma-algebra and does not usually correspond to a finite partition. To see that such a setting is unavoidable in the theory of Grebík and Rocha, observe that vertices inside one set Q_i above must have the same overall degree; on the other hand, it is easy to construct graphons such that any particular degree is achieved only on a nullset of Ω.[2] Even though the Grebík–Rocha theory is inherently infinitesimal, for the steps described in (C1), (C2), (G1) and (G2) it is essential that we work with a finite partition. To bridge the gap between the two settings, we employ Szemerédi's regularity lemma. Indeed, it is well-known that if $\mathcal{Q} = \{Q_i\}_{i\in[M]}$ is an ε-regular partition of U, then in particular we have an approximate version of (2) for most vertices $x \in Q_i$. All in all, we can essentially return to the finite setting as above, with many additional technical modifications to treat various imprecisions that are unavoidable in connection with the regularity lemma.

2. Preliminaries

2.1. **Notation.** Write $\mathbb{N}$ for the set of positive integers and $\mathbb{N}_0$ for the set $\mathbb{N} \cup \{0\}$. For $a \in \mathbb{N}_0$ write $[a]$ for the set $\{1, \dots, a\}$ and $[a]_0$ for $[a] \cup \{0\}$. For $\ell \in \mathbb{N}_0$ and a set S write $\binom{S}{\ell}$ for the set of all subsets of S of size ℓ. For a set S write $S^{(2)} = \{(i,j) \in S^2 : i \neq j\}$. For real numbers x and y, we write $x \pm y$ to mean an appropriate real number z satisfying $x - y \leq z \leq x + y$.

2.2. **Tools.** In this subsection we collect some tools. We first state a form of the Chernoff bounds.

[2]For example, take a graphon $W\colon [0,1]^2 \to [0,1]$ defined by $W(x,y) = (x+y)/2$.

Theorem 2.1 ([7, Theorem 2.1])**.** *Suppose that X is a sum of independent and identically distributed Bernoulli random variables. Then for $0 \leq \varepsilon \leq 3/2$ we have*

$$\mathbb{P}\left(X \geq (1+\varepsilon)\mathbb{E}[X]\right) \leq e^{-\varepsilon^2\mathbb{E}[X]/3} \text{ and } \mathbb{P}\left(X \leq (1-\varepsilon)\mathbb{E}[X]\right) \leq e^{-\varepsilon^2\mathbb{E}[X]/3}.$$

Next, we recall some tools related to degree sequences in graphs. Recall that a sequence $(c_i)_{i\in[n]}$ is *graphic* if there exists a graph whose degree sequence is $(c_i)_{i\in[n]}$. Graphic sequences are characterized by the Erdős–Gallai theorem.

Theorem 2.2 (Erdős–Gallai)**.** *A sequence $(d_i)_{i\in[n]}$ of nonnegative integers in nonincreasing order is graphic if and only if the following hold.*

(EG1) $\sum_{i=1}^n d_i$ is even.
(EG2) For each $k \in [n]$ we have $\sum_{i=1}^k d_i \leq k(k-1) + \sum_{i=k+1}^n \min\{d_i, k\}$.

Next, we state the Gale–Ryser theorem [5, 12], which characterizes the degree sequences that are realizable as bipartite graphs. More precisely, we say that two sequences $(a_i)_{i\in[n]}$ and $(b_i)_{i\in[m]}$ are *bigraphic* if there exists a bipartite graph with parts A and B with $|A| = n$ and $|B| = m$ so that the degree sequence in A matches $(a_i)_{i\in[n]}$ and the degree sequence in B matches $(b_i)_{i\in[m]}$.

Theorem 2.3 (Gale–Ryser)**.** *A pair $(a_i)_{i\in[n]}$ and $(b_i)_{i\in[m]}$ of sequences of nonnegative integers, with the former in nonincreasing order, is bigraphic if and only if the following hold.*

(GR1) We have $\sum_{i=1}^n a_i = \sum_{i=1}^m b_i$.
(GR2) For each $k \in [n]$ we have $\sum_{i=1}^k a_i \leq \sum_{i=1}^m \min\{b_i, k\}$.

2.3. **Graphons.** Our notation for graphons mostly follows [8]. Throughout the paper, $(\Omega, \mathcal{B})$ is a standard Borel space endowed with a Borel probability measure π. We emphasize that π may contain atoms. Write $\mathcal{W}(\Omega)$ for the set of all bounded symmetric measurable functions $W\colon \Omega^2 \to \mathbb{R}$ and call the elements of $\mathcal{W}$ *kernels*. *Graphons*, which are denoted by $\mathcal{W}_0(\Omega)$, are then those kernels whose range is contained in $[0,1]$.

Suppose that Λ is another probability space (with an implicit sigma-algebra and measure). Given a graphon $W \in \mathcal{W}_0(\Omega)$ and a measure-preserving map $\phi\colon \Lambda \to \Omega$, we define a graphon $W^\phi \in \mathcal{W}_0(\Lambda)$ by $W^\phi(x,y) = W(\phi(x), \phi(y))$. We say that two graphons $W \in \mathcal{W}_0(\Omega)$ and $U \in \mathcal{W}_0(\Lambda)$ are *isomorphic* if there exists a measure-preserving bijection $\phi\colon \Lambda \to \Omega$ such that $W^\phi = U$.

Assume that Ω is an atomless probability space. Given a graph F on vertex set V we can construct its *graphon representation* $W_F \in \mathcal{W}_0(\Omega)$ as follows. We partition Ω into sets $(\Omega_v)_{v\in V}$ of individual measures $1/|V|$. For each $(u,v) \in V \times V$ set the graphon W_F to be equal to 1 on $\Omega_u \times \Omega_v$ if $uv \in E(F)$ and 0 otherwise. Note that W_F depends on

the choice of the partition $(\Omega_v)_{v\in V}$ and hence is not unique. However, W_F is unique up to isomorphism.

By identifying kernels which are equal almost everywhere, we may view $\mathcal{W}(\Omega)$ as a subset of various normed spaces, including $L^\infty(\Omega^2)$, and consider corresponding notions of distance on $\mathcal{W}(\Omega)$. One of the most combinatorially interesting norms on $\mathcal{W}(\Omega)$ is the *cut norm*, which is given by

$$\|W\|_\square = \sup_{S,T\subseteq\Omega} \left| \int_{S\times T} W \,\mathrm{d}\pi^2 \right| , \tag{3}$$

where the supremum is taken over all measurable subsets $S, T \subseteq \Omega$. This gives rise to the *cut norm distance* on $\mathcal{W}(\Omega)$, which plays a central role in graph limit theory and is given by

$$d_\square(U,W) = \|U - W\|_\square \ . \tag{4}$$

The *cut distance* is inspired by the cut norm distance. Suppose that we have two probability measure spaces Ω and Λ, and let $W \in \mathcal{W}_0(\Omega)$ and $U \in \mathcal{W}_0(\Lambda)$. Then we set

$$\delta_\square(U,W) = \inf_\phi d_\square(U^\phi, W) \ , \tag{5}$$

where the infimum is taken over all measure-preserving bijections $\phi\colon \Omega \to \Lambda$. Using the concept of graphon representation, we can define $\delta_\square(U,G) := \delta_\square(U,W_G)$ and $\delta_\square(H,G) := \delta_\square(W_H,W_G)$ for graphs H and G.

A deletion of a small number of vertices does not significantly alter a graph in the cut distance. This is stated in the next lemma. This lemma is probably well-known but we could not find it in literature.

Lemma 2.4. *Suppose that G is a graph on n vertices and that $U \subset V(G)$ is a set of m vertices. Then we have $\delta_\square(G, G[U]) \le 2(1-\frac{m}{n})$.*

Proof. We shall compare a graphon representation of G on an atomless probability space Ω with one of $G[U]$ on the same probability space. Let us consider a partition $\Omega = \bigsqcup_{v\in V(G)} \Omega_v$ into sets of individual measures $\frac{1}{n}$ and a partition $\Omega\setminus\bigcup_{v\in U}\Omega_v = \bigsqcup_{w\in U}\Lambda_w$ into sets of individual measures $\frac{n-m}{nm}$. In particular, for each $w \in U$ we have $\pi(\Omega_w \cup \Lambda_w) = \frac{1}{m}$. Now let W_G be the graphon representation of G with respect to the partition $(\Omega_v)_{v\in V(G)}$, and let $W_{G[U]}$ be the graphon representation of $G[U]$ with respect to the partition $(\Omega_v \cup \Lambda_v)_{v\in U}$. In particular, W_G and $W_{G[U]}$ are equal everywhere except possibly on the set $\big(\Omega \times (\bigcup_{w\in U}\Lambda_w)\big) \cup \big((\bigcup_{w\in U}\Lambda_w)\times\Omega\big)$. This set has measure $2(1-\frac{m}{n}) - (1-\frac{m}{n})^2$. □

Let $(\Omega,\mathcal{B},\pi)$ be a standard probability space. The *generalized degree function* of a graphon W is the function $\deg_W\colon \Omega\times\mathcal{B}\to\mathbb{R}$ given by

$$\deg_W(x;Y) = \int_{y\in Y} W(x,y)\,\mathrm{d}\pi \ . \tag{6}$$

The last piece of background information regarding the standard theory of graphons concerns their spectral theory. In fact, we need only the basic concept of associating a self-adjoint Hilbert–Schmidt to each graphon, and even that we only need to introduce the concept of invariant sub-sigma-algebras of a graphon. These concepts are used in this paper to a large extent as blackboxes. Each graphon $W \in \mathcal{W}_0(\Omega)$ defines an operator $T_W \colon L^2(\Omega) \to L^2(\Omega)$ by

$$(T_W f)(x) = \int_{y \in \Omega} W(x,y) f(y) \, \mathrm{d}\pi \, ,$$

where $f \in L^2(\Omega)$ and $x \in \Omega$. See [8, Section 7.5] for details.

2.4. **Fractional isomorphism for graphons.** We recall some concepts related to the theory of fractional isomorphism for graphons developed by Grebík and Rocha in [6]. Some of these concepts may seem rather cryptic, especially for a reader whose background is not in analysis. That said, a superficial understanding of this theory should suffice to check most of our arguments.

Throughout this section, $(\Omega, \mathcal{B})$ is a standard Borel space with a Borel probability measure π. We say that $\mathcal{C} \subseteq \mathcal{B}$ is a *π-relatively complete sub-sigma-algebra* if it is a sub-sigma-algebra such that we have $\pi(S \triangle T) > 0$ for all $(S,T) \in (\mathcal{B} \setminus \mathcal{C}, \mathcal{C})$. Write Θ_π for the set of all π-relatively complete sub-sigma-algebras of $\mathcal{B}$. Given $\mathcal{C} \in \Theta_\pi$, we write $L^2(\Omega, \mathcal{C})$ for the set of all functions in $L^2(\Omega)$ which are $\mathcal{C}$-measurable. Let $W \in \mathcal{W}_0(\Omega)$ be a graphon. We say that $\mathcal{C} \in \Theta_\pi$ is *W-invariant* if $T_W(L^2(\Omega, \mathcal{C})) \subseteq L^2(\Omega, \mathcal{C})$. Write $\mathcal{C}(W)$ for the minimum W-invariant relatively complete sub-sigma-algebra of $\mathcal{B}$. The existence of $\mathcal{C}(W)$ is established by the combination of Definition 5.12 and Proposition 5.13 in [6].

Graphons on a standard Borel space have a sufficiently rich structure to admit notions of averaging and quotients.

Definition 2.5. *Let $W \in \mathcal{W}_0(\Omega)$ and $\mathcal{C} \in \Theta_\pi$ be W-invariant. We define $W_{\mathcal{C}} = \mathbb{E}(W \,|\, \mathcal{B} \times \mathcal{C})$.*

We have the following lemma about $W_{\mathcal{C}}$ from [6].

Lemma 2.6 (Claim 5.7 in [6])**.** *Suppose that $W \in \mathcal{W}_0(\Omega)$ and $\mathcal{C} \in \Theta_\pi$ is W-invariant. Then $W_{\mathcal{C}} = \mathbb{E}(W \,|\, \mathcal{B} \times \mathcal{C}) = \mathbb{E}(W \,|\, \mathcal{C} \times \mathcal{C}) \in \mathcal{W}_0(\Omega)$.*

The language of quotient spaces gives an alternative formulation for the notion of conditional expectation on graphons. Let us recall Theorem E.1. in [6]. It says that for each $\mathcal{C} \in \Theta_\pi$ there is a standard Borel space $(\Omega/\mathcal{C}, \mathcal{D})$, a Borel probability measure $\pi/\mathcal{C}$ on $\Omega/\mathcal{C}$ and a measurable surjection $q_{\mathcal{C}} \colon \Omega \to \Omega/\mathcal{C}$ such that $\pi/\mathcal{C}$ is the pushforward of π via $q_{\mathcal{C}}$. Furthermore, we have a linear isometry $I_{\mathcal{C}} \colon L^2(\Omega/\mathcal{C}, \pi/\mathcal{C}) \to L^2(\Omega, \pi)$ defined by $I_{\mathcal{C}} f(x) = f(q_{\mathcal{C}}(x))$ and its adjoint $S_{\mathcal{C}} \colon L^2(\Omega, \pi) \to L^2(\Omega/\mathcal{C}, \pi/\mathcal{C})$ satisfies $I_{\mathcal{C}} \circ S_{\mathcal{C}} = \mathbb{E}(- \,|\, \mathcal{C})$.

Definition 2.7. *Let $W \in \mathcal{W}_0(\Omega)$ and $\mathcal{C} \in \Theta_\pi$ be W-invariant. We define $W/\mathcal{C} = S_{\mathcal{C}\times\mathcal{C}}(W_{\mathcal{C}})$.*

$W/\mathcal{C}$ is formally defined on $(\Omega\times\Omega)/(\mathcal{C}\times\mathcal{C})$, but it is easy to verify that there is a measure-preserving bijection $i\colon (\Omega\times\Omega)/(\mathcal{C}\times\mathcal{C}) \to \Omega/\mathcal{C}\times\Omega/\mathcal{C}$ such that $i\circ q_{\mathcal{C}\times\mathcal{C}}(x,y) = (q_{\mathcal{C}}(x), q_{\mathcal{C}}(y))$ almost everywhere. Hence, we may identify $(\Omega\times\Omega)/(\mathcal{C}\times\mathcal{C})$ with $\Omega/\mathcal{C}\times\Omega/\mathcal{C}$ and assume that $W/\mathcal{C} \in \mathcal{W}_0(\Omega/\mathcal{C})$.

By Lemma 2.6, we have $I_{\mathcal{C}\times\mathcal{C}}(W/\mathcal{C}) = W_{\mathcal{C}}$ and

$$W_{\mathcal{C}}(x,y) = W/\mathcal{C}(q_{\mathcal{C}}(x), q_{\mathcal{C}}(y)) \tag{7}$$

for almost every $(x,y) \in \Omega\times\Omega$.

Now we state a definition of fractional isomorphism for graphons. It was shown in [6, Theorem 1.2] that this is one of several equivalent characterizations of this concept.

Definition 2.8. *We say that graphons $W \in \mathcal{W}_0(\Omega)$ and $U \in \mathcal{W}_0(\Lambda)$ are* fractionally isomorphic *if the quotient graphons $W/\mathcal{C}(W)$ and $U/\mathcal{C}(U)$ are isomorphic.*

2.5. **Sampling.** We recall the notion of an inhomogeneous random graph model $\mathbb{G}(n,W)$ on a given number of vertices n and for an underlying graphon $W \in \mathcal{W}_0(\Omega)$. This procedure generalizes the Erdős–Rényi binomial random graphs and plays a central role in the theory of graph limits. For details, see [8, Chapter 10]. Given $n \in \mathbb{N}$ and a graphon W, the random graph $\mathbb{G}(n,W)$ is defined as follows. The vertex set of $\mathbb{G}(n,W)$ is $[n]$. To determine the edge set, sample n independent π-random elements $x_1,\dots,x_n$ from Ω and connect each pair $\{i,j\} \in \binom{[n]}{2}$ of distinct vertices independently with probability $W(x_i,x_j)$. Note that the combination of this procedure with a partition of Ω naturally induces a (random) partition of $[n]$ as follows. Given a partition $\mathcal{P} = \{P_j\}_{j\in J}$ of Ω, the *partition* $\mathcal{X} = \{X_j\}_{j\in J}$ *induced by* $\mathcal{P}$ is the partition of $[n]$ in which each cell X_j of $\mathcal{X}$ is the set of all vertices $i \in [n]$ for which $x_i \in P_j$.

The following lemma tells us that for sufficiently large n the random graph $\mathbb{G}(n,W)$ is close to W with high probability.

Lemma 2.9 ([8, Lemma 10.16])**.** *For each $n \in \mathbb{N}\setminus\{1\}$ and each graphon W we have*

$$\mathbb{P}\left(\delta_\square(\mathbb{G}(n,W),W) \le \frac{22}{\sqrt{\log_2 n}}\right) \ge 1 - \exp\left(-\frac{n}{2\log_2 n}\right).$$

2.6. **Regularity.** Our proof involves determining in advance the parameters of the common equitable partition of the fractionally isomorphic graphs we aim to construct. A key idea is that we shall obtain the template for this from an application of a regularity lemma to the common isomorphic quotient $W/\mathcal{C}(W)$. Here we give a few definitions and state the regularity lemma we use.

Definition 2.10. *Let $W \in \mathcal{W}_0(\Omega)$ be a graphon and $C, D \subseteq \Omega$ be disjoint measurable sets of positive measure. The* density of (C, D) *in W is defined as*

$$d(C, D) = \frac{1}{\pi(C)\pi(D)} \int_{C\times D} W \, \mathrm{d}\pi^2 \, .$$

Let $\delta > 0$. We say that (C, D) is δ-regular *in W if for all measurable subsets $C' \subseteq C$ and $D' \subseteq D$ we have*

$$\left| \pi(C')\pi(D')d(C, D) - \int_{C'\times D'} W \, \mathrm{d}\pi^2 \right| \leq \delta\pi(C)\pi(D) \, . \tag{8}$$

The only property of regular pairs we need is the following well-known fact.

Fact 2.11. *Let $W \in \mathcal{W}_0(\Omega)$ be a graphon, $\delta > 0$, $d \in [0, 1]$, and (C, D) be a δ-regular pair of density d in W. Then for the sets*

$$R^+ := \left\{ x \in C : \deg_W(x; D) \leq (d - \sqrt{\delta})\pi(D) \right\} \text{ and}$$
$$R^- := \left\{ x \in C : \deg_W(x; D) \geq (d + \sqrt{\delta})\pi(D) \right\}$$

we have $\pi(R^+) \leq \sqrt{\delta}\pi(C)$ and $\pi(R^-) \leq \sqrt{\delta}\pi(C)$.

Here is a curious question not directly related to the rest of the paper. The only property of regular pairs we need in this paper is that we can control their degrees in the way stated in Fact 2.11. This property, which one might call 'degree-homogeneity', is much weaker than the original property of regularity. Could one directly prove a degree-homogeneity analogue of Szemerédi's regularity lemma with much better quantitative bounds than those in Szemerédi's regularity lemma?

In the setting of the regularity lemma, the edge distribution[3] is usually controlled only across pairs of clusters, that is, in the spirit of Definition 2.10. Here we need to introduce a notion that captures the property of having the same degree[4] *inside* a cluster.

Definition 2.12. *Let $W \in \mathcal{W}_0(\Omega)$ be a graphon and $C \subseteq \Omega$ be a set of positive measure. The* density inside C *in W is given by*

$$d(C) = \frac{1}{\pi(C)^2} \int_{C\times C} W \, \mathrm{d}\pi^2 \, .$$

We say that C is elegant *in W if we have $\deg_W(x; C) = \pi(C)d(C)$ for almost every $x \in C$. Otherwise, we say that C is* inelegant *in W. To be consistent with the concept of density of pairs, we sometimes write $d(C, C)$ instead of $d(C)$ and call this quantity the density of (C, C) in W.*

[3] in case of graphs; in the graphon setting, one should perhaps say 'value distribution'

[4] 'having the same degree' is normally called 'regularity', but such a notion would lead to a confusion in combination with Definition 2.10

Definition 2.13. *Let $W \in \mathcal{W}_0(\Omega)$ be a graphon and $\delta > 0$. We say that a partition $\mathcal{P} = \{P_i\}_{i\in[M]}$ of Ω is a* δ-regular partition *for W if we have*

$$\sum_{(*)} \pi(P_i)^2 + \sum_{(**)} \pi(P_i)\pi(P_j) \leq \delta \ ,$$

where $(*)$ denotes a summation over all $i \in [M]$ for which P_i is inelegant, and $(**)$ denotes a summation over all $(i,j) \in [M]^{(2)}$ for which (P_i, P_j) is not a δ-regular pair in W.

We call the vector $\mathbf{v}_\mathcal{P} = (\pi(P_i))_{i\in[M]}$ the *cluster footprint of* $\mathcal{P}$. Given a finite partition $\mathcal{P}$ of Ω, we shall often consider it ordered with respect to a fixed order, $\mathcal{P} = \{P_i\}_{i\in[M]}$.

Now we state the version of Szemerédi's regularity lemma for graphons which is tailored to our purposes.

Theorem 2.14. *For each $\delta > 0$ there exists $M^{\mathrm{reg}}(\delta) \in \mathbb{N}$ such that for every standard probability space $(\Omega, \mathcal{B}, \pi)$, every graphon $W \in \mathcal{W}_0(\Omega)$ has a δ-regular partition into at most $M^{\mathrm{reg}}(\delta)$ parts.*

Let us highlight two ways in which our version of the regularity lemma differs from the usual statements of the regularity lemma (for graphons). Firstly, our probability space may have atoms. We remark that this is not entirely novel because the traditional regularization of graphs does not split vertices, which can be thought of as atoms. Secondly, aside from controlling the degrees between clusters, we also control the degrees *inside* clusters using the notion of elegance. These two features come together in the sense that the (square of the) measure contained in the atoms of Ω may be very substantial in our applications, and hence we have to control them. A notable example of a space with large atoms is $\Omega = \Omega'/\mathcal{C}(W')$ for some regular graphon $W' \in \mathcal{W}_0(\Omega')$. Indeed, Ω is a singleton with the Dirac measure in this example.

Here we shall sketch the proof of Theorem 2.14. First, we enumerate the atoms in Ω which have individual measures exceeding $\delta/4$ as $A = \{a_1, \dots, a_m\}$. Then, we can partition $\Omega \setminus A$ into sets $Q_1, \dots, Q_\ell$ of individual measures between $\delta/4$ and $\delta/2$. Note that each cluster $\{a_i\}$ is automatically elegant. Hence, the sum of squares of measures of inelegant clusters for the partition $\mathcal{Q}_0 = A \cup \{Q_1, \dots, Q_\ell\}$ is at most $\delta/2$, and this stays true for any further refinement. Now we start the usual cluster-refinement procedure with the initial partition $\mathcal{Q}_0$. (Trivial note: atoms do not get split by the refinement procedure.) We repeatedly refine the clusters along irregular pairs until the total contribution of $\pi(C)\pi(D)$, taken over all irregular pairs (C, D), is less than $\delta/2$. The usual index-pumping argument tells us that the complexity of the final partition is bounded. All in all, we have an outcome as needed for Theorem 2.14.

3. Main lemmas and proof of Theorem 1.2

We begin with two technical definitions for our proof of Theorem 1.2. These definitions are used to control certain key quantities in our proof.

Definition 3.1. *For constants $\beta, \lambda, \delta, \alpha > 0$ we say that we have* $(\beta, \lambda, \delta, \alpha)$-Setup *if*

$$\beta \leq \frac{1}{1000}\,, \tag{9}$$

$$\lambda \leq \frac{\beta^3}{1000}\,, \tag{10}$$

$$\delta \leq \beta^2\lambda^2\,,\ \textit{and} \tag{11}$$

$$\beta - \alpha \in [20\lambda, \beta^3/5]\,. \tag{12}$$

For $M, m \in \mathbb{N}$ we say that we have $(\beta, \lambda, \delta, \alpha, M, m)$-Setup *if in addition to the above we have $m \geq \frac{10^4 M^2 2^{10^3\beta^{-2}}}{\beta^2\lambda^3\delta^4}$.*

Definition 3.2. *Suppose that $\beta > 0$ and $M \in \mathbb{N}$ are given. An $M \times M$ matrix $\mathcal{D} = (d_{i,j})_{i,j\in[M]}$ is a* β-robust density matrix *if it is symmetric and for all $i, j \in [M]^2$ we have $d_{i,j} \in [\beta, 1-\beta]$ or $d_{i,j} = 0$.*

We say that a vector $\mathbf{v} = (\mathbf{v}_i)_{i\in[M]} \in \mathbb{R}^M$ is *unit* if we have $\mathbf{v}_i \in [0,1]$ for all $i \in [M]$ and $\sum_{i\in[M]} \mathbf{v}_i = 1$.

Before stating our main lemmas, let us recall the structure of our proof of Theorem 1.2. The proof proceeds in three main steps. First, we tidy up each graphon according to a regular partition on the common isomorphic quotient. We set the density to zero in all inelegant clusters and on all pairs which are irregular or have extremely low density. We also impose a uniform density away from 1 on pairs with density very close to 1 and smooth over atypical degrees in regular pairs. This is the subject of Lemma 3.3. Next, we apply sampling procedure from Section 2.5 to each tidied up graphon with highly controlled degrees to generate graphs which are close approximations and have partitions that resemble common equitable partitions. This is the subject of Lemma 3.4. Finally, we carefully add vertices and add and/or delete edges so that we exactly attain the parameters of a common equitable partition. This is the subject of Lemma 3.5 and involves applications of Theorems 2.2 and 2.3.

Lemma 3.3. *Assume that $(\Omega, \mathcal{B}, \pi)$ is a standard Borel probability space. Assume a $(\beta, \lambda, \delta, \alpha)$-Setup. Given a graphon $W' \in \mathcal{W}_0(\Omega)$, there exists an integer $M \leq M^{\mathrm{reg}}(\delta)$, a unit vector $\mathbf{v} \in [0,1]^M$ and a β-robust density matrix $\mathcal{D} = (d_{i,j})_{i,j\in[M]}$ such that the following hold. Suppose that $W \in \mathcal{W}_0(\Omega)$ is a graphon such that $W/\mathcal{C}(W)$ is isomorphic to $W'/\mathcal{C}(W')$. Then there is a graphon $W_\beta \in \mathcal{W}_0(\Omega)$ with a partition $\mathcal{Q} = \{Q_i\}_{i\in[M]}$ of Ω with cluster footprint $\mathbf{v}$ such that the following hold.*

(GE1) For each $(i,j) \in [M]^2$ and each $x \in Q_i$ we have

$$\deg_{W_\beta}(x; Q_j) = (1 \pm 4\lambda) d_{i,j} \mathbf{v}_j \ .$$

(GE2) $\|W - W_\beta\|_\square \leq \|W - W_\beta\|_1 \leq 4\beta$.

Lemma 3.4. *Assume that $(\Omega, \mathcal{B}, \pi)$ is a standard Borel probability space. Assume a $(\beta, \lambda, \delta, \alpha, M, m)$-Setup. Suppose that we have a β-robust density matrix $\mathcal{D} = (d_{i,j})_{i,j\in[M]}$, a unit vector $\mathbf{v} \in [0,1]^M$, and a graphon $W \in \mathcal{W}_0(\Omega)$ with a partition $\mathcal{Q} = \{Q_i\}_{i\in[M]}$ of Ω with cluster footprint $\mathbf{v}$. Suppose further that for each $(i,j) \in [M]^2$ and each $x \in Q_i$ we have*

$$\deg_W(x; Q_j) = (1 \pm 4\lambda) d_{i,j} \mathbf{v}_j \ . \tag{13}$$

Then there is a graph F on m vertices with a partition $\mathcal{X} = \{X_i\}_{i\in[M]}$ of $V(F)$ such that the following hold.

(FS1) For each $i \in [M]$ for which $\mathbf{v}_i \geq \frac{\delta}{M}$ we have $|X_i| = (1 \pm \lambda)\mathbf{v}_i m$.
(FS2) For each $(i,j) \in [M]^2$ for which $\mathbf{v}_j \geq \frac{\delta}{M}$ and each $x \in X_i$ we have

$$\deg_F(x; X_j) = (1 \pm 6\lambda) d_{i,j} \mathbf{v}_j m \ .$$

(FS3) We have $\delta_\square(W, W_F) \leq \frac{22}{\sqrt{\log_2 m}}$.

As we advertised, in our last lemma we want to modify the graph F from Lemma 3.4 so that the approximate equalities from (FS1) and (FS2) become exact ones and in particular the bipartite graphs between different parts become biregular. This necessarily runs into issues of integrality. For example, it is not possible to construct a bipartite graph of nontrivial density (i.e. $\neq 0, 1$) for parts of coprime number of vertices. To circumvent this, we introduce the following rounding procedures. Let $\Gamma \in \mathbb{N}$. For a real number $x \in \mathbb{R}$, we set $[\![x]\!]_\Gamma := \Gamma \cdot \lfloor \frac{x}{\Gamma} \rfloor$. For a matrix $\mathcal{A} = (a_{i,j})_{i,j\in[M]}$, its Γ-*rounding* $\mathcal{B} = (b_{i,j})_{i,j\in[M]}$ is the matrix in which $b_{i,j} = \lceil \Gamma a_{i,j} \rceil / \Gamma$.

Lemma 3.5. *Assume a $(\beta, \lambda, \delta, \alpha, M, m)$-Setup. Assume that an even number $\Gamma \in \mathbb{N}$ satisfies $\Gamma \geq \delta^{-2}$ and that $m \geq \Gamma M/\delta^2$. Suppose that we have a β-robust density matrix $\mathcal{D} = (d_{i,j})_{i,j\in[M]}$, a unit vector $\mathbf{v} \in [0,1]^M$, and a graph F on m vertices with a partition $\mathcal{X} = \{X_i\}_{i\in[M]}$ of $V(F)$ such that (FS1) and (FS2) hold. Let $\mathcal{D}^* = (d^*_{i,j})_{i,j\in[M]}$ be the Γ-rounding of the matrix $\frac{1+\alpha}{1+\beta}\mathcal{D}$.*

Then for each $n \in \mathbb{N}$ satisfying $(1+\beta)m \leq n \leq (1+2\beta)m$ there is a graph G on n vertices with a partition $\mathcal{Z} = \{Z_i\}_{i\in[M]_0}$ of $V(G)$ such that the following hold.

(GF1) For each $i \in [M]$ we have $|Z_i| = [\![(1+\beta)m\mathbf{v}_i]\!]_\Gamma$.
(GF2) For each $(i,j) \in [M]^2$ for which $\mathbf{v}_i, \mathbf{v}_j \geq \frac{\delta}{M}$ and each $v \in Z_i$ we have

$$\deg_G(v; Z_j) = d^*_{i,j} |Z_j| \ . \tag{14}$$

(GF3) All the vertices in the cluster Z_0 and in each cluster Z_i with $\mathbf{v}_i < \frac{\delta}{M}$ are isolated.
(GF4) We have $\delta_\square(W_F, W_G) \le 5\beta$.

Now we apply our main lemmas to give the proof of Theorem 1.2.

Proof of Theorem 1.2. Let $\varepsilon > 0$. Without loss of generality, we may assume that $\varepsilon \le 10^{-3}$. Set $\beta = \varepsilon/10$, $\lambda = \beta^4$, $\delta = \beta^{10}$ and $\alpha = \beta - 20\lambda$. Set $n_1 = \left\lceil \frac{10^4 M^{\mathrm{reg}}(\delta)^2 2^{10^3\beta^{-2}}}{\beta^2\lambda^3\delta^4} \right\rceil$ and $n_0 = \lceil (1+\beta)n_1 \rceil$. We shall show that given $m \ge n_1$, for all $n \in \mathbb{N}$ satisfying $(1+\beta)m \le n \le (1+2\beta)m$ we have a family $\{H_U\}_{U\in\mathcal{U}}$ of fractionally isomorphic graphs on vertex set $[n]$ such that for any $U \in \mathcal{U}$ we have $\delta_\square(U, H_U) \le \varepsilon$. By taking $m = \left\lfloor \frac{n}{1+\beta} \right\rfloor$, this clearly implies the desired outcome, that is, for all $n \ge n_0$ we have a family $\{H_U\}_{U\in\mathcal{U}}$ of fractionally isomorphic graphs on vertex set $[n]$ such that for any $U \in \mathcal{U}$ we have $\delta_\square(U, H_U) \le \varepsilon$.

Let $m \ge n_1$. Fix an arbitrary $W' \in \mathcal{U}$. Lemma 3.3 returns a number $M \le M^{\mathrm{reg}}(\delta)$, a robust density matrix $\mathcal{D} = (d_{i,j})_{i,j\in[M]}$ and a unit vector $\mathbf{v} \in [0,1]^M$. Observe that we have a $(\beta, \lambda, \delta, \alpha, M, m)$-Setup. Let $n \in \mathbb{N}$ satisfy $(1+\beta)m \le n \le (1+2\beta)m$. Set $\Gamma = 2\left\lfloor \frac{\delta^2 n_1}{2M^{\mathrm{reg}}(\delta)} \right\rfloor$; clearly, Γ is an even positive integer satisfying $\delta^{-2} \le \Gamma \le \frac{\delta^2 m}{M}$. We shall produce graphs $\{G_W\}_{W\in\mathcal{U}}$ on n vertices. The key desired fact that all these graphs be fractionally isomorphic follows from that each of them has an equitable partition whose parameters are derived from the (unchanging) matrix $\mathcal{D}$ and vector $\mathbf{v}$ via (GF1), (GF2) and (GF3).

Let $W \in \mathcal{U}$. By definition, $W/\mathcal{C}(W)$ is isomorphic to $W'/\mathcal{C}(W')$. Hence, by Lemma 3.3 we have a graphon W_β with a partition $\mathcal{Q} = \{Q_i\}_{i\in[M]}$ with cluster footprint $\mathbf{v}$ such that (GE1) and (GE2) hold. By Lemma 3.4 applied with W_β, we have a graph F on m vertices with a partition $\mathcal{X} = \{X_i\}_{i\in[M]}$ of $V(F)$ such that (FS1)–(FS3) hold. By Lemma 3.5 applied with the graph F, there is a graph G_W on n vertices with a partition $\mathcal{Z} = \{Z_i\}_{i\in[M]}$ of $V(G_W)$ such that (GF1)–(GF4) hold.

The graphs in $\{G_W\}_{W\in\mathcal{U}}$ are pairwise fractionally isomorphic because they all satisfy (GF1), (GF2) and (GF3). Let $W \in \mathcal{U}$. By (GE2), (FS3) and (GF4) we have

$$\delta_\square(W, W_{G_W}) \le \delta_\square(W, W_\beta) + \delta_\square(W_\beta, W_F) + \delta_\square(W_F, W_{G_W}) \le 10\beta = \varepsilon\,.$$

This completes the proof. □

It remains to prove Lemmas 3.3–3.5.

3.1. **Proofs of main lemmas.** First we prove Lemma 3.3.

Proof of Lemma 3.3. Let W' be a graphon. Set $U' = W'/\mathcal{C}(W')$. In particular, U' is defined on a probability space $X' = \Omega/\mathcal{C}(W')$ with measure $\pi_{X'}$. We apply Theorem 2.14

to the graphon U' and obtain a δ-regular partition $\mathcal{P}' = \{P'_i\}_{i\in[M]}$ of the space X' for some $M \leq M^{\mathrm{reg}}(\delta)$. Let $\mathbf{v}$ be the cluster footprint of $\mathcal{P}'$.

For each pair $(i,j) \in [M]^2$ write $d'_{i,j}$ for the density of (P'_i, P'_j) in U'. Let S be the collection of pairs $(i,j) \in [M]^2$ where either $i \neq j$ and (P'_i, P'_j) is δ-regular in U', or $i = j$ and P'_i is an elegant cluster in U'. Let T be the collection of pairs $(i,j) \in S$ where $\beta \leq d'_{i,j} \leq 1-\beta$. Let T' be the collection of pairs $(i,j) \in S$ where $d'_{i,j} > 1-\beta$. For each $(i,j) \in [M]^2$ define the quantity

$$d_{i,j} = \begin{cases} d'_{i,j} & \text{if } (i,j) \in T, \\ 1-\beta & \text{if } (i,j) \in T', \\ 0 & \text{otherwise.} \end{cases}$$

Set $\mathcal{D} = (d_{i,j})_{i,j\in[M]}$ and note that it is a β-robust density matrix.

Now let W be a graphon such that $U = W/\mathcal{C}(W)$ is isomorphic to U'. In particular, U is defined on a probability space $X = \Omega/\mathcal{C}(W)$ with measure π_X. Furthermore, we have a measure-preserving quotient map $q_X : \Omega \to X$. By the fact that U and U' are isomorphic, we have a measure-preserving bijection $\phi : X' \to X$ so that $U' = U^\phi$.

Observe that the partition $\mathcal{P} = \{P_i\}_{i\in[M]}$ of X given by $P_i = \phi(P'_i)$ is δ-regular for U, and for each $(i,j) \in [M]^2$ the quantity $d'_{i,j}$ is the density of (P_i, P_j) in U. Furthermore, for each $(i,j) \in [M]^{(2)}$ the pair (P_i, P_j) is δ-regular in U if and only if the pair (P'_i, P'_j) is δ-regular in U', and for each $i \in [M]$ the cluster P_i is elegant in U if and only if the cluster P'_i is elegant in U'. For each $(i,j) \in T$ define the sets

$$\begin{aligned} R^-_{i,j} &= \left\{ x \in P_i : \deg_U(x; P_j) \leq (d_{i,j} - \sqrt{\delta})\pi_X(P_j) \right\}, \\ R^+_{i,j} &= \left\{ x \in P_i : \deg_U(x; P_j) \geq (d_{i,j} + \sqrt{\delta})\pi_X(P_j) \right\}, \\ R_{i,j} &= R^-_{i,j} \cup R^+_{i,j} \text{ and } P_{i,j} = P_i \setminus R_{i,j}. \end{aligned}$$

In the case where $i \neq j$, we apply Fact 2.11 to obtain

$$\pi_X(R^-_{i,j}), \pi_X(R^+_{i,j}) \leq \sqrt{\delta}\pi_X(P_i)\,. \tag{15}$$

Note that the same inequalities also hold in the case where $i = j$ because we have $\pi_X(R^-_{i,j}), \pi_X(R^+_{i,j}) = 0$. Define the graphon $U_\beta \in \mathcal{W}_0(X)$ by

$$U_\beta(x,y) = \begin{cases} U(x,y) & \text{if } (x,y) \in P_{i,j} \times P_{j,i} \text{ with } (i,j) \in T, \\ d_{i,j} & \text{if } (x,y) \in P_i \times P_j \setminus P_{i,j} \times P_{j,i} \text{ with } (i,j) \in T, \\ d_{i,j} & \text{if } (x,y) \in P_i \times P_j \text{ with } (i,j) \notin T. \end{cases}$$

Define the partition $\mathcal{Q} = \{Q_i\}_{i\in[M]}$ of Ω by $Q_i = q_X^{-1}(P_i)$. This definition immediately implies that

$$Q_i \text{ is } \mathcal{C}(W)\text{-measurable for each } i \in [M]\,. \tag{16}$$

For each $(i,j) \in T$ define the sets

$$\begin{aligned} S_{i,j}^- &= \Big\{x \in Q_i : \deg_W(x; Q_j) \le (d_{i,j} - 2\sqrt{\delta})\pi(Q_j)\Big\}\,,\\ S_{i,j}^+ &= \Big\{x \in Q_i : \deg_W(x; Q_j) \ge (d_{i,j} + 2\sqrt{\delta})\pi(Q_j)\Big\}\,,\\ S_{i,j} &= S_{i,j}^- \cup S_{i,j}^+ \quad \text{and} \quad Q_{i,j} = Q_i \setminus S_{i,j}. \end{aligned}$$

Define the graphon W_β on Ω as follows.

$$W_\beta(x,y) = \begin{cases} W(x,y) & \text{if } (x,y) \in Q_{i,j} \times Q_{j,i} \text{ with } (i,j) \in T,\\ U_\beta(q_X(x), q_X(y)) & \text{otherwise.} \end{cases}$$

The following claim helps us show that W_β satisfies (GE1) and (GE2).

Claim 3.6. *For each $(i,j) \in T$ the set $S_{i,j} \setminus q_X^{-1}(R_{i,j})$ has measure zero.*

Claim 3.7. *The proof of the claim does not use regularity or elegance but merely properties of the map q_X. Set $A_{i,j}^- = S_{i,j}^- \setminus q_X^{-1}(R_{i,j}^-)$. We have*

$$\begin{aligned} &\int_{x\in A_{i,j}^-} \deg_U(q_X(x); P_j)\,\mathrm{d}\pi\\ \boxed{\text{by (7)}} \quad &= \int_{x\in A_{i,j}^-} \deg_{W_{\mathcal{C}(W)}}(x; Q_j)\,\mathrm{d}\pi\\ \boxed{\text{by (16) and Definition 2.5}} \quad &= \int_{x\in A_{i,j}^-} \deg_W(x; Q_j)\,\mathrm{d}\pi\,. \end{aligned}$$

We shall make use of both ends of this identity. On the one hand, we have

$$\int_{A_{i,j}^-} \deg_W(x; Q_j) \le (d_{i,j} - 2\sqrt{\delta})\pi(A_{i,j}^-)\pi(Q_j) \tag{17}$$

because $A_{i,j}^- \subseteq S_{i,j}^-$. On the other hand, we have

$$\int_{A_{i,j}^-} \deg_U(q_X(x); P_j) \ge (d_{i,j} - \sqrt{\delta})\pi(A_{i,j}^-)\pi_X(P_j) \tag{18}$$

because $q_X(A_{i,j}^-) \cap R_{i,j}^- = \emptyset$. Since q_X is measure-preserving, we have $\pi(Q_j) = \pi_X(P_j)$. By comparing (17) and (18), we obtain $\pi(A_{i,j}^-) = 0$. Finally, note that an analogous argument for $A_{i,j}^+ = S_{i,j}^+ \setminus q_X^{-1}(R_{i,j}^+)$ implies $\pi(A_{i,j}^+) = 0$, so the claim follows.

Let us now verify that W_β satisfies (GE1). We shall consider two cases, which we denote by (♡) and (♠).

Case (♡): Suppose $(i,j) \in T$ and $x \in Q_{i,j}$. By the definition of T and (11) we have $d_{i,j} \geq \beta \geq \sqrt{\delta}/\lambda$. Hence, by the definition of $Q_{i,j}$ we have

$$\deg_W(x;Q_j) = d_{i,j}(1 \pm 2\lambda)\pi(Q_j)\,. \tag{19}$$

We also have

$$\pi(S_{j,i}) \overset{\text{C3.6}}{\leq} \pi_X(R_{j,i}) \overset{(15)}{\leq} 2\sqrt{\delta}\pi_X(P_j) \leq 2d_{i,j}\lambda\pi_X(P_j) = 2d_{i,j}\lambda\pi(Q_j)\,. \tag{20}$$

Since by definition $W(x,y)$ and $W_\beta(x,y)$ are equal for $y \in Q_{j,i}$, by the definition given in (6) we obtain

$$\left|\deg_W(x;Q_j) - \deg_{W_\beta}(x;Q_j)\right| = \left|\int_{y\in S_{j,i}} W(x,y) - W_\beta(x,y)\,\mathrm{d}\pi\right| \leq \pi(S_{j,i})\,.$$

Therefore, we have

$$\begin{aligned}\deg_{W_\beta}(x;Q_j) &= \deg_W(x;Q_j) \pm \pi(S_{j,i})\\ \boxed{\text{(19) and (20)}} \quad &= d_{i,j}(1 \pm 4\lambda)\pi(Q_j)\,,\end{aligned}$$

as required for (GE1) in this case.

Case (♠): Suppose that either $(i,j) \in T$ and $x \in S_{i,j}$, or $(i,j) \notin T$ and $x \in Q_i$. In this case, by Claim 3.6 and the definition of W_β we have

$$\deg_{W_\beta}(x;Q_j) = d_{i,j}\pi(Q_j),$$

as required for (GE1).

Now we shall verify that W_β satisfies (GE2). For each $(i,j) \in [M]^2$ define the quantities

$$K_{i,j} = \int_{(x,y)\in P_i\times P_j} |U(x,y) - U_\beta(x,y)|\,\mathrm{d}\pi_X^2 \quad \text{and}$$
$$L_{i,j} = \int_{(x,y)\in Q_i\times Q_j} |W(x,y) - W_\beta(x,y)|\,\mathrm{d}\pi^2\,.$$

By the definitions of T', U_β and W_β, we have

$$\sum_{(i,j)\in T'} L_{i,j} \leq 2\beta\,. \tag{21}$$

By Claim 3.6 and the definitions of U_β and W_β, we have

$$L_{i,j} \leq K_{i,j} \text{ for all } (i,j) \in [M]^2 \setminus T'\,. \tag{22}$$

Now by accounting for the editing needed to obtain U_β from U, we obtain

$$\sum_{(i,j)\in[M]^2\setminus T'} K_{i,j} \leq \delta + \beta + 4\sqrt{\delta}. \tag{23}$$

By definition we have

$$\|W - W_\beta\|_1 = \sum_{(i,j)\in[M]^2} L_{i,j} \ ,$$

so by combining with equations (21)–(23) we obtain

$$\|W - W_\beta\|_1 \leq \delta + \beta + 4\sqrt{\delta} + 2\beta \leq 4\beta$$

as required. □

Now we prove Lemma 3.4.

Proof of Lemma 3.4. Sample $F \sim \mathbb{G}(m, W)$. Let $\mathcal{X} = \{X_i\}_{i\in[M]}$ be the partition of $V(F)$ induced (in the sense of Section 2.5) by $\mathcal{Q}$. Let $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{E}_3$ be the events that (FS1), (FS2) and (FS3) respectively hold. We shall show that $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ holds with positive probability, thereby achieving the desired outcome.

We consider the failure probability of $\mathcal{E}_2$. Let $(i,j) \in [M]^2$ be such that $\mathbf{v}_j \geq \frac{\delta}{M}$. Since the failure probability on the pair (i,j) is trivially zero whenever $d_{i,j} = 0$, we may assume that $d_{i,j} \geq \beta$. Consider $h \in [m]$. Without loss of generality, we shall argue for $h = m$; the argument is entirely analogous for the other choices of h. For each $\ell \in [m-1]$ we set Y_ℓ to be the indicator function for the event that both $\ell \in X_j$ and $m\ell \in E(F)$ hold; set $Y = \sum_{h=1}^{m-1} Y_h = \deg_F(m; X_j)$. We work in the conditional probability space given $\{m \in X_i\}$. To obtain probability bounds in this conditional probability space, we further condition on $\{x_m = x\}$ for an arbitrary $x \in Q_i$. In the latter conditional probability space, the random variables $(Y_h)_{h\in[m-1]}$ are independent and identically distributed Bernoulli random variables with a common expectation $\mu = \mathbb{E}Y_h$ expressed by (13). Hence, by Theorem 2.1 the probability that $Y \neq (m-1)(1\pm 6\lambda)d_{i,j}\mathbf{v}_j$ is at most $2\exp\left(-\frac{\lambda^2 d_{i,j}\mathbf{v}_j m}{4}\right)$. It follows that the probability that $Y \neq (m-1)(1 \pm 6\lambda)d_{i,j}\mathbf{v}_j$ is at most $2\exp\left(-\frac{\lambda^2 d_{i,j}\mathbf{v}_j m}{4}\right)$ in the initial conditional probability space. A union bound over $(i,j) \in [M]^2$ and $h \in [m]$ tells us $\mathcal{E}_2$ does not hold with probability at most $2M^2 m \exp\left(-\frac{\lambda^2 d_{i,j}\mathbf{v}_j m}{4}\right) \leq 2M^2 m \exp\left(-\frac{\lambda^2 \beta \mathbf{v}_j m}{4}\right) < 0.1$.

Now we consider $\mathcal{E}_1$. For each $i \in [M]$ we have $\pi(Q_i) = \mathbf{v}_i$. Hence, by Theorem 2.1 and a union bound, the event $\mathcal{E}_1$ fails with probability at most $2M\exp\left(-\frac{\lambda^2 m\mathbf{v}_i}{3}\right) < 0.1$. Last, Lemma 2.9 tells us that the event $\mathcal{E}_3$ fails with probability at most $\exp\left(-\frac{m}{2\log_2 m}\right) < 0.1$. Hence, $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ fails to hold with probability at most 0.3, completing the proof. □

Finally, we prove Lemma 3.5.

Proof of Lemma 3.5. We construct $(G, \mathcal{Z})$ from $(F, \mathcal{X})$ as follows. We start with $(F, \mathcal{X})$. For each $i \in [M]$ we introduce a set Y_i of size

$$|Y_i| = [\![(1+\beta)m\mathbf{v}_i]\!]_\Gamma - |X_i| \tag{24}$$

and set $Z_i = X_i \sqcup Y_i$. We also introduce a set Z_0 of size $|Z_0| = n - \sum_{i\in[M]} |Z_i| \geq 0$.[5] The first edge-modification resolves (GF3) and is easy: we delete all the edges incident to vertices in clusters Z_i with $\mathbf{v}_i < \frac{\delta}{M}$. Let F' be the graph resulting from these edge deletions.

From now on, we shall work exclusively with pairs in $T = \{(i,j) \in [M]^2 : \mathbf{v}_i, \mathbf{v}_j \geq \frac{\delta}{M}\}$. For each such pair we set

$$D_{i,j} = d^*_{i,j}|Z_j| \,, \tag{25}$$

that is, $D_{i,j}$ is the target quantity on the right-hand side of (14). Note that $D_{i,j} \in \mathbb{N}$. Also, since the matrix $\mathcal{D}$ is symmetric, so is $\mathcal{D}^*$ and hence

$$D_{i,j}|Z_i| = d^*_{i,j} \cdot |Z_j| \cdot |Z_i| = d^*_{j,i} \cdot |Z_j| \cdot |Z_i| = D_{j,i}|Z_j| \,. \tag{26}$$

Since $\mathcal{D}^*$ is the Γ-rounding of $\frac{1+\alpha}{1+\beta}\mathcal{D}$ and $\Gamma^{-1} \leq \delta^2$, we have

$$\frac{1+\alpha}{1+\beta} d_{i,j} \leq d^*_{i,j} \leq (1+2\delta)\frac{1+\alpha}{1+\beta} d_{i,j} \,. \tag{27}$$

The definition of $[\![\cdot]\!]_\Gamma$ and the fact that $m \geq \Gamma M/\delta^2$ gives

$$(1-\delta)(1+\beta)m\mathbf{v}_i \leq |Z_i| = [\![(1+\beta)m\mathbf{v}_i]\!]_\Gamma \leq (1+\beta)m\mathbf{v}_i \,. \tag{28}$$

Let $\tilde{T} := \{\{i,j\} : (i,j) \in T \cap [M]^{(2)}\}$ be an off-diagonal unordered version of T. We add edges in two steps. In the first step, for each $(i,j) \in T$ we add edges between X_i and Y_j to achieve (14) for each $v \in X_i$. In the second step, for each $\{i,j\} \in \tilde{T}$ we add edges between Y_i and Y_j to achieve (14) for the vertices in Y_i and Y_j. Furthermore, for $(i,j) \in T$ with $i = j$ we add edges on Y_i to achieve (14) for the vertices in Y_i. See Figures 1 and 2 for illustrations of the construction in the cases $i \neq j$ and $i = j$, respectively. Let G be the resultant graph. In particular, (GF3) still holds in G.

We begin with the first step. For each $h \in X_i$ set $a_h^{i,j} = D_{i,j} - \deg_F(h; X_j)$. Set $A^{i,j} = \sum_{h\in X_i} a_h^{i,j}$. For each $\ell \in Y_j$ set $b_\ell^{i,j} \in \left\{ \left\lfloor \frac{A^{i,j}}{|Y_j|} \right\rfloor, \left\lceil \frac{A^{i,j}}{|Y_j|} \right\rceil \right\}$ so that $A^{i,j} = \sum_{\ell\in Y_j} b_\ell^{i,j}$. We shall apply the Gale–Ryser theorem (Theorem 2.3) to build a bipartite graph with parts X_i and Y_j whose degrees are given by $(a_h^{i,j})_{h\in X_i}$ and $(b_\ell^{i,j})_{\ell\in Y_j}$. Fix a non-increasing ordering of $\{a_h^{i,j}\}_{h\in X_i}$ and let $\mathcal{A}_k^{i,j}$ denote the set of the first k elements of this ordering. We first verify that $a_h^{i,j} \geq 0$ for all $h \in X_i$. Indeed, by the definition of $a_h^{i,j}$, (25), (27), (28) and (FS2), for each $h \in X_i$ we have

$$\begin{aligned} a_h^{i,j} &\geq (1-\delta)\frac{1+\alpha}{1+\beta} d_{i,j}(1+\beta)\mathbf{v}_j m - (1+6\lambda) d_{i,j}\mathbf{v}_j m \\ &\geq d_{i,j}\mathbf{v}_j m(\alpha - 7\lambda) \overset{(10),(12)}{\geq} 0 \,. \end{aligned} \tag{29}$$

[5] The only purpose of Z_0, which consists of isolated vertices, is to achieve a given order of the graph.

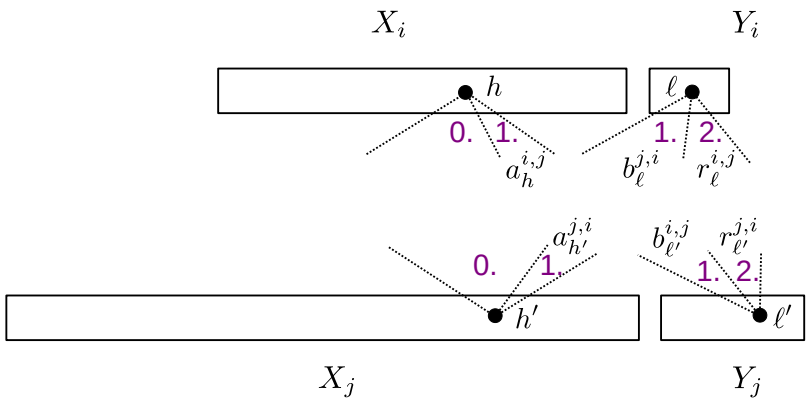


FIGURE 1. Adding edges in the bipartite graph $(X_i \cup Y_i, X_j \cup Y_j)$ for $i \neq j$. The order of steps is shown in purple (with 0. corresponding to already given edges). The quantities $a_h^{i,j}$, $b_\ell^{j,i}$, $r_\ell^{i,j}$, $a_{h'}^{j,i}$, $b_{\ell'}^{i,j}$ and $r_{\ell'}^{j,i}$ are degree sequences used in the construction.

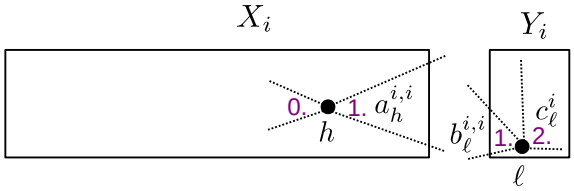


FIGURE 2. Adding edges in the graph $X_i \cup Y_i$. The order of steps is shown in purple (with 0. corresponding to already given edges). The quantities $a_h^{i,i}$, $b_\ell^{i,i}$, and c_ℓ^i are degree sequences used in the construction.

Let us now check the conditions of Theorem 2.3. Clearly (GR1) holds. Furthermore, this implies (GR2) for $k \geq \max\{b_\ell^{i,j}\}_{\ell \in Y_j}$. Now we focus on the case $k \leq \min\{b_\ell^{i,j}\}_{\ell \in Y_j}$; observe that this covers the remaining values of k because the quantities $\{b_\ell^{i,j}\}_{\ell \in Y_j}$ differ by at most 1. By the definition of $a_h^{i,j}$, (25), (27), (28) and (FS2), for each $h \in X_i$ we have

$$\begin{aligned} a_h^{i,j} &\leq (1+2\delta)\frac{1+\alpha}{1+\beta} d_{i,j}(1+\beta) m\mathbf{v}_j - (1-6\lambda) d_{i,j} m\mathbf{v}_j \\ &\leq d_{i,j}\mathbf{v}_j m(\alpha+7\lambda) \overset{(12)}{\leq} m\mathbf{v}_j(\beta-7\lambda) \overset{(FS1)}{\leq} |Y_j| \,. \end{aligned} \tag{30}$$

Hence, we have $\sum_{h \in \mathcal{A}_k^{i,j}} a_h^{i,j} \leq k|Y_j| \leq \sum_{\ell \in Y_j} \min\{k, b_\ell^{i,j}\}$. This completes the verification of (GR2). Then, by Theorem 2.3 we have a bipartite graph on (X_i, Y_j) whose edges we

add to F to achieve (14) for the vertices in X_i. Finally, let $F^\times$ be the resultant graph after finishing the first step for all $(i,j) \in T$.

Now we consider the second step. Let $\{i,j\} \in \tilde{T}$. For each $h \in Y_i$ set $r_h^{i,j} = D_{i,j} - \deg_{F^\times}(h; X_j)$. For each $\ell \in Y_j$ set $r_\ell^{j,i} = D_{j,i} - \deg_{F^\times}(\ell; X_i)$. We shall apply Theorem 2.3 to build a bipartite graph with parts Y_i and Y_j with degree sequences $(r_h^{i,j})_{h\in Y_i}$ and $(r_h^{j,i})_{h\in Y_j}$. Fix a non-increasing ordering of $\{r_h^{i,j}\}_{h\in Y_i}$ and let $\mathcal{B}_k^{i,j}$ denote the set of the first k elements of this ordering. We first verify that $r_h^{i,j}, r_\ell^{j,i} \geq 0$ for all $h \in Y_i$ and $\ell \in Y_j$. By symmetry, it suffices to prove only the first half of the statement. Every edge between $h \in Y_i$ and X_j was added only in the first step, so we have

$$\begin{aligned}\deg_{F^\times}(h; X_j) = b_h^{j,i} \leq \left\lceil \frac{A^{j,i}}{|Y_i|} \right\rceil &\overset{(30)}{\leq} 1 + \frac{d_{j,i}\mathbf{v}_i m(\alpha+7\lambda)\cdot|X_j|}{[\![(1+\beta)m\mathbf{v}_i]\!]_\Gamma - |X_i|} \\ &\overset{(FS1)}{\leq} \frac{d_{j,i}(\alpha+9\lambda)\mathbf{v}_j m}{\beta-6\lambda-\delta} \overset{(12)}{\leq} d_{i,j}\mathbf{v}_j m \leq D_{i,j}\ .\end{aligned} \tag{31}$$

This shows that $r_h^{i,j} \geq 0$. Now we verify the conditions of Theorem 2.3, starting with (GR1). We have

$$\begin{aligned}\sum_{h\in Y_i} r_h^{i,j} &= D_{i,j}|Y_i| - \sum_{h\in Y_i} b_h^{j,i} = D_{i,j}|Y_i| - \sum_{\ell\in X_j} a_\ell^{j,i} \\ &= D_{i,j}|Y_i| - \sum_{\ell\in X_j} \big(D_{j,i} - \deg_F(\ell; X_i)\big) \\ &= D_{i,j}|Y_i| - D_{j,i}|X_j| + e_F(X_i, X_j)\ .\end{aligned}$$

By symmetry, we also have $\sum_{h\in Y_j} r_h^{j,i} = D_{j,i}|Y_j| - D_{i,j}|X_i| + e_F(X_i,X_j)$. Now, crucially, recall that by (26) we have $D_{i,j}(|X_i|+|Y_i|) = D_{j,i}(|X_j|+|Y_j|)$. Hence, we obtain $\sum_{h\in Y_i} r_h^{i,j} = \sum_{h\in Y_j} r_h^{j,i}$ as needed for (GR1). Furthermore, this implies (GR2) for $k \geq \max\{r_\ell^{j,i}\}_{\ell\in Y_j}$. Now we focus on the case $k \leq \min\{r_\ell^{j,i}\}_{\ell\in Y_j}$. Observe that this covers the remaining values of k because we have $r_\ell^{j,i} = D_{j,i} - b_\ell^{i,j}$ and the fact that the quantities $\{b_\ell^{i,j}\}_{\ell\in Y_j}$ differ by at most 1 means that the quantities $\{r_\ell^{j,i}\}_{\ell\in Y_j}$ also differ by at most 1. For each $\ell \in Y_i$ we have

$$\begin{aligned}\deg_{F^\times}(\ell; X_j) = b_h^{j,i} \geq \left\lfloor \frac{A^{j,i}}{|Y_i|} \right\rfloor &\geq \frac{1}{|Y_i|}\sum_{h\in X_j} a_h^{j,i} - 1 \\ &\overset{(29)}{\geq} \frac{d_{j,i}\mathbf{v}_i m(\alpha-7\lambda)\cdot|X_j|}{[\![(1+\beta)m\mathbf{v}_i]\!]_\Gamma - |X_i|} - 1 \\ &\overset{(FS1)}{\geq} \frac{d_{j,i}(\alpha-9\lambda)\mathbf{v}_j m}{\beta+6\lambda} \\ &\overset{(12)}{\geq} [d_{i,j}(1+\alpha+2\lambda) - (\beta-7\lambda)]\mathbf{v}_j m \geq D_{i,j} - |Y_j|\ ,\end{aligned}$$

where the fact that $d_{i,j} \leq 1-\beta$ is used in the penultimate inequality and the definition of $D_{i,j}$ is utilized in the final inequality. Hence, we have $\sum_{h\in\mathcal{B}_k^{i,j}} r_h^{i,j} \leq k|Y_j| \leq \sum_{\ell\in Y_j} \min\{k, r_\ell^{j,i}\}$. This completes the verification of (GR2). Then, by Theorem 2.3 we have a bipartite graph on (Y_i, Y_j) whose edges we add to $F^\times$ to achieve (14) for the vertices in Y_i and Y_j.

Now we handle the second step for the case $i = j$. We need to achieve (14) in Y_i. To this end, we shall apply the Erdős–Gallai Theorem (Theorem 2.2) to build a graph on Y_i with degree sequence $\{c_\ell^i\}_{\ell\in Y_i}$ given by $c_\ell^i = D_{i,i} - b_\ell^{i,i}$ for each $\ell \in Y_i$. Observe that for $\Delta = \max\{c_\ell^i\}_{\ell\in Y_\ell} \in \mathbb{N}$ and for all $\ell \in Y_\ell$ we have $c_\ell^i \in \{\Delta-1, \Delta\}$. Fix a non-increasing ordering of $\{c_\ell^i\}_{\ell\in Y_i}$ and let $\mathcal{B}_k^{i,i}$ denote the set of the first k elements of this ordering. We first verify that $0 \leq c_\ell^i \leq |Y_i|-1$ for all $\ell \in Y_i$. Noting that every edge between $\ell \in Y_i$ and X_i was added only in the first step and repeating (31) mutatis mutandis, for each $\ell \in Y_i$ we have

$$\deg_{F^\times}(\ell; X_i) = b_h^{i,i} \leq \frac{d_{i,j}(\alpha+8\lambda)|X_j|}{\beta-6\lambda-\delta} \leq D_{i,i}\,. \tag{32}$$

This shows that $c_\ell^i \geq 0$. Furthermore, we have

$$\begin{aligned}\deg_{F^\times}(\ell; X_i) = b_h^{i,i} \geq \left\lfloor \frac{A^{i,i}}{|Y_i|} \right\rfloor &\overset{(29)}{\geq} \frac{d_{i,i}\mathbf{v}_i m(\alpha-7\lambda)\cdot|X_i|}{[\![(1+\beta)m\mathbf{v}_i]\!]_\Gamma - |X_i|} - 1\\ &\overset{(FS1)}{\geq} \frac{d_{i,j}(\alpha-8\lambda)|X_j|}{\beta+6\lambda}\\ &\overset{(12)}{\geq} [d_{i,i}(1+\alpha+2\lambda) - (\beta-7\lambda)]\mathbf{v}_i m \geq D_{i,i} - |Y_i| + 2\,,\end{aligned}$$

where the fact that $d_{i,i} \leq 1-\beta$ is used in the penultimate inequality and the definition of $D_{i,i}$ is utilized in the final inequality. This shows that $c_\ell^i \leq |Y_i| - 2$.

Let us now check the conditions of Theorem 2.2. First, we verify (EG1), that is, the quantity $\sum_{\ell\in Y_i} c_\ell^i$ is even. Since $\sum_{h\in X_i} \deg_F(h; X_i) = 2e_F(X_i)$ and $D_{i,i}$ are both even, the quantity

$$\begin{aligned}\sum_{\ell\in Y_i} c_\ell^i = D_{i,i}|Y_i| - \sum_{\ell\in Y_i} b_\ell^{i,i} &= D_{i,i}|Y_i| - \sum_{h\in X_i} a_h^{i,i}\\ &= D_{i,i}|Y_i| - D_{i,i}|X_i| + \sum_{h\in X_i} \deg_F(h; X_i)\end{aligned}$$

is even as needed. Now we turn to (EG2). We shall consider three different cases. First, we consider the case $k > \Delta$. In this case, we have

$$\sum_{\ell\in\mathcal{B}_k^{i,i}} c_\ell^i \leq k\Delta \leq k(k-1)\,,$$

which implies (EG2) as required. Second, we have the case $k < \Delta$. In this case, we have

$$\sum_{\ell \in \mathcal{B}_k^{i,i}} c_\ell^i \leq k(|Y_i| - 1) \leq k(k-1) + (|Y_i| - k)\min\{k, \Delta - 1\} ,$$

which implies (EG2) as required. Finally, we have the case $k = \Delta \leq |Y_i| - 2$. Observe that in this case we have

$$\sum_{\ell \in \mathcal{B}_k^{i,i}} c_\ell^i \leq k(k-1) + \min\{k, c_{k+1}\} + \min\{k, c_{k+2}\} ,$$

except in the case where $c_k > c_{k+1} = c_{k+2} = 0$. Observe that in the exceptional case we must have $k = \max\{c_\ell^i\} = 1$ and hence $\sum_{\ell \in Y_i} c_\ell^i = 1$, so (EG1) implies that it never holds. Hence, the inequality above holds, implying (EG2) as required. Then, by Theorem 2.2 we have a graph on Y_i whose edges we add to $F^\times$ to achieve (14) in Y_i.

To complete the proof, we shall check (GF4). Since $1 + \beta \leq \frac{n}{m} \leq 1 + 2\beta$ and $G[V(F)] = F'$, by Lemma 2.4 we have $\delta_\square(G, F') \leq 2\left(1 - \frac{m}{n}\right) \leq \frac{4\beta}{1+2\beta} \leq 4\beta$. Thus, it remains to prove that $\delta_\square(F, F') \leq \beta$. To do this, we recall that the only edge edits wholly contained in $V(F)$ are the deletions of edges incident to clusters X_i with $\mathbf{v}_i < \frac{\delta}{M}$. Each such cluster has at most $|X_i| = (1 \pm \lambda)\mathbf{v}_i m < \frac{2\delta n}{M}$ vertices, so at most $M \times \frac{2\delta n}{M} \times n < \beta n^2$ edges are deleted. This completes the proof. □

4. Regular case

As mentioned in the introduction, the proof of Theorem 1.2 given in Section 3 guarantees regularity of the approximating graphs in the case that the graphons are regular, which leads us to Theorem 1.3. Recall that a graphon is regular with degree d if and only if it is fractionally isomorphic to the constant-d graphon.

We shall walk through each of the three steps (Lemmas 3.3–3.5) of the proof of Theorem 1.2 and see that they indeed yield Theorem 1.3. Since the cases $d \in \{0, 1\}$ are trivial, we may assume that $d \in (0, 1)$. In particular, we may choose the constant $\beta > 0$ to satisfy $\beta \ll \min\{d, 1-d\}$. We may further assume that $C_d \in \mathcal{U}$. Let us first deal with Lemma 3.3. That is, we pick a representative of $\mathcal{U}$ to play the role of W' in Lemma 3.3; here we shall make the obvious choice of $W' = C_d$. In particular, $W'/\mathcal{C}(W')$ is a constant-d graphon defined on a singleton probability space $\Omega/\mathcal{C}(W')$. Hence, $M = 1$ and the 1×1-matrix $\mathcal{D}$ contains a single entry of value d. Furthermore, each $W \in \mathcal{U}$ satisfies (GE1) automatically. Now Lemma 3.4 is an intermediate step which yields for each $W \in \mathcal{U}$ a graph F_W on m vertices such that, by (FS2), each vertex $x \in V(F_W)$ satisfies

$$\deg_{F_W}(x) = (1 \pm \varepsilon)dm . \tag{33}$$

We pick $m \approx \frac{n}{1+\beta}$. Finally, we deal with Lemma 3.5. Here we create graphs G_W by adding some vertices and edges to F_W. More precisely, we fix an even integer $D \approx (1+\beta)dm$ and

add a little set Y_1 to the set $X_1 = V(F_W)$. Then, we add edges onto both the pair (X_1, Y_1) (one application of the Gale–Ryser Theorem) and inside the set Y_1 (one application of the Erdős–Gallai Theorem) so that the graph becomes regular with degree D. Observe that, in the proof of Lemma 3.5, the set Y_1 is created in (24) in a way that $|X_1 \cup Y_1|$ is divisible by Γ. However, this divisibility condition can be dropped in our setting as its only purpose[6] was to allow for biregularity between $X_1 \cup Y_1$ and other clusters $X_i \cup Y_i$, of which there are none in the current setting. Therefore, we are unconstrained by divisibility considerations and have enough flexibility to choose $|Y_1|$ so as to achieve $|X_1 \cup Y_1| = n$. Checking the conditions of the Gale–Ryser theorem and of the Erdős–Gallai Theorem works the same way as before.

In addition to the steps mentioned above, the proof of Theorem 1.2 involves two further steps in which edges incident to small clusters are deleted and a set Z_0 of isolated vertices is added. We remark that the former step does not apply in our setting. Furthermore, the addition of the set Z_0 serves the sole purpose of ensuring that the graph G has a specific number of vertices; this is achieved in the current setting by having more flexibility in the size of the set Y_1.

References

[1] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[2] A. Chirvasitu. Prescribed graphon symmetries and flavors of rigidity. arXiv:2206.11997.

[3] H. Dell, M. Grohe, and G. Rattan. Lovász meets Weisfeiler and Leman. In *45th International Colloquium on Automata, Languages, and Programming*, volume 107 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 40, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2018.

[4] Z. Dvořák. On recognizing graphs by numbers of homomorphisms. *J. Graph Theory*, 64(4):330–342, 2010.

[5] D. Gale. A theorem on flows in networks. *Pacific J. Math.*, 7:1073–1082, 1957.

[6] J. Grebík and I. Rocha. Fractional isomorphism of graphons. *Combinatorica*, 42(3):365–404, 2022.

[7] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[8] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[9] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[10] L. Lovász and B. Szegedy. The automorphism group of a graphon. *J. Algebra*, 421:136–166, 2015.

[11] M. V. Ramana, E. R. Scheinerman, and D. Ullman. Fractional isomorphism of graphs. *Discrete Math.*, 132(1-3):247–265, 1994.

[12] H. J. Ryser. Combinatorial properties of matrices of zeros and ones. *Canadian J. Math.*, 9:371–377, 1957.

[6] as explained after the statement of Lemma 3.4

[13] E. R. Scheinerman and D. H. Ullman. *Fractional graph theory*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., New York, 1997. A rational approach to the theory of graphs, With a foreword by Claude Berge, A Wiley-Interscience Publication.

[14] G. Tinhofer. Graph isomorphism and theorems of Birkhoff type. *Computing*, 36(4):285–300, 1986.

Institute of Computer Science of the Czech Academy of Sciences, Pod Vodárenskou věží 2, 182 07, Praha 8, Czech Republic

Part III

Inhomogeneous random structures from graphons

A LIMIT THEOREM FOR SMALL CLIQUES IN INHOMOGENEOUS RANDOM GRAPHS

JAN HLADKÝ, CHRISTOS PELEKIS, AND MATAS ŠILEIKIS

Abstract. The theory of graphons comes with a natural sampling procedure, which results in an inhomogeneous variant of the Erdős–Rényi random graph, called W-random graphs. We prove, via the method of moments, a limit theorem for the number of r-cliques in such random graphs. We show that, whereas in the case of dense Erdős–Rényi random graphs the fluctuations are normal of order n^{r-1}, the fluctuations in the setting of W-random graphs may be of order $0, n^{r-1}$, or $n^{r-0.5}$. Furthermore, when the fluctuations are of order $n^{r-0.5}$ they are normal, while when the fluctuations are of order n^{r-1} they exhibit either normal or a particular type of chi-square behavior whose parameters relate to spectral properties of W.

These results can also be deduced from a general setting [Janson and Nowicki, PTRF 1991], based on the projection method. In addition to providing alternative proofs, our approach makes direct links to the theory of graphons.

Keywords: graphons; inhomogeneous random graphs; limit theorems; subgraph counts; quasirandomness

1. Introduction

1.1. **Subgraph counts in random graphs.** The purpose of this work is to investigate the distribution of the number of fixed-size cliques in an inhomogeneous variant of the Erdős–Rényi random graph $\mathbb{G}(n,p)$. The study of the Erdős–Rényi random graph (see [11]) is over a half-century old. A central part in the development of the random graph theory concerns methods for understanding the distribution of subgraph counts. Even though "subgraphs" may be large-scale structures, like Hamilton cycles, here we are concerned with counting fixed-sized subgraphs. In particular, we want to describe the (bulk of the) distribution of the random variable that counts the number of copies of a fixed subgraph H as n tends to infinity or, in probabilistic language, to obtain a limit theorem for the distribution of subgraph counts of H.

This problem has many variants (all copies of H, induced copies of H, joint distribution for several subgraph counts, ...), and a variety of tools have been applied to tackle it, including Stein's method in [2], ideas from U-statistics in [16], and the method of moments in [19]. We refer the reader to [11] for an entire chapter devoted to the topic and for further references.

Given graphs H and G, let $N(H,G)$ denote the number of copies of H in G, i.e., the number of subgraphs of G that are isomorphic to H, and consider a random variable

$$X_n := N(H, \mathbb{G}(n)),$$

where $\mathbb{G}(n)$ is a model of random graphs on n vertices of interest. The asymptotic normality of X_n in the Erdős–Rényi random graph model $\mathbb{G}(n) = \mathbb{G}(n, p(n))$, has been fully described by Ruciński [19] already in 1988. When $p(n)$ is constant it first appeared as a special case of a result by Nowicki and Wierman [16] who showed that for any fixed graph H with at least one edge and any $p \in (0, 1)$,

$$\frac{X_n - \mathbb{E}[X_n]}{\sqrt{\mathrm{Var}[X_n]}} \xrightarrow{\mathrm{d}} Z \sim \mathcal{N}(0, 1).$$

(Here and below $Z_n \xrightarrow{\mathrm{d}} Z$ denotes convergence in distribution.) Our result extends this result in the case of cliques $H = K_r$, $r \geq 2$ and fixed p by generalizing $\mathbb{G}(n, p)$ to the random graph model $\mathbb{G}(n, W)$ which plays a key role in the theory of dense graph limits introduced in [7, 15]. This model, which we formally define in Section 1.2, can be described as follows: each vertex is assigned a random *type* and then each pair of vertices is independently included as an edge in $\mathbb{G}(n, W)$ with a probability that depends only on the types of the two vertices. The random graph $\mathbb{G}(n, p)$ corresponds to a single type and hence the same probability p for every pair of vertices.

Henceforth, we shall write $X_n(H, W) := N(H, \mathbb{G}(n, W))$. In Theorem 1.2, we state our main result, a limit theorem for $X_n(H, W)$ for each $H = K_r, r \geq 2$. However, since our main result Theorem 1.2 requires quite a few additional definitions, for a preview we state its implicit version, Theorem 1.1. We start the next subsection by the minimum number of definitions.

1.2. **Inhomogeneous random graphs and the simplified result statement.** A *graphon* is a symmetric Lebesgue measurable function $W : [0, 1]^2 \to [0, 1]$. Graphons arise as limits of sequences of large finite undirected graphs with respect to the so-called cut metric (see [14, Part 3]). Intuitively, graphons may be thought of as graphs on the vertex set $[0, 1]$ with infinitesimally small vertices and with a $W(x, y)$-proportion of all possible edges being present in the bipartite graph whose color classes are formed by a small neighbourhood of x and of y, respectively.

Graphons come with a natural sampling procedure, which results in an inhomogeneous variant of the Erdős–Rényi random graph. More precisely, given a graphon W, the random graph $\mathbb{G}(n, W)$ is a finite simple graph on n vertices, labelled by the set $[n] := \{1, \dots, n\}$, which is generated in two steps: in the first step we draw n numbers *types*) $U_1, \dots, U_n$ independently from the interval $[0, 1]$ according to the uniform distribution and we identify their index set with the labels of the vertex set of $\mathbb{G}(n, W)$; in the second step, each pair of vertices i and j in $\mathbb{G}(n, W)$ is connected independently with probability $W(U_i, U_j)$. Notice that if $W(x, y)$ is constant, say, $p \in [0, 1]$, then $\mathbb{G}(n, W)$ is the same as the Erdős–Rényi random graph $\mathbb{G}(n, p)$. Inhomogeneous random graphs $\mathbb{G}(n, W)$ provide substantial

additional challenges compared to $\mathbb{G}(n,p)$. For example, while a standard second moment argument shows that the clique number of $\mathbb{G}(n,p)$ satisfies $\omega(\mathbb{G}(n,p)) \sim \frac{2\log n}{\log(1/p)}$, extending this formula to $\mathbb{G}(n,W)$ required new techniques, [8]. Further work on inhomogeneous random graphs so far ([6, 10]) was done in a more general, possibly sparse, model which we mention in Section 5.

Corollary 10.4 in [14] implies that $X_n(H,W)$ obeys the law of large numbers, that is, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left\{X_n(H,W) = (1\pm\epsilon)\,\mathbb{E}[X_n(H,W)]\right\} = 1. \tag{1}$$

This is one of the key results in the theory of limits of dense graph sequences because it shows that each graphon can be approximated by finite graphs with similar subgraph densities. In this article we aim to understand the nature of fluctuations of $X_n(H,W)$ around its expectation.

Fix $r \geq 2$ and write $X_n = X_n(K_r,W)$ for the number of r-cliques in $\mathbb{G}(n,W)$. Since every r-set of vertices in $\mathbb{G}(n,W)$ induces a random graph distributed as $\mathbb{G}(r,W)$, every r-set of vertices induces a clique with probability

$$t(K_r,W) := \mathbb{P}\left\{\mathbb{G}(r,W) = K_r\right\}.$$

If we prescribe type x to one of the vertices in an r-set, then this r-set induces a clique with probability

$$t(x) = t_x(K_r,W) := \mathbb{P}\left\{\mathbb{G}(r,W) = K_r \;\middle|\; U_1 = x\right\}\ .$$

Clearly, $t(K_r,W) = \int_0^1 t(x)dx$.

For a measurable function $f : [0,1]^k \to \mathbb{R}$ we say f is *constant* and denote $f \equiv c$ whenever $f(x) = c$ for almost every $x \in [0,1]^k$.

We start with a simplified version of our main result. The additional information provided by the complete version, that is, Theorem 1.2, is the description of constants σ and $c_0, c_1, \ldots$ in terms of the graphon W.

Theorem 1.1. *For $t(K_r,W), t(x)$, and X_n defined above, the following holds.*

(a) If $t(K_r,W) = 0$ or $t(K_r,W) = 1$, then almost surely $X_n = 0$ or $X_n = \binom{n}{r}$, respectively.

(b) If $t(x)$ is not constant, then there is a constant $\sigma > 0$ such that

$$\frac{X_n - \mathbb{E}[X_n]}{n^{r-1/2}} \xrightarrow{\mathrm{d}} \sigma Z\ , \tag{2}$$

where Z is a standard normal random variable.

(c) If $t(x)$ is constant (but other than 0 or 1), then there are real numbers $c_0, c_1, \dots$, such that $\sum_i c_i^2 \in (0, \infty)$ and

$$\frac{X_n - \mathbb{E}[X_n]}{n^{r-1}} \xrightarrow{\mathrm{d}} c_0 Z_0 + \sum_{i \geq 1} c_i (Z_i^2 - 1) \,, \tag{3}$$

where $Z_0, Z_1, \dots$ are independent standard normal random variables. The series on the right-hand side of (3) converges a.s. and in L^2.

Proof. Theorem 1.2 implies all the claims of Theorem 1.1 except for the fact that $\sum_i c_i^2 > 0$, which follows from Remark 1.6. □

We note that non-normal limit theorem occurs even in $\mathbb{G}(n,p)$ with p fixed, but for *induced* subgraph counts, see [11, Theorem 6.52].

A toy instance of Theorem 1.1(c). The proof of Theorem 1.1(c) proceeds by the method of moments and thus in itself does not provide much intuition for the asserted non-normal limit fluctuations. This behavior is however suggested by the following simple example. Let $r \geq 2$ be arbitrary. Consider W which is a 'disjoint union of two equally-sized cliques', that is $W(x,y) = 1$ if $(x,y) \in [0,\frac{1}{2})^2 \cup [\frac{1}{2},1]^2$ and $W(x,y) = 0$ otherwise. It is easy to verify that $t(x) = 2^{-r+1}$ for each $x \in [0,1]$, in particular, $t(x)$ is constant. Now, the graph $\mathbb{G}(n,W)$ is determined already after the first step of the construction, that is, after sampling $U_1, \dots, U_n$. Indeed, $\mathbb{G}(n,W)$ will consist of two disconnected cliques, one consisting of the vertices $V_1 := \{i : U_i \in [0,\frac{1}{2})\}$ and the other one consisting of the vertices $V_2 := \{i : U_i \in [\frac{1}{2},1]\}$. By the central limit theorem, we have $|V_1| \approx \frac{n}{2} + \mathsf{N} \cdot \sqrt{n}$, where N is a normal random variable with mean 0 and variance $\frac{1}{4}$. Deterministically, $|V_2| = n - |V_1| \approx \frac{n}{2} - \mathsf{N} \cdot \sqrt{n}$. Hence, the number of K_r's in $\mathbb{G}(n,W)$ is equal to

$$\binom{|V_1|}{r} + \binom{|V_2|}{r} \approx \binom{\frac{n}{2} + \mathsf{N} \cdot \sqrt{n}}{r} + \binom{\frac{n}{2} - \mathsf{N} \cdot \sqrt{n}}{r} \approx \frac{1}{r!} \cdot \Big(\Big(\frac{n}{2} + \mathsf{N} \cdot \sqrt{n} \Big)^r + \Big(\frac{n}{2} - \mathsf{N} \cdot \sqrt{n} \Big)^r \Big) \,.$$

The first terms in the binomial expansions of these two summands add up to $\frac{2n^r}{2^r}$ which is non-random (and hence does not contribute to fluctuations). The second terms in the binomial expansions cancel out. Hence the main part of the fluctuations comes from the sum of the third terms,

$$\binom{r}{2} (\tfrac{n}{2})^{r-2} (\mathsf{N} \cdot \sqrt{n})^2 + \binom{r}{2} (\tfrac{n}{2})^{r-2} (-\mathsf{N} \cdot \sqrt{n})^2 = 2^{3-r} \binom{r}{2} \cdot \mathsf{N}^2 \cdot n^{r-1} \,.$$

In particular, we see that these fluctuations are of order n^{r-1} and involve the square of a normal random variable.

Connection to generalized U-statistics. During the post-submission revision the authors became aware of a result by Janson and Nowicki [12] on orthogonal decomposition of generalized U-statistics of independent random variables indexed by both vertices and edges of a graph (see also [13]). In particular, Theorem 2 in [12] implies Theorem 1.1. To see this, one has to realize that the clique count can be expressed as a U-statistic by generating independent uniform random variables $Y_{uv} \in [0,1]$ (independent of $U_1, ..., U_n$) and defining $\mathbb{G}(n, W)$ as the graph with edge set

$$\{uv : Y_{uv} \leq W(U_u, U_v)\},$$

where $U_v, v \in [n]$ are the random types as defined above.

The contribution of this paper is thus (i) a direct proof of limit distribution using the method of moments; (ii) explicit description of the limit distribution in terms of the graphon W (which admittedly also can be read out relatively easily from the result of Janson and Nowicki); (iii) establishing a link between asymptotic normality of clique counts and quasirandomness (see Conjecture 1.4 and the following comment).

Before stating the explicit version of Theorem 1.1 we need to introduce some definitions about spectra of graphons and more advanced concepts related to subgraph densities. On the way, we also recall facts that will be useful for the proof of the main theorem.

1.3. **Spectrum of a graphon and cycle densities.** Much of the spectral theory of graphs carries over to the dense graph limit setting, where graphons play the role that adjacency matrices of graphs play in algebraic graph theory. Spectral properties of graphons will be crucial in stating and proving the explicit version of Theorem 1.1(c).

In this section, we follow [14, Section 7.5], where details and proofs can be found. We work with the real Hilbert space $L^2[0,1]$. Suppose that $W : [0,1]^2 \to [0,1]$ is a graphon. Then we can associate with W its kernel operator $T_W : L^2[0,1] \to L^2[0,1]$ by setting

$$(T_W f)(x) = \int_0^1 W(x,y) f(y)\, dy$$

for each $f \in L^2[0,1]$. The operator T_W is a Hilbert–Schmidt operator and that has a discrete spectrum. That is, there exists a countable multiset, denoted $\mathrm{Spec}(W)$, of nonzero real *eigenvalues associated with* W. Moreover, we have that

$$\sum_{\lambda \in \mathrm{Spec}(W)} \lambda^2 = \int_{[0,1]^2} W(x,y)^2\, dx\, dy \leq 1. \tag{4}$$

The *degree function* of a graphon W is the function $\deg_W : [0,1] \to [0,1]$ defined as $\deg_W(x) = \int_y W(x,y) dy$. In order to gain some intuition about the degree function, the reader should note that if a vertex in $\mathbb{G}(n, W)$ is conditioned to have type x, it has expected

degree $(n-1)\cdot \deg_W(x)$. We say that W is *regular* if $\deg_W(x) \equiv d$ for some constant d. (Note that in such case $d = \int_{[0,1]^2} W(x,y)\,dx\,dy$.)

Observe that if W is regular, then $f \equiv 1$ is an eigenfunction of T_W with eigenvalue d. In this case, let $\mathrm{Spec}^-(W)$ be $\mathrm{Spec}(W)$ with the multiplicity of d decreased by 1. (It can also be shown that all eigenvalues are at most d in absolute value, but this is not necessary for our proof.)

One of the most useful properties of eigenvalues of a graphon is that they give a simple expression for cycle densities. Recall that C_ℓ is a cycle on ℓ vertices (with C_2 being a multigraph consisting of a double edge). We have (see [14, (7.22), (7.23)]) that

$$t(C_\ell, W) = \sum_{\lambda \in \mathrm{Spec}(W)} \lambda^\ell, \qquad \text{for } \ell \geq 2\,. \tag{5}$$

Note that for $\ell = 2$, the formula (5) is exactly (4).

The formula (5) plays a key part in the proof Theorem 1.2(c), and is the only fact that we use from the spectral theory of graphons.

1.4. **Conditional densities, K_r-regular graphons and $V_W^{(r)}$.** If H is a fixed multigraph and W is a graphon, the *density* (usually called *homomorphism density*) of H in W is defined as

$$t(H,W) := \mathbb{E} \prod_{\{i,j\}\in E(H)} W(U_i, U_j)\,. \tag{6}$$

(Notice that if the edge $\{i,j\}$ has multiplicity m in H, then the corresponding contribution to the density equals $W(U_i,U_j)^m$.) When H is a simple graph on k vertices, then the constant $t(H,W) \in [0,1]$ is the probability that a particular copy of H is present in $\mathbb{G}(n,W)$, which implies

$$\mathbb{E}[X_n(H,W)] = \frac{(n)_k}{\mathrm{aut}(H)} t(H,W),$$

where $\mathrm{aut}(H)$ is the number of automorphisms of H, and $(n)_k = n\cdot(n-1)\cdot\ldots\cdot(n-k+1)$. In particular we have that $\mathbb{E}X_n(K_r,W) = \binom{n}{r} t(K_r,W)$, a fact that will be used several times throughout the paper.

For a natural number k, we write $[k] := \{1,2,\ldots,k\}$. Given an integer $\ell \leq k$, let $\binom{[k]}{\ell}$ denote all ℓ-element subsets of $[k]$. Let $J \in \binom{[k]}{\ell}$ and suppose that H is a graph on the vertex set $[k]$ for which the vertices from the set J are considered as *marked*. Given a vector $\mathbf{x} = (x_j)_{j\in J} \in [0,1]^\ell$, we define

$$t_{\mathbf{x}}(H,W) = \mathbb{E}\left[\prod_{\{i,j\}\in E(H)} W(U_i,U_j) \,\middle|\, U_j = x_j : j \in J \right]. \tag{7}$$

Again, if H is simple with $|V(H)| = r$, then $t_{\mathbf{x}}(H, W)$ is the conditional probability that $\mathbb{G}(r, W) = H$ whenever vertex j is prescribed a type x_j, for $j \in J$. Note that, when $H = K_r$ is the r-clique, the function $\mathbf{x} \mapsto t_{\mathbf{x}}(K_r, W)$ depends only on the cardinality of J (and not on J itself). In this case, we write $K_r^\bullet$ and $K_r^{\bullet\bullet}$ for K_r with one, respectively two, marked vertices and denote the corresponding conditional densities by $t_x(K_r^\bullet, W)$ and $t_{x,y}(K_r^{\bullet\bullet}, W)$.

A graphon W is called *K_r-free* if $t(K_r, W) = 0$ and called *complete* if W equals 1 almost everywhere.

We say that W is *K_r-regular* if for almost every $x \in [0, 1]$ we have

$$t_x\left(K_r^\bullet, W\right) = t(K_r, W)\ .$$

In the case $r = 2$, we have $t_x\left(K_r^\bullet, W\right) = \deg_W(x)$, hence K_2-regularity coincides with the usual concept of regularity. For $r \geq 3$ condition of K_r-regularity means that in the random graph $\mathbb{G}(n, W)$ a vertex is expected to belong to the same number of copies of K_r, regardless of its type. Another important aspect of K_r-regularity of W is that any two particular copies of K_r that share exactly one vertex are uncorrelated, that is, in $\mathbb{G}(n, W)$ existence of one of the these two copies does not influence the probability of the existence of the other. As we will see in the proof, if W is *not* K_r-regular, then a pair of K_r-copies sharing one vertex are *positively* correlated (see (20)), intuitively causing larger variance of the K_r count.

As we will see in Theorem 1.2, the limit distribution of edges (i.e. case $r = 2$) can be described in terms of the function $\deg_W$ and the spectrum of W. For $r \geq 3$, however, we need to consider an auxiliary graphon $V_W^{(r)}$ defined below which encodes the information about the local clique densities in W.

Suppose that W is a graphon and $r \geq 2$. Then we define a graphon $V_W^{(r)} : [0, 1]^2 \to [0, 1]$ by setting

$$V_W^{(r)}(x, y) := t_{x,y}(K_r^{\bullet\bullet}, W). \tag{8}$$

So, $V_W^{(r)}(x, y)$ is intuitively the density of K_r's containing x and y. Note that $V_W^{(2)} = W$.

Suppose that we have two numbers $r \in \mathbb{N}$ and $j \in \{0, \ldots, r\}$. We write $K_r \ominus_j K_r$ for the (simple) graph on $2r - j$ vertices consisting of two copies of K_r sharing j vertices. For $j = 2$, we also denote by $K_r \oplus_2 K_r$ the multigraph obtained from $K_r \ominus_2 K_r$ by doubling the shared edge. In particular $K_2 \ominus_2 K_2 = K_2$ and $K_2 \oplus_2 K_2 = C_2$.

Denote by K_r' and K_r'' two copies of K_r that share exactly two vertices, which have labels 1 and 2. We have

$$\prod_{ij \in K_r \oplus K_r} W(U_i, U_j) = W(U_1, U_2) \prod_{ij \in K_r \ominus_2 K_r} W(U_i, U_j) = \prod_{ij \in K_r'} W(U_i, U_j) \prod_{ij \in K_r''} W(U_i, U_j)$$

Taking conditional expectation with respect to the event $U_1 = x, U_2 = y$ and noting that products $\prod_{ij\in K_r'} W(U_i, U_j)$ and $\prod_{ij\in K_r''} W(U_i, U_j)$ are conditionally independent, from definition (7) we obtain

$$t_{x,y}(K_r \oplus_2 K_r, W) = W(x,y)t_{x,y}(K_r \ominus_2 K_r, W) = (t_{x,y}(K_r^{\bullet\bullet}, W))^2 = \left(V_W^{(r)}(x,y)\right)^2 . \tag{9}$$

Let

$$\sigma_{r,W}^2 := \frac{1}{2((r-2)!)^2}\left(t(K_r \ominus_2 K_r, W) - t(K_r \oplus_2 K_r, W)\right) . \tag{10}$$

We have $\sigma_{r,W}^2 \geq 0$, since, assuming that the two shared vertices have labels 1 and 2,

$$\begin{aligned} t(K_r \ominus_2 K_r, W) - t(K_r \oplus_2 K_r, W) &= \int (t_{x,y}(K_r \ominus_2 K_r) - t_{x,y}(K_r \oplus_2 K_r))\, dx\, dy \\ &\overset{(9)}{=} \int [1 - W(x,y)] t_{x,y}(K_r \ominus_2 K_r)\, dx\, dy \\ &\geq \int 0\, dx\, dy = 0. \end{aligned} \tag{11}$$

Observe that

$$\deg_{V_W^{(r)}}(x) = \int_0^1 V_W^{(r)}(x,y)\, dy = \int_0^1 t_{x,y}(K_r^{\bullet\bullet}, W)\, dy = t_x(K_r^{\bullet}, W).$$

So W is K_r-regular if and only if $V_W^{(r)}$ is regular, with $\deg_{V_W^{(r)}} \equiv t_r := t(K_r, W)$. Hence, by the remark we made in Section 1.3, one of the eigenvalues associated with $V_W^{(r)}$ is t_r. In this case, $\mathrm{Spec}^-(V_W^{(r)})$ is $\mathrm{Spec}(V_W^{(r)})$ with the multiplicity of t_r decreased by 1.

1.5. **Statement of the main result.** We are now ready to state our main result.

Theorem 1.2. *Let W be a graphon. Fix $r \geq 2$ and set $t_r = t(K_r, W)$. Let $X_{n,r} = X_n(K_r, W)$ be the number of r-cliques in $\mathbb{G}(n, W)$. Then we have the following.*

(a) If W is K_r-free or complete then almost surely $X_{n,r} = 0$ or $X_{n,r} = \binom{n}{r}$, respectively.

(b) If W is not K_r-regular, then

$$\frac{X_{n,r} - \binom{n}{r} t_r}{n^{r-\frac{1}{2}}} \xrightarrow{\mathrm{d}} \hat{\sigma}_{r,W} \cdot Z\ , \tag{12}$$

where Z is a standard normal random variable and

$$\hat{\sigma}_{r,W} = \frac{1}{(r-1)!}\left(t(K_r \ominus_1 K_r, W) - t_r^2\right)^{1/2} > 0\ .$$

(c) If W is a K_r-regular graphon which is neither K_r-free nor complete, then

$$\frac{X_{n,r} - \binom{n}{r} t_r}{n^{r-1}} \xrightarrow{\mathrm{d}} \sigma_{r,W} \cdot Z + \frac{1}{2(r-2)!} \sum_{\lambda \in \mathrm{Spec}^-(V_W^{(r)})} \lambda \cdot (Z_\lambda^2 - 1)\ , \tag{13}$$

where Z and $(Z_\lambda)_{\lambda\in\mathrm{Spec}^-(V_W^{(r)})}$ are independent standard normal and $\sigma_{r,W}$ is defined in (10)*. (The series on the right-hand side of* (13) *converges a.s. and in L^2 by Lemma 2.1.)*

Let us comment briefly on Theorem 1.2. Part (a) is immediate. Part (b) tells us that in this setting we have a behaviour as in the central limit theorem; indeed we could have stated Part (b) as $(X_{n,r}-\mathbb{E}[X_{n,r}])/\sqrt{\mathrm{Var}[X_{n,r}]}\xrightarrow{\mathrm{d}} Z$. Finally, we will see in Remark 1.6 below that the limit in (13) is non-degenerate, implying the Theorem 1.2 gives a complete description of limit distributions of small cliques in $\mathbb{G}(n,W)$.

Let us mention that Part (b) has been recently reported in [9], using a framework of the so-called mod-Gaussian convergence, developed in that paper. This concept actually gives much more: firstly, the authors establish normal behaviour under conditions analogous to those in Part (b) also for other graphs than K_r. Secondly, they also prove a moderate deviation principle and a local limit theorem in this setting. So the reason we provide a proof of Part (b) is that ours — based on Stein's method — is much simpler (because we are proving a weaker statement). But the main emphasis of the paper is on Part (c), which is new and deals with a regime exhibiting a more exotic behaviour.

1.6. **When the distribution in Theorem 1.2(c) is normal or normal-free.** Recall that a chi-square distribution with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables. Therefore the series in (13) is a weighted infinite-dimensional variant of a chi-square distribution. By (4) and (17) below, this random variable has finite variance. Interestingly, very similar distributions appear in [4] and [5], also in connection with graph limits. That said, the particular setting of our paper seems to be substantially different from [4, 5].

Proposition 1.3. *In Theorem 1.2(c) the limit distribution is normal if and only if $V_W^{(r)}\equiv t_r$.*

Proof. In view of (13), a normally distributed limit implies that $\mathrm{Spec}^-(V_W^{(r)})=\emptyset$ and therefore $\mathrm{Spec}(V_W^{(r)})=\{t_r\}$. Recall that $V_W^{(r)}$ is regular with degree function being constant t_r. We claim that then $V_W^{(r)}\equiv t_r$. This claim can be viewed as a graphon version of a consequence of Chung–Graham–Wilson Theorem on quasirandom graph sequences[1]; here we give a short self-contained proof. Indeed, regularity of $V_W^{(r)}$ implies $\deg_{V_W^{(r)}}\equiv t_r$, so we

[1] more precisely, the part "each spot has the same density $\iff$ non-principal eigenvalues are small", denoted as P_4 and P_3 in [1, page 158]

have that

$$t_r^2 = \left(\int_{[0,1]} \deg_{V_W^{(r)}}(y)\, dy\right)^2 = \left(\int_{[0,1]^2} V_W^{(r)}(x,y)\, dx\, dy\right)^2$$

$$\boxed{\text{Jensen's inequality}} \quad \leq \int_{[0,1]^2} \left(V_W^{(r)}(x,y)\right)^2 dx\, dy \overset{(4)}{=} \sum_{\lambda \in \mathrm{Spec}(V_W^{(r)})} \lambda^2 = t_r^2\,.$$

Since the quadratic function is strictly convex, equality is Jensen's inequality is attained if and only if $V_W^{(r)}$ is constant, which implies $V_W^{(r)} \equiv t_r$. □

So, the question now is which graphons W lead to a constant graphon $V_W^{(r)}$. Since $V_W^{(2)} = W$, for $r = 2$ the answer is clearly given by constant graphons W. For $r \geq 3$, we put forward the following conjecture, which was first hinted in concluding remarks of [17].

Conjecture 1.4. *Suppose that $r \geq 3$ and $V_W^{(r)}$ is a constant-d graphon for some $d \in [0,1]$, that is, $t_{\cdot,\cdot}(K_r^{\bullet\bullet}, W) \equiv d$. Then W is K_r-free (when $d = 0$), or $W \equiv d^{1/\binom{r}{2}}$.*

In [17], the case $r = 3$ of the aforementioned conjecture was shown to be true. Therefore, we know that if W is a graphon which is K_3-regular and not K_3-free, then the only way we can get normal limit distribution in Theorem 1.2(c) is when W is a constant graphon. Conjecture 1.4 can also be rephrased as follows: among random graphs $\mathbb{G}(n, W)$, where W is a K_r-regular graphon, only $\mathbb{G}(n,p)$, $p \in (0,1)$, has asymptotically normal count of K_r.

Let us now comment on a complementary question: when is the normal term absent in (13)?

Proposition 1.5. *We have $\sigma_{r,W} = 0$ if and only if $W(x,y) = 1$ for almost every $(x,y) \in [0,1]^2$ for which $t_{x,y}(K_r^{\bullet\bullet}, W) > 0$.*

Proof. We first observe that $t_{x,y}(K_r^{\bullet\bullet}, W) > 0$ if and only if $t_{x,y}(K_r \ominus_2 K_r, W) > 0$. For $W(x,y) > 0$ this is immediate from the second equality in (9), while for $W(x,y) = 0$ from definition of $t_{x,y}(H,W)$ is is clear that both $t_{x,y}(K_r \ominus_2 K_r, W) = 0$ and $t_{x,y}(K_r^{\bullet\bullet}, W) = 0$. From (10) and (11), we have $\sigma_{r,W} = 0$ if and only only if $W(x,y) = 1$ for almost every $(x,y) \in [0,1]^2$ for which $t_{x,y}(K_r \ominus_2 K_r, W) > 0$. Together with the first observation this proves the proposition. □

For $r = 2$ the condition of Proposition 1.5 is equivalent to a condition that $W \in \{0,1\}$ almost everywhere. For $r \geq 3$ we have more freedom for constructions. For example, take $r = 3$, partition $[0,1]$ into 6 sets of measure $\frac{1}{6}$ each and put one copy of the complete 3-partite graphon on the first 3 sets and another copy on the last 3 sets. Make arbitrarily wild connections between the 1st and the 4th set, and set the rest of the connections

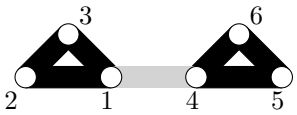


FIGURE 1. Example of K_3-regular graphon (depicted here as a graph) with no linear term in the non-normal limit.

between the first 3 and the last 3 sets to 0 (see Figure 1). Such a graphon W is K_3-regular but we have $\sigma_{r,W} = 0$.

Remark 1.6. *The limit in* (13) *is never degenerate. Indeed, if the non-normal part vanishes, then Proposition 1.3 implies* $V_W^{(r)} \equiv t_r \in (0,1)$. *Further, from* (9) *we obtain that almost everywhere*

$$W(x,y)t_{x,y}(K_r \ominus_2 K_r, W) = \left(V_W^{(r)}(x,y)\right)^2 = t_r^2 > 0.$$

In particular this implies $t_{x,y}(K_r \ominus_2 K_r, W) > 0$ *almost everywhere. Moreover, we have* $W < 1$ *on a set of positive measure (otherwise* $W \equiv 1$ *and we would be in the case (a)), which implies that the inequality in* (11) *is strict. Since the left-hand side of* (11) *is a multiple of* $\sigma_{r,W}^2$, *this implies that* $\sigma_{r,W}^2 > 0$.

1.7. **Acknowledgment.** We thank the anonymous referee for a very detailed report which helped us to improve the exposition.

2. Preliminaries

In this section we state definitions and facts needed in the proof of Theorem 1.2. It can be skipped at the first reading.

Asymptotic notation like $a_n = O(b_n)$ and $a_n \sim b_n$ (equivalently, $a_n = (1 + o(1))b_n$) is stated with respect to $n \to \infty$.

2.1. **Hypergraphs, associated graphs, and further spectral properties.** Given $r \geq 2$, an *r-uniform hypergraph* $\mathcal{H}$ on vertex set V is a family of r-element subsets (called *hyperedges*) of V. In this paper we assume that $\mathcal{H}$ is a multiset (even though a term *multihypergraph* would be a more standard term). We omit the words "r-uniform", when this is clear from the context. By $|\mathcal{H}|$ we denote the number of hyperedges, counting multiplicities. The *degree* of a vertex $v \in V$, denoted by $\deg_{\mathcal{H}}(v)$, is the number of hyperedges (counting multiplicities) of $\mathcal{H}$ containing v. We say that $\mathcal{H}$ is *spanning* if $V = \cup_{e\in\mathcal{H}} e$. Given a hypergraph $\mathcal{H}$, the *graph associated with* $\mathcal{H}$, sometimes also called the clique graph of $\mathcal{H}$, is a graph on the same vertex set, where each hyperedge S of $\mathcal{H}$ is replaced by a clique on S, with multiple edges being replaced by single ones.

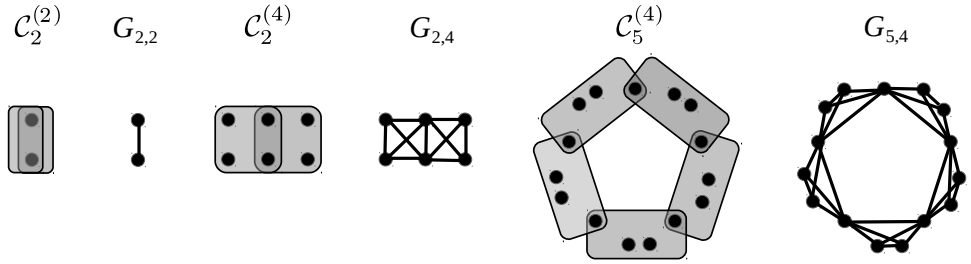


FIGURE 2. Examples of hypergraphs $\mathcal{C}_\ell^{(r)}$ and their associated graphs $G_{\ell,r}$.

We use a particular hypergraph version of cycles, known as loose cycles. For $\ell \geq 2$, let $\mathcal{C}_\ell^{(r)}$ be a hypergraph with ℓ edges, each of size r, created from a graph C_ℓ by inserting into each edge additional $r-2$ vertices, all $\ell(r-2)$ new vertices being distinct (hence $\mathcal{C}_2^{(r)}$ is a pair of r-sets sharing exactly 2 vertices). Finally, let $G_{\ell,r}$ be the graph associated with $\mathcal{C}_\ell^{(r)}$. Note that $G_{2,r} = K_r \ominus_2 K_r$. See Figure 2 for examples.

We can express the densities of cycles in the graphon $V_W^{(r)}$, defined in (8), in terms of the graphon W as follows. Assume that the vertices shared by consecutive r-cliques in $G_{\ell,r}$ have labels $1,\ldots,\ell$. Denoting $F(x_1,\ldots,x_r) = \prod_{1\leq i<j\leq r} W(x_i,x_j)$, we write

$$V_W^{(r)}(x_i,x_{i+1}) = \int F(x_i,x_{i+1},y_{i,3},\ldots,y_{i,r}) \prod_{j=3}^{r} dy_{i,r},$$

whence, writing indices of x_i's modulo r and using Fubini's theorem,

$$t(C_\ell, V_W^{(r)}) = \int \prod_{i=1}^{r} \left(\int F(x_i,x_{i+1},y_{i,3},\ldots,y_{i,r}) \prod_{j=3}^{r} dy_{i,j} \right) \prod_{i=1}^{r} dx_i$$

$$= \int \prod_{i=1}^{r} F(x_i,x_{i+1},y_{i,3},\ldots,y_{i,r})\; dx_1 \ldots\; dx_r\; dy_{1,3}\ldots\; dy_{1,r}\ldots\; dy_{r,3}\ldots\; dy_{r,r}. \tag{14}$$

For $\ell \geq 3$, a moment of thought reveals that the last integral is exactly $t(G_{\ell,r}, W)$, implying

$$t\left(C_\ell, V_W^{(r)}\right) = t\left(G_{\ell,r}, W\right), \qquad \text{for } \ell \geq 3 \text{ and } r \geq 2\,. \tag{15}$$

On the other hand, for $\ell = 2$ we have to be a bit more careful, noting that in the product (14) factor $W(x_1,x_2)$ appears twice, implying

$$t\left(C_2, V_W^{(r)}\right) = t\left(K_r \oplus_2 K_r, W\right), \qquad \text{for } r \geq 2\,, \tag{16}$$

where, recall, $K_r \oplus_2 K_r$ is a multigraph obtained by gluing two cliques along two vertices (with a double edge between these vertices).

2.2. **Moment generating functions.** As we noted above, the main result of [4] expresses a particular random variable as a sum of squares of independent normal random variables, which is very similar to the expression appearing in our Theorem 1.2(c). The following lemma asserts that such distributions are well-defined and gives a formula of their moment generating functions. Note that in [4] only the convergence of the series in L^1 is asserted, but it clearly converges in L^2 as well, since $S_n := \sum_{j=1}^{n} \lambda_j \left(Z_j^2 - 1\right)$ is easily seen to be a Cauchy sequence in the (complete) space L^2. In particular this implies that $\mathrm{Var}[S_n] \to \mathrm{Var}[S]$.

Lemma 2.1 (see [4], Proposition 7.1[2]). *Let $\{\lambda_j\}_j$ be a finite or countable sequence of real numbers such that $\sum_j \lambda_j^2 < +\infty$ and let $\{Z_j\}_j$ be independent standard normal random variables. Define a (possibly infinite) sum $S = \sum_j \lambda_j (Z_j^2 - 1)$. Then S converges almost surely and in L^2 and*

$$\mathrm{Var}[S] = 2 \sum_j \lambda_j^2. \tag{17}$$

Furthermore, the moment generating function $M_S(t) := \mathbb{E}\left[\exp(tS)\right]$ is finite for $|t| < \frac{1}{8}\left(\sum_j \lambda_j^2\right)^{-1/2}$ and equals

$$M_S(t) = \prod_j \frac{\exp(-\lambda_j t)}{\sqrt{1 - 2\lambda_j t}}\,.$$

2.3. **Stein's method and the Wasserstein distance.** Stein's method is one of the most powerful tools for obtaining limit theorems (in fact it gives much more, by providing bounds on approximation errors). It is particularly efficient for sums of dependent random variables, when each variable depends on a relatively small (but not necessarily bounded) number of other variables. Here, we follow a survey article by Ross [18, Section 3].

The dependency structure of a collection of random variables $(Y_i : i \in I)$ is encoded by a *dependency graph* $\mathcal{G}$ (on a vertex set I) if, for all $i \in I$, Y_i is independent of the random variables $\{Y_j\}_{j \notin N_i}$, where N_i is the neighborhood of i in $\mathcal{G}$ (including i itself). In general, a dependency graph is not uniquely determined, but in many scenarios, there exists a dependency graph, which naturally arises by capturing the obvious dependencies, and which is also minimal among all dependency graphs.

We shall work with the Wasserstein distance between two random variables, say X and Y, which we denote by $d_{\mathrm{Wass}}(X, Y)$. We do not need an exact definition — which the reader can find on page 214 of [18] — since the only property of the Wasserstein

[2]Note that there is an error in the arXiv version of Proposition 7.1 in [4].

distance that we will use is that for $Z \sim \mathcal{N}(0,1)$ and a sequence X_n of random variables $d_{\mathrm{Wass}}(X_n, Z) \to 0$ implies that $X_n \xrightarrow{\mathrm{d}} Z$ (see [18, Section 3]). We will use the following off-the-shelf bound for the Wasserstein distance based on Stein's method.

Theorem 2.2 (see Theorem 3.6 in [18]). *Let $(Y_i : i \in I)$ be a finite collection of random variables such that for every $i \in I$ we have $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}[Y_i^4] < \infty$. Writing $\sigma^2 := \mathrm{Var}\left[\sum_{i\in I} Y_i\right]$, define $Q = \sum_{i\in I} Y_i/\sigma$. Let $\mathcal{G}$ be a dependency graph of $(Y_i : i \in I)$. Writing $D = \max_{i\in I} |N_i|$, we have*

$$d_{\mathrm{Wass}}(Q, Z) \leq \frac{D^2}{\sigma^3} \cdot \sum_i \mathbb{E}\left[|Y_i|^3\right] + \frac{\sqrt{28} D^{3/2}}{\sqrt{\pi} \cdot \sigma^2} \cdot \sqrt{\sum_i \mathbb{E}\left[Y_i^4\right]}. \tag{18}$$

3. Proof of Theorem 1.2(b)

We will apply Theorem 2.2 to a collection $(Y_R : R \in \binom{[n]}{r})$ of random variables indexed by r-sets R of vertices. Writing I_R for the indicator that R induces a clique in $\mathbb{G}(n, W)$, we define $Y_R = I_R - \mathbb{E}[I_R] = I_R - t_r$. In this notation we have $X_{n,r} - \binom{n}{r} t_r = \sum_{R\in\binom{[n]}{r}} Y_R$.

We let $\mathcal{G}$ be the natural dependency graph of the collection $(Y_R : R \in \binom{[n]}{r})$ with edges corresponding to pairs $R_1 R_2$ such that $R_1 \cap R_2 \neq \emptyset$.

The proof consists of two steps. In the first one we bound the maximum degree of $\mathcal{G}$. This step is fairly straightforward. In the second and less elementary step we compute the asymptotics of the variance of $\sum_R Y_R$. As we will see, the leading term in the asymptotics comes from pairs of cliques that share exactly one vertex.

Let us proceeding with the details of the first step. Notice that in $\mathcal{G}$ every neighbourhood N_R has the same size D, namely,

$$D = \sum_{\ell=1}^{r} \binom{r}{\ell}\binom{n-r}{r-\ell} = O(n^{r-1}). \tag{19}$$

We now turn to the second step. Recall that $K_r \ominus_j K_r$ is a simple graph consisting of two r-cliques sharing j vertices. Set $d_j = t(K_r \ominus_j K_r, W)$, for $j \in [r]$. By Jensen's inequality we have that

$$t_r^2 = \left(\int_0^1 t_x(K_r^\bullet, W)\, dx\right)^2 < \int_0^1 t_x(K_r^\bullet, W)^2\, dx = d_1, \tag{20}$$

where the strict inequality follows from the assumption that W is not K_r-regular.

For disjoint R_1, R_2, variables Y_{R_1}, Y_{R_2} are independent, while $|R_1 \cap R_2| = \ell \geq 1$ implies $\mathbb{E}[Y_{R_1} Y_{R_2}] = d_\ell - t_r^2$, and (20) implies $\hat{\sigma}^2_{r,W} = (d_1 - t_r^2)/((r-1)!)^2 > 0$. Since $\mathbb{E}[Y_R] = 0$ for every R, we have $\sigma_n^2 := \mathrm{Var}\left[\sum_R Y_R\right] = \sum_{R_1,R_2} \mathbb{E}[Y_{R_1} Y_{R_2}]$. Since the number of ordered

pairs R_1, R_2 such that $|R_1 \cap R_2| = \ell$ is

$$\binom{n}{\ell}\binom{n-\ell}{r-\ell}\binom{n-r}{r-\ell} \sim \frac{n^{2r-l}}{\ell!(r-\ell)!^2},$$

we obtain

$$\begin{aligned}\sigma_n^2 &= \sum_{\ell=1}^{r} \frac{n^{2r-\ell}(1+o(1))}{\ell!(r-\ell)!^2}\left(d_\ell - t_r^2\right) \\ &= \frac{n^{2r-1}}{(r-1)!^2}\left(d_1 - t_r^2\right) + \sum_{\ell=2}^{r} O(n^{2r-\ell}) \\ &\sim \hat{\sigma}_{r,W}^2 n^{2r-1}\,. \qquad (21)\end{aligned}$$

Writing $Q_n = \sum_{R\in\binom{[n]}{r}} Y_R/\sigma_n$, we are ready to apply Theorem 2.2. Crudely bounding each of the sums of moments on the right-hand side of (18) by $\binom{n}{r} \le n^r$, and using (19) and (21), we obtain that $d_{\mathrm{Wass}}(Q_n, Z) = O(n^{-1/2}) \to 0$. By the remark we made just before Theorem 2.2, we conclude that $Q_n \xrightarrow{\mathrm{d}} Z$. In view of (21) and Slutsky's theorem,

$$\frac{\sum_{R\in\binom{[n]}{r}} Y_R}{n^{r-1/2}} = \frac{\sigma_n}{n^{r-1/2}} \cdot Q_n \xrightarrow{\mathrm{d}} \hat{\sigma}_{r,W} Z,$$

which completes the proof, in view of $\sum_{R\in\binom{[n]}{r}} Y_R = X_{n,r} - \binom{n}{r}t_r$.

4. Proof of Theorem 1.2(c)

Define a random variable as on the right-hand side of (13),

$$Y = \sigma_{r,W} \cdot Z + \frac{1}{2(r-2)!} \sum_{\lambda\in\mathrm{Spec}^-(V_W^{(r)})} \lambda \cdot (Z_\lambda^2 - 1). \qquad (22)$$

We employ the method of moments, in the way it is described in Section 6.1 of [11]. For this it is enough to show that the moments of the random variable $(X_{n,r} - \binom{n}{r}t_r)/n^{r-1}$ converge to the corresponding moments of the random variable Y, and to verify that the moment generating function $M_Y(t) = \mathbb{E}[e^{tY}]$ is finite in some neighbourhood of zero (so that the distribution of Y is determined by its moments).

Recall that for a standard normal random variable Z we have $M_Z(x) = \exp\left(x^2/2\right)$ and hence $M_{\sigma_{r,W}\cdot Z}(x) = \exp\left(\frac{\sigma_{r,W}^2 x^2}{2}\right)$. On the other hand, Lemma 2.1 tells us that the moment generating function of the second summand in (22) is

$$\prod_{\lambda\in\mathrm{Spec}^-(V_W^{(r)})} \exp\left(-\frac{\lambda x}{2(r-2)!}\right) \Big/ \sqrt{1 - \frac{\lambda x}{(r-2)!}}\,.$$

Since the moment generating function of a sum of independent random variables equals the product of the moment generating functions of individual generating functions, it follows that

$$M_Y(x) = \exp\left(\frac{\sigma_{r,W}^2 x^2}{2}\right) \prod_{\lambda \in \mathrm{Spec}^-(V_W^{(r)})} \frac{\exp\left(-\frac{\lambda x}{2(r-2)!}\right)}{\sqrt{1-\frac{\lambda x}{(r-2)!}}}\,. \tag{23}$$

On the other hand, to compute the moments of $(X_{n,r} - \binom{n}{r}t_r)/n^{r-1}$, we write

$$X_{n,r} - \binom{n}{r}t_r = \sum_{R \in \binom{n}{r}} (I_R - t_r)\,,$$

where, as in Section 3, I_R is the indicator of the event that the set of vertices R induces a clique in $\mathbb{G}(n,W)$. Given an m-tuple $(R_1,\dots,R_m)$ of (not necessary distinct) elements of $\binom{[n]}{r}$, let

$$\Delta(R_1,\dots,R_m) := \mathbb{E}\left[\prod_{i=1}^{m}(I_{R_i} - t_r)\right]\,. \tag{24}$$

Then

$$\mathbb{E}\left[\left(X_{n,r} - \binom{n}{r}t_r\right)^m\right] = \sum_{(R_1,\dots,R_m)\in\binom{[n]}{r}^m} \Delta(R_1,\dots,R_m)\,. \tag{25}$$

It is hence natural to analyze $\Delta(R_1,\dots,R_m)$ depending on the structure of the tuple $(R_1,\dots,R_m)$. Our plan is as follows. First, we introduce a certain family $\mathfrak{X}(n,r,m) \subset \binom{[n]}{r}^m$ and show that $\Delta(R_1,\dots,R_m) = 0$ for each $(R_1,\dots,R_m) \in \mathfrak{X}(n,r,m)$. Next, we define another family $\mathfrak{F}(n,r,m) \subset \binom{[n]}{r}^m \setminus \mathfrak{X}(n,r,m)$. In (30) we will show that the set $\binom{[n]}{r}^m \setminus (\mathfrak{X}(n,r,m) \cup \mathfrak{F}(n,r,m))$ has size $O(n^{(r-1)m-1})$, and hence the corresponding tuples have a contribution which is negligible with respect to the renormalization of (25) by $n^{-(r-1)m}$ (as, for example, in the statement of Claim 4.5). We then classify the tuples in $\mathfrak{F}(n,r,m)$ according to the pattern in which they overlap so that in each class every tuple the contribution $\Delta(R_1,\dots,R_m)$ is the same. By obtaining an explicit expression for this contribution (in Claim 4.3) and counting the number of tuples in each class (in Claim 4.4) we will arrive at a rather complicated asymptotic formula, which we interpret as a coefficient of some reasonably simple power series (see Claim 4.5). Finally, with some luck we discover that this power series is exactly the moment generating function M_Y (see Claim 4.6). Let us give details now.

Let $\mathfrak{X}(n,r,m) \subset \binom{[n]}{r}^m$ be those m-tuples $(R_1,\dots,R_m)$ for which we have $|R_i \cap (\cup_{j\neq i} R_j)| \leq 1$ for some $i \in [m]$. Suppose now that $(R_1,\dots,R_m) \in \mathfrak{X}(n,r,m)$. Without loss of generality, suppose that $|R_m \cap (\cup_{j=1}^{m-1} R_j)| \leq 1$. Assume first that $|R_m \cap (\cup_{j=1}^{m-1} R_j)| = 1$, say

$\{v\} = R_m \cap (\cup_{j=1}^{m-1} R_j)$. If we condition on $U_v = x$, the indicator I_{R_m} becomes independent of $\{I_{R_i} : i \in [m-1]\}$, and hence

$$\Delta(R_1, \dots, R_m) = \int_0^1 \mathbb{E}\left[I_{R_m} - t_r \,\middle|\, U_v = x\right] \cdot \mathbb{E}\left[\prod_{i=1}^{m-1} (I_{R_i} - t_r) \,\middle|\, U_v = x\right] dx\,. \tag{26}$$

Since W is K_r-regular, we have $\mathbb{E}\left[I_{R_m} - t_r \,\middle|\, U_v = x\right] = t_x(K_r^\bullet, W) - t_r = 0$ for almost every x. Therefore $\Delta(R_1, \dots, R_m) = 0$, as claimed. An even simpler calculation yields the same conclusion when $|R_m \cap (\cup_{j=1}^{m-1} R_j)| = 0$. Hence, we can rewrite (25) as

$$\mathbb{E}\left[\left(X_{n,r} - t_r\binom{n}{r}\right)^m\right] = \sum_{(R_1,\dots,R_m)\in\binom{[n]}{r}^m \setminus \mathfrak{X}(n,r,m)} \Delta(R_1, \dots, R_m)\,. \tag{27}$$

Every m-tuple $(R_1, \dots, R_m)$ can be identified with a spanning hypergraph henceforth denoted $\mathcal{H}(R_1, \dots, R_m)$, with vertex set $\cup_i R_i$ and hyperedge multiset $\{R_1, \dots, R_m\}$.

The following claim provides a sharp upper bound on the number of vertices in $\mathcal{H}(R_1, \dots, R_m)$, for $(R_1, \dots, R_m) \in \binom{[n]}{r}^m \setminus \mathfrak{X}(n, r, m)$.

Claim 4.1. *Suppose that $(R_1, \dots, R_m) \in \binom{[n]}{r}^m \setminus \mathfrak{X}(n, r, m)$. The number v of vertices in the hypergraph $\mathcal{H} = \mathcal{H}(R_1, \dots, R_m)$ satisfies $v \le (r-1)m$. The equality is attained if and only if each hyperedge in $\mathcal{H}$ contains exactly 2 vertices of degree 2 and all other vertices have degree 1.*

Proof of Claim 4.1. Let k be the number of vertices in $\mathcal{H}$ of degree 1. Since $(R_1, \dots, R_m) \notin \mathfrak{X}(n, r, m)$ we have that

$$k \le (r-2)m\,. \tag{28}$$

Since $\mathcal{H}$ is spanning and $v - k$ vertices have degree at least 2, it follows that

$$k + 2(v-k) \le \sum_{v\in V(\mathcal{H})} \deg(v) = rm\,. \tag{29}$$

Therefore

$$v \overset{(29)}{\le} \frac{rm-k}{2} + k = \frac{rm+k}{2} \overset{(28)}{\le} \frac{rm+(r-2)m}{2} = (r-1)m\,,$$

and the first statement follows. To prove the second statement, suppose that the number of vertices in $\mathcal{H} = \mathcal{H}(R_1, \dots, R_m)$ satisfies $v = (r-1)m$. Then (28) and (29) are both equalities and so $\mathcal{H}$ consists of $(r-2)m$ vertices of degree 1, and m vertices of degree 2. Since, by assumption, every R_i contains at least two vertices of degree 2 it readily follows that it contains exactly two such vertices. The other implication is immediate. □

Let $\mathfrak{F}(n, r, m)$ be those $(R_1, \dots, R_m) \in \binom{[n]}{r}^m \setminus \mathfrak{X}(n, r, m)$ for which the corresponding hypergraph $\mathcal{H}(R_1, \dots, R_m)$ has $(r-1)m$ vertices. Since for each $(R_1, \dots, R_m) \in \binom{[n]}{r}^m \setminus (\mathfrak{X}(n, r, m) \cup \mathfrak{F}(n, r, m))$ we have $|\cup_i R_i| \le (r-1)m - 1$, we can record each element of

$\binom{[n]}{r}^m \setminus (\mathfrak{X}(n,r,m) \cup \mathfrak{F}(n,r,m))$ by an $((r-1)m-1)$-set of $[n]$, and then by specifying to which of the sets $R_1, \dots, R_m$ each element of that $((r-1)m-1)$-set is an element of. Thus,

(30)
$$\left| \binom{[n]}{r}^m \setminus (\mathfrak{X}(n,r,m) \cup \mathfrak{F}(n,r,m)) \right| \leq \binom{n}{(r-1)m-1} \cdot (2^m)^{(r-1)m-1} = O(n^{(r-1)m-1}) \,.$$

Since $|\Delta(R_1, \dots, R_m)| \leq 1$, from (27) and (30) we infer

$$\mathbb{E}\left[\left(X_{n,r} - t_r \binom{n}{r} \right)^m \right] = \sum_{(R_1,\dots,R_m) \in \mathfrak{F}(n,r,m)} \Delta(R_1, \dots, R_m) + O(n^{(r-1)m-1}) \,. \tag{31}$$

Now, fix $(R_1, \dots, R_m) \in \mathfrak{F}(n,r,m)$ and consider the hypergraph $\mathcal{H} = \mathcal{H}(R_1, \dots, R_m)$. Notice that when $r = 2$ then some edges in $\mathcal{H}$ may be double edges, but all hyperedges are simple when $r \geq 3$. Now, replace every r-edge, say R, of $\mathcal{H}$ by a 2-edge that consists of the vertices of R having degree 2 and notice that this results in a 2-regular multigraph, that is, a union of vertex-disjoint cycles and double edges. In particular, this implies that the hypergraph $\mathcal{H}$ is a union of vertex-disjoint loose cycles.

We now need to deal with the right-hand side of (24) for tuples in $\mathfrak{F}(n,r,m)$, that is for tuples corresponding to unions of loose cycles. We first note that we can factor it over cycles, namely if there is a partition $[m] = V_1 \cup \cdots \cup V_k$ such that each $\mathcal{H}(R_i : i \in V_j)$ is a loose cycle, then the random variables $\prod_{i \in V_j} (I_{R_i} - t_r), j = 1, \dots, k$ are independent and hence

$$\Delta(R_1, \dots, R_m) = \prod_j \Delta(R_i : i \in V_j). \tag{32}$$

In order to calculate the individual factors in (32), the following claim will be useful.

Claim 4.2. *For each $i, r \geq 2$, for any proper subhypergraph $\mathcal{C} \subset \mathcal{C}_i^{(r)}$, we have $\mathbb{E}\left[\prod_{R \in \mathcal{C}} I_R\right] = t_r^{|\mathcal{C}|}$.*

Proof of Claim 4.2. We proceed by induction on the number of hyperedges of $\mathcal{C}$. The case when $\mathcal{C} = \emptyset$ is trivial. So suppose that $\mathcal{C} \neq \emptyset$.

Since $\mathcal{C}$ is a proper subhypergraph of $\mathcal{C}_i^{(r)}$, it contains a hyperedge $S \in \mathcal{C}$ such that for $\mathcal{C}^- := \mathcal{C} \setminus \{S\}$ we have $|S \cap \bigcup \mathcal{C}^-| \leq 1$ (here and below $\bigcup \mathcal{H}$ stands for the union of the hyperedges of $\mathcal{H}$). See Figure 3. Let us deal first with the case $|S \cap \bigcup \mathcal{C}^-| = 1$, and let v be the vertex shared by S and $\bigcup \mathcal{C}^-$. By the same argument as in (26), we have

$$\mathbb{E}\left[\prod_{R \in \mathcal{C}} I_R\right] = \int_0^1 \mathbb{E}\left[I_S \,|\, U_v = x\right] \cdot \mathbb{E}\left[\prod_{R \in \mathcal{C}^-} I_R \,|\, U_v = x\right] \, dx,$$

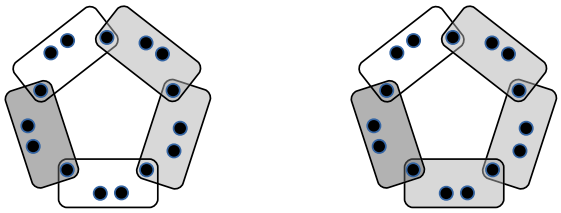


FIGURE 3. Two examples of a hypergraph $\mathcal{C}^-$ (light grey) and an edge S (dark grey) in a hypergraph $\mathcal{C}_5^{(4)}$ in the proof of Claim 4.2.

By the K_r-regularity, we have $\mathbb{E}[I_S \mid U_v = \cdot] \equiv t_r$. Thus, using the induction hypothesis on $\mathcal{C}^-$, we conclude that

$$\mathbb{E}\left[\prod_{R\in\mathcal{C}} I_R\right] = \int_0^1 t_r\, \mathbb{E}\left[\prod_{R\in\mathcal{C}^-} I_R \mid U_v = x\right]\, dx = t_r\, \mathbb{E}\left[\prod_{R\in\mathcal{C}^-} I_R\right] = t_r \cdot t_r^{|\mathcal{C}^-|}\,,$$

as was needed. The case $|S \cap \bigcup \mathcal{C}^-| = 0$ is even simpler:

$$\mathbb{E}\left[\prod_{R\in\mathcal{C}} I_R\right] = \mathbb{E}\left[I_S\right] \cdot \mathbb{E}\left[\prod_{R\in\mathcal{C}^-} I_R\right] = t_r \cdot t_r^{|\mathcal{C}^-|}\,.$$

□

Recall that for each $(R_1, \ldots, R_m) \in \mathfrak{F}(n, r, m)$, the hypergraph $\mathcal{H}(R_1, \ldots, R_m)$ is a union of loose cycles. Isomorphism classes of such hypergraphs can be encoded by a vector whose entry at position i is the number of loose cycles of length i. More precisely, let us consider the following set of $(m-1)$-dimensional vectors,

$$\mathcal{V}_m := \left\{ \mathbf{k} = (k_2, \ldots, k_m) \in \mathbb{N}_0^{m-1} : \sum_{i=2}^{m} i k_i = m \right\}. \tag{33}$$

Suppose that $\mathbf{k} \in \mathcal{V}_m$. Let $\mathcal{H}_{\mathbf{k}}^{(r)}$ denote the hypergraph formed by k_i copies of $\mathcal{C}_i^{(r)}$ for each $i = 2, \ldots, m$.

Claim 4.3. *Suppose that $(R_1, \ldots, R_m) \in \mathfrak{F}(n, r, m)$ is an m-tuple for which $\mathcal{H}(R_1, \ldots, R_m)$ is isomorphic to $\mathcal{H}_{\mathbf{k}}^{(r)}$, for some $\mathbf{k} \in \mathcal{V}_m$. Then*

$$\Delta(R_1, \ldots, R_m) = \prod_{\ell=2}^{m} \left(t\left(G_{\ell,r}, W\right) - t_r^{\ell} \right)^{k_\ell}, \tag{34}$$

where $G_{\ell,r}$ is the graph associated to $\mathcal{C}_\ell^{(r)}$, as defined in Section 2.1.

Proof of Claim 4.3. In view of (32), it is enough to prove (34) for an ℓ-tuple $(Q_1,\ldots,Q_\ell)\in\mathfrak{F}(n,r,\ell)$ for which $\mathcal{H}(Q_1,\ldots,Q_\ell)$ is isomorphic to $\mathcal{C}_\ell^{(r)}$. We have

$$\begin{aligned}\Delta(Q_1,\ldots,Q_\ell) &= \sum_{A\subseteq[\ell]}(-t_r)^{\ell-|A|}\cdot\mathbb{E}\left[\prod_{j\in A}I_{Q_j}\right]\\ \boxed{\text{Claim 4.2}} &= \mathbb{E}\left[\prod_{j\in[\ell]}I_{Q_j}\right]+\sum_{A\subsetneq[\ell]}(-t_r)^{\ell-|A|}\,t_r^{|A|}\\ &= \mathbb{E}\left[\prod_{j\in[\ell]}I_{Q_j}\right]+(t_r-t_r)^\ell-t_r^\ell\\ &= t\,(G_{\ell,r},W)-t_r^\ell\,.\end{aligned}$$

□

Claim 4.4. *Fix $\mathbf{k}\in\mathcal{V}_m$. Then the number of m-tuples $(R_1,\ldots,R_m)\in\binom{[n]}{r}^m$ for which $\mathcal{H}(R_1,\ldots,R_m)$ is isomorphic to $\mathcal{H}_{\mathbf{k}}^{(r)}$ is equal to*

$$A(n,r,\mathbf{k}):=\frac{m!\cdot(n)_{(r-1)m}}{\prod_{\ell=2}^m(2\ell((r-2)!)^\ell)^{k_\ell}\ \cdot\ k_\ell!}. \tag{35}$$

Proof of Claim 4.4. Suppose first that $r\geq 3$. Notice that the number of automorphisms of $\mathcal{C}_\ell^{(r)}$ equals $2\ell\cdot((r-2)!)^\ell$, and therefore the number of automorphisms of $\mathcal{H}_{\mathbf{k}}^{(r)}$ satisfies

$$\mathrm{aut}(\mathcal{H}_{\mathbf{k}}^{(r)})=\prod_{\ell=2}^m(2\ell((r-2)!)^\ell)^{k_\ell}\ \cdot\ k_\ell!\,.$$

As there are $\frac{(n)_{(r-1)m}}{\mathrm{aut}(\mathcal{H}_{\mathbf{k}}^{(r)})}$ copies of $\mathcal{H}_{\mathbf{k}}^{(r)}$ on n vertices and each copy corresponds to $m!$ many m-tuples $(R_1,\ldots,R_m)$, the proof of the case $r\geq 3$ is complete.

The case $r=2$ is similar; the only difference being that the number of automorphisms of $\mathcal{C}_2^{(2)}$ equals 2 and that each copy of $\mathcal{H}_{\mathbf{k}}^{(2)}$ corresponds to $\frac{m!}{2^{k_2}}$ many m-tuples. The details are left to the reader. □

We now resume expressing $\mathbb{E}\left[\left(X_{n,r}-t_r\binom{n}{r}\right)^m\right]$, which we abandoned at (31). Recall that $G_{\ell,r}$ is the graph associated with $\mathcal{C}_\ell^{(r)}$. Adding (34) and (35), we get

$$\begin{aligned}\mathbb{E}\left[\left(X_{n,r}-t_r\tbinom{n}{r}\right)^m\right] &= \sum_{\mathbf{k}\in\mathcal{V}_m}A(n,r,\mathbf{k})\prod_{\ell=2}^m(t(G_{\ell,r},W)-t_r^\ell)^{k_\ell}+O(n^{(r-1)m-1})\\ &= n^{(r-1)m}m!\sum_{\mathbf{k}\in\mathcal{V}_m}\prod_{\ell=2}^m\left(\frac{t(G_{\ell,r},W)-t_r^\ell}{2\ell((r-2)!)^\ell}\right)^{k_\ell}\frac{1}{k_\ell!}+O(n^{(r-1)m-1})\,. \qquad (36)\end{aligned}$$

For $\ell = 2, 3, \ldots$, let us write

$$d_\ell := \frac{t(G_{\ell,r}, W) - t_r^\ell}{2\ell((r-2)!)^\ell} \,. \tag{37}$$

Treating $g(x) := \sum_{\ell=2}^{\infty} d_\ell x^\ell$ as a formal power series, and substituting $y = g(x)$ in $e^y = \sum_{j=0}^{\infty} y^j/j!$ we obtain another power series (since the free coefficient of g is zero),

$$f(x) := \exp\left(g(x)\right) = \exp\left(\sum_{\ell=2}^{\infty} d_\ell x^\ell\right) \,. \tag{38}$$

The following two claims are needed to show that $f(x)$ is the moment generating function of the limit of $(X_{n,r} - \binom{n}{r} t_r)/n^{r-1}$.

Claim 4.5. *For each $m \in \mathbb{N}$, as $n \to \infty$, the quantity $\frac{1}{m!} \cdot \mathbb{E}[\left(X_{n,r} - t_r\binom{n}{r}\right)^m / (n^{(r-1)m})]$ converges to $[\![x^m]\!] f(x)$, the coefficient of x^m in the (formal) power series $f(x)$.*

Proof. We have

$$\begin{aligned}
[\![x^m]\!] f(x) &= [\![x^m]\!] \left(\exp\left(\sum_{\ell=2}^{\infty} d_\ell x^\ell\right)\right) = [\![x^m]\!] \left(\sum_{j=1}^{\infty} \frac{1}{j!} \cdot \left(\sum_{\ell=2}^{\infty} d_\ell x^\ell\right)^j\right) \\
&\boxed{\text{multinomial theorem}} \quad = \sum_{j=1}^{m} \frac{1}{j!} \sum_{\mathbf{k} \in \mathcal{V}_m : k_2 + \ldots + k_m = j} \binom{j}{k_2, \cdots, k_m} \cdot \prod_{\ell=2}^{m} d_\ell^{k_\ell} \\
&\boxed{\text{definition of the multinomial coefficient}} \quad = \sum_{j=1}^{m} \sum_{\mathbf{k} \in \mathcal{V}_m : k_2 + \ldots + k_m = j} \prod_{\ell=2}^{m} \frac{d_\ell^{k_\ell}}{k_\ell!} \\
&= \sum_{\mathbf{k} \in \mathcal{V}_m} \prod_{\ell=2}^{m} d_\ell^{k_\ell} \frac{1}{k_\ell!} \\
&\boxed{\text{by (36)}} \quad = \lim_{n \to \infty} \frac{1}{m!} \cdot \frac{\mathbb{E}\left[\left(X_{n,r} - t_r\binom{n}{r}\right)^m\right]}{n^{(r-1)m}} \,.
\end{aligned}$$

□

Since $|d_i| \leq ((r-2)!)^{-i}$, in particular, the sequence $|d_i|, i \geq 2$ is bounded. Therefore the series $\sum_i d_i x^i$ has positive radius of convergence, and $f(x)$ can be expanded as its Taylor series around zero. In the next claim, we show that the function f equals the moment-generating function M_Y defined in (23).

Claim 4.6. *In some neighbourhood of zero we have $M_Y(x) = f(x)$.*

Proof of Claim 4.6. Recall the definition (37) of d_ℓ. For $\ell = 2$ we have

$$\begin{aligned} 4((r-2)!)^2 \cdot d_2 &= t(K_r \ominus_2 K_r, W) - t_r^2 \\ &= t(K_r \ominus_2 K_r, W) - t(C_2, V_W^{(r)}) + t(C_2, V_W^{(r)}) - t_r^2 \\ \boxed{(16)} \quad &= t(K_r \ominus_2 K_r, W) - t(K_r \oplus_2 K_r, W) + t(C_2, V_W^{(r)}) - t_r^2 \\ \boxed{(5)} \quad &= t(K_r \ominus_2 K_r, W) - t(K_r \oplus_2 K_r, W) + \sum_{\lambda \in \mathrm{Spec}^-(V_W^{(r)})} \lambda^2, \end{aligned}$$

which, in view of the definition (10) of $\sigma^2_{r,W}$, implies

$$d_2 = \frac{\sigma^2_{r,W}}{2} + \frac{1}{4} \sum_{\lambda \in \mathrm{Spec}^-(V_W^{(r)})} \left(\frac{\lambda}{(r-2)!} \right)^2. \tag{39}$$

For $\ell \geq 3$, from (15) and (5) it follows

$$\begin{aligned} d_\ell &= \frac{t(G_{\ell,r}, W) - t_r^\ell}{2\ell((r-2)!)^\ell} \\ &= \frac{t(C_\ell, V_W^{(r)}) - t_r^\ell}{2\ell((r-2)!)^\ell} = \frac{1}{2\ell} \sum_{\lambda \in \mathrm{Spec}^-(V_W^{(r)})} \left(\frac{\lambda}{(r-2)!} \right)^\ell. \end{aligned} \tag{40}$$

We are now ready to relate f and M_Y. In view of (4), we have

$$\sum_{\lambda \in \mathrm{Spec}^-(V_W^{(r)})} \lambda^2 < \infty \tag{41}$$

and in particular

$$\sup_{\lambda \in \mathrm{Spec}^-(V_W^{(r)})} |\lambda| < \infty. \tag{42}$$

Substituting (39) and (40) into (38), we obtain

$$\log f(x) = \sum_{\ell=2}^{\infty} d_\ell x^\ell = \frac{\sigma^2_{r,W} x^2}{2} + \sum_{\ell=2}^{\infty} \sum_{\lambda \in \mathrm{Spec}^-(V_W^{(r)})} \frac{1}{2\ell} \left(\frac{\lambda x}{(r-2)!} \right)^\ell. \tag{43}$$

In order to interchange the order of summation in (43), we check a condition that allows applying Fubini's theorem, namely that (43) remains finite, if we replace all summands by their absolute values. By (42), for $|x|$ small enough, the sequence $a_\lambda := \frac{\lambda x}{(r-2)!}$ satisfies $c := \sup_\lambda |a_\lambda| < 1$. Therefore, for each $\ell \geq 2$,

$$\sum_\lambda |a_\lambda|^\ell \leq \sum_\lambda c^{\ell-2} a_\lambda^2 \overset{(41)}{=} O(c^\ell),$$

which implies $\sum_\ell \sum_\lambda |a_\lambda|^\ell/(2\ell) = \sum_\ell O(c^\ell) < \infty$, verifying the desired condition. So, changing the order of summation in (43), we obtain

$$\begin{aligned}\log f(x) = \sum_{\ell=2}^{\infty} d_\ell x^\ell &= \frac{\sigma_{r,W}^2 x^2}{2} + \sum_{\lambda \in \mathrm{Spec}^-(V_W^{(r)})} \sum_{\ell=2}^{\infty} \frac{1}{2\ell} \left(\frac{\lambda x}{(r-2)!}\right)^\ell \\ \boxed{\text{Taylor's series}} \quad &= \frac{\sigma_{r,W}^2 x^2}{2} - \frac{1}{2} \sum_{\lambda \in \mathrm{Spec}^-(V_W^{(r)})} \left(\ln\left(1 - \frac{\lambda x}{(r-2)!}\right) + \frac{\lambda x}{(r-2)!}\right) .\end{aligned}$$

By exponentiating the above expression we easily obtain (23), thus completing the proof. □

We are now finished with the proof of Theorem 1.2(c). Indeed, Claims 4.5 and 4.6 imply that the mth moment of $(X_{n,r} - \binom{n}{r}t_r)/n^{r-1}$ converges to $m![\![x^m]\!]f(x) = m![\![x^m]\!]M_Y(x) = \mathbb{E}[Y^m]$ for every m. As argued at the beginning of Section 4, this implies that $\frac{X_{n,r}-\binom{n}{r}t_r}{n^{r-1}} \xrightarrow{\mathrm{d}} Y$.

5. Concluding remarks

In this paper, we initiated the study of limit theorems for complete subgraph counts in $\mathbb{G}(n, W)$. However, the results in this paper should be considered just first steps, and the area offers several obvious open problems.

- Extend Theorem 1.2 to sparser regimes. Recall that the central limit theorem for the count of K_r in $\mathbb{G}(n,p)$ holds for $p = p(n)$ as small as $p(n) \gg n^{-2/r-1}$, that is, as long as the expected number of K_r's tends to infinity.
- To model a sparse inhomogeneous random graph, fix a scaling factor $p = p(n) \to 0$. Then $\mathbb{G}(n, p \cdot W)$ is a sparse inhomogeneous random graph model. Note that then the assumption that W is bounded from above by 1 can be relaxed somewhat. For example, the giant component of $\mathbb{G}(n, W/n)$ is studied in [6]. So, we suggest to obtain limit theorems for the count of K_r (or other graphs) in $\mathbb{G}(n, p \cdot W)$.
- To strengthen the limit theorem obtained here to a local limit theorem. Even in the case of $\mathbb{G}(n,p)$ this is a very difficult problem which was resolved only recently for cliques [3] and even more recently for general connected graphs [20]. Note that such a local limit theorem would have additional restrictions. For example, if W is a graphon consisting of two constant-1 components of measure $\frac{1}{2}$ each, then $X_{n,2}$ is of the form $\binom{k}{2} + \binom{n-k}{2}$, $k \in \mathbb{N}$, that is, not all integer values can be achieved (including those in the bulk of the distribution).

References

[1] N. Alon, J. H. Spencer, *The Probabilistic Method*, Fourth edition, John Wiley and Sons, 2015.

[2] A.D. Barbour, M. Karoński, A. Ruciński, *A central limit theorem for decomposable random variables with applications to random graphs*, Journal of Combinatorial Theory, Series B **47** (1989) 125–145.

[3] R. Berkowitz, *A local limit theorem for cliques in $G(n,p)$*, (preprint) arXiv:1811.03527.

[4] B.B. Bhattacharya, P. Diaconis, S. Mukherjee, *Universal limit theorems in graph coloring problems with connections to extremal combinatorics*, The Annals of Applied Probability **27** (1) (2017) 337–394.

[5] B.B. Bhattacharya, S. Mukherjee, *Monochromatic subgraphs in randomly colored graphons*, European Journal of Combinatorics, 81, 328-353.

[6] B. Bollobás, S. Janson, O. Riordan, *The phase transition in inhomogeneous random graphs*, Random Structures & Algorithms, **31** (1) (2007) 3–122.

[7] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, K. Vesztergombi, K., *Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing*, Advances in Mathematics, **219** (6) (2008) 1801–1851.

[8] M. Doležal, J. Hladký, A. Máthé, *Cliques in dense inhomogeneous random graphs*, Random Structures & Algorithms, **51** (2) (2017) 275–314.

[9] V. Féray, P.-L. Méliot, A. Nikeghbali, *Graphons, permutons and the Thoma simplex: three mod-Gaussian moduli spaces*, Proc. Lond. Math. Soc. **121** (2020), no. 4, 876-926.

[10] N. Fraiman, D. Mitsche, *The diameter of inhomogeneous random graphs*, Random Structures & Algorithms, **53** (2) (2018) 308–326.

[11] S. Janson, T. Łuczak, A. Ruciński, *Random Graphs*, John Wiley and Sons, 2011.

[12] S. Janson, K. Nowicki, *The asymptotic distributions of generalized U-statistics with applications to random graphs*, Probability theory and related fields, **90** (3) (1991), 341–375.

[13] G. Kaur, A. Röllin. *Higher-order fluctuations in dense random graph models*, (preprint) arXiv:2006.15805.

[14] L. Lovász, *Large networks and graph limits*, Vol. 60 of American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI, 2012.

[15] L. Lovász, B. Szegedy, *Limits of dense graph sequences*, Journal of Combinatorial Theory, Series B, **96** (6) (2006), 933–957.

[16] K. Nowicki, J.C. Wierman, *Subgraph Counts in Random Graphs Using Incomplete U-statistics methods*, Annals of Discrete Mathematics, **38** (1988) 299–310.

[17] C. Reiher, M. Schacht, *Forcing quasirandomness with triangles*, Forum of Mathematics, Sigma **7** (2019) E9. doi:10.1017/fms.2019.7

[18] N. Ross, *Fundamentals of Stein's method*, Probability Surveys **8** (2011), 210–293.

[19] A. Ruciński, *When Are Small Subgraphs of a Random Graph Normally Distributed?*, Probability Theory and Related Fields **78** (1988) 1–10.

[20] A. Sah, M. Sawhney, *Local limit theorems for subgraph counts*, (preprint) arXiv:2006.11369.

Institute of Mathematics of the Czech Academy of Sciences, Žitná 25, 115 67, Praha 1, Czechia. Research supported by the Alexander von Humboldt Foundation, GAČR project GJ18-01472Y and RVO: 67985840.

Institute of Mathematics of the Czech Academy of Sciences, Žitná 25, 115 67, Praha 1, Czechia. Research supported by GAČR project GJ18-01472Y and RVO: 67985840

Institute of Computer Science of the Czech Academy of Sciences, Pod Vodárenskou věží 2, 182 07 Prague, Czechia. Research partially supported by the Czech Science Foundation, grants GJ16-07822Y, GA19-08740S and GJ20-27757Y. With institutional support RVO:67985807

CLIQUES IN DENSE INHOMOGENEOUS RANDOM GRAPHS

MARTIN DOLEŽAL, JAN HLADKÝ, AND ANDRÁS MÁTHÉ

Abstract. The theory of dense graph limits comes with a natural sampling process which yields an inhomogeneous variant $\mathbb{G}(n, W)$ of the Erdős–Rényi random graph. Here we study the clique number of these random graphs. We establish the concentration of the clique number of $\mathbb{G}(n, W)$ for each fixed n, and give examples of graphons for which $\mathbb{G}(n, W)$ exhibits wild long-term behavior. Our main result is an asymptotic formula which gives the almost sure clique number of these random graphs. We obtain a similar result for the bipartite version of the problem. We also make an observation that might be of independent interest: Every graphon avoiding a fixed graph is countably-partite.

1. Introduction

The Erdős–Rényi random graph $\mathbb{G}(n, p)$ is a random graph with vertex set $[n] = \{1, \dots, n\}$, where each edge is included independently with probability p. Since Gilbert, and independently Erdős and Rényi introduced the model in 1959, this has been arguably the most studied random discrete structure. Here, we recall facts about cliques in $\mathbb{G}(n, p)$; these were among the first properties studied in the model. The key question in the area concerns the order of the largest clique. We write $\omega(G)$ for the order of the largest clique in a graph G. Matula [21], and independently Grimmett and McDiarmid [11] have shown that when $p \in (0, 1)$ is fixed, and $\varepsilon > 0$ is arbitrary, we have

$$\lim_{n\to\infty} \mathbf{P}\left[\frac{\omega(\mathbb{G}(n,p))}{\log n} = (1 \pm \varepsilon)\frac{2}{\log(1/p)}\right] = 1\,. \tag{1}$$

Here, as well as in the rest of the paper, we use log for the natural logarithm. The actual result is stronger in two directions: firstly, it extends also to sequences of probabilities $p_n \to 0$, and secondly, Matula, Grimmett and McDiarmid proved an asymptotic concentration of $\omega(\mathbb{G}(n, p))$ on two consecutive values for which they provided an explicit formula.

Our aim however was extending (1) in a different direction. That direction is motivated by the theory of limits of dense graph sequences. Let us introduce the key concepts of the theory; for a thorough treatise we refer to Lovász's book [18]. The key object of that

2010 *Mathematics Subject Classification.* 05C80.

Key words and phrases. random graphs, graph limits, clique number.

MD: Supported by GA ČR Grant 16-07378S and RVO: 67985840.

JH: The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement umber 628974.

AM was supported by a Leverhulme Trust Early Career Fellowship and by the Hungarian National Research, Development and Innovation Office – NKFIH, 104178.

theory are graphons. A *graphon* is a symmetric measurable function $W : \Omega^2 \to [0,1]$, where Ω is a probability space. In their foundational work, Lovász and Szegedy [19] proved that each sequence of finite graphs contains a subsequence that converges — in the so-called *cut metric* — to a graphon. This itself does not justify the graphons as limit objects; it still could be that the space of graphons is unnecessarily big. In other words, one would like to know that every graphon W is attained as a limit of finite graphs. To this end, Lovász and Szegedy introduced the random graph model $\mathbb{G}(n, W)$. The set of vertices of $G \sim \mathbb{G}(n, W)$ is the set $[n]$. To sample G, first generate n independent elements $x_1, \ldots, x_n \in \Omega$ according to the law of Ω. Then, connect i and j by an edge with probability $W(x_i, x_j)$ (independently of other choices). Lovász and Szegedy proved that with probability one, the sequence of samples from $\mathbb{G}(n, W)$ converges to W.

The strength of the theory of graph limits is that convergence in the cut metric implies convergence of many key statistics of graphs (or graphons). These include frequencies of appearances of small subgraphs, and normalized MAX-CUT-like properties. An important direction of research is to understand which other parameters are continuous with respect to the cut metric; those parameters can then be defined even on graphons. In Example 1.1 below we show that the clique number can be very discontinuous with respect to the cut metric. Not all is lost even in such a situation. One can still study discontinuous parameters of a graphon W via the sampling procedure $\mathbb{G}(n, W)$. The random sampling will suppress pathological counterexamples such as those in Example 1.1, and thus allow us to associate limit information even to these limit parameters.

Example 1.1. Let $f : \mathbb{N} \to \mathbb{N}$ be a function that tends to infinity very slowly.

- Let us consider a sequence of n-vertex graphs consisting of a clique of order $\lceil n/f(n) \rceil$ and $n - \lceil n/f(n) \rceil$ isolated vertices. This sequence converges to the 0-graphon, the smallest graphon in the graphon space. Yet, the clique numbers are unusually high, almost of order $\Theta(n)$.
- Let us consider a sequence of n-vertex $f(n)$-partite Turán graphs. This sequence converges to the 1-graphon, the largest graphon in the graphon space. Yet, the clique numbers tend to infinity very slowly. □

By the previous example, we see that

- there are sequences of finite graphs with clique numbers growing much faster than logarithmic with a limit graphon $W \equiv 0$,
- while there are other sequences of finite graphs with clique numbers growing much slower than logarithmic with a limit graphon $W \equiv 1$.

So, suppose that we have suppressed these pathological examples by looking at "typical graphs that are close to W" rather than "all graphs that are close to W", and let us see what value motivated by the clique number can be associated to W. To this end, suppose

that $W\colon \Omega^2 \to [0,1]$ is such a graphon that $W(x,y) \in [p_1, p_2]$ for every $x, y \in \Omega$, where $0 < p_1 \leqslant p_2 < 1$ are fixed. Then the edges of $\mathbb{G}(n,W)$ are stochastically between $\mathbb{G}(n,p_1)$ and $\mathbb{G}(n,p_2)$. Thus, (1) tells us that the clique number $\omega(\mathbb{G}(n,W))$ asymptotically almost surely satisfies

$$(1-o(1))\frac{2}{\log(1/p_1)} \leqslant \frac{\omega(\mathbb{G}(n,W))}{\log n} \leqslant (1+o(1))\frac{2}{\log(1/p_2)}\,. \tag{2}$$

Thus, it is actually plausible to believe that

$$\frac{\omega(\mathbb{G}(n,W))}{\log n} \tag{3}$$

converges in probability. In this paper, we study this and related questions.

1.1. **Related literature on inhomogenous random graphs.** Inhomogeneous random graphs allow one to express different intensities of bonds between the corresponding parts of the base space. This is obviously useful in modeling phenomena in biology, sociology, computer science, physics, and other settings. The price one has to pay for this flexibility of these models is in extra difficulties in mathematical analysis of their properties. This is one of the reasons why literature on $\mathbb{G}(n,W)$ is fairly scarce, compared to $\mathbb{G}(n,p)$; another reason apparently being that the inhomogeneous model is much more recent. Actually, in the inhomogeneous model, most work was done in the *sparse regime*, which we shall introduce now. To get a counterpart to sparse random graphs $\mathbb{G}(n,p_n)$, $p_n \to 0$, one introduces rescaling $\mathbb{G}(n,p_n \cdot W)$. In this setting, W need not be bounded from above anymore (even though the question of "how unbounded" W can be is rather subtle and we neglect it here).

The most impressive example of work concerning sparse inhomogeneous random graphs is [2] in which the existence and the size of the giant component in $\mathbb{G}(n, \frac{1}{n} \cdot W)$ was determined. This work has initiated a big amount of further work on $\mathbb{G}(n, \frac{1}{n} \cdot W)$, such as [26, 27], as well as on related percolation models [1].

The threshold for connectivity of inhomogeneous random graphs was investigated in [6]. The diameter of inhomogeneous random graphs was studied in [9].

A particular subclass of the random graph models $\mathbb{G}(n,W)$ are the so-called *stochastic block models* introduced already in 1980's in the field of mathematical sociology [13]. They are used extensively in many areas of mathematics, computer science, and physics. In our language, (the dense version of) stochastic block models correspond to the case when W is a step-function with finitely many or countably many steps. The stochastic block model is mathematically much more tractable. For example, the study of criticality in stochastic block models in [14] seems to be much more tractable than in the case of general inhomogeneous random graphs.

2. Our contribution

In this section we present our main results. The notation in this section is standard. We refer the reader to Section 3 for formal definitions.

We saw that for many natural graphons W, $\omega(\mathbb{G}(n,W))$ grows logarithmically. It is easy to construct a graphon for which $\omega(\mathbb{G}(n,W))$ grows for example as $\log\log n$, or another graphon for which $\omega(\mathbb{G}(n,W))$ grows for example as $n^{0.99}$. More surprisingly, our next proposition shows that we can have an oscillation between these two regimes even for one graphon.

Proposition 2.1. *For an arbitrary function $f:\mathbb{N}\to\mathbb{R}_+$ with $\lim_{n\to\infty} f(n)=+\infty$ there exists a graphon W and a sequence of integers $1=\ell_0<k_1<\ell_1<k_2<\ell_2<\dots$ such that asymptotically almost surely,*

$$\omega(\mathbb{G}(k_i,W)) < f(k_i)\ , \text{ and} \tag{4}$$

$$\omega(\mathbb{G}(\ell_i,W)) > \frac{\ell_i}{f(\ell_i)}\ . \tag{5}$$

While Proposition 2.1 shows that the long-term behavior of $\omega(\mathbb{G}(n,W))$ can be quite wild, for a fixed (but large) n, the distribution of $\omega(\mathbb{G}(n,W))$ is concentrated.

Theorem 2.2. *For each graphon W and each n, we have that for each $\varepsilon>0$,*

$$\lim_{n\to\infty} \mathbf{P}\left[\left|\frac{\omega(\mathbb{G}(n,W))}{\mathbf{E}[\omega(\mathbb{G}(n,W))]}-1\right|>\varepsilon\right]=0\ .$$

The proofs of Proposition 2.1 and Theorem 2.2 are given in Section 4. In the proof of Theorem 2.2, we need to consider the case of graphons W for which $\mathbf{E}[\omega(\mathbb{G}(n,W))]$ is bounded (as $n\to\infty$) separately. Investigation of such graphons led to a result that is of independent interest. Let us give the details. Clearly, for each graphon W, and each $n\in\mathbb{N}$, $\omega(\mathbb{G}(n,W))$ is stochastically dominated by $\omega(\mathbb{G}(n+1,W))$. As a consequence, the sequence

$$\mathbf{E}[\omega(\mathbb{G}(1,W))],\mathbf{E}[\omega(\mathbb{G}(2,W))],\mathbf{E}[\omega(\mathbb{G}(3,W))],\dots$$

is nondecreasing. We say that W *has a bounded clique number* if $\lim_{n\to\infty}\mathbf{E}[\omega(\mathbb{G}(n,W))]<+\infty$. Note that one example of graphons of bounded clique numbers are graphons W which have zero homomorphism density of H (see (11) for the definition) for some finite graph H. A subclass of these are *k-partite graphons*. These are graphons $W:\Omega^2\to[0,1]$ for which there exists a measurable partition $\Omega=\Omega_1\dot\cup\Omega_2\dot\cup\dots\dot\cup\Omega_k$ such that for each $i\in[k]$, $W\restriction_{\Omega_i\times\Omega_i}=0$ almost everywhere. In the following example, we show that the structure of graphons with a bounded clique number can be more complicated. We consider a sequence of triangle-free graphs $G_1,G_2,\dots$ whose chromatic numbers tend to infinity (it is a standard exercise that such graphs indeed exist). Let $W_1,W_2,\dots$ be their graphon

representations. We now glue these graphons into one graphon W. Clearly, $\omega(\mathbb{G}(n,W)) \leqslant 2$ with probability one, but W is not k-partite for any k. Here, we show that the structure of graphons with a bounded clique number cannot be much more complicated than in the example above. We call a graphon $W : \Omega^2 \to [0,1]$ *countably-partite*, if there exists a measurable partition $\Omega = \Omega_1 \dot\cup \Omega_2 \dot\cup \dots$ such that for each $i \in \mathbb{N}$, $W \restriction_{\Omega_i \times \Omega_i} = 0$ almost everywhere.

Theorem 2.3. *Every graphon with a bounded clique number is countably-partite.*

The proof of Theorem 2.3 is given in Section 4.3.

Let us turn our attention to the main subject of the paper, that is, to the behavior of the clique number in $\mathbb{G}(n,W)$ scaled as in (3). As a warm-up for studying (3), we first deal with its bipartite counterpart. To this end, we shall work with *bigraphons*. Bigraphons, introduced first in [20], arise as limits of balanced bipartite graphs. A bigraphon is a measurable function $U : \Omega_1 \times \Omega_2 \to [0,1]$. Here, Ω_1 and Ω_2 are probability spaces which represent the two partition classes, and the value $U(x,y)$ represents the edge intensity between the parts corresponding to x and y. This suggests the sampling procedure for generating the inhomogeneous random bipartite graph $\mathbb{B}(n,U)$: we sample uniformly and independently at random points $x_1,\dots,x_n$ from Ω_1 and $y_1,\dots,y_n$ from Ω_2. In the bipartite graph $\mathbb{B}(n,U)$ with colour classes $\{a_i\}_{i=1}^n$ and $\{b_j\}_{j=1}^n$, we connect a_i with b_j with probability $U(x_i,y_j)$. We define the natural bipartite counterpart to the clique number. Given a bipartite graph $G = (A,B;E)$, we define its *biclique number* as the largest ℓ such that there exist sets $X \subseteq A$, $Y \subseteq B$, $|X| = |Y| = \ell$, that induce a complete bipartite graph. We denote the biclique number of G by $\omega_2(G)$.

The main result concerning the biclique number is the following.

Theorem 2.4. *Let $U : \Omega_1 \times \Omega_2 \to [0,1]$ be a bigraphon whose essential supremum $p = \operatorname{ess\,sup} U$ is strictly between zero and one. Then we asymptotically almost surely have*

$$\omega_2(\mathbb{B}(n,U)) = (1 \pm o(1)) \cdot \frac{2}{\log 1/p} \cdot \log n \,.$$

The proof of Theorem 2.4 is given in Section 5.

We turn to our main result which determines the quantity (3). Suppose that W is a graphon with strictly positive essential infimum. Define

(6)
$$\kappa(W) = \sup\left\{ \frac{2\|h\|_1^2}{\int_{(x,y)\in\Omega^2} h(x)h(y)\log\left(1/W(x,y)\right)\mathsf{d}(\nu^2)} \;:\; h \text{ is a nonnegative } L^1\text{-function on } \Omega \right\}.$$

Here, we set $\frac{0}{0} = 0$ and $\frac{a}{0} = +\infty$ for $a \in \mathbb{R} \setminus \{0\}$.

We can now state our main result.

Theorem 2.5. *Suppose that W is a graphon whose essential infimum is strictly positive. Then*

- *if $\kappa(W) < +\infty$ then a.a.s. $\omega(\mathbb{G}(n, W)) = (1 + o(1)) \cdot \kappa(W) \cdot \log n$, and*
- *if $\kappa(W) = +\infty$ then a.a.s. $\omega(\mathbb{G}(n, W)) \gg \log n$.*

Theorem 2.5 is consistent with (1). Indeed, suppose that $W \equiv p \in (0, 1)$ is a constant graphon. Then for any h in (6) (which is not constant zero), we have that

$$\frac{2\|h\|_1^2}{\int_{(x,y)\in\Omega^2} h(x)h(y)\log\left(1/W(x,y)\right)\mathsf{d}(\nu^2)} = \frac{2}{\log 1/p}\,.$$

We provide heuristics for Theorem 2.5 for more complicated graphons in Section 6. Unfortunately, we were unable to turn these relatively natural heuristics into a rigorous proof. The actual proof of Theorem 2.5 is given in Section 8, building on tools from Section 7.

There are several alternative ways of expressing $\kappa(W)$. For example, when we heuristically derive Theorem 2.5, we make use of the following identity.

Fact 2.6. *We have*

$$\kappa(W) = \sup\left\{\|f\|_1 \,:\, f \text{ is a nonnegative } L^1\text{-function on } \Omega,\ \Gamma(f, W) \geqslant 0\right\}\,, \tag{7}$$

where

$$\Gamma(f, W) = \underbrace{\int_{x\in\Omega} f(x)\mathsf{d}(\nu)}_{(*)} + \underbrace{\frac{1}{2}\int_{(x,y)\in\Omega^2} f(x)f(y)\log W(x,y)\mathsf{d}(\nu^2)}_{(**)}\,. \tag{8}$$

Other expressions of $\kappa(W)$ are given in Proposition 3.8.

Remark 2.7. While we formulate all the problems in terms of cliques, we could have worked with the complementary notion of independent sets instead. Indeed, investigating one of these notions with respect to $\mathbb{G}(n, W)$ is equivalent to investigating the other with respect to $\mathbb{G}(n, 1 - W)$.

Using this observation, we get from Theorem 2.5 the following corollary for the size of the maximum independent set $\alpha(\mathbb{G}(n, W))$.

Corollary 2.8. *Suppose that W is a graphon whose essential supremum is strictly less than 1. Then*

- *if $\kappa(1 - W) < +\infty$ then a.a.s. $\alpha(\mathbb{G}(n, W)) = (1 + o(1)) \cdot \kappa(1 - W) \cdot \log n$, and*
- *if $\kappa(1 - W) = +\infty$ then a.a.s. $\alpha(\mathbb{G}(n, W)) \gg \log n$.*

In our language, (the dense version of) stochastic block models correspond to the case when W is a step-function with finitely many or countably many steps. There exists a conceptually simpler proof of Theorem 2.5 when restricted to stochastic block models.

This simplification occurs both on the real-analytic side (i.e., general measurable functions versus step-functions) and on the combinatorial side. See our remark in Section 6.3.

3. Preliminaries

3.1. **Notation.** For $n \in \mathbb{N}$, we write $[n] = \{1, \dots, n\}$, and $[n]_0 = \{0, 1, \dots, n\}$. As always, we denote by $\binom{n}{k}$ the binomial coefficient $\frac{n!}{k!(n-k)!}$. For the multinomial coefficients of higher orders, we employ the notation $\binom{n}{k_1|k_2|\dots|k_l} = \frac{n!}{k_1!k_2!\dots k_l!(n-\sum k_i)!}$ (here we suppose $k_1 + k_2 + \dots + k_l \leqslant n$). We omit rounding symbols where it does not affect correctness of the calculations.

We shall always assume that Ω is a standard Borel probability space without atoms. We always write ν for the probability measure associated with Ω.

We shall sometimes make use of tools from real analysis which are available only for $\mathbb{R}$ and $\mathbb{R}^d$. The Lebesgue measure will be denoted by λ. It should be always clear from the context whether we mean the one-dimensional Lebesgue measure on $\mathbb{R}$, two dimensional Lebesgue measure on $\mathbb{R}^2$ or any higher dimensional Lebesgue measure.

We write $\|\cdot\|_1$ to denote the L^1-norm of functions (or vectors in a finite-dimensional space). Non-negative vectors, and non-negative L^1-functions[1] are called *histograms* (see our explanation at the end of Section 6.1). For a histogram f, we write $\mathrm{Box}(f)$ for all histograms g for which $g \leqslant f$ (pointwise). We say that a histogram is *non-trivial* if it is not almost everywhere zero.

We recall the notions of *essential supremum* and *essential infimum*. Suppose that Ω is a space equipped with a measure ν. For a measurable function $f\colon X \to \mathbb{R}$ we define $\operatorname{ess\,sup} f$ as the least number a such that $\nu(\{x \in \Omega : f(x) > a\}) = 0$. The quantity $\operatorname{ess\,inf} f$ is defined analogously.

3.2. **Random graphs $\mathbb{H}(n, W)$.** There is a natural intermediate step when obtaining the random graph $\mathbb{G}(n, W)$ from a graphon W which is often denoted by $\mathbb{H}(n, W)$. To obtain $\mathbb{H}(n, W)$ we sample n random independent points $x_1, \dots, x_n$ from the probability space underlying W. The random graph has the vertex set $[n]$. The edge-set is an edge-set of a complete graph equipped with edge-weights. The weight of the edge ij is $W(x_i, x_j)$. Self-loops are not included.

3.3. **Graphons.** The above crash course in graph limits almost suffices for the purposes of this paper, and we need only a handful of additional standard definitions. See [18] for further references.

All (non-discrete) probability spaces in this paper are standard Borel probability spaces without atoms. Recall that the Isomorphism Theorem (see e.g. [15, Theorem 17.41]) tells

[1] the vector-/function-space will be clear from the context

us that there is a measure-preserving isomorphism between each two such spaces (i.e. a bijection between the spaces such that this function and its inverse are measurable and preserve measures). In particular, suppose that $W : \Omega^2 \to [0,1]$ is a graphon defined on a probability space Ω, and let X be another probability space. Let us fix an isomorphism $\psi : X \to \Omega$ between X and Ω. By a *representation of W on X* we mean the graphon $W' : X^2 \to [0,1]$, $(x,y) \mapsto W(\psi(x),\psi(y))$. Of course, the representation depends on the actual choice of the isomorphism ψ. Note however that the distribution of $\mathbb{G}(n,W')$ does not depend on the choice of ψ as it is the same as the distribution of $\mathbb{G}(n,W)$. Note also that from the Isomorphism Theorem above we get the following fact.

Fact 3.1. *Each graphon can be represented on the open unit interval* $(0,1)$.

We shall need to "zoom in" on a certain part of a graphon. The next definition is used to this end.

Definition 3.2. Suppose that $W : \Omega^2 \to [0,1]$ is a graphon on a probability space Ω with a measure ν. By a *subgraphon of W obtained by restricting to a set $A \subseteq \Omega$* of positive measure we mean a function $U : A^2 \to [0,1]$ which is simply the restriction $W \restriction_{A\times A}$. When working with this notion, we need to turn A to a probability space. That is, we view U as a graphon on the probability space A endowed with measure $\nu_A(B) := \frac{\nu(B)}{\nu(A)}$ for every measurable set $B \subseteq A$.

Observe that in the above setting for every $B \subseteq A$ of positive measure we have

$$\frac{1}{\nu(B)^2}\int_{(x,y)\in B\times B} \log(1/W(x,y))\mathsf{d}(\nu^2) = \frac{1}{\nu_A(B)^2}\int_{(x,y)\in B\times B} \log(1/W(x,y))\mathsf{d}(\nu_A^2)\ . \tag{9}$$

Note that a lower bound on $\omega(\mathbb{G}(n,W\restriction_{A\times A}))$ provides readily a lower bound on $\omega(\mathbb{G}(n,W))$. More precisely, suppose that we can show that asymptotically almost surely, $\omega(\mathbb{G}(n,W\restriction_{A\times A})) \geqslant c\log n$. Consider now sampling the random graph $\mathbb{G}(n,W)$. By the Law of Large Numbers, out of the n sampled points $x_1,\ldots,x_n \in \Omega$, there will be $(\nu(A)-o(1))n > \frac{1}{2}\nu(A)n$ many of them contained in A. In other words, there is a coupling of $G = \mathbb{G}(n,W)$ and $G' = \mathbb{G}(\frac{1}{2}\nu(A)n, W\restriction_{A\times A})$ such that with high probability, G' is contained in G as a subgraph. We conclude that

$$\omega(\mathbb{G}(n,W)) \overset{\text{a.a.s.}}{\geqslant} \omega\left(\mathbb{G}(\tfrac{1}{2}\nu(A)n, W\restriction_{A\times A})\right) \overset{\text{a.a.s.}}{\geqslant} c\log\left(\tfrac{1}{2}\nu(A)n\right) = (c-o(1))\log n\ . \tag{10}$$

The *homomorphism density* of a graph $H = (\{v_1,\ldots,v_\ell\},E)$ in a graphon W is defined by

$$t(H,W) = \int_{(x_1,\ldots,x_\ell)\in\Omega^\ell} \prod_{i<j:v_iv_j\in E} W(x_i,x_j)\ \mathsf{d}(\nu^\ell)\ . \tag{11}$$

Suppose that Ω is an atomless standard Borel probability space. Let $W_1, W_2 : \Omega^2 \to [0,1]$ be two graphons. We then define the *cut-norm distance* of W_1 and W_2 by

$$d_\square(W_1, W_2) = \sup_{S,T} \left| \int_{(x,y)\in S\times T} (W_1(x,y) - W_2(x,y)) \ \mathsf{d}(\nu^2) \right|,$$

where S and T range over all measurable subsets of Ω. Strictly speaking, $d_\square$ is only a pseudometric since two graphons differing on a set of measure zero have zero distance. Based on the cut-norm distance we can define the key notion of *cut distance* by

$$\delta_\square(W_1, W_2) = \inf_{\varphi} d_\square(W_1^\varphi, W_2)\ , \tag{12}$$

where $\varphi : \Omega \to \Omega$ ranges through all measure preserving automorphisms of Ω, and W_1^φ stands for a graphon defined by $W_1^\varphi(x,y) = W_1(\varphi(x), \varphi(y))$. Then $\delta_\square$ is also a pseudometric.

Suppose that $H = (\{v_1, \dots, v_\ell\}, E)$ is a graph (which is allowed to have self-loops), and let Ω be an arbitrary atomless standard Borel probability space. By a *graphon representation W_H of H* we mean the following construction. We consider an arbitrary partition $\Omega = A_1 \dot\cup A_2 \dot\cup \dots \dot\cup A_\ell$ into sets of measure $\frac{1}{\ell}$ each. We then define the graphon W_H as 1 or 0 on each square $A_i \times A_j$, depending on whether $v_i v_j$ forms an edge or not. Note that W_H is not unique since it depends on the choice of the sets $A_1, \dots, A_\ell$. However, all the possible graphons W_H are at zero distance in the $\delta_\square$-pseudometric. So, when writing W_H we refer to any representative of the above class. With this in mind, we can also define the cut distance of H and any graphon $W\colon \Omega^2 \to [0,1]$, denoted by $\delta_\square(H, W)$, as $\delta_\square(W_H, W)$. Also, all of this extends in a straightforward way to weighted graphs with a weight function $w : E \to [0,1]$.

Remark 3.3. Suppose that H is a finite graph and I is the unit interval (open or closed). Then in the above construction we can take the sets $A_i \subseteq I$ to be intervals.

The key property of the sampling procedures described earlier (both $\mathbb{G}(n, W)$ and $\mathbb{H}(n, W)$) is that the samples are typically very close to the original graphon W in the cut distance. Let us state this fact (for the sampling procedure $\mathbb{H}(n, W)$), proven first in [4], formally.

Lemma 3.4 ([18, Lemma 10.16])**.** *Let W be an arbitrary graphon. Then for each $k \in \mathbb{N}$ with probability at least $1 - \exp(-\frac{k}{2\log k})$, we have $\delta_\square\big(\mathbb{H}(k, W), W\big) \leqslant \frac{20}{\sqrt{\log k}}$.* □

In some situations, we shall consider a wider class of kernels, the so-called L^∞-graphons. These are just symmetric L^∞-functions $W : \Omega^2 \to \mathbb{R}_+$. That is, we do not require L^∞-graphons to be bounded by 1 but rather by an arbitrary constant. The random graph

$\mathbb{H}(n,W)$ makes sense even for L^∞-graphons.[2] By simple rescaling, we get the following from Lemma 3.4.

Corollary 3.5. *Let W an arbitrary L^∞-graphon. Then for each $k \in \mathbb{N}$ with probability at least $1-\exp(-\frac{k}{2\log k})$, we have $\delta_\square\big(\mathbb{H}(k,W),W\big) \leqslant \frac{20\|W\|_\infty}{\sqrt{\log k}}$.*

Let us note that the proof of Lemma 3.4 is quite involved.

3.4. **Lebesgue points and approximate continuity.** Let $f\colon (0,1)^2 \to \mathbb{R}$ be an integrable function. Recall that $(x,y) \in (0,1)^2$ is a called a *Lebesgue point* of f if we have

$$\lim_{r\to 0+} \frac{1}{\lambda(M_r(x,y))} \int_{(s,t)\in M_r} |f(s,t)-f(x,y)|\mathsf{d}(\lambda^2) = 0, \tag{13}$$

where $M_r(x,y) = [x-r,x+r]\times[y-r,y+r]$. Recall that $(x,y) \in \mathbb{R}^2$ is a *point of density* of a set $A \subseteq \mathbb{R}^2$ if

$$\lim_{r\to 0+} \frac{\lambda(M_r(x,y)\setminus A)}{\lambda(M_r(x,y))} = 0. \tag{14}$$

Recall also that a function $f\colon (0,1)^2 \to \mathbb{R}$ is said to be *approximately continuous* at $(x,y) \in (0,1)^2$ if for every $\varepsilon > 0$, the point (x,y) is a point of density of the set $\{(s,t) \in (0,1)^2\colon |f(s,t)-f(x,y)| < \varepsilon\}$.

We will use the following classical result (see e.g. [23, Theorem 7.7]).

Theorem 3.6. *Let $f\colon (0,1)^2 \to \mathbb{R}$ be an integrable function. Then almost every point of $(0,1)^2$ is a Lebesgue point of f. In particular, we have*

$$\lim_{r\to 0+} \frac{1}{\lambda(M_r(x,y))} \int_{(s,t)\in M_r} f(s,t)\mathsf{d}(\lambda^2) = f(x,y)$$

for almost every $(x,y) \in (0,1)^2$. □

An easy consequence of the previous theorem is also the following classical result.

Theorem 3.7. *Let $f\colon (0,1)^2 \to \mathbb{R}$ be a measurable function. Then f is approximately continuous at almost every point of $(0,1)^2$.* □

3.5. **Alternative formulas for $\kappa(W)$.** First, let us prove Fact 2.6 which gives our first identity for $\kappa(W)$.

Proof of Fact 2.6. If there exists a nonzero histogram h such that

$$\int_{(u,v)\in\Omega^2} h(u)h(v)\log W(u,v)\mathsf{d}(\nu^2) = 0$$

then clearly both suprema in (6) and (7) are $+\infty$.

[2] But $\mathbb{G}(n,W)$ does not make sense.

Let h be an arbitrary histogram in (6) and suppose that

$$\int_{(u,v)\in\Omega^2} h(u)h(v)\log\left(1/W(u,v)\right)\mathsf{d}(\nu^2) > 0\,.$$

We take

$$f = \frac{2\|h\|_1\cdot h}{\int_{(u,v)\in\Omega^2} h(u)h(v)\log\left(1/W(u,v)\right)\ \mathsf{d}(\nu^2)}\,.$$

This way, the function f is a histogram on Ω, and $\|f\|_1$ equals to the term in the supremum in (6). So, to justify that the right-hand side of (7) is at least as big as that of (6), we only need to show that $\Gamma(f,W)\geqslant 0$. Indeed,

$$\begin{aligned}\Gamma(f,W) &= \tfrac{2\|h\|_1}{\int_{(u,v)\in\Omega^2} h(u)h(v)\log(1/W(u,v))\mathsf{d}(\nu^2)}\cdot\int_{x\in\Omega} h(x)\mathsf{d}(\nu)\\ &\quad+\frac{1}{2}\cdot\left(\tfrac{2\|h\|_1}{\int_{(u,v)\in\Omega^2} h(u)h(v)\log(1/W(u,v))\mathsf{d}(\nu^2)}\right)^2\cdot\int_{(x,y)\in\Omega^2} h(x)h(y)\log W(x,y)\mathsf{d}(\nu^2)\\ &= 0\,.\end{aligned}$$

On the other hand, let f be an arbitrary histogram appearing in (7) such that

$$\int_{(u,v)\in\Omega^2} f(u)f(v)\log\left(1/W(u,v)\right)\mathsf{d}(\nu^2) > 0\,.$$

For $c\in\mathbb{R}$, let us denote by cf the c-multiple of the function f. Then the map $c\mapsto\Gamma(cf,W)$ is clearly a quadratic function with the limit $-\infty$ at $+\infty$. And since $\Gamma(f,W)\geqslant 0$, there is $c_0\geqslant 1$ such that $\Gamma(c_0f,W)=0$. Now if we define $h=c_0f$ then the corresponding term in (6) equals to $\|h\|_1=c_0\|f\|_1\geqslant\|f\|_1$. □

In Fact 2.6 we gave an alternative formula for $\kappa(W)$. In Proposition 3.8 we give two further expressions. These expressions require the graphon W to be changed on a nullset. Note that we have the liberty of making such a change as the distribution of the model $\mathbb{G}(n,W)$ remains unaltered.

Proposition 3.8. *Suppose that $W:(0,1)^2\to[0,1]$ is an arbitrary graphon. Then W can be changed on a nullset in such a way that the following holds:*

We have

(15)

$$\exp\left(-\frac{2}{\kappa(W)}\right) = \lim_{r\to\infty} P_r \quad \textit{where} \quad P_r = \sup_{F\subseteq(0,1),|F|=r}\left(\prod_{x,y\in F,x<y} W(x,y)\right)^{2/(|F|^2-|F|)} \quad \textit{and}$$

$$\frac{2}{\kappa(W)} = \inf_A \frac{1}{\lambda(A)^2}\int_{(x,y)\in A^2}\log\left(1/W(x,y)\right)\mathsf{d}(\lambda^2)\,, \tag{16}$$

where the infimum ranges over all measurable sets $A\subseteq(0,1)$ of positive measure.

Moreover,

$$\kappa(W) = \sup_{a\in\mathbb{R}_+, A\subseteq\Omega} \{a\lambda(A) \,:\, \Gamma(a\cdot\mathbf{1}_A, W) \geqslant 0\} \ . \tag{17}$$

More precisely, for each $\varepsilon > 0$ and each set A from (16) *satisfying*

$$(1+\varepsilon)\frac{2}{\kappa(W)} \geqslant \frac{1}{\lambda(A)^2}\int_{(x,y)\in A^2} \log(1/W(x,y))\mathsf{d}(\lambda^2)$$

for the number $a = (1-\varepsilon)\frac{\kappa(W)}{\lambda(A)}$ we have that $\Gamma(a\cdot\mathbf{1}_A, W) \geqslant 0$.

Remark. We can rewrite (16) as $\frac{2}{\kappa(W)} = \inf_h \frac{1}{\|h\|_1^2}\int_{(x,y)\in\Omega^2} \log\left(1/W(x,y)\right) h(x)h(y)\mathsf{d}(\lambda^2)$, where h ranges over all indicator functions on $(0,1)$. Our proof of Proposition 3.8 could be easily modified to show that in the infimum, we can range over all histograms h instead. Since the Radon–Nikodym theorem gives us a one-to-one correspondence between non-negative L^1-functions and absolutely continuous measures, we get that

$$\frac{2}{\kappa(W)} = \inf_{\pi} \int_{(x,y)\in(0,1)^2} \left(\log 1/W(x,y)\right) \mathsf{d}(\pi^2) \ ,$$

where the infimum ranges over all probability measures π on $(0,1)$ that are absolutely continuous with respect to the Lebesgue measure. While we shall not need this identity we remark that it can be used to derive a heuristic — slightly different from that given in Section 6 — for Theorem 2.5.

Proof of Proposition 3.8. Let us replace the value of W at every point $(x,y) \in (0,1)^2$ that is not a point of approximate continuity by c. This is a change of measure zero by Theorem 3.7.

Let us deal with the first part of the statement postponing (17) to later.

Claim 3.8.1. *For each $r \in \{2,3,\ldots\}$ we have $\log P_r \geqslant -\frac{2}{\kappa(W)}$.*

Proof of Claim 3.8.1. Let h be an arbitrary function appearing in (6) (not constant zero). Fix $r \in \{2,3,\ldots\}$, and let $F \subseteq (0,1)$ be a random set consisting of r independent points sampled from $(0,1)$ according to the density function $d = h/\|h\|_1$. Then by linearity of expectation we have

$$\begin{aligned}&\mathbf{E}_F\left[\tfrac{2}{r(r-1)}\sum_{x,y\in F, x<y} \log W(x,y)\right]\\ &= \mathbf{E}_{x,y\sim d}\left[\log W(x,y)\right] = \frac{1}{\|h\|_1^2}\int_{(x,y)\in\Omega^2} h(x)h(y)\log W(x,y)\ \mathsf{d}(\lambda^2)\ .\end{aligned}$$

This shows that there exists a deterministic r-element set F for which

$$\frac{2}{r(r-1)}\sum_{x,y\in F, x<y} \log W(x,y) \geqslant \frac{1}{\|h\|_1^2}\int_{(x,y)\in(0,1)^2} h(x)h(y)\log W(x,y)\mathsf{d}(\lambda^2)\ .$$

This concludes the proof of Claim 3.8.1. □

Let us denote the right-hand side of (16) as $-P$.

Claim 3.8.2. *For each $r \in \{2,3,\ldots\}$, we have that $P \geqslant \frac{r-1}{r}\log P_r + \frac{\log c}{r}$.*

Proof of Claim 3.8.2. Suppose that $r \in \{2,3,\ldots\}$ is given, and let $F = \{x_1 < x_2 < \ldots < x_r\}$ be an arbitrary set of points in $(0,1)$ as in (15). Let $\varepsilon \in (0, c/2)$ be arbitrary. Let $C = \log c - \log(c-\varepsilon)$. Firstly, note that the concavity of the logarithm gives that

$$\log(a-\varepsilon) \geqslant \log a - C \tag{18}$$

for each $a \in [c,\infty)$. Secondly, note that $C \searrow 0$ as $\varepsilon \searrow 0$.

Let us take $\delta > 0$ such that for each $i \in [r]$, we have that the sets $S_i = [x_i - \delta, x_i + \delta]$ are pairwise disjoint, and such that the measure of each of the sets $D_{ij} = \{(x,y) \in S_i \times S_j : W(x_i,x_j) - \varepsilon > W(x,y)\}$ is at most $\varepsilon(2\delta)^2$. The latter property can be achieved since each point (x_i,x_j) is either a Lebesgue point of W, or it is a point attaining the infimum of W. Let $A = \bigcup_{i=1}^r S_i$. Then,

$$\begin{aligned}
&\frac{1}{\lambda(A)^2}\int\limits_{(x,y)\in A^2} \log W(x,y)\mathsf{d}(\lambda^2)\\
&\quad= \frac{1}{(2\delta r)^2}\sum_i\sum_j\left(\int\limits_{(x,y)\in(S_i\times S_j)\setminus D_{ij}} \log W(x,y)\ \mathsf{d}(\lambda^2) + \int\limits_{(x,y)\in D_{ij}} \log W(x,y)\ \mathsf{d}(\lambda^2)\right)\\
&\quad\geqslant \frac{1}{(2\delta r)^2}\sum_i\sum_j\left(\int\limits_{(x,y)\in(S_i\times S_j)\setminus D_{ij}} \log(W(x_i,x_j)-\varepsilon)\ \mathsf{d}(\lambda^2) + \int\limits_{(x,y)\in D_{ij}} \log c\ \mathsf{d}(\lambda^2)\right).
\end{aligned}$$

Now, let us use (18) for the first term and the fact that $\lambda(D_{ij}) \leqslant \varepsilon(2\delta)^2$ for the second term. Thus,

$$\begin{aligned}
&\frac{1}{\lambda(A)^2}\int_{(x,y)\in A^2} \log W(x,y)\mathsf{d}(\lambda^2)\\
&\quad\geqslant \frac{1}{r^2}\sum_i\sum_j \log W(x_i,x_j)\ - C + \varepsilon\log c\\
&\quad\geqslant \frac{r-1}{r}\cdot\frac{2}{r(r-1)}\sum_{i,j:i<j}\log W(x_i,x_j)\ + \frac{1}{r^2}\sum_i \log W(x_i,x_i) - C + \varepsilon\log c\\
&\quad\geqslant \frac{r-1}{r}\cdot\frac{2}{r(r-1)}\sum_{i,j:i<j}\log W(x_i,x_j)\ - C + (\varepsilon+\tfrac{1}{r})\log c\,.
\end{aligned}$$

Letting $\varepsilon \searrow 0$ (which means that also $C \searrow 0$), we get the claim. □

By Claim 3.8.1, we have $\liminf_{r\to\infty} \log P_r \geqslant -\frac{2}{\kappa(W)}$. By Claim 3.8.2, we have $P \geqslant \limsup_{r\to\infty} \log P_r$. Further, it is obvious that $-\frac{2}{\kappa(W)} \geqslant P$: Indeed, the supremum in (6) ranges over all histograms, of which indicators of measurable sets are just a particular case. The combination of the three above inequalities proves the fact.

So, it remains to deal with (17). Positive multiples of indicator functions are histograms, so (7) tells us that $\kappa(W) \geqslant \sup_{a\in\mathbb{R}_+, A\subseteq\Omega} \{a\lambda(A) : \Gamma(a\cdot\mathbf{1}_A, W) \geqslant 0\}$. It remains to deal with the opposite inequality. We shall prove this in the "more precisely" form. Let $\varepsilon > 0$ be arbitrary and take A such that $(1+\varepsilon)\frac{2}{\kappa(W)} \geqslant \frac{1}{\lambda(A)^2}\int_{(x,y)\in A^2} \log(1/W(x,y))\mathsf{d}(\lambda^2)$. Set $a = (1-\varepsilon)\frac{\kappa(W)}{\lambda(A)}$. We claim that the pair (a, A) is admissible for the supremum in (17). Indeed,

$$\begin{aligned}\Gamma(a\cdot\mathbf{1}_A, W) &\overset{(8)}{=} a\lambda(A) - \tfrac{1}{2}a^2\int_{(x,y)\in A^2} -\log W(x,y)\mathsf{d}(\lambda^2)\\ &= (1-\varepsilon)\kappa(W)\left(1 - \tfrac{1}{2}(1-\varepsilon)\kappa(W)\frac{1}{\lambda(A)^2}\int_{(x,y)\in A^2}\log(1/W(x,y))\mathsf{d}(\lambda^2)\right)\\ &\geqslant (1-\varepsilon)\kappa(W)\left(1 - \tfrac{1}{2}(1-\varepsilon)\kappa(W)(1+\varepsilon)\frac{2}{\kappa(W)}\right) = \varepsilon^2(1-\varepsilon)\kappa(W) > 0\,.\end{aligned}$$

Since $a\cdot\lambda(A) = (1-\varepsilon)\kappa(W)$, and since $\varepsilon > 0$ was arbitrary, this finishes the proof. □

3.6. **Exhaustion principle.** We recall the principle of exhaustion (see e.g. [8, Lemma 11.12] for a more general statement).

Lemma 3.9. *Let $\mathcal{C}$ be a collection of measurable subsets of $(0,1)$ with positive Lebesgue measure. Suppose that for every $A \subseteq (0,1)$ with positive Lebesgue measure, there is $C \in \mathcal{C}$ such that $C \subseteq A$. Then there is an at most countable subcollection $\mathcal{B}$ of $\mathcal{C}$ of pairwise disjoint sets such that $\sum_{B\in\mathcal{B}} \lambda(B) = 1$.* □

4. Concentration and oscillation: Proofs of Proposition 2.1 and Theorem 2.2

4.1. **Proof of Proposition 2.1.** We shall need the following well-known crude bound on the minimum difference between uniformly random points.

Fact 4.1. *Suppose that the numbers $x_1, \ldots, x_n$ are uniformly sampled from the interval $(0,1)$. Then asymptotically almost surely, $\min_{i\neq j} |x_i - x_j| > n^{-3}$.*

Consider a sequence of positive numbers $1 = a_1 > a_2 > a_3 > \ldots > 0$, with $\lim_{n\to\infty} a_n = 0$, to be determined later. Define a graphon $W : (0,1)^2 \to [0,1]$ as

$$W(x,y) = \begin{cases} 0 & \text{if } a_{2i-1} \geqslant |x-y| > a_{2i},\\ 1 & \text{if } a_{2i} \geqslant |x-y| > a_{2i+1}.\end{cases}$$

Let us show how to achieve (4). Suppose that numbers $a_1, \dots, a_{2i-1}$ were already set. Fix an arbitrary number n large enough such that $n^{-3} < a_{2i-1}$ and $f(n) > 1 + 1/a_{2i-1}$. Then, set $a_{2i} := n^{-3}$. We claim that with high probability, there is no set of $f(n)$ vertices in $\mathbb{G}(n, W)$ forming a clique. Indeed, consider the representation of the vertices of $\mathbb{G}(n, W)$ in the interval $[0, 1]$. By Fact 4.1 we can assume that the mutual distances between these points are more than a_{2i}. Consider an arbitrary set $S \subseteq (0, 1)$ of these points of size $f(n)$. By the pigeonhole principle there are two points $x, y \in S$ with $|x - y| \leqslant 1/(f(n) - 1) < a_{2i-1}$. On the other hand, $|x - y| > a_{2i}$. We conclude that $W(x, y) = 0$, and thus S does not induce a clique.

Next, let us show how to achieve (5). Suppose that numbers $a_1, \dots, a_{2i}$ were already set. Fix a large number n. In particular, suppose that $n^{-3} < a_{2i}$ and $f(n) > \frac{2}{a_{2i}}$, and let $a_{2i+1} := n^{-3}$. Now, consider the process of generating vertices in $\mathbb{G}(n, W)$. By the Law of Large Numbers, out of n vertices, with high probability, at least $\frac{1}{2}a_{2i}n$ vertices fall in the interval $(\frac{1}{2} - \frac{a_{2i}}{2}, \frac{1}{2} + \frac{a_{2i}}{2})$. By Fact 4.1, with high probability, the differences of pairs of these points are bigger than a_{2i+1}. In particular, the said set of vertices forms a clique of order at least $\frac{1}{2}a_{2i}n > \frac{n}{f(n)}$, as needed for (5). □

Remark 4.2. It may seem that by replacing the values 0 and 1 by some constants $0 < p_1 < p_2 < 1$ in the construction in the proof of Proposition 2.1 we get an oscillation between $c_1 \log n$ and $c_2 \log n$. Theorem 2.5 tells us however that this is not the case: the clique number normalized by $\log n$ will converge in probability. □

4.2. **Proof of Theorem 2.2.** The proof of Theorem 2.2 was suggested to us by Lutz Warnke.

First, we handle the case when $\mathbf{E}[\omega(\mathbb{G}(n, W))]$ is bounded.

Lemma 4.3. *Let W be a graphon, and $L = \lim_{n\to\infty} \mathbf{E}[\omega(\mathbb{G}(n, W))]$. Then $L = \sup\{k \in \mathbb{N} : t(K_k, W) > 0\}$. In addition, if L is finite then $\lim_{n\to\infty} \mathbf{P}[\omega(\mathbb{G}(n, W)) = L] = 1$.*

Proof. The statement follows from the following claim. Suppose that W is a graphon. Then for each $\ell \in \mathbb{N}$ we have

$$\lim_{n\to\infty} \mathbf{P}[\omega(\mathbb{G}(n, W)) \geqslant \ell] \in \{0, 1\} .$$

Indeed, suppose that for some ℓ and n we have that $\mathbf{P}[\omega(\mathbb{G}(n, W)) \geqslant \ell] = \delta > 0$. Then, for each k, we have that $\mathbf{P}[\omega(\mathbb{G}(kn, W)) \geqslant \ell] \geqslant 1 - (1 - \delta)^k$. Consequently, $\lim_{n\to\infty} \mathbf{P}[\omega(\mathbb{G}(n, W)) \geqslant \ell] = 1$. □

Suppose that W is represented on a probability space Ω. By Lemma 4.3, we can assume that $\mathbf{E}[\omega(\mathbb{G}(n, W))] \to \infty$.

To prove the concentration, we shall use Talagrand's inequality. For this, we need to represent $\mathbb{G}(n, W)$ on a suitable product space J. It turns out that the right product

space corresponds to "vertex-exposure" technique known in the theory of random graphs. Let $J := \prod_{i=1}^{n} J_i$, where $J_i = \Omega \times [0,1]^{i-1}$. This indeed is a "vertex-exposure model" of $\mathbb{G}(n, W)$. To see this, consider an arbitrary element $\mathbf{x} \in J$. We can write

$$\mathbf{x} = (x_1, (\,); x_2, (p_{1,2}); x_3, \begin{pmatrix} p_{1,3} \\ p_{2,3} \end{pmatrix}; \ldots; x_n, \begin{pmatrix} p_{1,n} \\ p_{2,n} \\ \cdots \\ p_{n-1,n} \end{pmatrix}) \,,$$

where $x_i \in \Omega$, and $p_{i,j} \in [0,1]$. In the instance of $\mathbb{G}(n, W)$ corresponding to $\mathbf{x}$, vertices i and j are connected if and only if $W(x_i, x_j) \geqslant p_{i,j}$. It is straightforward to check that this gives the right distribution on $\mathbb{G}(n, W)$.

Consider the clique number, this time on the domain J. That is, we have a function $\Psi : J \to \mathbb{R}$, where $\Psi(\mathbf{x})$ is the clique number of the graph corresponding to $\mathbf{x}$. Then Ψ is a (discrete) 1-Lipschitz function. That is if $\mathbf{x}, \mathbf{y} \in J$ are such that they differ in one coordinate, then $|\Psi(\mathbf{x}) - \Psi(\mathbf{y})| \leqslant 1$.[3] Further, Ψ satisfies the so-called *small certificates condition*. This means that whenever $\Psi(\mathbf{x}) \geqslant \ell$, there exists a set C of at most ℓ many coordinates such that $\Psi(\mathbf{y}) \geqslant \ell$ for each $\mathbf{y} \in J$ which agrees with $\mathbf{x}$ on each coordinate from C. (In other words, the values of $\mathbf{x}$ on coordinates from C alone certify that $\Psi(\mathbf{x}) \geqslant \ell$.) Indeed, it is enough just to reveal the values at the indices of one maximum clique. Talagrand's inequality (see [22, Remark 2 following Talagrand's Inequality II, p. 81])[4] then states that there exists an absolute constant $\beta > 0$ such that for $t_n = (\mathbf{E}[\omega(\mathbb{G}(n, W))])^{\frac{3}{4}}$, we have (for every large enough n) that

$$\begin{aligned} \mathbf{P}\big[|\omega(\mathbb{G}(n,W)) - \mathbf{E}[\omega(\mathbb{G}(n,W))]| > t_n\big] &\leqslant 2\exp\left(-\frac{\beta t_n^2}{\mathbf{E}[\omega(\mathbb{G}(n,W))]}\right) \\ &= 2\exp\left(-\beta\sqrt{\mathbf{E}[\omega(\mathbb{G}(n,W))]}\right) \,. \end{aligned}$$

The conclusion immediately follows by letting n go to infinity. □

4.3. **Graphons with a bounded clique number.** Lemma 4.3 provides some information about graphons with a bounded clique number. In this section, we prove Theorem 2.3 which gives a much more explicit description.

Proof of Theorem 2.3. Let $W\colon \Omega^2 \to [0,1]$ be a graphon with a bounded clique number. By Lemma 4.3 we know that $L = \sup\{k \in \mathbb{N} : t(K_k, W) > 0\}$ is a finite (natural) number.

[3] One could consider a stronger notion of 1-Lipschitzness, namely, to require that changing $\mathbf{x}$ on an E-coordinate by ε would change the value of our function by at most ε. This clearly is not true for Ψ. However, the weaker version is sufficient for our purposes.

[4] Actually, as was communicated to us by Lutz Warnke and Mike Molloy, there is a typo in [22]. The effect of this typo, however, is only the value of the constant β below. Since we do not make β explicit, this typo is irrelevant.

First of all, we will show that there is a set $B \subseteq \Omega$ of positive measure such that $W \restriction_{B\times B} = 0$ almost everywhere. We may assume that $L \geqslant 2$ (the case $L = 1$ is trivial). For every $(L-1)$-tuple $\boldsymbol{x} = (x_1, \dots, x_{L-1}) \in \Omega^{L-1}$, let us denote $Q_{\boldsymbol{x}} = \{y \in \Omega\colon \prod_{i=1}^{L-1} W(x_i, y) > 0\}$. It follows from the equality $t(K_{L+1}, W) = 0$ that for every (up to a set of measure zero) $\boldsymbol{x} \in \Omega^{L-1}$ such that $\prod_{i<j} W(x_i, x_j) > 0$, we have $W \restriction_{Q_{\boldsymbol{x}}\times Q_{\boldsymbol{x}}} = 0$ almost everywhere. But since $t(K_L, W) > 0$, the set of all $\boldsymbol{x} \in \Omega^{L-1}$ such that $\prod_{i<j} W(x_i, x_j) > 0$ and $\nu(Q_{\boldsymbol{x}}) > 0$, has positive measure. So it suffices to set $B = Q_{\boldsymbol{x}}$ for a suitable $\boldsymbol{x} \in \Omega^{L-1}$.

Next, observe that for every $A \subseteq \Omega$ of positive measure, there is $B \subseteq A$ of positive measure such that $W \restriction_{B\times B} = 0$ almost everywhere. This follows by the previous considerations applied on the subgraphon $W^* = W \restriction_{A\times A}$ (for which we still have that $\sup\{k \in \mathbb{N} : t(K_k, W^*) > 0\} < +\infty$).

Finally, let $W'\colon (0,1)^2 \to [0,1]$ be a representation of the graphon W on $(0,1)$. Then the statement follows by an application of Lemma 3.9. □

5. Biclique number

In this section, we prove Theorem 2.4. First, we introduce some additional notation. When we refer to a bipartite graph $H = (V, W; E)$ as a *bigraph*, we consider a distinguished order of the colour classes $V = \{v_1, \dots, v_p\}$ and $W = \{w_1, \dots, w_q\}$. In such a case we define the *bipartite density* of H in a bigraphon $U : \Omega_1 \times \Omega_2 \to [0,1]$ by

$$t_{\mathrm{B}}(H, U) = \int_{(x_1,\dots,x_p)\in\Omega_1^p} \int_{(y_1,\dots,y_q)\in\Omega_2^q} \prod_{ij:v_iw_j\in E} U(x_i, y_j)\, \mathsf{d}(\nu_2^q)\mathsf{d}(\nu_1^p)\,. \tag{19}$$

Note that for a bigraph $H = (V, W; E)$ and its conjugate $H' = (W, V; E)$ the quantities $t_{\mathrm{B}}(H, U)$ and $t_{\mathrm{B}}(H', U)$ are not necessarily equal.

As we will see, the upper bound in Theorem 2.4 is trivial. For the lower bound, we need to make a small detour to Sidorenko's conjecture.

5.1. **Sidorenko's conjecture.** A famous conjecture of Simonovits and Sidorenko, often referred to as "Sidorenko's conjecture", [25, 24] asserts that among all graphs of a given (large) order n and fixed edge density d, the density of a fixed bipartite graph is minimized for a typical sample from $\mathbb{G}(n, d)$. The conjecture can be particularly neatly phrased in the language of graphons — as observed already by Sidorenko a decade before the notion of graphons itself — in which case it asserts that

$$t_{\mathrm{B}}(H, U) \geqslant \left(\int_{x\in\Omega_1} \int_{y\in\Omega_2} U(x, y)\, \mathsf{d}(\nu_2)\mathsf{d}(\nu_1) \right)^{e(H)}, \tag{20}$$

for each bigraphon $U : \Omega_1 \times \Omega_2 \to [0,1]$ and each bigraph H. Despite recent results [12, 5, 17, 16], the conjecture is wide open. We shall need the solution of Sidorenko's Conjecture

for $H = K_{n,m}$ which was observed already by Sidorenko. We give a short self-contained and unoriginal proof here.

Proposition 5.1. *Suppose that $U : \Omega_1 \times \Omega_2 \to [0,1]$ is an arbitrary bigraphon and $n, m \in \mathbb{N}$ are arbitrary. Then for the complete bigraph $K_{n,m}$ we have*

$$t_{\mathrm{B}}(K_{n,m}, U) \geqslant \Big(\int_{x \in \Omega_1} \int_{y \in \Omega_2} U(x,y) \, \mathsf{d}(\nu_2) \mathsf{d}(\nu_1) \Big)^{nm} .$$

Proof. We have

$$\begin{aligned}
\int_{x_1,\dots,x_n} \int_{y_1,\dots,y_m} \prod_{i \in [n], j \in [m]} U(x_i, y_j) &= \int_{x_1,\dots,x_n} \Big(\int_y \prod_{i \in [n]} U(x_i, y) \Big)^m \\
&\geqslant \Big(\int_{x_1,\dots,x_n} \int_y \prod_{i \in [n]} U(x_i, y) \Big)^m \\
&= \Big(\int_y \int_{x_1,\dots,x_n} \prod_{i \in [n]} U(x_i, y) \Big)^m \\
&= \Big(\int_y \Big(\int_x U(x,y) \Big)^n \Big)^m \\
&\geqslant \Big(\int_x \int_y U(x,y) \Big)^{nm} ,
\end{aligned}$$

where both inequalities follow by applications of Hölder's inequality. □

5.2. **Bicliques in almost constant bigraphons.** As a preliminary step for our proof of Theorem 2.4, we study bicliques in random bipartite graphs sampled from almost constant bigraphons. This condition is formalized by the following definition.

Definition 5.2. A bigraphon $U : \Omega_1 \times \Omega_2 \to [0,1]$ is (d, ε)-*constant* if

$$\int_{(x,y) \in \Omega_1 \times \Omega_2} U(x,y) \mathsf{d}(\nu_1 \times \nu_2) \geqslant d$$

and $\operatorname{ess\,sup} U \leqslant d + \varepsilon$.

Proposition 5.3. *Let $0 < d_1 < d_2 < 1$ be given. Then for every $\alpha \in (0,1)$ there exists $\varepsilon \in (0,1)$ such that the following holds: Whenever we have $d \in (d_1, d_2)$ and a (d, ε)-constant bigraphon $U : \Omega_1 \times \Omega_2 \to [0,1]$ then for $G \sim \mathbb{B}(k, U)$ we have a.a.s. that*

$$\omega_2(G) \geqslant (1 - \alpha) \cdot \frac{2}{\log 1/d} \cdot \log k .$$

Proof. Let $\alpha \in (0,1)$ be arbitrary. Suppose that $\varepsilon > 0$ is sufficiently small (we will make this precise later), $d \in (d_1, d_2)$ and $U : \Omega_1 \times \Omega_2 \to [0,1]$ is (d, ε)-constant. Suppose further that k is large.

Let X_k be the number of bicliques in $\mathbb{B}(k,U)$ whose two colour classes have size $\ell = (1-\alpha)\cdot\frac{2}{\log 1/d}\cdot\log k$. Multiplicities caused by automorphisms of $K_{\ell,\ell}$ are not counted. By Proposition 5.1 we have

$$\mathbf{E}[X_k] = \binom{k}{\ell}^2 \cdot t_{\mathrm{B}}(K_{\ell,\ell},U) \geqslant \binom{k}{\ell}^2 d^{\ell^2} \geqslant \left(\frac{k}{\ell}\right)^{2\ell} d^{\ell^2} = \left(\frac{k^{2\alpha}}{\ell^2}\right)^{\ell} . \tag{21}$$

Next, we are going to show by a second moment argument that $X_k \approx \mathbf{E}[X_k]$ a.a.s. For $p,q = 0,1,\dots,\ell$, we define the bigraph $K_{[\ell,p,q]}$ as a result of gluing two copies of $K_{\ell,\ell}$ along p vertices in the first colour class and q vertices in the second colour class. Alternatively, $K_{[\ell,p,q]}$ can be obtained by erasing edges of two disjoint copies of the bigraph $K_{\ell-p,\ell-q}$ from $K_{2\ell-p,2\ell-q}$. We have

$$e(K_{[\ell,p,q]}) = 2\ell^2 - pq\ . \tag{22}$$

We have

$$\mathbf{E}[X_k^2] = \sum_{p=0}^{\ell}\sum_{q=0}^{\ell}\mathbf{E}[Y_{p,q}]\ , \tag{23}$$

where $Y_{p,q}$ counts the number of bigraphs $K_{[\ell,p,q]}$ which preserve the order of the colour classes. We expand the second moment as

$$\mathbf{E}[Y_{p,q}] = \binom{k}{\ell-p\,|\,\ell-p\,|\,p}\cdot\binom{k}{\ell-q\,|\,\ell-q\,|\,q}\cdot t_{\mathrm{B}}(K_{[\ell,p,q]},U)\ . \tag{24}$$

Claim 5.3.1. *For every $c>0$ there exists $\varepsilon_1>0$ such that the following holds: Whenever $d\in(d_1,d_2)$ and $U:\Omega_1\times\Omega_2\to[0,1]$ is (d,ε_1)-constant then for each $p,q\in[\ell]_0$, $p+q\geqslant c\log k$, we have*

$$\mathbf{E}[X_k]^2 \geqslant \log^3 k\cdot\mathbf{E}[Y_{p,q}]\ .$$

Proof of Claim 5.3.1. Let $c>0$ be arbitrary. Suppose that $\varepsilon_1>0$ is sufficiently small, $d\in(d_1,d_2)$ and $U:\Omega_1\times\Omega_2\to[0,1]$ is (d,ε_1)-constant. Let $p,q\in[\ell]_0$ be such that $p+q\geqslant c\log k$ (and of course $p+q\leqslant 2\ell$). Upper-bounding the terms in (24) (while

bearing in mind that $\ell = \Theta(\log k)$), we get

$$\begin{aligned}
\mathbf{E}[Y_{p,q}] &\leqslant k^{4\ell-p-q}(d+\varepsilon_1)^{e(K_{[\ell,p,q]})} \\
&\overset{(22)}{=} k^{4\ell-p-q}(d+\varepsilon_1)^{2\ell^2-pq} \\
\boxed{\text{AM-GM Ineq., } d+\varepsilon_1<1} \quad &\leqslant \left(k^2(d+\varepsilon_1)^{\ell}\right)^{2\ell}\left(k(d+\varepsilon_1)^{\frac{p+q}{4}}\right)^{-(p+q)} \\
\boxed{p+q\leqslant 2\ell \text{ and } d+\varepsilon_1<1} \quad &\leqslant \left(k^2(d+\varepsilon_1)^{\ell}\right)^{2\ell}\left(k(d+\varepsilon_1)^{\frac{\ell}{2}}\right)^{-(p+q)} \qquad (25)\\
\boxed{k(d+\varepsilon_1)^{\frac{\ell}{2}}\geqslant k^{\alpha}} \quad &\leqslant \exp\left(4\ell\log k\left(1+(1-\alpha)\tfrac{\log(d+\varepsilon_1)}{\log 1/d}\right)\right)k^{-c\alpha\log k} \\
&= k^{4\alpha\ell}\exp\left(4\ell\log k\,(1-\alpha)\left(1-\tfrac{\log(d+\varepsilon_1)}{\log d}\right)-c\alpha\log^2 k\right) \\
\boxed{\varepsilon_1 \text{ sufficiently small}} \quad &\leqslant k^{4\alpha\ell}\exp\left(-\tfrac{1}{2}c\alpha\log^2 k\right)
\end{aligned}$$

It is now enough to compare this with (21). □

Claim 5.3.2. *There exist numbers $C, \varepsilon_2 > 0$ such that the following holds: Whenever $d \in (d_1, d_2)$ and $U : \Omega_1 \times \Omega_2 \to [0,1]$ is (d, ε_2)-constant then for each $p, q \in [\ell]_0$, $1 \leqslant p+q < C\log k$, we have*

$$\mathbf{E}[Y_{0,0}] \geqslant k^{\frac{1}{2}}\mathbf{E}[Y_{p,q}] \,.$$

Proof of Claim 5.3.2. Suppose that $\varepsilon_2 > 0$ is sufficiently small, $d \in (d_1, d_2)$ and $U : \Omega_1 \times \Omega_2 \to [0,1]$ is (d, ε_2)-constant. Let us compare the combinatorial coefficients corresponding to $\mathbf{E}[Y_{0,0}]$ and $\mathbf{E}[Y_{p,q}]$ in (24). We have

$$\frac{\binom{k}{\ell\,|\,\ell}^2}{\binom{k}{\ell-p\,|\,\ell-p\,|\,p}\cdot\binom{k}{\ell-q\,|\,\ell-q\,|\,q}} = k^{(1+o(1))(p+q)} \,, \qquad (26)$$

where $o(1) \to 0$ as $k \to \infty$ uniformly for any choice of p and q.

It remains to compare the terms $t_{\mathrm{B}}(K_{[\ell,0,0]}, U)$ and $t_{\mathrm{B}}(K_{[\ell,p,q]}, U)$.

First, we claim that for each $i, j, h \in \mathbb{N}$ we have

$$t_{\mathrm{B}}(K_{i,h}, U)t_{\mathrm{B}}(K_{j,h}, U) \leqslant t_{\mathrm{B}}(K_{i+j,h}, U) \,. \qquad (27)$$

Indeed, by Hölder's inequality, we have

$$t_{\mathrm{B}}(K_{i,h}, U) = \int_T \left(\int_x \deg(x,T)\right)^i \leqslant \left(\int_T \left(\int_x \deg(x,T)\right)^{i+j}\right)^{\frac{i}{i+j}} \,,$$

$$t_{\mathrm{B}}(K_{j,h}, U) = \int_T \left(\int_x \deg(x,T)\right)^j \leqslant \left(\int_T \left(\int_x \deg(x,T)\right)^{i+j}\right)^{\frac{j}{i+j}} \,,$$

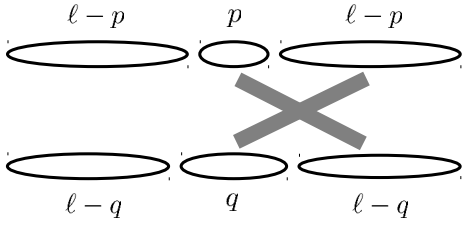


FIGURE 1. The $p\ell + q\ell - 2pq$ edges of $K_{[\ell,p,q]}$ corresponding to factors bounded from above by $d + \varepsilon_2$ in (19).

where the integrations are over $T = (t_1, \dots, t_h) \in (\Omega_2)^h$, and $x \in \Omega_1$, and $\deg(x, T) = \prod_{r=1}^{h} U(x, t_r)$. Thus

$$t_{\mathrm{B}}(K_{i,h}, U) t_{\mathrm{B}}(K_{j,h}, U) \leqslant \int_T \left(\int_x \deg(x, T) \right)^{i+j} = t_{\mathrm{B}}(K_{i+j,h}, U) \,,$$

as we wanted.

A double application of (27) followed by an application of Proposition 5.1 gives

$$\begin{aligned} t_{\mathrm{B}}(K_{\ell,\ell}, U) &\geqslant t_{\mathrm{B}}(K_{\ell-p,\ell}, U) t_{\mathrm{B}}(K_{p,\ell}, U) \geqslant t_{\mathrm{B}}(K_{\ell-p,\ell-q}, U) t_{\mathrm{B}}(K_{\ell-p,q}, U) t_{\mathrm{B}}(K_{p,\ell}, U) \\ &\geqslant t_{\mathrm{B}}(K_{\ell-p,\ell-q}, U) d^{q\ell-pq} d^{p\ell} \,. \end{aligned} \tag{28}$$

In the defining formula (19) for $t_{\mathrm{B}}(K_{[\ell,p,q]}, U)$ we show that $d + \varepsilon_2$ are upper bounds on some factors in $\prod U(x_i, y_i)$, as in Figure 1. Observe that after removal of the $p\ell + q\ell - 2pq$ edges indicated in Figure 1, the graph $K_{[\ell,p,q]}$ decomposes into a disjoint union of $K_{\ell,\ell}$ and $K_{\ell-p,\ell-q}$. Note that for a disjoint union $H_1 \oplus H_2$ of two bigraphs H_1 and H_2 we have $t_{\mathrm{B}}(H_1 \oplus H_2, U) = t_{\mathrm{B}}(H_1, U) t_{\mathrm{B}}(H_2, U)$. Thus,

$$t_{\mathrm{B}}(K_{[\ell,p,q]}, U) \leqslant (d + \varepsilon_2)^{p\ell+q\ell-2pq} t_{\mathrm{B}}(K_{\ell,\ell}, U) t_{\mathrm{B}}(K_{\ell-p,\ell-q}, U). \tag{29}$$

Therefore,

(30)

$$\begin{aligned} t_{\mathrm{B}}(K_{[\ell,0,0]}, U) &= t_{\mathrm{B}}(K_{\ell,\ell}, U) t_{\mathrm{B}}(K_{\ell,\ell}, U) \\ &\overset{(28)}{\geqslant} t_{\mathrm{B}}(K_{\ell,\ell}, U) t_{\mathrm{B}}(K_{\ell-p,\ell-q}, U) d^{p\ell+q\ell-pq} \\ &\overset{(29)}{\geqslant} t_{\mathrm{B}}(K_{[\ell,p,q]}, U) \left(\frac{d}{d+\varepsilon_2} \right)^{p\ell+q\ell} \left(\frac{(d+\varepsilon_2)^2}{d} \right)^{pq} \\ &\boxed{\text{AM-GM Ineq., } \varepsilon_2 \text{ such that } \tfrac{(d+\varepsilon_2)^2}{d} \leqslant 1} \quad \geqslant t_{\mathrm{B}}(K_{[\ell,p,q]}, U) \left(1 - \frac{\varepsilon_2}{d+\varepsilon_2} \right)^{p\ell+q\ell} \left(\frac{(d+\varepsilon_2)^2}{d} \right)^{(p+q)\cdot\frac{p+q}{4}} \\ &\geqslant t_{\mathrm{B}}(K_{[\ell,p,q]}, U) \left(1 - \frac{\varepsilon_2}{d+\varepsilon_2} \right)^{p\ell+q\ell} d^{(p+q)\cdot\frac{C\log k}{4}} \,. \end{aligned}$$

Substituting (26) and (30) into (24) we get

$$\begin{aligned}
\frac{\mathbf{E}[Y_{0,0}]}{\mathbf{E}[Y_{p,q}]} &\geqslant \left(k^{1+o(1)}\left(1-\frac{\varepsilon_2}{d+\varepsilon_2}\right)^{\ell} d^{\frac{C\log k}{4}}\right)^{p+q} \\
\boxed{\text{for } \zeta \in (0,\tfrac{1}{2}) \text{ we have } 1-\zeta \geqslant \exp(-2\zeta)} \quad &\geqslant \left(k^{1+o(1)}\exp\left(-(1-\alpha)\frac{2\varepsilon_2}{d+\varepsilon_2}\cdot\frac{2}{\log 1/d}\cdot\log k\right) d^{\frac{C\log k}{4}}\right)^{p+q} \\
&= \left(k^{1+o(1)}\cdot k^{(1-\alpha)\frac{2\varepsilon_2}{d+\varepsilon_2}\cdot\frac{2}{\log d}}\cdot k^{\frac{C\log d}{4}}\right)^{p+q} \\
&\geqslant k^{\frac{1}{2}(p+q)} \geqslant k^{\frac{1}{2}} ,
\end{aligned}$$

for $C = 1/\log(1/d_1)$, and $\varepsilon_2 < \frac{1}{20} d_1 \log(1/d_2)$. □

Let $C > 0$ and $\varepsilon_2 > 0$ be given by Claim 5.3.2. Let $\varepsilon_1 > 0$ be given by Claim 5.3.1 for $c = C$. Now if $\varepsilon < \min(\varepsilon_1, \varepsilon_2)$, we are ready provide upper bounds on the summands in (23). Note that we have $\Theta(\log^2 k)$ many of these summands. We get

$$\begin{aligned}
\mathbf{E}[X_k^2] &= \mathbf{E}[Y_{0,0}] + \sum_{\substack{p,q\in[l]_0 \\ 1\leqslant p+q<C\log k}} \mathbf{E}[Y_{p,q}] + \sum_{\substack{p,q\in[l]_0 \\ p+q\geqslant C\log k}} \mathbf{E}[Y_{p,q}] \\
&\leqslant \mathbf{E}[Y_{0,0}] + \Theta(\log^2 k)\mathbf{E}[Y_{0,0}]k^{-\frac{1}{2}} + \Theta(\log^2 k)\mathbf{E}[X_k]^2\log^{-3} k \\
\boxed{\mathbf{E}[Y_{0,0}] = (1+o(1))\mathbf{E}[X_k]^2} \quad &= (1+o(1))\mathbf{E}[X_k]^2.
\end{aligned}$$

Therefore we have $\mathbf{E}[X_k^2] = (1+o(1))\mathbf{E}[X_k]^2$, and it follows by Chebyshev's inequality that $\mathbf{P}[X_k > 0] = 1 - o(1)$. □

5.3. **Proof of Theorem 2.4.** The upper bound is easy, since it claims that $\omega_2(\mathbb{B}(n,U))$ is typically not bigger than the biclique number in the balanced bipartite Erdős–Rényi random graph $\mathbb{B}(n,p)$ (which clearly stochastically dominates $\mathbb{B}(n,U)$). For completeness, we include the calculations. We write $Y_k(\mathbb{B}(n,U))$ for the number of complete balanced bipartite graphs on $k+k$ vertices inside $\mathbb{B}(n,U)$. For $k = (1+\varepsilon)\cdot\frac{2}{\log 1/p}\cdot\log n$, we have,

$$\mathbf{E}[Y_k] \leqslant \binom{n}{k}\cdot\binom{n}{k}\cdot p^{k^2} \leqslant n^{2k}p^{k^2} = (n^2p^k)^k . \tag{31}$$

The statement now follows from Markov's Inequality, provided that we can show that $n^2p^k \to 0$. Indeed,

$$n^2p^k = n^2 p^{\frac{2\log n}{\log 1/p}} p^{\frac{2\varepsilon\log n}{\log 1/p}} = p^{\frac{2\varepsilon\log n}{\log 1/p}} \to 0 .$$

We now turn to the lower bound. Let $\alpha \in (0,1)$ be arbitrary and let $\varepsilon_0 > 0$ be given by Proposition 5.3 for $d_1 = \frac{p}{2}$ and $d_2 = p$. Let $\varepsilon < \min\left(\varepsilon_0, \frac{p}{2}\right)$ be arbitrary. We denote by ν_i the measure given on Ω_i, $i = 1,2$. The definition of the essential supremum, together

with Theorem 3.6, gives that there exist two measurable sets $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$ such that $\nu_1(A), \nu_2(B) > 0$ and

$$\int_{(x,y)\in\Omega_1\times\Omega_2} U(x,y)\mathsf{d}(\nu_1 \times \nu_2) \geqslant (p-\varepsilon)\nu_1(A)\nu_2(B) \ .$$

We put $\delta = \min(\nu_1(A), \nu_2(B))$. By rescaling the measures ν_1, ν_2, we get probability measures ν_1^* on A and ν_2^* on B. Then we can view $U \restriction_{A\times B}$ as a bigraphon, which we denote by U^*. Note that U^* is $(p-\varepsilon, \varepsilon)$-constant (and thus also $(p-\varepsilon, \varepsilon_0)$-constant as $\varepsilon < \varepsilon_0$).

Consider now the sampling process to generate $B \sim \mathbb{B}(n, U)$ as described above. A standard concentration argument gives that with high probability, at least $\frac{\delta n}{2}$ points x_i sampled in Ω_1 lie in the set A, and at least $\frac{\delta n}{2}$ points y_j sampled in Ω_2 lie in the set B. In other words, with high probability we can find a copy of $\mathbb{B}(\frac{\delta n}{2}, U^*)$ inside $\mathbb{B}(n, U)$. Looking at the biclique number, we get that for $\ell = (1-\alpha)\cdot\frac{2}{\log 1/(p-\varepsilon)}\cdot\log(\delta n/2)$, we have

$$\mathbf{P}[\omega_2(\mathbb{B}(n,U)) \geqslant \ell] \geqslant \mathbf{P}\left[\omega_2(\mathbb{B}(\tfrac{\delta n}{2}, U^*)) \geqslant \ell\right] - o(1) \geqslant 1 - o(1) \ ,$$

where the last inequality follows from Proposition 5.3. Since $\log(\delta n/2) = (1+o(1))\log n$, and since $\alpha \in (0,1)$ and $\varepsilon < \min\left(\varepsilon_0, \frac{p}{2}\right)$ were arbitrary, the claim follows. □

6. Formula for graphs with logarithmic clique number

In this section we try to informally justify Theorem 2.5. While we believe that the derivation here captures the essence of the problem, the actual proof, presented in Section 8, is quite different. At the end of this section we comment on what fails in turning the heuristics into a rigorous argument.

6.1. **First moment for a 2-step graphon.** Let us try to gain some intuition on Theorem 2.5 by looking at one of the simplest non-constant graphons. Let $W : \Omega^2 \to [0,1]$ be represented on the unit interval Ω (c.f. Fact 3.1) and defined by

$$W(x,y) = \begin{cases} p_{11} & \text{if } x,y \in \Omega_1, \\ p_{22} & \text{if } x,y \in \Omega_2, \text{ and} \\ p_{12} & \text{otherwise.} \end{cases} \tag{32}$$

Here, $p_{11}, p_{22}, p_{12} \in (0,1)$ are arbitrary, and Ω is partitioned arbitrarily into two measurable sets Ω_1 and Ω_2 of positive measures β_1 and β_2.

Our aim is to determine for which numbers $c \in \mathbb{R}^+$ there typically exists a clique of order $c\log n$ in $G \sim \mathbb{G}(n, W)$, and for which c's there is typically none. So let us fix $c \in \mathbb{R}^+$ and let X_n count the number of cliques of order $c\log n$ in $G \sim \mathbb{G}(n, W)$. By Markov's Inequality, $\omega(G)$ will be typically smaller than $c\log n$ in the regime when $\mathbf{E}[X_n] \to 0$. On the other hand, it is plausible that a second moment argument will give that typically

$\omega(G) \geqslant c \log n$ when $\mathbf{E}[X_n] \to +\infty$. With this belief — which is supported by the success of a second-moment argument in the proof of Theorem 2.4 — let us estimate $\mathbf{E}[X_n]$. Actually, we rather look at a refined quantity $Y_n^{\alpha_1,\alpha_2}(G)$ which is defined as the number of cliques in G that consist of $\alpha_1 \log n$ vertices whose representation on Ω lies in Ω_1, and $\alpha_2 \log n$ vertices that are represented in Ω_2. We have

$$\mathbf{E}[X_n] = \sum_{m=0}^{c \log n} \mathbf{E}\left[Y_n^{m/\log n, c-m/\log n} \right] .$$

We expect the quantities $Y_n^{\alpha_1,\alpha_2}$ to be either super-polynomially small or super-polynomially large. Since the sum above has only $\Theta(\log n)$-many summands, we expect that

(33)
$$\mathbf{E}[X_n] \to +\infty \quad \text{if and only if} \quad \exists \alpha_1, \alpha_2 \geqslant 0 \text{ such that } \alpha_1 + \alpha_2 = c \text{ and } \mathbf{E}[Y_n^{\alpha_1,\alpha_2}] \to +\infty .$$

For a clique that contributes to $Y_n^{\alpha_1,\alpha_2}(G)$ to be present, $\binom{\alpha_1 \log n}{2}$ edges in the $(\Omega_1 \times \Omega_1)$-part of W, $\binom{\alpha_2 \log n}{2}$ edges in the $(\Omega_2 \times \Omega_2)$-part, and $\alpha_1\alpha_2 \log^2 n$ edges in the $(\Omega_1 \times \Omega_2)$-part must be present in the specific locations of perspective cliques or complete bipartite graphs. By the Law of Large Numbers, approximately $\beta_1 n$ points in the sampling process for $\mathbb{G}(n, W)$ end up in Ω_1 and approximately $\beta_2 n$ points end up in Ω_2. We get

(34)
$$\begin{aligned} \mathbf{E}[Y_n^{\alpha_1,\alpha_2}] &\approx \binom{\beta_1 n}{\alpha_1 \log n} \binom{\beta_2 n}{\alpha_2 \log n} p_{11}^{\binom{\alpha_1 \log n}{2}} p_{22}^{\binom{\alpha_2 \log n}{2}} p_{12}^{\alpha_1\alpha_2 \log^2 n} \\ &= \exp\left((1+o(1)) \log^2 n \left(\alpha_1 + \alpha_2 + \frac{\alpha_1^2}{2} \log p_{11} + \frac{\alpha_2^2}{2} \log p_{22} + \alpha_1\alpha_2 \log p_{12} \right) \right) . \end{aligned}$$

Thus, whether $\mathbf{E}[Y_n^{\alpha_1,\alpha_2}] \to 0$ or $\mathbf{E}[Y_n^{\alpha_1,\alpha_2}] \to +\infty$ depends on whether

$$\underbrace{\alpha_1 + \alpha_2}_{(*)} + \underbrace{\frac{1}{2}\left(\alpha_1^2 \log p_{11} + \alpha_2^2 \log p_{22} + \alpha_1\alpha_2 \log p_{12} + \alpha_2\alpha_1 \log p_{12} \right)}_{(**)} \tag{35}$$

is negative or positive, respectively. It is straightforward to generalize this formula to graphons with more steps. Observe also that the values of β_1 and β_2 get lost in the transition between the first and the second line of (34), and are immaterial in (35) consequently (provided that they are positive). In particular, the step sizes β_i could have been "infinitesimally small". Thus, we can see a direct correspondence between $(*)$ and $(**)$ in (8) and (35), where the integration corresponds to passing to infinitesimal steps. In view of this, we will denote the term in (35) by $\Gamma(\alpha, W)$, where $\alpha = (\alpha_1, \alpha_2)$. The optimization over α_1 and α_2 in (33) corresponds to taking the supremum in (7).

This is why we call the functions f in (7) (or vectors, in case of step-graphons) histograms: they specify the densities of the vertices of the anticipated cliques over the space

Ω. Also, motivated by (7) and its interpretation above, we say that a histogram f is *admissible* for a graphon W if $\Gamma(f, W) \geqslant 0$.

Last, let us note that a physicist might refer to $(*)$ as the "entropy contribution", as it comes from the choice of the vertices of a clique, while $(**)$ could be referred to as the "energy" needed to include all required edges of that clique.

6.2. **Introducing the second moment to the example.** So far, our prediction was based on a first moment argument. Combined with Markov's Inequality this gives readily an upper bound on the typical clique number of $\mathbb{G}(n, W)$. We now want to complement the upper bound with a lower bound based on a second moment argument. Let us first recall the situation in the setting of the Erdős–Rényi random graphs $\mathbb{G}(n, p)$. There, a straightforward calculation for the random variable X_n counting cliques of order $c \log n$ (where $c > 0$ is fixed) gives that $\mathbf{E}[X_n]^2 = (1 + o(1))\mathbf{E}[X_n^2]$ if and only if $\mathbf{E}[X_n] \to +\infty$. Thus, the first and the second moment start working together at the same time.

The situation is more complicated in the model $\mathbb{G}(n, W)$. We will illustrate this on the graphon W from (32). Suppose that $\alpha_1, \alpha_2 \geqslant 0$ are such that (35) is positive, and we ask in hope whether

$$\mathbf{E}[Y_n^{\alpha_1,\alpha_2}]^2 = (1 + o(1))\mathbf{E}[(Y_n^{\alpha_1,\alpha_2})^2] \tag{36}$$

In (34), we provided asymptotics for $\mathbf{E}[Y_n^{\alpha_1,\alpha_2}]$. Thus, to understand whether we have (36), we need to calculate $\mathbf{E}[(Y_n^{\alpha_1,\alpha_2})^2]$. We have

$$\mathbf{E}[(Y_n^{\alpha_1,\alpha_2})^2] = \mathbf{E}\left[\left(\sum_K \mathbf{1}_{K \text{ induces a clique}}\right)^2\right] = \sum_{K,L} \mathbf{P}[K \text{ and } L \text{ induce cliques}]\,, \tag{37}$$

where K and L range over all sets of vertices with $\alpha_1 \log n$ vertices represented in Ω_1 and $\alpha_2 \log n$ vertices represented in Ω_2. Let K_1 be the vertices of K represented in Ω_1, and let K_2, L_1, and L_2 be defined analogously. It is clear that $|K_1 \setminus L_1| = |L_1 \setminus K_1|$ and $|K_2 \setminus L_2| = |L_2 \setminus K_2|$. So for fixed sets K and L, we have that the probability that K and L induce cliques is equal to

$$p_{11}^{\binom{|K_1 \cup L_1|}{2} - |K_1 \setminus L_1|^2} \cdot p_{22}^{\binom{|K_2 \cup L_2|}{2} - |K_2 \setminus L_2|^2} \cdot p_{12}^{|K_1 \cup L_1| \cdot |K_2 \cup L_2| - 2|K_1 \setminus L_1||K_2 \setminus L_2|}\,.$$

Thus, grouping (37) depending on the values of $m_1 = |K_1 \cap L_1|$ and $m_2 = |K_2 \cap L_2|$ we get

$$\begin{aligned}\mathbf{E}[(Y_n^{\alpha_1,\alpha_2})^2] \approx \sum_{m_1=0}^{\alpha_1 \log n} \sum_{m_2=0}^{\alpha_2 \log n} & \binom{\beta_1 n}{m_1 \mid \alpha_1 \log n - m_1 \mid \alpha_1 \log n - m_1} \\ & \times \binom{\beta_2 n}{m_2 \mid \alpha_2 \log n - m_2 \mid \alpha_2 \log n - m_2} \\ & \times p_{11}^{\binom{2\alpha_1 \log n - m_1}{2} - (\alpha_1 \log n - m_1)^2} \cdot p_{22}^{\binom{2\alpha_2 \log n - m_2}{2} - (\alpha_2 \log n - m_2)^2} \\ & \times p_{12}^{(2\alpha_1 \log n - m_1)\cdot(2\alpha_2 \log n - m_2) - 2(\alpha_1 \log n - m_1)(\alpha_2 \log n - m_2)},\end{aligned} \tag{38}$$

where the approximate equality represents the fact that we assumed on the right-hand side that exactly $\beta_i n$ vertices are represented in Ω_i. Let us write $\gamma_i = m_i / \log n$, and let us write $z_n^{\gamma_1,\gamma_2}$ for the individual summands on the right-hand side of (38). Routine manipulations give that

$$\begin{aligned}\frac{\log z_n^{\gamma_1,\gamma_2}}{\log^2 n} \approx{}& \gamma_1 + \gamma_2 + 2(\alpha_1 - \gamma_1) + 2(\alpha_2 - \gamma_2) \\ & + (\alpha_1^2 - \tfrac{1}{2}\gamma_1^2) \log p_{11} + (\alpha_2^2 - \tfrac{1}{2}\gamma_2^2) \log p_{22} + (2\alpha_1\alpha_2 - \gamma_1\gamma_2) \log p_{12} \,.\end{aligned}$$

Thus, if we want the second moment (36) to work then comparing the calculations above with (34), we must have for each $\gamma_1 \in [0, \alpha_1]$ and $\gamma_2 \in [0, \alpha_2]$ that

$$\begin{aligned}& 2\left(\alpha_1 + \alpha_2 + \frac{\alpha_1^2}{2} \log p_{11} + \frac{\alpha_2^2}{2} \log p_{22} + \alpha_1\alpha_2 \log p_{12}\right) \\ & \gtrapprox \gamma_1 + \gamma_2 + 2(\alpha_1 - \gamma_1) + 2(\alpha_2 - \gamma_2) + (\alpha_1^2 - \tfrac{1}{2}\gamma_1^2) \log p_{11} + \\ & \qquad (\alpha_2^2 - \tfrac{1}{2}\gamma_2^2) \log p_{22} + (2\alpha_1\alpha_2 - \gamma_1\gamma_2) \log p_{12} \,,\end{aligned}$$

which rewrites as $\Gamma(\gamma, W) \gtrapprox 0$. To summarize, to justify (7) (at least for step-functions), it would suffice to have the following.

Dream Lemma. *If W is a step-graphon with k steps and $\alpha \in [0, +\infty)^k$ is a vector with $\Gamma(\alpha, W) > 0$, then for all $\gamma \in \mathrm{Box}(\alpha)$ we have $\Gamma(\gamma, W) \geqslant 0$.*

This, however, does not hold in general. Indeed, take for example

$$p_{11} = e^{-3}, p_{12} = p_{22} = e^{-\frac{1}{4}}, \quad \alpha_1 = 1, \alpha_2 = 1, \quad \gamma_1 = 1, \gamma_2 = 0.$$

We have $\Gamma(\alpha, W) = \frac{1}{8} > 0$ but $\Gamma(\gamma, W) = -\frac{1}{2} < 0$. It is worth explaining what is happening in the example in words. The parameters p_{11} and α_1 are set so that asymptotically almost surely, $\mathbb{G}(\frac{n}{2}, p_{11})$ contains no cliques of order $\alpha_1 \log n$. However, in the rare cases when $\mathbb{G}(\frac{n}{2}, p_{11})$ (viewed as a subgraph of $\mathbb{G}(n, W)$) does contain such a clique, there are typically

many ways of extending it on the Ω_2-part by $\alpha_2 \log n$ additional vertices, thus inflating $\mathbf{E}[Y_n^{\alpha_1,\alpha_2}]$ substantially.

The above suggests a correction for (7) in that we should range only over those histograms f for which $\Gamma(g,W) \geqslant 0$ for all histograms $g \in \mathrm{Box}(f)$. Note also that the *necessity* of testing the admissibility condition over all sub-histograms of f has a clear combinatorial interpretation: If cliques with a given histogram typically appear, then for each given sub-histogram cliques with that sub-histogram must appear (just because a subset of a clique again induces a clique).

Now, after all the arguing why (7) should seem wrong, let us explain why it is actually all right. We show in Lemma 7.2 that for any histogram f attaining the supremum in (7) we automatically have that all sub-histograms are admissible (recall that a histogram h is admissible for a graphon W if $\Gamma(h,W) \geqslant 0$). If the supremum is not attained then for each histogram f almost attaining the supremum, we have $\Gamma(g,W) \gtrapprox 0$ for all sub-histograms g, which is sufficient for the argument.

6.3. **Failure of turning the above heuristics into a rigorous argument.** There are two types of errors that we introduced in the above argument. Firstly, the "little imprecisions" when we replaced a sum by its maximal term (such as in (33)) or when we used the $\lessapprox$-symbol. Each such step introduces an error of $o(1)$ to the quantities $\frac{\log X_n}{\log^2 n}$ and $\frac{\log Y_n^{\alpha_1,\alpha_2}}{\log^2 n}$. That means, that actually we can only conclude that

$$\mathbf{E}[Y_n^{\alpha_1,\alpha_2}]^2 = \exp\big(o(\log^2 n)\big)\,\mathbf{E}[(Y_n^{\alpha_1,\alpha_2})^2]\,,$$

which is too crude for the second moment argument to work.

Secondly, the notion of a "set of vertices following a certain histogram" makes sense only in the stochastic block model, but not when we have a finite set of vertices in an uncountable probability space. Let us jump ahead and note that in the rigorous proof in Section 8 we, in a sense, are able to make use of histograms in the continuous setting. Namely, Lemma 3.4 allows us to discretize a given graphon in an appropriate sense, after which it does make sense to talk about histograms.

Let us remark that for the stochastic block model the first issue (which is the only in that case) can be dealt with by pedestrian calculations, thus yielding a routine proof of Theorem 2.5 for the special class of stochastic block models.

7. Tools for the proof of Theorem 2.5

In this section we prepare tools for the lower bound in Theorem 2.5. In Section 7.1 we state and prove Lemma 7.2 which asserts that if f^* is a histogram almost attaining the supremum in (7) then $\Gamma(f,W)$ is almost positive for all subhistograms $f \leqslant f^*$. The need for this lemma was motivated in Section 6.2. In Section 7.2 we introduce a new graphon

parameter $\xi(W)$. This parameter is motivated by controlling the second moment of the number of cliques of a given size. All the work in Section 7 steers towards deriving the two main results of this section, Lemma 7.7 and Lemma 7.9. The former lemma asserts that each graphon W contains a subgraphon U with $\xi(U) \approx \frac{1}{\kappa(W)}$. The latter asserts that $\omega(\mathbb{G}(n,U)) \gtrsim \frac{1}{\xi(U)} \cdot \log n$. These two lemmas combine easily to give the proof of the lower bound in Theorem 2.5 (as is shown in Section 8).

7.1. **Subhistograms of optimal histograms are admissible.** The main result of this section, Lemma 7.2, tells us that if f^* is a histogram almost attaining the supremum in (7) then $\Gamma(f,W)$ is almost positive for all subhistograms $f \leqslant f^*$. We showed that this particular case of the (false, in general) Dream Lemma is needed for the second moment to work. The proof of Lemma 7.2 is technical, building on Lemma 7.1. It turns out that in those situations when the supremum in (7) is attained, Lemma 7.2 has a much shorter (but conceptually the same) proof. We offer this simplified statement in Lemma A.1 in the Appendix.

Lemma 7.1. *Suppose that W is an arbitrary graphon with $0 < \operatorname{ess\,inf} W \leqslant \operatorname{ess\,sup} W < 1$. Then there is a constant $K > 0$ depending only on the graphon W such that the following holds: Let g be an arbitrary histogram admissible for W and let $\delta \in (0,1)$. Suppose that $a \in (0,1)$ and that $g = g' + g''$ for some non-trivial histograms g' and g'' such that $\|g'\|_1 < \|g\|_1 - \delta$. Then either $\Gamma(g') \geqslant -a$, or there exists a histogram g^* which is admissible for W and for which we have*

$$\|g^*\|_1 \geqslant \|g\|_1 + K\delta^3 a^{\frac{5}{2}} \,. \tag{39}$$

Proof. Since we shall work exclusively with the graphon W, we write $\Gamma(\cdot)$ as a shortcut for $\Gamma(\cdot, W)$. Let us write $m^- = \operatorname{ess\,inf} W$ and $m^+ = \operatorname{ess\,sup} W$.

Let us fix numbers $a, \delta \in (0,1)$ and a decomposition $g = g' + g''$ of g into non-trivial histograms g', g'' such that $\|g'\|_1 < \|g\|_1 - \delta$. For $\varepsilon_1 \in (0,1)$ and $\varepsilon_2 > 0$, let us write $g^*(\varepsilon_1, \varepsilon_2) = (1-\varepsilon_1)g' + (1+\varepsilon_2)g''$. Let us also write

$$\begin{aligned}
A &= \|g'\|_1 \,, \\
B &= \|g''\|_1 \,, \\
C &= -\frac{1}{2}\int_{(x,y)\in\Omega^2} g'(x)g'(y)\log W(x,y)\ \mathsf{d}(\nu^2) \,, \\
D &= -\frac{1}{2}\int_{(x,y)\in\Omega^2} g''(x)g''(y)\log W(x,y)\ \mathsf{d}(\nu^2), \\
E &= -\int_{(x,y)\in\Omega^2} g'(x)g''(y)\log W(x,y)\ \mathsf{d}(\nu^2) \,,
\end{aligned} \tag{40}$$

and note that $A, B, C, D, E > 0$. We have $\Gamma(g') = A - C$ and $\Gamma(g) = A + B - C - D - E$. There is nothing to prove when $\Gamma(g') \geqslant -a$. Thus, assume otherwise. Then

$$A < C - a \,. \tag{41}$$

Upper-bounding C by $\frac{1}{2}\|g'\|_1^2 \log(1/m^-)$ and using that $C > a$, we get

$$\|g'\|_1 > \sqrt{\frac{2a}{\log\left(1/m^-\right)}} \,. \tag{42}$$

For each $\varepsilon_1 \in (0,1)$ and $\varepsilon_2 > 0$, the difference $\Gamma(g^*(\varepsilon_1, \varepsilon_2)) - \Gamma(g)$ can be expressed as

$$(1-\varepsilon_1)A + (1+\varepsilon_2)B - (1-\varepsilon_1)^2 C - (1+\varepsilon_2)^2 D - (1-\varepsilon_1)(1+\varepsilon_2)E - (A + B - C - D - E)$$

$$= \varepsilon_1(-A + 2C + E) + \varepsilon_2(B - 2D - E) - \varepsilon_1^2 C - \varepsilon_2^2 D + \varepsilon_1\varepsilon_2 E.$$

In particular, if $\varepsilon_2 = (1+\beta)\frac{A}{B}\varepsilon_1$ (where $\varepsilon_1 \in (0,1)$ and $\beta > 0$ will be determined later) then we have

(43)

$$\begin{aligned}
&\Gamma\left(g^*(\varepsilon_1, (1+\beta)\tfrac{A}{B}\varepsilon_1)\right) - \Gamma(g) \\
=&\varepsilon_1\left(2C + E - \tfrac{2AD}{B} - \tfrac{AE}{B}\right) + \varepsilon_1 \cdot \varepsilon_1\left(-C - \tfrac{A^2 D}{B^2} + \tfrac{AE}{B} - 2\tfrac{A^2}{B^2}\beta D + \beta\tfrac{AE}{B} - \tfrac{A^2}{B^2}\beta^2 D\right) + \\
&\quad + \varepsilon_1\beta(A - \tfrac{2AD}{B} - \tfrac{AE}{B}) \\
\geqslant&\varepsilon_1 \underbrace{\left(2C + E - \tfrac{2AD}{B} - \tfrac{AE}{B}\right)}_{\texttt{T1}} - \varepsilon_1 \cdot \underbrace{\varepsilon_1\left(C + \tfrac{A^2 D}{B^2} + 2\tfrac{A^2}{B^2}\beta D + \tfrac{A^2}{B^2}\beta^2 D\right)}_{\texttt{T2}} - \varepsilon_1 \underbrace{\beta(\tfrac{2AD}{B} + \tfrac{AE}{B})}_{\texttt{T3}} \,.
\end{aligned}$$

Let us expand the term `T1`.

$$\begin{aligned}
2C + E - \frac{2AD}{B} - \frac{AE}{B} &\overset{(41)}{>} 2A - \frac{2AD}{B} - \frac{AE}{B} + 2a \\
&> 2\frac{A}{B}(B - D - E) + 2a \\
&> 2\frac{A}{B}\left(B - D - E + (A - C)\right) + 2a \\
&= 2\frac{A}{B}\Gamma(g) + 2a \\
&\geqslant 2a.
\end{aligned}$$

Now, set $\varepsilon_1 = \frac{a}{4}\min(\frac{1}{C}, \frac{B^2}{2A^2 D})$ and $\beta = \min(1, \frac{aB}{4AD}, \frac{aB}{2AE})$. Routine calculations give that each of the terms `T2` and `T3` is smaller than a. Plugging the bounds above in (43), we get

$$\Gamma\left(g^*(\varepsilon_1, (1+\beta)\tfrac{A}{B}\varepsilon_1)\right) \geqslant \Gamma(g) \geqslant 0. \tag{44}$$

We have

$$\begin{aligned} C &= \frac{1}{2}\int_{(x,y)\in\Omega^2} g'(x)g'(y)\log\left(1/W(x,y)\right)\ \mathsf{d}(\nu^2) \\ &\leqslant \frac{1}{2}\int_{(x,y)\in\Omega^2} g(x)g(y)\log\left(1/m^-\right)\ \mathsf{d}(\nu^2) \leqslant \frac{1}{2}\kappa(W)^2\log\frac{1}{m^-} \end{aligned}$$

and similarly $D \leqslant \frac{1}{2}\kappa(W)^2\log\left(1/m^-\right)$ and $E \leqslant \kappa(W)^2\log\left(1/m^-\right)$. We also have

$$\frac{B}{A} = \frac{\|g-g'\|_1}{\|g'\|_1} \geqslant \frac{\delta}{\|g\|_1-\delta} \geqslant \frac{\delta}{\kappa(W)-\delta} \geqslant \frac{\delta}{\kappa(W)}\ .$$

Therefore

(45)

$$\begin{aligned} \varepsilon_1\beta &\geqslant \min\left(\frac{a}{4C}, \frac{aB^2}{8A^2D}, \frac{a^2B}{16ACD}, \frac{a^2B^3}{32A^3D^2}, \frac{a^2B}{8ACE}, \frac{a^2B^3}{16A^3DE}\right) \\ &\geqslant a^2\min\left(\frac{1}{4C}, \frac{B^2}{8A^2D}, \frac{B}{16ACD}, \frac{B^3}{32A^3D^2}, \frac{B}{8ACE}, \frac{B^3}{16A^3DE}\right) \\ &\geqslant a^2\frac{1}{\log^2\left(1/m^-\right)}\min\left(\frac{\log\left(1/m^-\right)}{2\kappa(W)^2}, \frac{\delta^2\log\left(1/m^-\right)}{4\kappa(W)^4}, \frac{\delta}{4\kappa(W)^5}, \frac{\delta^3}{8\kappa(W)^7}, \frac{\delta}{4\kappa(W)^5}, \frac{\delta^3}{8\kappa(W)^7}\right) \\ &\geqslant a^2\delta^3\frac{1}{\log^2\left(1/m^-\right)}\min\left(\frac{\log\left(1/m^-\right)}{2\kappa(W)^2}, \frac{\log\left(1/m^-\right)}{4\kappa(W)^4}, \frac{1}{4\kappa(W)^5}, \frac{1}{8\kappa(W)^7}, \frac{1}{4\kappa(W)^5}, \frac{1}{8\kappa(W)^7}\right)\ . \end{aligned}$$

It follows that

$$\begin{aligned} \|g^*\left(\varepsilon_1,\varepsilon_2\right)\|_1 &= (1-\varepsilon_1)A + (1+(1+\beta)\tfrac{A}{B}\varepsilon_1)B = \|g\|_1 + \beta\varepsilon_1\|g'\|_1 \\ &\overset{(42),(45)}{>} \|g\|_1 + a^{\frac{5}{2}}\delta^3\min\left(\frac{\log\left(1/m^-\right)}{2\kappa(W)^2}, \frac{\log\left(1/m^-\right)}{4\kappa(W)^4}, \frac{1}{4\kappa(W)^5}, \frac{1}{8\kappa(W)^7}, \frac{1}{4\kappa(W)^5}, \frac{1}{8\kappa(W)^7}\right)\sqrt{\frac{2}{\log^5\left(1/m^-\right)}}. \end{aligned}$$

This finishes the proof. □

Lemma 7.2. *Suppose that W is an arbitrary graphon with $0 < \operatorname{ess\,inf} W \leqslant \operatorname{ess\,sup} W < 1$. Then there exists a number $\gamma_0 > 0$ and a function $q : (0,\gamma_0) \to \mathbb{R}_+$ with $\lim_{\gamma\searrow 0} q(\gamma) = 0$ such that the following holds. Let f^* be an admissible histogram for W and let $\gamma \in (0,\gamma_0)$. Suppose that $\|f^*\|_1 \geqslant \kappa(W) - \gamma$. Then for every $f \in \operatorname{Box}(f^*)$ we have $\Gamma(f,W) \geqslant -q(\gamma)$.*

Proof. Let K be the constant from Lemma 7.1. Set γ_0 so that $\gamma_0 < K^2$. Suppose that $\gamma \in (0,\gamma_0)$ and f^* is an admissible histogram with $\|f^*\|_1 \geqslant \kappa(W) - \gamma$.

We write $\Gamma(\cdot)$ as a shortcut for $\Gamma(\cdot, W)$.

Let $f \in \mathrm{Box}(f^*)$ be non-trivial. Suppose first that $\|f\|_1 \geqslant \|f^*\|_1 - \sqrt[9]{\gamma}$. Using (8) we get

$$\begin{aligned}\Gamma(f) &= \int_{x\in\Omega} f(x)\,\mathsf{d}(\nu) + \frac{1}{2}\int_{(x,y)\in\Omega^2} f(x)f(y)\log W(x,y)\,\mathsf{d}(\nu^2)\\ &\geqslant \int_{x\in\Omega} f^*(x)\,\mathsf{d}(\nu) - \sqrt[9]{\gamma} + \frac{1}{2}\int_{(x,y)\in\Omega^2} f^*(x)f^*(y)\log W(x,y)\,\mathsf{d}(\nu^2)\\ &= \Gamma(f^*) - \sqrt[9]{\gamma} \geqslant -\sqrt[9]{\gamma}\,.\end{aligned} \tag{46}$$

Suppose next that $\|f\|_1 < \|f^*\|_1 - \sqrt[9]{\gamma}$. We apply Lemma 7.1 to

$$g = f^*, \quad g' = f, \quad g'' = f^* - f, \quad \delta = \sqrt[9]{\gamma}, \quad a = \frac{\sqrt[5]{\gamma}}{K^{\frac{2}{5}}}\,.$$

Then there is no histogram g^* admissible for W such that $\|g^*\|_1 \geqslant \|f^*\|_1 + K\delta^3 a^{\frac{5}{2}}$ as the right hand side equals $\|f^*\|_1 + \gamma^{\frac{5}{6}} > \|f^*\|_1 + \gamma \geqslant \kappa(W)$. So the lemma tells us that

$$\Gamma(f) \geqslant -\frac{\sqrt[5]{\gamma}}{K^{\frac{2}{5}}}\,. \tag{47}$$

Combining (46) with (47), it suffices to define the function q by

$$q(\gamma) = \max\left(\sqrt[9]{\gamma}, \frac{\sqrt[5]{\gamma}}{K^{\frac{2}{5}}}\right)\,, \quad q > 0\,,$$

since then it is clear that $\lim_{\gamma\searrow 0} q(\gamma) = 0$. □

7.2. **The graphon parameter $\xi(\cdot)$.** In Section 6.2 we outlined why the second moment argument for counting cliques should go through. (Recall that the second moment argument is needed to prove the lower bound in Theorem 2.5, which is the more difficult half of the statement.) For the actual execution of this step, however, we need to introduce a new graphon parameter. This parameter is a version of the cut norm with an exotic scaling. Given an arbitrary graphon W represented on a probability space Ω we define

$$\xi(W) = \sup_{B\subseteq\Omega,\nu(B)>0} \frac{1}{2\nu(B)}\int_{(x,y)\in B\times B} \log(1/W(x,y))\,\mathsf{d}(\nu^2)\,. \tag{48}$$

The key feature of $\xi(W)$, which we prove in Lemma 7.9, is that the second moment argument for counting cliques of order almost $\frac{1}{\xi(W)}\log n$ works. More precisely, in the proof of Lemma 7.9, we set up a random variable Y which essentially counts the number of individually-weighted cliques of the said order.[5] We show that $\mathbf{E}[Y^2] \approx \mathbf{E}^2[Y]$. This allows us to conclude that there must be at least one clique of such an order.

Of course, Lemma 7.9 itself is not enough to establish the lower bound in Theorem 2.5: we need to connect the new quantity $\xi(W)$ to the original quantity $\kappa(W)$. Given

[5]The weighting of the particular cliques is a technical but important subtlety, see (58) below.



FIGURE 2. The scheme of the proofs of Lemmas 7.9 and 7.7.

Lemma 7.9 described above, we would hope that $\kappa(W) = \frac{1}{\xi(W)}$. Unfortunately, in general, we only have $\kappa(W) \geqslant \frac{1}{\xi(W)}$, see Fact 7.3. Not all is lost though. In Lemma 7.7 we prove that every graphon W contains a subgraphon U for which $\frac{1}{\xi(U)} > \kappa(W) - \varepsilon$ (here, $\varepsilon > 0$ is arbitrarily small). After picking a subgraphon U for which $\frac{1}{\xi(U)} \approx \kappa(W)$ we continue with the proof of the lower bound in Theorem 2.5 as follows. We prove in Lemma 7.9 that asymptotically almost surely $\omega(\mathbb{G}(n,U)) \gtrsim \frac{1}{\xi(U)} \log n$. As described in (10), the above combination of Lemma 7.7 and Lemma 7.9 will conclude the desired proof of the lower bound in Theorem 2.5.

Now, let us state and prove the already advertised Fact 7.3. This fact will not be needed in our proof of Theorem 2.5. However, since it is so basic we record it here.

Fact 7.3. *Let W be an arbitrary graphon on a probability space Ω, and let $U = W\restriction_{A\times A}$ be a subgraphon obtained by restricting W to a set A of positive measure. Then $\frac{1}{\xi(U)} \leqslant \kappa(W)$.*

Proof. By considering the set $B = A$ in (48) we see that

$$\xi(U) \geqslant \frac{1}{2}\int_{(x,y)\in A^2} \log(1/W(x,y))\, \mathsf{d}(\nu_A^2) \overset{(9)}{=} \frac{1}{2}\frac{1}{\nu(A)^2}\int_{(x,y)\in A^2} \log(1/W(x,y))\, \mathsf{d}(\nu^2) \overset{(16)}{\geqslant} \frac{1}{\kappa(W)}\,.$$

□

In the rest of this section we prove Lemmas 7.9 and 7.7. The paths towards these lemmas are shown in Figure 2.

For the next lemma, note that if G is a finite graph then the value of $\xi(W_G)$ does not depend on the particular representation W_G of G.

Lemma 7.4. *Suppose that $c \in (0,1]$. Let W be a graphon with $\operatorname{ess\,inf} W \geqslant c$ and G an edge-weighted complete graph with all edge-weights $w(i,j)$ in the interval $[c,1]$. Consider the "negative logarithms of W and G", that is, an L^∞-graphon $W'(x,y) := \log(1/W(x,y))$ and a weighted graph G' with $V(G') = V(G)$ and weight function $w'(i,j) = \log(1/w(i,j))$. Then for an arbitrary $\gamma \in (0,1]$ we have*

$$|\xi(W) - \xi(W_G)| \leqslant \max\left(\tfrac{\gamma}{2}\log\tfrac{1}{c}, \tfrac{1}{\gamma}\delta_\square(W',G')\right)\,.$$

Proof. We shall prove the upper bound only for $\xi(W) - \xi(W_G)$. The upper bound on $\xi(W_G) - \xi(W)$ is done completely analogously. Suppose that W is represented on a probability space Ω. Looking at definition (48), we need to provide an upper bound

$$\underbrace{\frac{1}{\nu(A)} \int_{(x,y)\in A^2, x<y} \log\left(1/W(x,y)\right) \, \mathsf{d}(\nu^2)}_{\texttt{S1}} - \underbrace{\xi(W_G)}_{\texttt{S2}} \tag{49}$$

for each set $A \subseteq \Omega$ of positive measure. If $\nu(A) \leqslant \gamma$ then the integral is over a set of measure at most $\frac{1}{2}\nu^2(A) \leqslant \frac{\gamma}{2}\nu(A)$. Thus, the term S1 can be bounded from above by $-\frac{\gamma}{2}\log(\operatorname{ess\,inf} W) \leqslant \frac{\gamma}{2}\log\frac{1}{c}$, as needed.

Suppose next that $\nu(A) > \gamma$. Suppose first that $\delta_\square(W', G') = 0$. Using the invertible measure preserving maps from (12), we know that for each $\varepsilon > 0$ there exists a graphon representation $W_{G'}$ of G' on Ω such that $d_\square(W', W_{G'}) < \varepsilon\nu(A)$. Then

$$\begin{aligned}
&\frac{1}{\nu(A)} \int_{(x,y)\in A^2, x<y} \log\left(1/W(x,y)\right) \, \mathsf{d}(\nu^2) - \frac{1}{\nu(A)} \int_{(x,y)\in A^2, x<y} W_{G'}(x,y) \, \mathsf{d}(\nu^2) \\
&= \frac{1}{2\nu(A)} \left(\int_{(x,y)\in A^2} \left(W'(x,y) - W_{G'}(x,y)\right) \, \mathsf{d}(\nu^2) \right) \\
&\leqslant \frac{1}{\nu(A)} \cdot d_\square(W', W_{G'}) \leqslant \varepsilon \,,
\end{aligned}$$

as was needed.

Suppose next that $\delta_\square(W', G') > 0$. Using the invertible measure preserving maps from (12), we know that for each $\varepsilon > 0$ there exists a graphon representation $W_{G'}$ of G' on Ω such that $d_\square(W', W_{G'}) < (1+\varepsilon)\delta_\square(W', W_{G'})$. We shall fix such a representation $W_{G'}$ for $\varepsilon = \frac{\nu(A)}{\gamma} - 1$. Then (49) can be bounded from above by

$$\begin{aligned}
&\frac{1}{\nu(A)} \int_{(x,y)\in A^2, x<y} \log\tfrac{1}{W(x,y)} \, \mathsf{d}(\nu^2) - \frac{1}{\nu(A)} \int_{(x,y)\in A^2, x<y} W_{G'}(x,y) \, \mathsf{d}(\nu^2) \\
&= \frac{1}{2\nu(A)} \left(\int_{(x,y)\in A^2} \left(W'(x,y) - W_{G'}(x,y)\right) \, \mathsf{d}(\nu^2) \right) \\
&\leqslant \frac{1}{\nu(A)} \cdot d_\square(W', W_{G'}) \leqslant \frac{1}{\gamma} \cdot \delta_\square(W', W_{G'}) \,,
\end{aligned}$$

as was needed. □

Lemma 7.5. *Suppose that $c \in (0,1]$. Let W be a graphon with $\operatorname{ess\,inf} W \geqslant c$. Suppose that a sequence of integers $(k_n)_{n=1}^\infty$ has the property that $\sqrt{\log n} \leqslant k_n \leqslant \sqrt[3]{n}$. Suppose that $\varepsilon > 0$ is arbitrary.*

In the weighted random graph $G \sim \mathbb{H}(n,W)$ consider the family $\mathcal{H}$ of all sets $X \subseteq V(G)$ of size k_n which have the property that $|\xi(W)-\xi(W_{G[X]})| \geqslant \varepsilon$. Then asymptotically almost surely (as $n \to +\infty$) we have that $|\mathcal{H}| \leqslant \varepsilon\binom{n}{k_n}$.

Actually, the assertion of Lemma 7.5 is violated only with probability at most $\exp(-\frac{n}{2\log n})$, as can be seen from the proof of Lemma 7.5. We shall not need this refinement, though. For the proof of Lemma 7.5 we shall need the following well-known fact which we include here for the reader's convenience.

Fact 7.6. *Let us place m balls independently at random into one of n bins. If $n \geqslant m^3$ then with probability at least $1-2n^{-1/3}$ each bin contains at most one ball.*

Proof. Let us first bound the probability that one distinguished ball is placed into a bin which contains some other balls. Recall that for each $n \geqslant 2$,

$$1-\tfrac{1}{n} \geqslant \exp(-\tfrac{2}{n})\ . \tag{50}$$

The mentioned probability is exactly

$$1-(1-\tfrac{1}{n})^{m-1} \overset{(50)}{\leqslant} 1-\exp(-\tfrac{2(m-1)}{n}) \leqslant 1-\exp(-\tfrac{2m}{n}) \leqslant 1-\exp(-2n^{-2/3}) \leqslant 2n^{-2/3}\ .$$

The claim then follows by summing this error probability over all $m \leqslant n^{1/3}$ balls. □

Proof of Lemma 7.5. Let Ω be the probability space underlying W. Let $W' = \log 1/W$ be the negative logarithm of W. Note that W' is bounded from above by $\log 1/c$. Sampling the random graph $G \sim \mathbb{H}(n,W)$ can be naturally coupled with sampling a random graph $G' \sim \mathbb{H}(n,W')$. So, for the first part of the argument, we shall analyze the graph G'.

Suppose first that n is fixed. Corollary 3.5 implies that with probability at least $1-\exp(-\frac{n}{2\log n}) = 1-o(1)$ we have $\delta_\square(G',W') \leqslant \frac{20\log 1/c}{\sqrt{\log n}}$. We shall prove the statement for each weighted graph G' satisfying this property (provided that n is sufficiently large). That means that we assume that G' is fixed, and G is its exponentiated version. In particular, all the probabilistic calculations below are only with respect to later randomized steps. Let $\mathcal{K}$ be the family of all subsets X of $V(G') = V(G)$ of size k_n for which $\delta_\square(G',G'[X]) \geqslant \frac{20\log 1/c}{\sqrt{\log k_n}}$.

Consider the graphon representation $W_{G'}$ of G' represented on a partition $A_1 \dot\cup A_2 \dot\cup \ldots \dot\cup A_n = \Omega$. Sample the graph $H \sim \mathbb{H}(k_n, W_{G'})$. If we condition on the event $\mathcal{E}$ that the k_n representatives of the vertices of H in the sampling procedure were selected from pairwise distinct "bins" $A_1 \dot\cup A_2 \dot\cup \ldots \dot\cup A_n$ then H is a uniformly random subgraph of G' of order k_n. Fact 7.6 gives that $\mathbf{P}[\mathcal{E}] \geqslant 1-2n^{-1/3}$. Thus,

$$\mathbf{P}\left[\delta_\square(H,W_{G'}) \geqslant \tfrac{20\log 1/c}{\sqrt{\log k_n}}\right] \geqslant \mathbf{P}\left[\ \delta_\square(H,W_{G'}) \geqslant \tfrac{20\log 1/c}{\sqrt{\log k_n}} \mid \mathcal{E}\right] \mathbf{P}[\mathcal{E}] \geqslant \frac{|\mathcal{K}|}{\binom{n}{k_n}}(1-2n^{-1/3})\ .$$

Another application of Corollary 3.5 gives that $\mathbf{P}[\delta_\square(H, W_{G'}) \geqslant \frac{20\log 1/c}{\sqrt{\log k_n}}] < \exp(-\frac{k_n}{2\log k_n})$. Thus, we get (for n sufficiently large) that $|\mathcal{K}| \leqslant \varepsilon\binom{n}{k_n}$. So, the lemma will follow provided that we prove that $\mathcal{H} \subseteq \mathcal{K}$, which we prove next.

Indeed, let $X \notin \mathcal{K}$ be an arbitrary vertex set of size k_n. Then $\delta_\square(G'[X], W') \leqslant \delta_\square(G'[X], G') + \delta_\square(W_{G'}, W') \leqslant \frac{20\log 1/c}{\sqrt{\log k_n}} + \frac{20\log 1/c}{\sqrt{\log n}} \leqslant \frac{40\log 1/c}{\sqrt{\log k_n}}$. Then Lemma 7.4 tells us that for each $\gamma \in (0,1)$,

$$|\xi(W) - \xi(W_{G[X]})| \leqslant \max\left(\tfrac{\gamma}{2}\log 1/c, \frac{1}{\gamma}\cdot\frac{40\log 1/c}{\sqrt{\log k_n}}\right) .$$

We take $\gamma = 1/\sqrt[4]{\log k_n}$, and see that the right-hand side is, for large enough n, smaller than ε. This proves that $X \notin \mathcal{H}$ and consequently concludes the lemma. □

Our next two lemmas are crucial in proving the lower bound in Theorem 2.5. The first lemma, Lemma 7.7, tells us that in every graphon W there exists a subgraphon U of W for which we have $\frac{1}{\xi(U)} \gtrapprox \kappa(W)$. The second lemma, Lemma 7.9, then tells us that in $\mathbb{G}(n,U)$ we can typically find cliques of order almost $\frac{1}{\xi(U)}\log n$.

Lemma 7.7. *Suppose that $W : \Omega^2 \to [0,1]$ is a graphon with $0 < \operatorname{ess\,inf} W \leqslant \operatorname{ess\,sup} W < 1$. Then for every $\varepsilon > 0$ there exists a set $A \subseteq \Omega$ of positive measure such that for the subgraphon $U = W\restriction_{A\times A}$ we have $\frac{1}{\xi(U)} \geqslant \kappa(W) - \varepsilon$.*

Proof. Let us write $m = \log(1/\operatorname{ess\,inf} W)$. Consider the number $\gamma_0 > 0$ and the function $q : (0,\gamma_0) \to \mathbb{R}_+$ given by Lemma 7.2 for the graphon W.

Let $\delta > 0$ be fixed such that $\delta\kappa(W) < \gamma_0$. We use (16) to find a set A of positive measure such that

$$(1+\delta)\frac{2}{\kappa(W)} \geqslant \frac{1}{\nu(A)^2}\int_{(x,y)\in A^2} \log(1/W(x,y))\,\mathsf{d}(\nu^2) . \tag{51}$$

Consider now the subgraphon $U = W\restriction_{A\times A}$ on the probability space A endowed with the measure ν_A.

We now turn to obtaining the bound on $\xi(U)$. To this end we want to control each term in (48).

Claim 7.7.1. *Suppose that $B \subseteq A$ is an arbitrary set of positive measure. We have*

$$\frac{1}{1+\delta}\cdot\frac{1}{\nu_A(B)}\int_{(x,y)\in B^2}\log(1/W(x,y))\mathsf{d}(\nu_A^2) \leqslant \frac{2}{\kappa(W)} + \sqrt{\frac{q(\delta\kappa(W))}{(1-\delta)\kappa(W)}}\cdot m . \tag{52}$$

Proof of Claim 7.7.1. In the following, we abbreviate $q = q(\delta\kappa(W))$. Suppose first that

$$\nu_A(B) < \sqrt{\frac{q}{(1-\delta)\kappa(W)}} .$$

Then

$$\frac{1}{\nu_A(B)} \int_{(x,y)\in B^2} \log(1/W(x,y))\, \mathsf{d}(\nu_A^2) \leqslant \nu_A(B)m < \sqrt{\frac{q}{(1-\delta)\kappa(W)}} \cdot m \,,$$

as needed.

So, it remains to consider the case

$$\nu_A(B) \geqslant \sqrt{\frac{q}{(1-\delta)\kappa(W)}} \,. \tag{53}$$

Let $c > 0$ be maximum such that $\Gamma(c \cdot \mathbf{1}_A, W) \geqslant 0$. That is, we have $\Gamma(c \cdot \mathbf{1}_A, W) = 0$ which can be rewritten using (8) as $0 = c\nu(A) + c^2 \frac{1}{2} \int_{(x,y)\in A^2} \log W(x,y)\, \mathsf{d}(\nu^2)$. Thus,

$$c = \frac{2\nu(A)}{\int_{(x,y)\in A^2} \log(1/W(x,y))\, \mathsf{d}(\nu^2)} \,. \tag{54}$$

Now the assumption of Proposition 3.8 is satisfied by (51) (after a change of W on a null set).[6] The "more precisely" part of this proposition tells us that

$$c \geqslant (1-\delta)\frac{\kappa(W)}{\nu(A)} \,. \tag{55}$$

Consider now the function $f = c \cdot \mathbf{1}_A$. We have $\|f\|_1 \geqslant (1-\delta)\kappa(W)$. Thus Lemma 7.2 tells us that $\Gamma(g, W) \geqslant -q$ for each subhistogram g of f. Let us apply this to the function $g := c \cdot \mathbf{1}_B$. Then

$$\Gamma(g, W) = c\nu(B) - \frac{1}{2}c^2 \int_{(x,y)\in B^2} \log(1/W(x,y))\, \mathsf{d}(\nu^2) \geqslant -q \,,$$

yielding

$$\nu(B) \geqslant \frac{1}{2}c \int_{(x,y)\in B^2} \log(1/W(x,y))\, \mathsf{d}(\nu^2) - \frac{q}{c} \overset{(54)}{=} \nu(A)\frac{\int_{(x,y)\in B^2} \log(1/W(x,y))\, \mathsf{d}(\nu^2)}{\int_{(x,y)\in A^2} \log(1/W(x,y))\, \mathsf{d}(\nu^2)} - \frac{q}{c} \,.$$

[6] Proposition 3.8 was formulated only for graphons represented on the unit interval. However, we can use Fact 3.1 to represent W on the unit interval, and then we can apply Proposition 3.8.

This can be rewritten as

$$\begin{aligned}
&\frac{1}{\nu(A)}\int_{(x,y)\in A^2}\log(1/W(x,y))\,\mathsf{d}(\nu^2)\\
&\quad\geqslant \frac{1}{\nu(B)}\int_{(x,y)\in B^2}\log(1/W(x,y))\,\mathsf{d}(\nu^2)-\frac{q\int_{(x,y)\in A^2}\log(1/W(x,y))\,\mathsf{d}(\nu^2)}{c\nu(A)\nu(B)}\\
&\quad\geqslant \frac{1}{\nu(B)}\int_{(x,y)\in B^2}\log(1/W(x,y))\,\mathsf{d}(\nu^2)-\frac{qm\nu(A)}{c\nu(B)} \qquad (56)\\
&\quad\overset{(55)}{\geqslant} \frac{1}{\nu(B)}\int_{(x,y)\in B^2}\log(1/W(x,y))\,\mathsf{d}(\nu^2)-\frac{qm\nu(A)}{(1-\delta)\kappa(W)\nu_A(B)}\\
&\quad\overset{(53)}{\geqslant} \frac{1}{\nu(B)}\int_{(x,y)\in B^2}\log(1/W(x,y))\mathsf{d}(\nu^2)-\sqrt{\frac{q}{(1-\delta)\kappa(W)}}\cdot m\cdot\nu(A)\,.
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{2}{\kappa(W)}&\overset{(51)}{\geqslant}\frac{1}{1+\delta}\frac{1}{\nu(A)^2}\int_{(x,y)\in A^2}\log(1/W(x,y))\,\mathsf{d}(\nu^2)\\
&\overset{(56)}{\geqslant}\frac{1}{1+\delta}\frac{1}{\nu(A)\nu(B)}\int_{(x,y)\in B^2}\log(1/W(x,y))\,\mathsf{d}(\nu^2)-\frac{1}{1+\delta}\sqrt{\frac{q}{(1-\delta)\kappa(W)}}\cdot m\\
&\overset{(9)}{\geqslant}\frac{1}{1+\delta}\frac{1}{\nu_A(B)}\int_{(x,y)\in B^2}\log(1/W(x,y))\,\mathsf{d}(\nu_A^2)-\sqrt{\frac{q}{(1-\delta)\kappa(W)}}\cdot m\,,
\end{aligned}$$

as required. □

The term $\sqrt{\frac{q(\delta\kappa(W))}{(1-\delta)\kappa(W)}}$ in (52) does not depend on the choice of the set A. Thus, it tends to zero as we let $\delta\searrow 0$. We conclude that for $\delta>0$ sufficiently small, if we select A as in (51), we have

$$\frac{1}{2\nu_A(B)}\int_{(x,y)\in B^2}\log(1/W(x,y))\mathsf{d}(\nu_A^2)\leqslant\frac{1+\delta}{\kappa(W)}+\frac{1+\delta}{2}\cdot\sqrt{\frac{q(\delta\kappa(W))}{(1-\delta)\kappa(W)}}\cdot m$$

for each $B\subseteq A$ of positive measure. By (48), we have

$$\xi(U)\leqslant\frac{1+\delta}{\kappa(W)}+\frac{1+\delta}{2}\cdot\sqrt{\frac{q(\delta\kappa(W))}{(1-\delta)\kappa(W)}}\cdot m\,.$$

If $\delta>0$ is sufficiently small then the right hand side (which tends to $\frac{1}{\kappa(W)}$ as $\delta\searrow 0$) is smaller than $\frac{1}{\kappa(W)-\varepsilon}$, and so

$$\frac{1}{\xi(U)}\geqslant\kappa(W)-\varepsilon\,,$$

as was needed. □

We shall need the following observation.

Fact 7.8. *Let G be a finite edge-weighted complete graph with vertex set $[n]$ whose symmetric weight function $w : V(G)^2 \to [0,1]$ puts weight 1 on all self-loops. Then for every $C \subseteq [n]$ and for the graphon representation W_G of G we have*

$$\prod_{\substack{i,j\in C\\ i<j}} w(i,j) \geqslant \exp(-\xi(W_G)n|C|) \,.$$

Proof. Suppose that W_G is a representation of G on the unit interval I. Suppose further that each vertex $i \in [n]$ is represented by an interval $D_i \subseteq I$ (c.f. Remark 3.3), and that for each $1 \leqslant i < j \leqslant n$ the interval D_i lies to the left of the interval D_j. Of course, such a change of representation of W_G does not change $\xi(W_G)$.

The case $C = \emptyset$ is trivial, so assume $C \neq \emptyset$. Consider the set $X = \bigcup_{i\in C} D_i$. Definition (48) gives that

$$\xi(W_G) \geqslant \frac{1}{\lambda(X)} \int_{(x,y)\in X^2, x<y} \log(1/W_G(x,y)) \, \mathsf{d}(\lambda^2).$$

Let us split the integration above according to the partition $(D_i \times D_j)_{i,j\in[n]}$. We can neglect the terms for which $i = j$ since then the integrand is $\log(1/1) = 0$. So, suppose that $i < j$. Then for each $x \in D_i$, $y \in D_j$ we have $W_G(x,y) = w(i,j)$. Thus, in this case $\int_{(x,y)\in D_i\times D_j, x<y} \log(1/W_G(x,y)) \, \mathsf{d}(\lambda^2) = \frac{1}{n^2} \log(1/w(i,j))$. We conclude that

$$\xi(W_G) \geqslant \frac{n}{|C|} \cdot \frac{1}{n^2} \sum_{\substack{i,j\in C\times C\\ i<j}} \log(1/w(i,j))$$

The lemma follows after exponentiation. □

Lemma 7.9. *Suppose that W is a graphon with* $\operatorname{ess\,inf} W > 0$. *Suppose that $\alpha < 1/\xi(W)$. Then asymptotically almost surely, $\mathbb{G}(n,W)$ contains a clique of order $\alpha \log n$.*

Proof. Choose $\delta > 0$ such that $\alpha(\xi(W) + \delta) < 1$ and let us write

$$\gamma = 1 - \alpha(\xi(W) + \delta) \,. \tag{57}$$

Let $H \sim \mathbb{H}(n,W)$. Set $k := \alpha \log n$. Let $\mathcal{A}$ be be the family of all sets $A \subseteq V(H)$ of size k for which $|\xi(W) - \xi(W_{H[A]})| < \delta$. Lemma 7.5 tells us that with high probability, the graph H has the property that $|\mathcal{A}| \geqslant (1-\delta)\binom{n}{k}$. Condition on this event, and fix a realization of the weighted graph H with a weight function $w : \binom{V(H)}{2} \to [0,1]$ having the above property.

We shall now obtain from H an unweighted graph G by including each edge ij with probability $w(i,j)$. It is our task to show that with high probability, G contains a clique of order k. (Recall that this probability is only with respect to obtaining G from H.)

For each $A \in \mathcal{A}$ set up the indicator X_A of the event that $G[A]$ is a clique. Define

$$Y_A := \frac{X_A}{\mathbf{E}[X_A]} \tag{58}$$

(note that the denominator is not zero because $\operatorname{ess\,inf} W > 0$). Let $Y = \sum_{A \in \mathcal{A}} Y_A$. To conclude the proof, we want to prove that for each $\varepsilon > 0$ (which we now consider fixed), we have

$$Y > 0 \text{ with probability at least } 1 - \varepsilon, \tag{59}$$

provided that n is sufficiently large.

We have $\mathbf{E}[Y_A] = 1$ for each $A \in \mathcal{A}$, and consequently $\mathbf{E}[Y] = |\mathcal{A}|$, which tends to infinity with $n \to +\infty$. Below we shall prove that $\mathbf{E}[Y^2] \leqslant (1+\varepsilon)\mathbf{E}^2[Y]$, which will establish (59) via the usual second-moment argument.

Claim 7.9.1. *We have* $\mathbf{E}[Y^2] < (1+\varepsilon)\mathbf{E}^2[Y]$.

Proof of Claim 7.9.1. For $\ell = 0, 1, \dots, k$, let us write

$$M_\ell = \sum_{\substack{A,B \in \mathcal{A} \\ |A \cap B| = \ell}} \mathbf{E}[Y_A Y_B] \, .$$

Then we have $\mathbf{E}[Y^2] = \sum_{\ell=0}^{k} M_\ell$. So, it is our goal to bound each of the numbers M_ℓ. We have

$$M_0 \leqslant \binom{n}{k}^2 . \tag{60}$$

For $\ell > 0$ we have[7]

$$M_\ell = \sum_{\substack{A,B \in \mathcal{A} \\ |A \cap B| = \ell}} \mathbf{E}[Y_A Y_B] = \sum_{C \in \binom{V(G)}{\ell}} \sum_{\substack{A,B \in \mathcal{A} \\ A \cap B = C}} \frac{\mathbf{E}[X_A X_B]}{\mathbf{E}[X_A]\mathbf{E}[X_B]} \, .$$

Given two sets $A, B \in \mathcal{A}$, it is easy to see that

$$\mathbf{E}[X_A X_B] = \frac{\mathbf{E}[X_A]\mathbf{E}[X_B]}{\prod_{ij \in \binom{A \cap B}{2}} w(i,j)} \, .$$

[7] the calculations below are also valid in the case $\ell = 0$, but we shall not use them in that case

Thus,

(61)

$$\begin{aligned}
M_\ell &= \sum_{C\in\binom{V(G)}{\ell}} \sum_{\substack{A,B\in\mathcal{A}\\ A\cap B=C}} \prod_{ij\in\binom{C}{2}} w^{-1}(i,j) \\
&\boxed{\text{Fact 7.8 applied on } H[A] \text{ and } C} \quad \leqslant \sum_{C\in\binom{V(G)}{\ell}} \sum_{\substack{A,B\in\mathcal{A}\\ A\cap B=C}} \exp\left((\xi(W)+\delta)k\ell\right) \\
&\leqslant \binom{n}{\ell \mid k-\ell \mid k-\ell} \exp\left((\xi(W)+\delta)k\ell\right) \\
&= \frac{n!(n-k)!^2}{(n-2k+\ell)!n!^2}\cdot\frac{k!^2}{\ell!(k-\ell)!^2}\cdot\binom{n}{k}^2 \exp\left((\xi(W)+\delta)k\ell\right) \\
&\boxed{n \text{ is sufficiently large}} \quad \leqslant \left(\frac{2}{n}\right)^\ell \cdot k^{2\ell}\cdot\binom{n}{k}^2 \exp\left((\xi(W)+\delta)k\ell\right) \\
&= \left(\frac{2k^2\exp\left((\xi(W)+\delta)k\right)}{n}\right)^\ell \cdot \binom{n}{k}^2 \\
&\overset{(57)}{\leqslant} \left(2^\ell k^{2\ell}\exp\left(-\gamma\ell\log n\right)\right)\cdot\binom{n}{k}^2 \\
&\boxed{n \gg k} \quad \leqslant \exp\left(-\tfrac{\gamma\ell\log n}{2}\right)\cdot\binom{n}{k}^2 .
\end{aligned}$$

Recall that $\mathbf{E}[Y] = |\mathcal{A}| \geqslant (1-\delta)\binom{n}{k}$. Thus,

$$\begin{aligned}
\frac{\mathbf{E}[Y^2]}{\mathbf{E}^2[Y]} &= \frac{\sum_{\ell=0}^k M_\ell}{\mathbf{E}^2[Y]} \\
&\overset{(60)}{\leqslant} \frac{1}{(1-\delta)^2} + \frac{\sum_{\ell=1}^k M_\ell}{(1-\delta)^2\binom{n}{k}^2} \\
&\overset{(61)}{\leqslant} \frac{1}{(1-\delta)^2} + \frac{1}{(1-\delta)^2}\sum_{\ell=1}^k \exp\left(-\tfrac{\gamma\ell\log n}{2}\right) \\
&\leqslant \frac{1}{(1-\delta)^2}\sum_{\ell=0}^\infty \exp\left(-\tfrac{\gamma\ell\log n}{2}\right) .
\end{aligned}$$

Note that the last expression is a geometric series, and its quotient $\exp\left(-\frac{\gamma\log n}{2}\right)$ tends to 0 as $n\to\infty$. Therefore the sum of the series tends to 1. Thus for sufficiently large n and for sufficiently small $\delta > 0$ we get $\frac{\mathbf{E}[Y^2]}{\mathbf{E}^2[Y]} < 1+\varepsilon$, as was needed. □

Claim 7.9.1 tells us that $\mathbf{Var}[Y] \leqslant \varepsilon\mathbf{E}^2[Y]$. Therefore, (59) follows from Chebyshev's Inequality. □

8. Proof of Theorem 2.5

Let $c = \operatorname{ess\,inf} W$. Suppose that W is represented on the unit interval $I = (0,1)$ equipped with the Lebesgue measure λ. Let us replace the value of W in every point $(x,y) \in (0,1)^2$ that is not a point of approximate continuity by c. This is a change of measure zero by Fact 3.7. In particular, $\kappa(W)$ does not change, nor does the distribution of the model $\mathbb{G}(n,W)$.

8.1. **Upper bound.** Let $\varepsilon \in (0, \kappa(W)/4)$ be arbitrary. Let n be sufficiently large. We want to show that a.a.s. $G \sim \mathbb{G}(n,W)$ contains no clique of order $k = (\kappa(W)+\varepsilon)\log n$. Let $X_n(G)$ count such cliques. We have

$$\mathbf{E}[X_n(G)] = \int_{(x_1,x_2,\ldots,x_n)\in I^n} \sum_{A\in\binom{[n]}{k}} \prod_{i,j\in A, i<j} W(x_i,x_j)\, \mathsf{d}(\lambda^n)\ .$$

This summation has $\binom{n}{k} < n^k = \exp\left((\kappa(W)+\varepsilon)\log^2 n\right)$ terms. By (15), each of these terms is bounded by $P_k^{\frac{k(k-1)}{2}}$ where $\lim_{k\to\infty} P_k = \exp\left(-\frac{2}{\kappa(W)}\right)$. So if n is sufficiently large then each term is bounded by

$$\exp\left(-\frac{2}{\kappa(W)}\cdot\frac{k(k-1)}{2}+\varepsilon\right)\ .$$

Thus,

$$\begin{aligned}\mathbf{E}[X_n(G)] &\leqslant \exp\left((\kappa(W)+\varepsilon)\log^2 n - \frac{2}{\kappa(W)}\cdot\frac{k(k-1)}{2}+\varepsilon\right)\\ &= \exp\left(\left(-\varepsilon-\frac{\varepsilon^2}{\kappa(W)}\right)\log^2 n + \left(1+\frac{\varepsilon}{\kappa(W)}\right)\log n+\varepsilon\right)\to 0\ ,\end{aligned}$$

as n goes to infinity. Markov's inequality concludes the proof.

8.2. **Lower bound.** We shall assume that $\operatorname{ess\,sup} W < 1$. Let us justify this step. Suppose that W is an arbitrary graphon. We can then take a sequence of graphons $W_1, W_2, \ldots$, where $W_j = \min(W, 1-\frac{1}{j})$ (pointwise). Then (16) tells us that $\kappa(W_j)\to\kappa(W)$ (even in the case $\kappa(W)=+\infty$). Thus, it suffices to prove a lower bound for each of the graphons W_j.

Let $\varepsilon > 0$ be arbitrary. We apply Lemma 7.7 to find a set $A \subseteq \Omega$ of positive measure such that for the subgraphon $U = W\restriction_{A\times A}$ we have $\frac{1}{\xi(U)} \geqslant \kappa(W)-\varepsilon$. Lemma 7.9 then tells us that asymptotically almost surely, $\omega(\mathbb{G}(n,U)) \geqslant (\kappa(W)-2\varepsilon)\log n$. Since there is a coupling of $G=\mathbb{G}(n,W)$ and $G'=\mathbb{G}(\frac{\lambda(A)n}{2},U)$ such that G asymptotically almost surely contains a copy of G', we obtain that (cf. (10)),

$$\omega(\mathbb{G}(n,W)) \geqslant (\kappa(W)-3\varepsilon)\log n \text{ asymptotically almost surely.}$$

Since $\varepsilon > 0$ was arbitrary, this completes the proof of Theorem 2.5.

9. Concluding remarks

Our concluding remarks concern possibilities of extending the main result, Theorem 2.5.

9.1. **Sharpening the results.** As mentioned in Section 1, Matula, Grimmett and McDiarmid proved for $p \in (0,1)$ an asymptotic concentration of $\omega(\mathbb{G}(n,p))$ on two consecutive values for which they provided an explicit formula. It is possible that when, say, $0 < \operatorname{ess\,inf} W \leqslant \operatorname{ess\,sup} W < 1$, then $\omega(\mathbb{G}(n,W))$ is asymptotically concentrated on two consecutive values.

9.2. **Sparse inhomogeneous random graphs.** Let us look at our set of problems for $\mathbb{G}(n, p_n \cdot W)$, where $p_n \to 0$, i.e., the model introduced in Section 1.1. Note that Remark 2.7 is no longer valid: the problem of maximum clique and maximum independent set in $\mathbb{G}(n, p_n \cdot W)$ is genuinely different. It turns out that the more interesting problem is that of the independent set. For the Erdős–Rényi random graph $\mathbb{G}(n,p_n)$, the problem of determining the independence number is essentially solved by the above mentioned work [21, 11], and by the work of Frieze [10] down to the range $p_n \gg \frac{1}{n}$. Note that the regime $p_n \ll \frac{1}{\sqrt{n}}$ is more subtle as the second moment argument does not work, and indeed Frieze's contribution was in establishing concentration of the count of large independent sets by alternative means. The regime $p_n = C/n$ seems to require methods from statistical physics. In the related model of random regular graphs, these methods have already provided an answer [7].

It would be of interest to see whether the methods we developed in this paper can give an answer also for the independence number in sparser inhomogeneous random graphs. It seems that the two core ingredients of our proof, Lemma 3.4 and the second moment argument do have sparse counterparts.

- The sparse counterpart to Lemma 3.4 is [3, Theorem 2.14] which says that if $p_n \gg \frac{1}{n}$ then the sequence $\frac{1}{p_n} \cdot \mathbb{G}(n, p_n \cdot W)$ (here, the factor $\frac{1}{p_n}$ in front of the random graph $\mathbb{G}(n, p_n \cdot W)$ denotes edge weighting; this is the natural way to deal with the scaling in this situation) converges to W in the cut-distance almost surely.
- Our second moment argument is complicated but it builds on the seminal work [21, 11] which works down to the range $p_n = \Theta(\frac{1}{\sqrt{n}})$. Thus, at least when $W \in L^\infty(\Omega^2)$, our methods possibly extend to this range. The situation when $W \in L^p(\Omega^2)$ for some general p is probably more subtle.

  Of course, one might ask whether the methods used by Frieze [10] could be extended to the inhomogeneous setting, thus possibly giving results even for $\frac{1}{n} \ll p_n < \frac{1}{\sqrt{n}}$. This however goes beyond the scope of this paper.

Acknowledgments

JH thanks Lutz Warnke for suggesting the idea of the proof of Theorem 2.2. We thank the referees for their helpful comments.

The contents of this publication reflects only the authors' views and not necessarily the views of the European Commission of the European Union.

References

[1] B. Bollobás, Ch. Borgs, J. Chayes, and O. Riordan. Percolation on dense graph sequences. *Ann. Probab.*, 38(1):150–183, 2010.

[2] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, 31(1):3–122, 2007.

[3] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao. An L^p theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions, 2014. arXiv:1401.2906.

[4] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[5] D. Conlon, J. Fox, and B. Sudakov. An approximate version of Sidorenko's conjecture. *Geom. Funct. Anal.*, 20(6):1354–1366, 2010.

[6] L. Devroye and N. Fraiman. Connectivity of inhomogeneous random graphs. *Random Structures Algorithms*, 45(3):408–420, 2014.

[7] J. Ding, A. Sly, and N. Sun. Maximum independent sets on random regular graphs, 2013.

[8] M. Fabian, P. Habala, P. Hájek, V. Montesinos, and V. Zizler. *Banach space theory*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011. The basis for linear and nonlinear analysis.

[9] N. Fraiman and D. Mitsche. The diameter of inhomogeneous random graphs, 2015. arXiv:1510.08882.

[10] A. M. Frieze. On the independence number of random graphs. *Discrete Math.*, 81(2):171–175, 1990.

[11] G. R. Grimmett and C. J. H. McDiarmid. On colouring random graphs. *Math. Proc. Cambridge Philos. Soc.*, 77:313–324, 1975.

[12] H. Hatami. Graph norms and Sidorenko's conjecture. *Israel J. Math.*, 175:125–150, 2010.

[13] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: first steps. *Social Networks*, 5(2):109–137, 1983.

[14] M. Kang, Ch. Koch, and A. Pachón. The phase transition in multitype binomial random graphs. *SIAM J. Discrete Math.*, 29(2):1042–1064, 2015.

[15] A. S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[16] J. L. X. Li and B. Szegedy. On the logarithimic calculus and Sidorenko's conjecture. Preprint (arXiv:1107.1153).

[17] L. Lovász. Subgraph densities in signed graphons and the local Simonovits-Sidorenko conjecture. *Electron. J. Combin.*, 18(1):Paper 127, 21, 2011.

[18] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[19] L. Lovász and B. Szegedy. Szemerédi's Lemma for the analyst. *J. Geom. and Func. Anal*, 17:252–270, 2007.

[20] L. Lovász and B. Szegedy. Regularity partitions and the topology of graphons. In *An irregular mind*, volume 21 of *Bolyai Soc. Math. Stud.*, pages 415–446. János Bolyai Math. Soc., Budapest, 2010.

[21] D. W. Matula. The largest clique size in a random graph. Technical report, Department of Computer Science, Southern Methodist University, 1976.

[22] M. Molloy and B. Reed. *Graph colouring and the probabilistic method*, volume 23 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2002.

[23] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[24] A. Sidorenko. A correlation inequality for bipartite graphs. *Graphs Combin.*, 9(2):201–204, 1993.

[25] M. Simonovits. Extremal graph problems, degenerate extremal problems, and supersaturated graphs. In *Progress in graph theory (Waterloo, Ont., 1982)*, pages 419–437. Academic Press, Toronto, ON, 1984.

[26] T. S. Turova. The largest component in subcritical inhomogeneous random graphs. *Combin. Probab. Comput.*, 20(1):131–154, 2011.

[27] R. van der Hofstad. Critical behavior in inhomogeneous random graphs. *Random Structures Algorithms*, 42(4):480–508, 2013.

Appendix A. A simplified version of Lemma 7.2

Here we provide a weaker version of Lemma 7.2 which deals with the case that the supremum in (7) is attained. While this version is not sufficient for our purposes we decided to offer it to the reader because its proof is based on the same idea yet is stripped off technicalities.

Lemma A.1. *Suppose that W is an arbitrary graphon, and suppose that f^* is an admissible histogram for W for which $\|f^*\|_1 = \kappa(W)$. Then every subhistogram of f^* is admissible for W.*

Proof. The statement follows immediately from Claim A.1.1 (which is a simplified version of Lemma 7.1).

We abbreviate $\Gamma(\cdot, W)$ as $\Gamma(\cdot)$. Also, when we say "admissible", we mean with respect to W.

Claim A.1.1. *Assume that g is an arbitrary admissible histogram. Suppose further that $g = g' + g''$ for some histograms g' and g''. Then either g' is admissible, or there exist $\varepsilon_1, \varepsilon_2 \in (0,1)$ such that for $g^* = (1-\varepsilon_1)g' + (1+\varepsilon_2)g''$ we have that g^* is admissible, and $\|g^*\|_1 > \|g\|_1$.*

Proof of Claim A.1.1. For $\varepsilon_1, \varepsilon_2 \in (0,1)$, let us write $g^*(\varepsilon_1, \varepsilon_2) = (1-\varepsilon_1)g' + (1+\varepsilon_2)g''$. We define numbers A, B, C, D, and E as in (40). Note that $A, B, C, D, E \geqslant 0$. For any $\varepsilon_1, \varepsilon_2 \in (0,1)$, the difference $\Gamma(g^*(\varepsilon_1, \varepsilon_2)) - \Gamma(g)$ can be expressed as

$$(1-\varepsilon_1)A + (1+\varepsilon_2)B - (1-\varepsilon_1)^2 C - (1+\varepsilon_2)^2 D - (1-\varepsilon_1)(1+\varepsilon_2)E - (A + B - C - D - E)$$
$$= \varepsilon_1(-A + 2C + E) + \varepsilon_2(B - 2D - E) - \varepsilon_1^2 C - \varepsilon_2^2 D + \varepsilon_1\varepsilon_2 E.$$

In particular, if $\varepsilon_2 = \frac{A}{B}\varepsilon_1$ then we have

(62)
$$\Gamma\left(g^*\left(\varepsilon_1, \frac{A}{B}\varepsilon_1\right)\right) - \Gamma(g) = \varepsilon_1\left(2C + E - \frac{2AD}{B} - \frac{AE}{B}\right) + \varepsilon_1^2\left(-C - \frac{A^2D}{B^2} + \frac{AE}{B}\right).$$

Now let us assume that $\Gamma(g') < 0$, i.e. $A < C$. Then we have

(63)
$$\begin{aligned} 2C + E - \frac{2AD}{B} - \frac{AE}{B} &> 2A - \frac{2AD}{B} - \frac{AE}{B} \\ &> 2\frac{A}{B}(B - D - E) \\ &> 2\frac{A}{B}\left(B - D - E + (A - C)\right) \\ &= 2\frac{A}{B}\Gamma(g) \\ &\geqslant 0. \end{aligned}$$

By (63), the right hand side of (62) is a quadratic expression (in the variable ε_1) with a positive linear coeficient. Therefore there is $\varepsilon_1 > 0$ (which we fix now) small enough such that $\varepsilon_1, \frac{A}{B}\varepsilon_1 \in (0,1)$ and

(64)
$$\Gamma\left(g^*\left(\varepsilon_1, \frac{A}{B}\varepsilon_1\right)\right) > \Gamma(g)\,(\geqslant 0)\,.$$

Since the function $(\varepsilon_1, \varepsilon_2) \mapsto \Gamma\left(g^*\left(\varepsilon_1, \varepsilon_2\right)\right)$ is obviously continuous, we can find $\varepsilon_2 \in \left(\frac{A}{B}\varepsilon_1, 1\right)$ such that $\Gamma\left(g^*\left(\varepsilon_1, \varepsilon_2\right)\right)$ is still nonnegative. Then we have

$$\|g^*\left(\varepsilon_1, \varepsilon_2\right)\|_1 = (1 - \varepsilon_1)A + (1 + \varepsilon_2)B > (1 - \varepsilon_1)A + \left(1 + \frac{A}{B}\varepsilon_1\right)B = A + B = \|g\|_1.$$

This finishes the proof. □

□

Institute of Mathematics, Czech Academy of Sciences. Žitná 25, 110 00, Praha, Czech Republic. The Institute of Mathematics of the Czech Academy of Sciences is supported by RVO:67985840.

The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, 182 07 Prague, Czech Republic. With institutional support RVO:67985807. *Address at time of submission:* Institute of Mathematics, Czech Academy of Sciences. Žitná 25, 110 00, Praha, Czech Republic. The Institute of Mathematics of the Czech Academy of Sciences is supported by RVO:67985840.

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK

THE LOCAL LIMIT OF THE UNIFORM SPANNING TREE ON DENSE GRAPHS

JAN HLADKÝ, ASAF NACHMIAS, AND TUAN TRAN

Abstract. Let G be a connected graph in which almost all vertices have linear degrees and let $\mathcal{T}$ be a uniform spanning tree of G. For any fixed rooted tree F of height r we compute the asymptotic density of vertices v for which the r-ball around v in $\mathcal{T}$ is isomorphic to F. We deduce from this that if $\{G_n\}$ is a sequence of such graphs converging to a graphon W, then the uniform spanning tree of G_n locally converges to a multi-type branching process defined in terms of W.

As an application, we prove that in a graph with linear minimum degree, with high probability, the density of leaves in a uniform spanning tree is at least $e^{-1} - \mathsf{o}(1)$, the density of vertices of degree 2 is at most $e^{-1} + \mathsf{o}(1)$ and the density of vertices of degree $k \geqslant 3$ is at most $\frac{(k-2)^{k-2}}{(k-1)!e^{k-2}} + \mathsf{o}(1)$. These bounds are sharp.

1. Introduction

It is a classical fact [9, 11] that a uniformly chosen tree from the set of n^{n-2} trees on n vertices, viewed from an independently chosen uniform random vertex, is locally distributed as a $\mathsf{Poisson}(1)$ Galton–Watson branching process conditioned to live forever when n is large. One can use this result to find the distribution of a certain "local" structures. For instance, it follows that the degree distribution of a uniformly chosen vertex of a uniformly chosen tree on n vertices converges to the law of a $\mathsf{Poisson}(1) + 1$ random variable.

A uniformly chosen tree on n vertices is a uniform spanning tree (UST) of the complete graph on n vertices. Our goal in this paper is to explicitly describe the local structure of the UST of any dense graph or, equivalently, of a sequence of dense graphs converging to a given graphon. Let us first present this result.

Given a connected graph G we write $\mathcal{T}$ for a uniformly drawn spanning tree of G and $B_{\mathcal{T}}(v, r)$ for the graph-distance ball of radius r in $\mathcal{T}$ around the vertex $v \in V(G)$. Our goal is to describe the asymptotic distribution of $B_{\mathcal{T}}(X, r)$, viewed up to graph isomorphism, where X is a uniformly chosen random vertex of G. To that aim, let $\Omega = [0, 1]$ and μ be the Lebesgue measure on Ω. For a given graphon $W : \Omega^2 \to [0, 1]$ (see Section 1.1 for a brief introduction to graphons) and $\omega \in \Omega$ we write $\deg(\omega) = \int_{y\in\Omega} W(\omega, y)\mathrm{d}y$, and call this number the *degree* of ω. A graphon is called *nondegenerate* if for almost every $\omega \in \Omega$ we have $\deg(\omega) > 0$.

Let T be a fixed rooted tree with $\ell \geqslant 2$ vertices and of height $r \geqslant 1$. We write by Stab_T the set of graph automorphisms of T that preserve the root. In what follows we denote the vertices of T by the numbers $\{1, \dots, \ell\}$ such that the vertices $p, \dots, \ell$ are the vertices at height r of T and $p \in \{2, \dots, \ell\}$. Given a nondegenerate graphon W and a tree T as

above we define

(1)
$$\mathrm{Freq}(T;W) := \frac{1}{|\mathrm{Stab}_T|} \int\limits_{\omega_1,\dots,\omega_\ell} \exp\left(-\sum_{j=1}^{p-1} b_W(\omega_j)\right) \frac{\sum_{j=p}^{\ell}\deg(\omega_j)}{\prod_{j=1}^{\ell}\deg(\omega_j)} \prod_{(i,j)\in E(T)} W(\omega_i,\omega_j)\mathrm{d}\omega_1\cdots\mathrm{d}\omega_\ell\,,$$

where

(2)
$$b_W(\omega) = \int_{y\in\Omega} \frac{W(\omega,y)}{\deg(y)}\mathrm{d}\mu\,.$$

Theorem 1.1. *Let T be a fixed rooted tree as above, and $W:\Omega^2\to[0,1]$ be a nondegenerate graphon. Then for any $\varepsilon>0$ there exists $\xi=\xi(\varepsilon,W,T)>0$ such that if G is a connected simple graph on at least ξ^{-1} vertices that is ξ-close to W in the cut-distance, then with probability at least $1-\varepsilon$ a uniformly chosen spanning tree $\mathcal{T}$ of G satisfies*

$$\left|\mathbb{P}(B_{\mathcal{T}}(X,r)\cong T) - \mathrm{Freq}(T;W)\right| \leqslant \varepsilon\,,$$

where X is an independently and uniformly chosen vertex of G, and by $B_{\mathcal{T}}(X,r)\cong T$ we mean that between the two rooted trees there is a graph-isomorphism preserving the root.

Theorem 1.1 is a natural statement in the context of limits of graph sequences. It asserts that if $\{G_n\}$ is a sequence of connected graphs converging to a graphon W, then the UST of G_n converges locally to a certain multi-type branching process that is defined in terms of W. This interpretation involve two graph limit procedures: a dense graph limit [16, 7] to describe the limit of the dense graph sequence G_n, and sparse graph limit to describe the random limit of the UST of G_n, known as Benjamini–Schramm convergence [4]. We refer the reader to Sections 1.1 and 1.2 for an introduction to these limiting procedure, and begin by describing the limiting branching process.

Definition 1.2. *Given a nondegenerate graphon $W:\Omega^2\to[0,1]$ we define a multi-type branching process κ_W. The process has continuum many types that are either (anc,ω) or (oth,ω), where $\omega\in\Omega$. Here, "anc" stands for "ancestral" and "oth" stands for "other".*

(1) The initial particle (i.e., the root) has type (anc,ω), where the distribution of ω is μ.

(2) If a particle has type (oth,ω), then its progeny is $\{(\mathsf{oth},\omega_1),\dots,(\mathsf{oth},\omega_k))\}$ where $\{\omega_1,\dots,\omega_k\}$ are a Poisson point process on Ω with intensity $\frac{W(\omega,\omega')}{\deg(\omega')}$ at $\omega'\in\Omega$. In particular, k has distribution $\mathsf{Poisson}\,(b_W(\omega))$, where $b_W(\cdot)$ is defined in (2).

(3) If a particle has type (anc,ω) the progeny is $\{(\mathsf{anc},\omega_0),(\mathsf{oth},\omega_1),\dots,(\mathsf{oth},\omega_k))\}$ where $\{\omega_1,\dots,\omega_k\}$ are a Poisson point process on Ω with the same intensity as above and ω_0 is an independent new particle of Ω which is distributed according to the probability measure on Ω that has density $\frac{W(\omega,\cdot)}{\deg(\omega)}$.

We note that an ancestral vertex (i.e., of type anc) always has at least 1 progeny and since the initial particle is ancestral the process κ_W survives forever with probability 1. The following theorem is quickly deduced from Theorem 1.1 at Section 4.2.

Theorem 1.3. *Suppose that $\{G_n\}$ is a sequence of simple connected graphs of growing orders that converge to a nondegenerate graphon W in the cut-distance and let T_n be a UST of G_n. Then the sequence $\{T_n\}$ almost surely converge in the Benjamini–Schramm sense to κ_W.*

Since the L^1-norm (corresponding to the edit distance in graph theory) is finer than the cut-norm, we in particular obtain the following. Suppose that G is a large connected simple n-vertex graph with minimum degree $\Omega(n)$. Suppose that we add and/or delete $\mathsf{o}(n^2)$ edges in a way that the resulting graph G' stays connected. Then the structure of a typical UST of G and of G' is very similar in the Benjamini–Schramm sense. Even this statement seems to be new.

When G_n are a sequence of dense *regular* graphs, then the function $b_W : \Omega \to (0, \infty)$ defined in (2) is identically 1 and we obtain the following.

Corollary 1.4. *Suppose that $\{G_n\}$ is a sequence of simple connected regular graphs of growing orders that converge to a nondegenerate graphon W in the cut-distance and let T_n be a UST of G_n. Then the sequence $\{T_n\}$ almost surely converge in the Benjamini–Schramm sense to a $\mathsf{Poisson}(1)$ Galton–Watson branching process conditioned to live forever.*

Theorem 1.3 allows us to deduce several extremal properties of the UST on dense graphs. The number of vertices of degree $k \geqslant 1$ in the UST of the complete graph K_n is $(e^{-1}/(k-1)! + \mathsf{o}(1))n$. In fact, using Prüfer codes one can establish that the degree of the a vertex in the UST of K_n has distribution $1 + \mathsf{Bin}(n-2, \frac{1}{n}) \approx 1 + \mathsf{Poisson}(1)$.[1] Using Theorem 1.1 we are able to find the extremal values of the number of vertices of degree k in a general dense graph. In the following theorem we show that the complete graph (or any other *regular* dense graph) is the minimizer of the number of leaves and the maximizer of the number of vertices of degrees 2 and 3 among the class of dense graphs. Somewhat surprisingly, the maximizer for the number of vertices of degree $k \geqslant 4$ is a different dense graph (see Section 5).

Theorem 1.5. *For any $k \geqslant 1$ we denote by L_k the random variable counting the number of vertices of degree k in a UST of a simple connected graph G. For every $\varepsilon, \delta > 0$ there*

[1] Prüfer codes, see e.g. [19, page 245], provide a standard bijection between spanning trees of K_n and and words of length $n-2$ over the alphabet $V(K_n)$. A quick look at this bijection shows the number of occurrences of any letter $v \in V(K_n)$ in a Prüfer code is the degree of the vertex v in the corresponding spanning tree decreased by 1. Therefore, the degree of v in a uniform spanning tree has indeed distribution $1 + \mathsf{Bin}(n-2, \frac{1}{n})$.

exist numbers $n_0 \in \mathbb{N}$ *and* $\gamma > 0$ *such that the following holds: Whenever* G *is a graph on* $n \geqslant n_0$ *vertices with at least* $(1-\gamma)n$ *vertices of degrees at least* δn *then*

$$\mathbb{P}\big(L_1 \leqslant (e^{-1}-\varepsilon)n\big) \leqslant \varepsilon\,, \tag{3}$$

$$\mathbb{P}\big(L_2 \geqslant (e^{-1}+\varepsilon)n\big) \leqslant \varepsilon\,, \tag{4}$$

and for any $k \geqslant 3$ *we have*

$$\mathbb{P}\left(L_k \geqslant \left(\frac{1}{(k-1)!}\frac{(k-2)^{k-2}}{e^{k-2}}+\varepsilon\right)n\right) \leqslant \varepsilon\,. \tag{5}$$

We derive Theorem 1.5 in Section 5. The rest of this section is organized as follows. We first give the formal definitions and background of dense and sparse graph limits necessary for the reader to parse the statement of Theorems 1.1 and 1.3. Next we discuss some aspects of the theorem, such as the necessity of its assumptions. We end this section with a discussion of related results in the literature.

1.1. **Dense graph limits.** Graphons were introduced by Borgs, Chayes, Lovász, Sós, Szegedy, and Vesztergombi [16, 7] as limit objects to sequences of dense graphs. Here, we review basic facts and we refer the reader to [15, Part II] for a thorough treatment of the subject. A *graphon* is a symmetric Lebesgue measurable function $W : \Omega^2 \to [0,1]$, where Ω is any standard atomless probability space.[2] The underlying measure on Ω will be always denoted by μ.

Given a measurable function $U : \Omega^2 \to [-1,1]$ we define its *cut-norm* by

$$\|U\|_\square = \sup_{S,T}\left|\int_{x\in S}\int_{y\in T} U(x,y)\right|, \tag{6}$$

where S and T range over all measurable subsets of Ω. Now, we can define the key notion of *cut-distance*. For two graphons $W_1, W_1 : \Omega^2 \to [0,1]$, we define

$$\delta_\square(W_1,W_2) = \inf_{\varphi}\|W_1^\varphi - W_2\|_\square\,, \tag{7}$$

where $\varphi : \Omega \to \Omega$ ranges through all measure preserving automorphisms of Ω, and W_1^φ stands for a graphon defined by $W_1^\varphi(x,y) = W_1(\varphi(x),\varphi(y))$. Note that the definition of $\delta_\square(W_1,W_2)$ extends in a straightforward way if W_2 lives on some other standard atomless probability space Λ. In that case, φ ranges of all measure preserving bijections from Λ to Ω.

Suppose that G is an n-vertex graph. Then we can consider a *graphon representation* of G. To this end, partition an atomless standard probability space Ω into n sets, each set of measure $\frac{1}{n}$, $\Omega = \bigsqcup_{v\in V(G)}\Omega_v$. Then define a graphon W_G to be 1 on $\Omega_u \times \Omega_v$ for each edge $uv \in E(G)$, and 0 otherwise. Note that W_G is not unique as it depends on the choice

[2]Recall that all such probability spaces are isomorphic.

of the partition $\{\Omega_v\}$. With the notion of a graphon representation, we can define distance between a graph and a graphon. Namely, if $W : \Omega^2 \to [0,1]$ is a graphon and G is a graph, we define $\delta_\square(W,G) := \delta_\square(W,W_G)$. This definition does not depend on the choice of the representation W_G. We say that G is α-*close* to W if $\delta_\square(W,G) \leqslant \alpha$. Though through much of the paper, we shall work with loopless multigraphs (introduced in Section 2.1), we always restrict ourselves to graphs when representing as graphons, or when referring to the cut-distance.

Throughout the paper, all sets and functions are tacitly assumed to be measurable. Conversely all our constructions of auxiliary sets and functions are measurable, too.

We say that a sequence $(G_n)_n$ of simple graphs *converges* to a graphon W if and only if $\delta_\square(W,G_n) \to 0$. Now, it is possible to understand the first sentence of Theorem 1.3. The following statement, proved first in [16], is the core of the theory of graphons.

Theorem 1.6. *For every sequence of simple graphs of increasing orders there exists a subsequence which converges to a graphon.*

1.2. **Sparse graph limits.** Let $\mathcal{G}_\bullet$ denote the space of rooted locally finite graphs viewed up to root-preserving graph isomorphisms. That is, each element of $\mathcal{G}_\bullet$ is (G,ϱ) where G is a graph and ϱ is a vertex of it, and two such elements (G_1,ϱ_1) and (G_2,ϱ_2) are considered equivalent if and only if there exists a graph automorphism $\varphi : G_1 \to G_2$ such that $\varphi(\varrho_1) = \varrho_2$. Given a rooted graph (G,ϱ) and an integer $r \geqslant 1$ we write $B_G(\varrho,r)$ for the graph-distance ball of radius r around ϱ in G, that is, $B_G(\varrho,r) \in \mathcal{G}_\bullet$ is a finite graph rooted at ϱ on the set of vertices of graphs distance at most r from ϱ in G together with all the edges induced from G. There is a natural notion of a metric on $\mathcal{G}_\bullet$. The distance between two elements $(G_1,\varrho_1),(G_2,\varrho_2) \in \mathcal{G}_\bullet$ is defined to be 2^{-R} where $R \geqslant 0$ is the largest number such that there is a root-preserving graph isomorphism between $B_{G_1}(\varrho_1,R)$ and $B_{G_2}(\varrho_2,R)$. Having defined the metric we can consider probability spaces on $\mathcal{G}_\bullet$ with respect to the Borel σ-algebra.

We say that the law of a random element $(G,\varrho) \in \mathcal{G}_\bullet$ is the *Benjamini–Schramm limit* of a (possibly random) sequence of finite graphs G_n, if and only if, for any fixed integer $r \geqslant 1$ the random variable $B_{G_n}(\varrho_n,r)$ converges in distribution to $B_G(\varrho,r)$ where ϱ_n is an independently chosen uniform random vertex of G_n. In this case we say that the sequence G_n Benjamini–Schramm converges to (G,ϱ). Note that by putting $r = 1$ in the definition we deduce that in this convergence the degree of the random root ϱ_n must converge to the degree of ϱ which explains why this limiting procedure is best suited for sparse graphs. See further discussion in [4].

We have finished defining all the needed terminology required to parse Theorems 1.1 and 1.3.

1.3. **Necessity of the assumptions.** The assumptions in Theorem 1.3 are the minimal necessary. First, we obviously need the assumption that G_n are connected in order for spanning trees to exist.

Next we claim that the local structure of a uniform spanning tree cannot be determined from a degenerate graphon W. Indeed, suppose that a graphon W is given, and let Ω^0 be the elements of Ω that have zero degree W and $\Omega^+ = \Omega \backslash \Omega^0$. Assume that W is degenerate so that $\mu(\Omega^0) = \delta > 0$.

We can now construct two graph sequences that converge to W. We start with dense graphs G_n of size $(1-\delta)n$ that converge[3] to $W^+ = W_{|\Omega^+}$ and in the first sequence we attach to G_n a path of length δn at an arbitrary vertex and in the second sequence we attach δn edges arbitrary to an vertex of G_n creating δn new vertices of degree 1. It is clear that both sequences converge to W. However, the USTs on the two sequences have different Benjamini–Schramm limits. Indeed, let p_1 denote the probability that in κ_{W^+} the root is a leaf. Then the probability that a randomly chosen vertex is a leaf in the first sequence tends to $(1-\delta)p_1$ and in the second sequence this probability tends to $(1-\delta)p_1 + \delta$.

1.4. **Discussion.** Theorem 1.3 shows that the local structure of the UST is continuous on the space of dense graphs with the cut-metric (7) and describes this local structure explicitly. As mentioned earlier, the only instance in the literature of Theorem 1.3 that we are aware of is the case of the UST of the complete graph K_n. In this case Grimmett [9] showed that the limiting object is an infinite path upon which we to each vertex an independent $\mathsf{Poisson}(1)$ branching process — that is, a $\mathsf{Poisson}(1)$ branching process conditioned to survive forever. This is precisely κ_W where corresponding graphon W is just $W \equiv 1$.

The analogous continuity result for sparse graphs is also true, though in this case it is typically harder to describe explicitly the limiting object.

Theorem 1.7 ([1, Proposition 7.1])**.** *Suppose that $\{G_n\}$ is a Benjamini–Schramm convergent sequence of connected graphs and let T_n be a UST of G_n. Then there exists a random rooted tree (T, ϱ) such that T_n Benjamini–Schramm converges to (T, ϱ).*

The limiting object in the above theorem (T, ϱ) is the *wired uniform spanning forest* of the Benjamini–Schramm limit (G, ϱ) of the graphs $\{G_n\}$, see [18].

One can also ask whether the normalized number of spanning trees is continuous with respect to taking graph limits. Given a graph G, let $t(G)$ be the number of spanning trees in G. In the bounded-degree model, Lyons [17, Theorem 3.2] proved that $n^{-1} \log t(G)$

[3]Such a sequence can be obtained for example by taking typical inhomogeneous random graphs $\mathbb{G}(n, W^+)$, see [15, Lemma 10.16].

is continuous in the Benjamini–Schramm topology. In the dense model, the natural normalization of $t(G)$ is $n^{-1}t(G)^{1/n}$. For example when $G = K_n$ then by Cayley's formula, $n^{-1}t(G)^{1/n}$ tends to 1, and when G is a typical Erdős–Rényi random graph $\mathbb{G}(n,p)$ for $p \in (0,1)$ fixed, then $n^{-1}t(G)^{1/n}$ tends to p. However, one cannot expect that with no further assumptions continuity of this parameter with respect to the cut-metric will hold. Indeed, let G_n formed by a clique of order $n - \frac{n}{\sqrt{\log n}}$ and a path of length $\frac{n}{\sqrt{\log n}}$ attached to it at an arbitrary vertex. The sequence $\{G_n\}$ converges to the complete graphon $W \equiv 1$ but the number of spanning is substantially lower than for complete graphs; indeed, in this case it can easily by seen using Cayley's formula that $n^{-1}t(G_n)^{1/n}$ tends to 0.

However, when we impose a minimum degree condition on the graphs, we can infer the asymptotic normalized number of spanning trees from the limit graphon.

Theorem 1.8. *Let $\delta > 0$. Suppose that $G_1, G_2, \ldots$ is a sequence of simple connected graphs that converge to a graphon W. Suppose that the order of G_n is n and the minimum degree is at least δn. Then the number of spanning trees satisfies*

$$\lim_{n\to\infty} \frac{\sqrt[n]{t(G_n)}}{n} = \exp\left(\int_x \log(\deg_W(x))\right) .$$

□

Theorem 1.8 follows almost immediately from a result due to Kostochka [12] which states that if $1 < d_1 \leqslant d_2 \leqslant \ldots \leqslant d_n$ is the degree sequence of a simple connected graph G then for some absolute constant $C > 0$ we have

$$\frac{\prod_i d_i}{d_1^{(Cn\log d_1)/d_1}} \leqslant t(G) \leqslant \frac{\prod_i d_i}{n-1} . \tag{8}$$

To see that (8) yields Theorem 1.8, it is enough to recall that the degree distribution of a limit graphon is inherited from degree distribution of graphs that converge to it (see Lemma 5.1(1)). In the case of regular graphs, Theorem 1.8 can also be derived from [2, 8].

Lastly, let us mention a result of a similar flavor to ours in the context of percolation [5]. There, the authors show that the critical percolation probability of a dense graph is $\frac{1}{\lambda_n}$ where λ_n is the largest eigenvalue of the adjacency matrix. In particular it follows that if two dense graphs are close in the cut-metric, then this threshold is also close. They also describe the limiting local structure of bond percolation on dense graph in terms of a branching process on the limiting graphon. While there is some resemblance to the branching process of Theorem 1.3, they are quite different.

1.5. **About the proof of Theorem 1.1.** Our proof combines, for the first time to the best of our knowledge, two seemingly unrelated mathematical areas, Kirchhoff's electric

network theory and Szemerédi's regularity lemma-like graph partitioning techniques. These two are shown here to work seamlessly together.

Kirchhoff's theory of electric networks [10] allows to compute the probability that a given edge $e = xy$ is in a UST of a connected graph. This probability is precisely the *effective electric resistance* between x and y, when we consider the graph as an electric network and let current flow from x to y, see Section 3.1.1 and (24). Since there is an edge connecting x to y, this quantity is always a number in $[0, 1]$. This is the starting point of our proof.

Next we use partition theory (Section 2) to decompose our graph G into a bounded number of dense expanders so that different expanders of the decomposition are connected by $\mathsf{o}(n^2)$ edges. Heuristically, the UST of G is close the union of independent USTs on each of these dense expanders.

Thus, it is natural to study electric theory on dense expanders. It is intuitive (and easy to prove) that if x and y are two vertices in graph, then the effective resistance between x and y is at least $\frac{1}{\deg(x)} + \frac{1}{\deg(y)}$ since at the most efficient scenario the electric current splits equally from x to all its neighbors and arrives to y equally from all of y's neighbors. Of course this lower bound is not sharp — the graph could be the disjoint union of two large stars around x and y and an edge connecting x and y, in which case the resistance between x and y is 1.

However, when the graph is a dense expander, one can use random walks estimates and employ the fascinating and classical connection between random walks and electric networks, to deduce a corresponding approximate upper bound, see Corollary 3.3. The random walk estimate we prove (Lemma 3.2) states that if one starts a random walk on a dense expander from some vertex that is not x or y, then the probability that x is visited before y is $\frac{(1+\mathsf{o}(1))\deg(x)}{\deg(x)+\deg(y)}$.

It is now quite pleasant to observe that Rayleigh's monotonicity (21), which states that the electric resistance can only decrease by enlarging a network, shows that this upper bound on the resistance holds in each expander in the decomposition of G, and the matching lower bound holds for *most* edges of G since there are $\mathsf{o}(n^2)$ edges between components, see Lemma 3.3.

This explains why for most edges in G the probability that they are exhibited in the UST is the sum of the degree reciprocals. An iterative argument is presented in Section 3.3, employing the spatial Markov property of the UST (Proposition 3.1), to control the probability of events such as $B_{\mathcal{T}}(v, r) \cong T$. There are some delicate technicalities to overcome involving the "outside" effects of the decomposition. Once these are overcome, one reaches the discrete version $\mathrm{Freq}(T; G)$ of the parameter $\mathrm{Freq}(T; W)$ of Theorem 1.1 and we show that this parameter approximates the desired probability (Lemma 3.12). In Section 4 it is

shown that $\mathrm{Freq}(T;G)$ is close to $\mathrm{Freq}(T;W)$, its continuous counterpart, if G is close to W in the cut-distance, concluding the proof.

1.6. **Organisation of the paper.** We tried to write the paper so that it can be read by probabilists and graph theorists alike. For this reason we recall even concepts relatively well known to one of the communities in a pedestrian manner. Also, at places we try to convey an idea of a proof even when this idea is standard, but in only one of the two communities.

In Section 2 we introduce a suitable decomposition of dense graphs, into so-called linear expanders. In Section 3 we prove the discrete estimate approximating the probability of the event $\mathbb{P}(B_{\mathcal{T}}(X,r)=T)$ by the discrete parameter $\mathrm{Freq}(T;G)$. In Section 4 we show that $\mathrm{Freq}(T;G)$ and $\mathrm{Freq}(T;W)$ are close whenever G and W are close in the cut-metric. Lastly, in Section 5 we derive Theorem 1.5.

2. Decomposing dense graphs into linear expanders

2.1. **Dense expanders.** Informally, a graph is a dense expander if whenever a vertex set and its complement are of linear size (in the order of the graph) then there are quadratically many edges between these two parts. So, a primal example of a dense graph that is *not* an expander is a disjoint union of two cliques of order $\frac{n}{2}$ with a perfect matching connecting them.

We give our definition of expansion for *loopless multigraphs*. That is, self-loops are not allowed, and two vertices may be connected by several edges. The quantity $e(A,B)$ counts all ordered pairs ab that form an edge, $a\in A$ and $b\in B$, *including multiplicities*. Note that each edge with both endvertices in $A\cap B$ contributes twice to $e(A,B)$. For a vertex v and a vertex set A, we write $\deg(v,A):=e(\{v\},A)$. We write $\deg(v):=\deg(v,V)$, where V is the vertex set of the (multi)graph. Note that with these conventions, we have $2e(G)=\sum_{v\in V}\deg(v)$, and $\sum_{a\in A}\deg(a,B)=e(A,B)=\sum_{b\in B}\deg(b,A)$, for each vertex sets A and B.

Definition 2.1. *We say that a loopless multigraph H is a* γ-expander *if for each $U\subseteq V(H)$, we have $e(U,V(H)\backslash U)\geqslant\gamma|U|(v(H)-|U|)$.*

We will later use a simple observation. Removing edges from an expander can obviously render its expansion properties. However, if one removes edges touching only one vertex while leaving the degree of the vertex high, the expansion properties are not damaged by too much as the following simple proposition states.

Proposition 2.2. *Let G be a γ-expander loopless multigraph on m vertices and let $v\in V(G)$ be a vertex. Assume that the maximal number of edges between any two vertices in $V(G)\backslash\{v\}$ is at most $\ell\in\mathbb{N}$. Consider the graph G' obtained from G by erasing some set*

of edges emanating from v of size at most ℓm so that the degree of v in G' is at least γm. Assume also that $8\ell^2\gamma^{-2} \leqslant m$. Then G' is a $\frac{\gamma}{2}$-expander.

Proof. We need to prove that for any $U \subseteq V(G)$,

$$e_{G'}(U, V(G)\backslash U) \geqslant \frac{\gamma}{2}|U|(m - |U|)\,. \tag{9}$$

By symmetry it is enough to prove it when $|U| \leqslant m/2$. We proceed by considering two cases. In the first case we assume $|U| \geqslant 4\gamma^{-1}\ell$. Then, since we erased at most ℓm edges, we have

$$e_{G'}(U, V(G)\backslash U) \geqslant \gamma|U|(m - |U|) - \ell m \geqslant \frac{\gamma}{2}|U|(m - |U|)\,,$$

since in this case $\gamma|U|(m - |U|) \geqslant 2\ell m$.

In the second case we assume that $|U| \leqslant 4\gamma^{-1}\ell$. If $U = \{v\}$, then (9) follows since the degree of v in G' is at least γm. If $v \notin U$, then the maximum number of edges that could have been erased from $e_G(U, V(G)\backslash U)$ is at most $\ell|U|$, hence

$$e_{G'}(U, V(G)\backslash U) \geqslant \gamma|U|(m - |U|) - 4\ell^2\gamma^{-1} \geqslant \frac{\gamma}{2}|U|(m - |U|)\,,$$

since $\gamma|U|(m - |U|) \geqslant \gamma m$ and $8\ell^2\gamma^{-2} \leqslant m$. Lastly, if $v \in U$ and $|U| > 1$, then

$$e_{G'}(U, V(G)\backslash U) \geqslant e_G(U\backslash\{v\}, V(G)\backslash U \cup \{v\}) - \ell|U| \geqslant \frac{\gamma}{2}|U|(m - |U|)\,,$$

by the same logic as above, concluding the proof. □

2.2. **Expander decomposition of dense graphs.** The main result of this section, Theorem 2.7, asserts that each graph that is close to a nondegenerate graphon can be decomposed into a bounded number expanders that are almost isolated from each other. In Definition 2.6 below we describe the expander decomposition that we actually use. Let us recall that the need of such a decomposition (rather than a single expander) stems from examples such as that of a disjoint union of two cliques of order $\frac{n}{2}$ with a perfect matching connecting them mentioned at the beginning of Section 2.1. Passing to a limit we see that in the graphon perspective, the perfect matching vanishes and we are left with two components. Therefore, we now introduce graphon counterparts to the notion of graph connectivity and components, and give their basic properties.

Definition 2.3. *A graphon W on a ground space (Ω, μ) is* disconnected *if either $W = 0$ a.e. or there exists a subset $A \subseteq \Omega$ with $0 < \mu(A) < 1$ such that $W = 0$ a.e. on $A \times A^c$; otherwise W is* connected*.*

We shall require a result of Bollobás, Janson and Riordan [6, Lemma 5.17] which enables us to decompose a graphon into (at most) countably many connected components.

Lemma 2.4. *Let $W : \Omega \times \Omega \to [0,1]$ be a graphon. Then there exists a partition $\Omega = \bigcup_{i=0}^{N} \Omega_i$ into measurable subsets with $0 \leqslant N \leqslant \infty$ such that $\mu(\Omega_i) > 0$ for $i \geqslant 1$, the restriction of W to $\Omega_i \times \Omega_i$ is connected for each $i \geqslant 1$, and $W = 0$ a.e. on $(\Omega \times \Omega) \setminus \bigcup_{i=1}^{N} (\Omega_i \times \Omega_i)$.*

Bollobás, Borgs, Chayes and Riordan [5, Lemma 7] showed that connectivity implies an apparently stronger statement.

Lemma 2.5. *Let $W : \Omega^2 \to [0,1]$ be a connected graphon, and let $0 < \alpha < \frac{1}{2}$ be given. There is some constant $\beta = \beta(W, \alpha) > 0$ such that $\int_{A \times A^c} W \geqslant \beta$ for every measurable subset $A \subseteq \Omega$ with $\alpha \leqslant \mu(A) \leqslant \frac{1}{2}$.*

We can now give our definition of expander decomposition.

Definition 2.6. *Suppose that G is a loopless multigraph of order n. We say that $V(G) = V_0 \sqcup V_1 \sqcup \ldots \sqcup V_k$ is a* $(\gamma, \eta, \varepsilon)$-expander decomposition *if*

(G1) $|V_0| \leqslant \varepsilon n$,
(G2) *for each $i \in [k]$ we have that $e(V_i, V \setminus V_i) \leqslant \eta |V_i| n$,*
(G3) *for each $i \in [k]$ and each $U \subseteq V_i$, we have $e(U, V_i \setminus U) \geqslant \gamma |U| |V_i \setminus U|$.*

Theorem 2.7. *Suppose that $W : \Omega^2 \to [0,1]$ is a nondegenerate graphon. Then for every $\varepsilon, \eta > 0$ there exist positive constants $\gamma = \gamma(W, \varepsilon, \eta)$, $\xi = \xi(W, \varepsilon, \eta)$ and $n_0 = n_0(W, \varepsilon, \eta)$ such that if G is a graph with $v(G) > n_0$ and $\delta_\square(G, W) < \xi$ then G admits a $(\gamma, \eta, \varepsilon)$-expander decomposition.*

For the proof of Theorem 2.7, we shall need the following result. While we were not able to find an explicit reference, we consider this result folklore. To this end, we need the notion of subgraphons which we introduce now. Suppose that $W : \Omega^2 \to [0,1]$ is a graphon on a probability space (μ, Ω). Similarly to the graph case, for a set $\Lambda \subseteq \Omega$ we have the notion of a *subgraphon of W induced by Λ*. This is the restricted function $W[\Lambda] := W_{\restriction \Lambda \times \Lambda}$. In order for $W[\Lambda]$ to be a graphon, we always have to consider it together with the renormalized probability space $(\frac{\mu(\cdot)}{\mu(\Lambda)}, \Lambda)$.

Lemma 2.8. *Suppose that $W : \Omega^2 \to [0,1]$ is a nondegenerate graphon. Suppose that we have a partition $\Omega = \Omega^* \sqcup \bigsqcup_{i=1}^{k} \Omega_i$ such that for each $i \in [k]$ we have that W is zero almost everywhere on $\Omega_i \times (\Omega \setminus \Omega_i)$. Then for every $\lambda > 0$ there exists a number $\xi > 0$ so that we have the following. If G is an n-vertex graph with $\delta_\square(G, W) < \xi$ then there exists a partition $V(G) = \bigsqcup_{i=0}^{k} V_i$ such that for each $i \in [k]$,*

(a) $|V_0| = \mu(\Omega^*) n \pm \lambda n$, and $|V_i| = \mu(\Omega_i) n \pm \lambda n$,
(b) $e_G(V_i, V(G) \setminus V_i) < \lambda n^2$,
(c) $\delta_\square(G[V_i], W[\Omega_i]) < \lambda$.

Proof of Theorem 2.7. Suppose that we are given a graphon $W : \Omega^2 \to [0,1]$ and two parameters $\varepsilon, \eta > 0$. By Lemma 2.4, there exists a partition $\Omega = \bigsqcup_{i=1}^{K} \Omega_i$ (with $K \leqslant \infty$) into components of W. Let $k \in \mathbb{N}$ be such that $\mu\left(\bigcup_{i=0}^{k} \Omega_i\right) \geqslant 1 - \frac{\varepsilon}{4}$. Set $\Omega^* := \bigcup_{i=k+1}^{K} \Omega_i$. Let $\alpha = \min\left\{\frac{\varepsilon}{6k}, \min_{i\in[k]} \frac{\mu(\Omega_i)}{40}\right\}$. Lemma 2.5 shows the existence of a positive constant $\beta = \beta(W, \alpha)$ such that

$$\iint\limits_{A\times(\Omega_i\setminus A)} W \geqslant \beta \text{ for all } i \in [k] \text{ and all } A \subseteq \Omega_i \text{ with } \alpha\mu(\Omega_i) \leqslant \mu(A) \leqslant \mu(\Omega_i)/2. \tag{10}$$

Let γ, ξ and λ satisfy

$$\gamma = \min\left\{\frac{\mu(\Omega_i)\eta}{48}, \frac{\mu(\Omega_i)\beta}{500}\right\} \quad \text{and} \quad 0 < \xi \ll \lambda \ll \min\left\{\beta, \min_{i\in[k]} \mu(\Omega_i)\eta\right\}. \tag{11}$$

Suppose that G is a graph given at the input of the proposition.

Let $V(G) = V_0' \sqcup V_1' \sqcup \ldots \sqcup V_k'$ be a partition satisfying properties of Lemma 2.8 for the graphon W and its partition $\Omega = \Omega^* \sqcup \bigsqcup_{i=1}^{k} \Omega_i$, together with input error parameter ξ and output error parameter λ. We will modify this partition to obtain a $(\gamma, \eta, \varepsilon)$-expander decomposition of G.

Lemma 2.8(a) gives that

$$|V_0'| \leqslant \tfrac{1}{2}\varepsilon n. \tag{12}$$

For each $i \in [k]$, we perform the following cleaning procedure. Let $U : \Omega_i^2 \to [0,1]$ be a graphon representation of $G[V_i']$ on the (renormalized) probability space Ω_i such that we have $\|W[\Omega_i] - U\|_\square < \lambda$. Such a representation exists by Lemma 2.8(c).

Let $P_i^0 := V_i'$ and $Q_i^0 := \varnothing$. Now, for $j = 1, 2, 3, \ldots$ we proceed as follows. If there exists at least one set $X_i^j \subseteq P_i^{j-1}$ of size at most $\frac{3}{5}|P_i^{j-1}|$ with $e(X_i^j, P_i^{j-1}\setminus X_i^j) < \gamma|X_i^j|n$, then we take this set, and let $Q_i^j := X_i^1 \cup X_i^2 \cup \ldots \cup X_i^j$, $P_i^j := V_i'\setminus Q_i^j$, and proceed with $j+1$. If no set X_i^j exists, then we set $j(i) := j-1$, $V_i := P_i^{j(i)}$, and terminate. Since the sets X_i^j $(j = 1, 2, \ldots, j(i)-1)$ are nonempty, we will stop eventually.

For every $i \in [k]$ and every $j \in [j(i)]$, we have

$$e(Q_i^j, P_i^j) = \sum_{\ell=1}^{j} e(X_i^\ell, P_i^j) \leqslant \sum_{\ell=1}^{j} e(X_i^\ell, P_i^{\ell-1}\setminus X_i^\ell) \leqslant \gamma n \sum_{\ell=1}^{j} |X_i^\ell| = \gamma|Q_i^j|n. \tag{13}$$

Claim 2.7.1. *For each $i \in [k]$,*

$$|V_i'\setminus V_i| \leqslant 3\alpha n. \tag{14}$$

Proof. Suppose to the contrary that $|V_i'\setminus V_i| \geqslant 3\alpha n$. Let $j \in \{0, 1, 2, \ldots, j(i)\}$ be the largest index for which

$$|P_i^j| \geqslant \tfrac{1}{4}|V_i'|. \tag{15}$$

Now, there are two cases to consider. Either $|P_i^{j+1}| < \frac{1}{4}|V_i'|$ and then we have $|P_i^j| \leqslant (\frac{1}{4} + \frac{3}{5})|V_i'|$ by the way we chose the set X_i^{j+1}. Another case is that $j = j(i)$, that is, we terminated in the step j. Then, by our assumption, $|V_i'\backslash P_i^j| = |V_i'\backslash V_i| \geqslant 3\alpha n$. Put together,

$$|Q_i^j| = |V_i'\backslash P_i^j| \geqslant \min\big\{(1 - \tfrac{1}{4} - \tfrac{3}{5})|V_i'|, 3\alpha n\big\} = 3\alpha n, \tag{16}$$

as $|V_i'| \geqslant \frac{1}{2}\mu(\Omega_i)n$ (by Lemma 2.8(a)) and $\alpha \leqslant \frac{1}{40}\mu(\Omega_i)$ (by (11)).

We learn from (13) that

$$e(Q_i^j, P_i^j) \leqslant \gamma|Q_i^j|n \overset{(15)}{\leqslant} 4\gamma|Q_i^j||P_i^j| \cdot \frac{n}{|V_i'|} \leqslant 5\gamma|Q_i^j||P_i^j| \cdot \frac{1}{\mu(\Omega_i)} \overset{(11)}{<} \tfrac{\beta}{100}|Q_i^j||P_i^j|. \tag{17}$$

Let $\Lambda \subseteq \Omega_i$ represent the vertices of P_i^j. We have $\mu(\Lambda) \geqslant \frac{1}{4}\mu(\Omega_i) - \lambda \geqslant \frac{1}{5}\mu(\Omega_i)$, due to Lemma 2.8(a) and (15). Similarly, (16) gives $\mu(\Omega_i\backslash\Lambda) \geqslant \alpha\mu(\Omega_i)$. Thus, (10) applies. We have

$$\int_\Lambda \int_{\Omega_i\backslash\Lambda} U \geqslant \int_\Lambda \int_{\Omega_i\backslash\Lambda} W - \|W[\Omega_i] - U\|_\square \overset{(10)}{\geqslant} \beta - \lambda\,,$$

which contradicts (17). □

We have defined the sets $V_1, V_2, \ldots, V_k$. Set $V_0 := V_0' \cup \bigcup_{i=1}^k (V_i'\backslash V_i)$. Let us now check that $V(G) = V_0 \sqcup V_1 \sqcup \ldots \sqcup V_k$ is indeed a desired expander decomposition.

As for property (G1), we have

$$|V_0| = |V_0'| + \sum_{i=1}^k |V_i'\backslash V_i| \overset{(12),\ \text{Cl2.7.1}}{\leqslant} \tfrac{1}{2}\varepsilon n + 3\alpha k n \leqslant \varepsilon n.$$

For (G2), we first notice that

$$|V_i| \overset{(14)}{\geqslant} |V_i'| - 3\alpha n \geqslant (\mu(\Omega_i) - \lambda)n - 3\alpha n \geqslant \tfrac{1}{2}\mu(\Omega_i)n,$$

as $\alpha \leqslant \frac{1}{12}\mu(\Omega_i)$. Thus we find, as required,

$$e(V_i, V\backslash V_i) \leqslant e(V_i', V\backslash V_i') + e(V_i'\backslash V_i, V_i) \overset{(13)}{\leqslant} \lambda n^2 + 4\gamma|V_i'\backslash V_i|n \overset{\text{Cl2.7.1}}{\leqslant} \lambda n^2 + 12\gamma\alpha n^2 \leqslant \eta|V_i|n,$$

where the last inequality holds since $|V_i| \geqslant \frac{1}{2}\mu(\Omega_i)n$, $\lambda \ll \mu(\Omega_i)\eta$, and $\gamma \leqslant \frac{\mu(\Omega_i)\eta}{48}$.

Finally, property (G3) follows immediately from the stopping condition. This completes our proof of Theorem 2.7. □

2.3. **Properties of the expander decomposition.** For the proof of Theorem 1.1 we argue that the majority of vertices do not "see" much beyond the component in the expander decomposition they belong to. This is formalized in the following definitions.

Definition 2.9. *Suppose that G is a loopless multigraph of order n. Assume that $V(G) = V_0 \sqcup V_1 \sqcup \ldots \sqcup V_k$ is some expander decomposition of G. For a vertex v we write $i(v)$ for the unique $i \in \{0, 1, \ldots, k\}$ such that $v \in V_i$. Given $\alpha > 0$ and $\varepsilon > 0$ we say that $v \in V \backslash V_0$ is* (α, ε)-good *with respect to the decomposition if the following hold:*

(a) $\deg(v) \geqslant \Omega(\varepsilon n)$,
(b) $\deg(v; V_{i(v)}) \geqslant (1 - \mathsf{O}(\varepsilon^2)) \deg(v)$,
(c) $\sum_{u \in V_{i(v)}, u \sim v} \big(\frac{1}{\deg(u; V_{i(v)})} - \frac{1}{\deg(u)}\big) \leqslant \mathsf{O}(\alpha^{1/2})$,
(d) $\sum_{u \in V_{i(v)}, u \sim v} \frac{1}{\deg(u; V_{i(v)})} \leqslant \mathsf{O}(\alpha^{-1/4})$.

Definition 2.10. *Suppose that G is a loopless multigraph of order n. Given numbers $\beta, \alpha, \gamma > 0$ and $\varepsilon \in (0, \alpha)$, we say that G has an* $(\beta, \alpha, \gamma, \varepsilon)$-good-decomposition *if*

(1) G admits a $(\gamma, \varepsilon^5, \varepsilon^5)$-expander decomposition $V(G) = V_0 \sqcup V_1 \sqcup \ldots \sqcup V_k$, and
(2) At least $(1 - \mathsf{O}(\alpha^{1/4}))n$ vertices of G are (α, ε)-good.
(3) At least $(1 - \mathsf{O}(\beta))n$ vertices of G have degree at least $\Omega(\alpha^{1/10} n)$.

Next we refine Theorem 2.7.

Lemma 2.11. *For any $\beta > 0$ and any nondegenerate graphon $W : \Omega^2 \to [0, 1]$, there exist $\alpha, \varepsilon, \gamma, \xi > 0$ with $\beta \gg \alpha \gg \varepsilon \gg \gamma \gg \xi$ such that if G is a simple graph on $n \geqslant \xi^{-1}$ vertices with $d_\square(G, W) \leqslant \xi$, then G has a $(\beta, \alpha, \varepsilon, \gamma)$-good-decomposition.*

Proof. Let β and W be given. Since W is nondegenerate, there exists $\alpha > 0$ such that any m-vertex graph (m is arbitrary) that is ξ_1-close (for $\xi_1 > 0$ sufficiently small) to W has at least $(1 - \beta)m$ of degrees at least $\alpha^{1/10} m$, so requirement (3) of Definition 2.10 holds. Similarly, we can find constants $\varepsilon \in (0, \alpha^{20})$ and $\xi_2 > 0$ such that any m-vertex graph (m is arbitrary) which is ξ_2-close to W has at most αm vertices of degrees at most εm. We apply Theorem 2.7 with input ε^5 and $\eta = \varepsilon^5$ and retrieve $\gamma > 0$ and $\xi_3 > 0$. We set $\xi := \min(\xi_1, \xi_2, \xi_3)$. Suppose now that $G = (V, E)$ is a graph satisfying the assumptions of the lemma. Theorem 2.7 readily gives item (1) of Definition 2.10.

To show item (2), we first note that by property (G2) of the expander decomposition we have that $e(V_i, V \backslash V_i) \leqslant \varepsilon^5 n |V_i|$ for all $i \in [k]$. By summing over i we deduce that

$$\sum_{i=1}^{k} e(V_i, V \backslash V_i) \leqslant \varepsilon^5 n^2 \,. \tag{18}$$

Denote by S the set of vertices of $V\backslash V_0$ violating (b) of Definition 2.9 using 1 as the implicit constant in the term $\mathsf{O}(\varepsilon^2)$,[4] that is,

$$S = \left\{v \in V\backslash V_0 : \deg(v; V\backslash V_{i(v)}) \geqslant \varepsilon^2 \deg(v)\right\}.$$

Then by (18) we have that

$$\varepsilon^5 n^2 \geqslant \sum_{v\in S} \deg(v; V\backslash V_{i(u)}) \geqslant \varepsilon^2 \sum_{v\in S} \deg(v)\,.$$

Therefore $\sum_{v\in S} \deg(v) \leqslant \varepsilon^3 n^2$, from which we learn that

$$|S \cap \{v : \deg(v) \geqslant \varepsilon n\}| \leqslant \varepsilon^2 n\,. \tag{19}$$

We deduce that

$$\left|\left\{v \in V\backslash V_0 : \deg(v) \leqslant \varepsilon n \text{ or } \deg(v; V\backslash V_{i(v)}) \geqslant \varepsilon^2 \deg(v)\right\}\right| \leqslant (\alpha + \varepsilon^2)n\,. \tag{20}$$

Next, for $i \in [k]$ we write

$$\sum_{v\in V_i} \sum_{u\in V_i, u\sim v} \Big(\frac{1}{\deg(u; V_i)} - \frac{1}{\deg(u)}\Big) = |V_i| - \sum_{u\in V_i} \frac{\deg(u; V_i)}{\deg(u)} = \sum_{u\in V_i} \frac{\deg(u; V\backslash V_i)}{\deg(u)}\,.$$

We sum this over $i \in [k]$ and get that

$$\sum_{v\in V\backslash V_0} \sum_{u\in V_{i(v)}, u\sim v} \Big(\frac{1}{\deg(u; V_i)} - \frac{1}{\deg(u)}\Big) = \sum_{u\in V\backslash V_0} \frac{\deg(u; V\backslash V_{i(u)})}{\deg(u)} \leqslant (\alpha + \varepsilon^2)n + \varepsilon^2 n = \mathsf{O}(\alpha n)\,,$$

where we bounded the ratio by 1 for those vertices counted in (20), and by ε^2 for the vertices that were not. From the last inequality we deduce that there cannot be more than $\Omega(\alpha^{1/2}n)$ vertices $v \in V\backslash V_0$ such that requirement (c) in the definition of (α, ε)-good is not satisfied.

Lastly, to show (d), we have that

$$\sum_{i=1}^{k} \sum_{v\in V_i} \sum_{u\in V_i : u\sim v} \frac{1}{\deg(u; V_i)} = \sum_{i=1}^{k} |V_i| \leqslant n\,,$$

therefore there cannot be more than $\Omega(\alpha^{1/4}n)$ vertices $v \in V\backslash V_0$ such that (d) is violated. This concludes our proof. □

Part (2) of Definition 2.10 asserts that there are many (α, ε)-good vertices in the graph, yet there could still be components of the decomposition V_i in which the majority of their vertices are not (α, ε)-good. These cannot occupy too much of the mass. Indeed, for some $i \in [k]$, we say the set V_i is (α, ε)-*big* if at least $(1 - \mathsf{O}(\alpha^{1/8}))|V_i|$ of its vertices are (α, ε)-good, and $e(G[V_i]) \geqslant \Omega(\alpha^{1/9}|V_i|n)$.

[4]Note that later, in the proof of Lemma 3.10, we shall be forced to use larger implicit constants in (b) of Definition 2.9.

Proposition 2.12. *Suppose that G is a graph with n vertices that has a $(\beta,\alpha,\gamma,\varepsilon)$-good-decomposition (as in Definition 2.10). Then*

$$\sum_{i\in[k]:V_i \text{ is } (\alpha,\varepsilon)-\text{big}} |V_i| \geqslant (1-\mathsf{O}(\beta^{1/8}))n\,.$$

Proof. Let I_1 be the indices $i\in[k]$ such that V_i has $\Omega(\alpha^{1/8}n)$ vertices that are not (α,ε)-good. Since this is a $(\beta,\alpha,\gamma,\varepsilon)$-good-decomposition we have that the total number of not (α,ε)-good vertices is $\mathsf{O}(\alpha^{1/4}n)$. Hence,

$$\sum_{i\in I_1}|V_i| \leqslant \mathsf{O}(\alpha^{1/8}n) = \mathsf{O}(\beta^{1/8}n)\,.$$

Next, let I_2 be the indices $i\in[k]$ such that $e(G[V_i]) = O(\alpha^{1/9}|V_i|n)$. Put $V' = \bigcup_{i\in I_2} V_i$. We have

$$\sum_{v\in V'}\deg(v) \leqslant 2\sum_{i\in I_2} e(G[V_i]) + \sum_{i\in I_2} e(V_i, V\backslash V_i) \leqslant O(\alpha^{1/9}|V'|n) + \varepsilon^5|V'|n = O(\alpha^{1/9}|V'|n)\,.$$

If $|V'| = \Omega(\beta^{1/8}n)$, then by property (3) in Definition 2.10, we may bound $\sum_{v\in V'}\deg(v)$ from below by $\Omega(\alpha^{1/10}|V'|n)$, giving a contradiction to the last estimate. The proof is concluded since $\sum_{i\in[k]}|V_i| \geqslant (1-\varepsilon^5)n$. □

3. Local neighborhoods of the UST via electric networks

3.1. **Preliminaries.** Suppose that $G=(V,E)$ is a loopless multigraph. We denote by $\{X_t\}_{t\geqslant 0}$ the simple random walk starting from some (possibly random) vertex X_0. That is, $\{X_t\}_{t\geqslant 0}$ is a Markov chain with state space V and transition matrix $p(x,y) = \frac{e(\{x\},\{y\})}{\deg(x)}$. We denote by $\mathbb{P}_v$ the probability measure of the simple random walk started at a vertex v. We will frequently use two stopping times: the hitting time of a vertex v is the random variable $\tau_v := \min\{t\geqslant 0 : X(t) = u\}$ and the hitting time after zero of a vertex v is the random variable $\tau_v^+ := \min\{t>0 : X(t)=v\}$. Clearly when $X_0\neq v$ these two random times are equal.

3.1.1. *Effective resistance.* Our analysis relies on the relation between random walks, USTs and the theory of electrical networks. We briefly recall here the basic theory we will use and refer the reader to [18, Chapter 2] for a comprehensive study. Given a graph $G=(V,E)$ we write $\overrightarrow{E}$ for the set of directed edges of size $2|E|$ which contain each edge of E in both direction. Given two distinct vertices u,v we say that an antisymmetric function $f:\overrightarrow{E}\to\mathbb{R}$ is a *flow* from u to v if for each vertex $w\notin\{u,v\}$ the sum of f over edges outgoing from w is zero. A flow is called *unit* if the sum of f over edges outgoing from u is 1. The *effective resistance* $\mathcal{R}_{\text{eff}}(u\leftrightarrow v;G)$ between u and v is defined as the minimum energy $\mathcal{E}(f) = \sum_{e\in E} f(e)^2$ of any unit flow f from u to v. If u and v are not in the same connected component, we define effective resistance between them to be ∞. When it is

clear what the underlying graph G is we simply write $\mathcal{R}_{\text{eff}}(u \leftrightarrow v)$. From this definition it is immediate that if G' is a subgraph of G, then

$$\mathcal{R}_{\text{eff}}(u \leftrightarrow v; G) \leqslant \mathcal{R}_{\text{eff}}(u \leftrightarrow v; G') \,. \tag{21}$$

The latter inequality is also known as Rayleigh's monotonicity law. The *discrete Dirichlet's principle* gives a dual definition of the effective resistance in terms of functions on the vertices. It states that

$$\mathcal{R}_{\text{eff}}(u \leftrightarrow v)^{-1} = \inf \left\{ \sum_{(x,y)\in E} (h(x) - h(y))^2 : h : V \to \mathbb{R} \,,\, h(u) = 0 \,,\, h(v) = 1 \right\} , \tag{22}$$

see [18, Exercise 2.13]. We will also a basic probabilistic interpretation of the effective resistance which can be found in [18, Chapter 2]:

$$\mathbb{P}_u \left[\tau_v < \tau_u^+ \right] = \frac{1}{\deg(u) \mathcal{R}_{\text{eff}}(u \leftrightarrow v)} \,. \tag{23}$$

3.1.2. *Uniform spanning trees.* There is a fundamental connection between the uniform spanning tree and electric networks due to Kirchhoff [10]. Let G be connected loopless multigraph and $e = xy$ be an edge of the graph. As before we denote by $\mathcal{T}$ a UST of G. Kirchhoff's formula [10] (see also [18, Chapter 4]) states that for any edge $e = (x, y)$ of G we have

$$\mathbb{P}(e \in \mathcal{T}) = \mathcal{R}_{\text{eff}}(x \leftrightarrow y) \,. \tag{24}$$

Let S be a subset of $E(G)$. We would like to condition on events of the form $S \subseteq \mathcal{T}$ or $S \cap \mathcal{T} = \varnothing$. We denote by G/S the loopless multigraph obtained from G by contracting the edges of S and erasing any loops that has been formed, and by $G - S$ the graph G with the edges of S erased. The following is an easy and classical observation, see [18, Chapter 4].

Proposition 3.1. *Let G be a connected loopless multigraph and S a subset of edges of G.*

(1) If $G - S$ is connected, then the UST $\mathcal{T}$ of G conditioned on $S \cap \mathcal{T} = \varnothing$ has the distribution of the UST on $G - S$.

(2) If S does not contain a cycle, then the UST $\mathcal{T}$ of G conditioned on $S \subseteq \mathcal{T}$ has the distribution of $S \cup \mathcal{T}_{G/S}$ where $T_{G/S}$ is a UST of G/S.

3.1.3. *Mixing time.* Let $G = (V, E)$ be a finite connected loopless multigraph and consider the lazy simple random walk on it, that is, the Markov chain on the vertex set V with transition probability $p(x, y) = \frac{e(\{x\},\{y\})}{2 \deg(x)}$ whenever $x \neq y$ and $p(x, x) = 1/2$ for any vertex

x. Let π be the stationary distribution $\pi(x) = \deg(x)/2|E|$ and for each two disjoint subsets of vertices A, B we write

$$Q(A,B) = \sum_{x\in A, y\in B} \pi(x)p(x,y) = e(A,B)/4|E|\,.$$

The Cheeger constant Φ_* is defined as

$$\Phi_* = \min_{S:\pi(S)\leqslant\frac{1}{2}} \frac{Q(S, V\backslash S)}{\pi(S)}\,,$$

where $\pi(S) = \sum_{x\in S}\pi(x)$. This "bottleneck" ratio is frequently used to control the spectral gap of the lazy random walk from which we may bound its mixing time. Let $1 = \lambda_1 > \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ be the eigenvalues of the transition matrix p (we have $\lambda_1 > \lambda_2$ since G is connected, and $\lambda_n \geqslant 0$ since the chain is lazy, see [14]). A result by Jerrum and Sinclair [20], Lawler and Sokal [13] and Alon and Milman [3] states that

$$\Phi_*^2/2 \leqslant 1 - \lambda_2 \leqslant 2\Phi_*. \tag{25}$$

Assume now that G is a γ-expander as in Definition 2.1 and that the number of parallel edges between any two vertices is at most $f \geqslant 1$. Then the degree of each vertex is at most fn, hence $\pi(S) \leqslant f|S|n/2|E|$ for any $S \subseteq V$. Similarly, for any $S \subseteq V$ we have that $\pi(V\backslash S) \leqslant fn|V\backslash S|/2|E|$, so if $\pi(V\backslash S) \geqslant 1/2$ we get that $|V\backslash S| \geqslant |E|/fn$. Since the minimum degree in a γ-expander is at least $\gamma(n-1)$ we get that if $\pi(V\backslash S) \geqslant 1/2$, then $|V\backslash S| \geqslant \gamma(n-1)/2f$. Putting all this together we get that if G is a γ-expander and $n \geqslant 2$, then

$$\Phi_* \geqslant \frac{\gamma|S||V\backslash S|}{2|S|fn} \geqslant \frac{\gamma^2}{8f^2}\,,$$

from which we get by (25) that

$$1 - \lambda_2 \geqslant \frac{c\gamma^4}{f^4}\,, \tag{26}$$

where $c = 1/128$. Recall that the *total variation* distance $\|\mu-\nu\|_{\mathrm{TV}}$ between two probability measures μ and ν on the same probability space is defined to be $\sup_A |\mu(A)-\nu(A)|$, where the sup is ranging over all events A. For $\varepsilon > 0$ the ε-mixing-time $T_{\mathrm{mix}}(\varepsilon)$ of the chain is defined as

$$T_{\mathrm{mix}}(\varepsilon) = \min\left\{t : \|p^t(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \leqslant \varepsilon \text{ for all } x \in V\right\}.$$

The mixing time and the spectral gap are related via the following statement, see [14, Theorem 12.4],

$$T_{\mathrm{mix}}(\varepsilon) \leqslant \frac{1}{1-\lambda_2}\Big[\frac{1}{2}\log\frac{1}{\min_{x\in V}\pi(x)} + \log(1/2\varepsilon)\Big].$$

Since $f \geqslant 1$ bounds the maximal number of parallel edges between any two vertices, we get that $|E| \leqslant fn^2$ and hence $\pi(x) \geqslant \gamma(n-1)/2|E| \geqslant c\gamma/(fn)$ for some universal constant $c > 0$. We deduce from this, the above bound on $T_{\mathrm{mix}}(\varepsilon)$ and (26) that

$$T_{\mathrm{mix}}(\varepsilon) \leqslant Cf^4\gamma^{-4}\big[\log n + \log f/\gamma + \log \varepsilon^{-1}\big], \tag{27}$$

for some universal constant $C > 0$.

3.2. **Random walks on dense expanders.**

Lemma 3.2. *Suppose that H is a loopless multigraph on n vertices that is a γ-expander and that the number of parallel edges between any two vertices is at most $f \geqslant 1$. Then for any two distinct nodes u and v and a node $w \neq v$ we have that*

$$\mathbb{P}_w\left[\tau_v < \tau_u^+\right] = \frac{\deg(v)}{\deg(u) + \deg(v)} + \mathsf{O}(\gamma^{-7} f^7 n^{-1} \log(n)).$$

Proof. Let us assume that there are no edges between u and v — this can only matter for the assertion of the statement when $w = u$ and in this case affect the estimate by the probability that this edge is traversed on the first step of the random walk; the latter probability is at most $\mathsf{O}(\gamma^{-1} f/n)$ the assumption, since $\deg(u) \geqslant \gamma n$. This error is swallowed in the error estimate of the lemma.

We denote by $\mathbb{P}^{\mathrm{lazy}}$ the lazy random walk on the graph, that is, the random walk that with probability $1/2$ stays put and otherwise jumps to a uniformly chosen neighbor. It is clear that if w is a vertex such that $w \notin \{u, v\}$ then $\mathbb{P}_w^{\mathrm{lazy}}(\tau_v < \tau_u^+) = \mathbb{P}_w(\tau_v < \tau_u^+)$. We will first show, via a coupling argument, that for any two vertices $w_1, w_2 \notin \{u, v\}$ we have

$$\mathbb{P}_{w_1}^{\mathrm{lazy}}(\tau_v < \tau_u^+) = \mathbb{P}_{w_2}^{\mathrm{lazy}}(\tau_v < \tau_u^+) + \mathsf{O}(\gamma^{-6} f^6 n^{-1} \log(n)). \tag{28}$$

Indeed, let $\{X_t\}_{t \geqslant 0}$ and $\{Y_t\}_{t \geqslant 0}$ be two lazy simple random walks starting at w_1 and w_2, respectively. We put $\varepsilon = n^{-1}$ in (27) and bound $\log(f/\gamma)$ by f/γ and get that if $T = C\gamma^{-5} f^5 \log(n)$, then $\|X_T - \pi\|_{\mathrm{TV}} \leqslant n^{-1}$ and the same estimate holds for Y_T, hence $\|X_T - Y_T\|_{\mathrm{TV}} \leqslant 2n^{-1}$ (where by $\|X_T - Y_t\|_{\mathrm{TV}}$ we mean the total variation distance between the *laws* of X_T and Y_T).

By [14, Proposition 4.7] we deduce that we can couple the walks $\{X_t\}$ and $\{Y_t\}$ so that $X_T = Y_T$ with probability at least $1 - 2n^{-1}$. If this occurs we continue the coupling so that $X_t = Y_t$ for all $t \geqslant T$ by using the same random neighbor at each step of the walk. Thus, if both walks have not visited u or v between time 1 and T, then the event $\{\tau_v < \tau_u^+\}$ occurs for $\{X_t\}$ if and only if it occurs for $\{Y_t\}$. Since the minimal degree of H is at least $\gamma(n-1)$ and the maximal number of parallel edges between any two vertices is at most f we learn that the probability of visiting u or v between time 1 and T is at most $Tf/\gamma(n-1) \leqslant 2C\gamma^{-6} f^6 n^{-1} \log(n)$, concluding the proof of (28).

We now continue the proof of the lemma for the case that $w = u$. If the walker starts at u and $\tau_v < \tau_u^+$, then it cannot be lazy in the first step. Hence

$$\mathbb{P}_u^{\text{lazy}}(\tau_v < \tau_u^+) = \frac{1}{2}\mathbb{P}_u(\tau_v < \tau_u^+)\,. \tag{29}$$

Consider now the Markov chain on the two states $\{u, v\}$ with transition probabilities

$$p(u,v) = \mathbb{P}_u^{\text{lazy}}(\tau_v < \tau_u^+)\,, \qquad p(v,u) = \mathbb{P}_v^{\text{lazy}}(\tau_u < \tau_v^+)\,,$$

with $p(u,u) = 1-p(u,v)$ and $p(v,v) = 1-p(v,u)$. This is the lazy random walk "watched" on the vertices u and v. By summing over paths it is immediate that

$$\deg(u)p(u,v) = \deg(v)p(v,u)\,. \tag{30}$$

In order to visit v before returning to u, the lazy walker must walk to a random neighbor in the first step, so

$$p(u,v) = \frac{1}{2}\mathbb{P}_{N(u)}^{\text{lazy}}(\tau_v < \tau_u^+)\,, \tag{31}$$

where $\mathbb{P}_{N(u)}^{\text{lazy}}$ indicates a uniform starting position from the set $N(u)$ of neighbors of u. Similarly, when starting from v, in order to return to v before visiting u, the lazy walker can either stay put on the first step, or jump to a uniform neighbor of v and from there visit v before u, thus

$$p(v,v) = \frac{1}{2} + \frac{1}{2}\mathbb{P}_{N(v)}^{\text{lazy}}(\tau_v < \tau_u^+)\,. \tag{32}$$

Since we assumed there are no edges between u and v, by (28) we have that

$$\mathbb{P}_{N(v)}^{\text{lazy}}(\tau_v < \tau_u^+) = \mathbb{P}_{N(u)}^{\text{lazy}}(\tau_v < \tau_u^+) + \mathsf{O}(\gamma^{-6}f^6 n^{-1}\log(n))\,.$$

This together with (31) and (32) gives that $p(u,v) + p(v,u) = 1/2 + \mathsf{O}(\gamma^{-6}f^6n^{-1}\log(n))$. Together with (30) and the fact that all degrees are at least $\gamma(n-1)$ and at most fn gives that

$$p(u,v) = \frac{\deg(v)}{2(\deg(u)+\deg(v))} + \mathsf{O}(\gamma^{-7}f^7n^{-1}\log(n))\,,$$

concluding the proof of lemma when $w = u$ by (29). The proof for any $w \notin \{u, v\}$ can now be completed easily. By (28) and our assumption that there are no edges between u and v shows that

$$\mathbb{P}_w^{\text{lazy}}(\tau_v < \tau_u^+) = \mathbb{P}_{N(u)}^{\text{lazy}}(\tau_v < \tau_u^+) + \mathsf{O}(\gamma^{-6}f^6n^{-1}\log n)\,,$$

and so the lemma follows by (31). □

Corollary 3.3. *Suppose that H is a loopless multigraph on n vertices that is a γ-expander and that the number of parallel edges between any two vertices is at most $f \geqslant 1$. Then for any two vertices $u \neq v$*

$$\mathcal{R}_{\text{eff}}(u \leftrightarrow v) = (1 + \mathsf{O}(\gamma^{-8} f^8 n^{-1} \log(n)))\Big(\frac{1}{\deg(u)} + \frac{1}{\deg(v)}\Big)\,.$$

Proof. Follows immediately by (23) and Lemma 3.2 together with the fact that all degrees are at least $\gamma(n-1)$ and at most fn. □

We now extend Corollary 3.3 to the setting of a general dense graph.

Lemma 3.4. *Suppose that G is a loopless multigraph with n vertices given together with a $(\gamma, \varepsilon^5, \varepsilon^5)$-expander decomposition $V(G) = V_0 \sqcup V_1 \sqcup \ldots \sqcup V_k$. Assume further that the maximal number of parallel edges among any two pairs of vertices is at most $f \geqslant 1$. Assume that $\gamma^{-8} f^8 n^{-1} \log n \leqslant \varepsilon$. Let $i \in [k]$ and $u \neq v$ be two distinct vertices of V_i. Then*

$$(1 - \mathsf{O}(\varepsilon))\Big(\frac{1}{\deg(u)} + \frac{1}{\deg(v)}\Big) \leqslant \mathcal{R}_{\text{eff}}(u \leftrightarrow v) \leqslant (1 - \mathsf{O}(\varepsilon))\Big(\frac{1}{\deg(u; V_i)} + \frac{1}{\deg(v; V_i)}\Big)\,,$$

and if in addition u and v are (α, ε)-good, then

$$\mathcal{R}_{\text{eff}}(u \leftrightarrow v) = (1 + \mathsf{O}(\varepsilon)))\Big(\frac{1}{\deg(u)} + \frac{1}{\deg(v)}\Big)\,.$$

Proof. Since $G[V_i]$ is a γ-expander on at least γn vertices, by Corollary 3.3 together with Rayleigh's monotonicity (21) we have

$$\mathcal{R}_{\text{eff}}(u \leftrightarrow v) \quad \leqslant \quad (1 + \mathsf{O}(\gamma^{-8} f^8 n^{-1} \log(n))\Big(\frac{1}{\deg(u; V_i)} + \frac{1}{\deg(v; V_i)}\Big)\,,$$

giving the upper bound of the first assertion of the lemma. The upper bound of the second assertion immediately follows using the part (b) of Definition 2.9.

For the lower bound we will use Dirichlet's principle (22) and let $h : V(G) \to [0,1]$ be the function assigning $h(v) = 1$, $h(u) = 0$ and for any vertex $x \notin \{u, v\}$ we put $h(x) = \deg(v)/(\deg(u) + \deg(v))$. By our assumption there are at most f edges (x, y) such that $x = u$ and $y = v$ in which $h(y) - h(x) = 1$. Next, there are at most $\deg(u)$ edges (x, y) for which $x = u$ and $h(y) - h(x) = \deg(v)/(\deg(u) + \deg(v))$. Similarly, there are at most $\deg(v)$ edges (x, y) for which $y = v$ and $h(y) - h(x) = \deg(u)/(\deg(u) + \deg(v))$. All other edges (x, y) of the graph have $h(x) - h(y) = 0$. Hence by (22) we have

$$\mathcal{R}_{\text{eff}}(u \leftrightarrow v)^{-1} \leqslant f + \frac{\deg(u)\deg(v)^2 + \deg(v)\deg(u)^2}{(\deg(u) + \deg(v))^2} \leqslant \frac{\deg(u)\deg(v)}{\deg(u) + \deg(v)}(1 + \gamma^{-4} f^2 n^{-1})\,,$$

where we used the fact that since each V_i is a γ-expander, its cardinality must be $\Omega(\gamma n)$ and hence $\deg(u)$ and $\deg(v)$ can be bounded below by $\Omega(\gamma^2 n)$ and above by fn. This gives the required lower bound. □

3.3. **The density of fixed trees in the UST.** The main result of this section, Lemma 3.12, allows express the frequency of a given fixed rooted tree T in a graph using a discrete analogue of the parameter Freq, see Definition 3.6 below.

Let us introduce some definitions and a setting that will be used throughout this section. In what follows we are always given an arbitrary $\beta > 0$ and a nondegenerate graphon W. We then apply Lemma 2.11 and extract the corresponding $\alpha, \varepsilon, \gamma$ and ξ so that if G is a connected graph on n vertices and n is sufficiently large (as a function of α, ε and γ) and $\delta_{\square}(G, W) \leqslant \xi$, then G has a $(\beta, \alpha, \gamma, \varepsilon)$-good decomposition as in Definition 2.10. We denote the given expander decomposition of G by $V(G) = V_0 \sqcup V_1 \sqcup \ldots \sqcup V_k$. In light of the quantification of Lemma 2.11 we may assume that

$$\beta \gg \alpha \gg \varepsilon \gg \gamma \gg \xi \gg n^{-1} .$$

Next, let T be a finite rooted tree of height r and ℓ vertices denoted by $1, \ldots, \ell$ so that 1 is the root of T. Given a graph G we say that ℓ distinct vertices $(v_1, \ldots, v_\ell)$ of G are *compatible with* T if the pairs

$$T(v_1, \ldots, v_\ell) := \big\{(v_q, v_t) : (q, t) \in E(T)\big\} ,$$

are all edges of G. Without loss of generality we may assume that the numbering $\{1, \ldots, \ell\}$ of the vertices of T is such that there exists some $p \in \{2, \ldots, \ell\}$ such that the vertices of distance r from the root (which all must be leaves) are $p, \ldots, \ell$.

Definition 3.5. *Assume the setting as introduced at the beginning of Section 3.3. Given some fixed $\ell \geqslant 1$ and $i \in [k]$ we say that an ℓ-tuple of distinct vertices of G are* i-pure *if each vertex in the ℓ-tuple belongs to V_i and is (α, ε)-good.*

Next we define $\mathrm{Freq}(T; G)$ which is the discrete analogue of $\mathrm{Freq}(T; W)$. Note the notation $\mathrm{Freq}(T; G)$ does not reflect the fact that this parameter depends on the partition, and not just on the graph G.

Definition 3.6. *Assume the setting as introduced at the beginning of Section 3.3. For each $i \in [k]$ we define*

$$\mathrm{Freq}(T; G, i) := |\mathrm{Stab}_T|^{-1} \sum_{\substack{(v_1, \ldots, v_\ell) \\ i\text{-pure} \\ \text{compatible with } T}} |V_i|^{-1} \exp\left(- \sum_{j=1}^{p-1} b(v_j) \right) \frac{\sum_{j=p}^{\ell} \deg(v_j)}{\prod_{j=1}^{\ell} \deg(v_j)} ,$$

where

$$b(v) = \sum_{u \in V_{i(v)}, u \sim v} \frac{1}{\deg(u)} . \tag{33}$$

Note that (33) *is a graph counterpart to the graphon quantity defined in* (2).

Finally, let

$$\mathrm{Freq}(T;G) := \sum_{i\in[k]:V_i \text{ is } (\alpha,\varepsilon)-\text{big}} \frac{|V_i|}{n}\cdot \mathrm{Freq}(T;G,i)\,. \tag{34}$$

We denote by $\mathcal{T}$ a sample of the UST of G and by $B_{\mathcal{T}}(v,r)$ the graph-distance ball in $\mathcal{T}$ of radius r around $v\in V(G)$ and we think about it as a subset of edges of $\mathcal{T}$. We will begin our proof with estimating the probability that $B_{\mathcal{T}}(v,r)$ is manifested on an i-pure ℓ-tuple of vertices (as in Definition 3.5) that are compatible with T; later we will see that that all other manifestations of $B_{\mathcal{T}}(v,r)$ are negligible. For an i-pure ℓ-tuple that is compatible with T we write $B_{\mathcal{T}}(v_1,r)\overset{\text{pure}}{\cong} T(v_1,\dots,v_\ell)$ for the event

- the edges $T(v_1,\dots,v_\ell)$ are in $\mathcal{T}$, and,
- for each $1\leqslant j\leqslant p-1$, the edges of emanating from v_j that are not in $T(v_1,\dots,v_\ell)$ and have both endpoints in V_i are not in $\mathcal{T}$.

Lemma 3.7. *Assume the setting as introduced at the beginning of Section 3.3, and let $\mathcal{T}$ be a UST of G. Let T be a fixed rooted tree with $\ell\geqslant 2$ vertices $1,\dots,\ell$ and height $r\geqslant 1$ (as usual T is rooted at 1). Assume that the vertices at height r of T are $\{p,\dots,\ell\}$ for some $2\leqslant p\leqslant \ell$. Then for any $i\in[k]$ and any i-pure ℓ-tuple $(v_1,\dots,v_\ell)$ that is compatible with T we have that*

$$\mathbb{P}\big(B_{\mathcal{T}}(v_1,r)\overset{\text{pure}}{\cong} T(v_1,\dots,v_\ell)\big) = (1+\mathsf{O}(\ell\alpha^{1/4}))\exp\left(-\sum_{j=1}^{p-1} b(v_j)\right)\cdot\frac{\sum_{j=p}^{\ell}\deg(v_j)}{\prod_{j=1}^{\ell}\deg(v_j)}\,,$$

where $b(v)$ is defined in (33).

Proof. Assume that n is large enough as in Lemma 3.4. We first show by induction that the probability that $T(v_1,\dots,v_\ell)\subseteq E(\mathcal{T})$ is

$$(1+\mathsf{O}(\ell\varepsilon))\frac{\sum_{j=1}^{\ell}\deg(v_j)}{\prod_{j=1}^{\ell}\deg(v_j)}\,. \tag{35}$$

Indeed, when T contains only one edge this statement follows immediately from Kirchoff's formula (24) and Lemma 3.4. If T has more than one edge, assume without loss of generality that ℓ is a leaf in T at distance r from the root and that (q,ℓ) is an edge of T. We then use the induction hypothesis on the tree $T\backslash\{\ell\}$, which has $\ell-1$ vertices. We condition on the event $T(v_1,\dots,v_\ell)\backslash\{(v_q,v_\ell)\}\subseteq E(\mathcal{T})$ and contract these $\ell-1$ vertices to a single vertex of degree $\deg(v_1)+\ldots+\deg(v_{\ell-1})$. We shall make use of Proposition 3.1 when working with the contracted graph. Since the contracted multigraph is still a γ-expander with at most ℓ parallel edges between any two vertices, and the vertices $(v_1,\dots,v_\ell)$ are all (α,ε)-good, we may apply Lemma 3.4 and Kirchoff's formula (24) to get that the probability that the

edge (v_q, v_ℓ) of G is in $\mathcal{T}$ is

$$(1-\mathsf{O}(\varepsilon))\Big(\frac{1}{\deg(v_\ell)}+\frac{1}{\deg(v_1)+\ldots+\deg(v_{\ell-1})}\Big)\,.$$

By our induction hypothesis we get that the required probability is

$$(1-\mathsf{O}((\ell-1)\varepsilon))(1-\mathsf{O}(\varepsilon))\frac{\sum_{q=1}^{\ell-1}\deg(v_q)}{\prod_{q=1}^{\ell-1}\deg(v_q)}\Big(\frac{1}{\deg(v_\ell)}+\frac{1}{\deg(v_1)+\ldots+\deg(v_{\ell-1})}\Big)\,,$$

concluding the proof of (35).

We now condition on the event $T(v_1,\ldots,v_\ell)\subseteq E(\mathcal{T})$ and turn to compute the probability that all other edges of G emanating from $v_1,\ldots,v_{p-1}$ which have both endpoints in V_i are not in $\mathcal{T}$. That is, the event that $E_{v_j}\cap\mathcal{T}=\varnothing$ for all $1\leqslant j\leqslant p-1$, where E_{v_j} are all the edges of G between v_j and $V_i\backslash\{v_1,\ldots v_\ell\}$.

Let j be an integer $1\leqslant j\leqslant p-1$ and condition additionally on all the edges of $(E_{v_1}\cup\cdots\cup E_{v_{j-1}})\cap\mathcal{T}=\varnothing$ (if $j=1$ there is no further conditioning). We will prove that under this conditioning the probability that $E_{v_j}\cap\mathcal{T}=\varnothing$ is

$$(1+\mathsf{O}(\alpha^{1/4}))\exp\left(-b(v_j)\right)\cdot\frac{\sum_{q=j+1}^{\ell}\deg(v_q)}{\sum_{q=j}^{\ell}\deg(v_q)}\,. \tag{36}$$

Thus, once (36) is proved, by multiplying it over $j=1,\ldots,p-1$, we get that conditioned on $\mathcal{O}$ and on $E_G(T)\subseteq E(\mathcal{T})$, the probability that all the required edges are not in $\mathcal{T}$ is

$$(1+\mathsf{O}(\ell\alpha^{1/4}))\exp\left(-\sum_{j=1}^{p-1}b(v_j)\right)\cdot\frac{\deg(v_p)+\ldots+\deg(v_\ell)}{\deg(v_1)+\ldots+\deg(v_\ell)}\,.$$

We multiply (35) by this and get the required assertion of the lemma.

We are left to prove (36). Enumerate the neighbors of v_j which are not in $\{v_1,\ldots,v_\ell\}$ by $u_1,\ldots,u_{\deg(v_j;V_i)}$. For each $1\leqslant m\leqslant\deg(v_j;V_i)$ we condition on the first $m-1$ edges being closed, by Proposition 3.1 we remove these edges from the graph, and after this removal the graph remains an $\frac{\gamma}{2}$-expander by Proposition 2.2. Thus, by Lemma 3.4 we have that in this conditioned graph the resistance on the m-th edge is

$$(1+\mathsf{O}(\varepsilon))\Big(r_m+\frac{1}{\sum_{q=j}^{\ell}\deg(v_q)-(m-1)}\Big)\,, \tag{37}$$

where r_m satisfies

$$\frac{1}{\deg(u_m)}\leqslant r_m\leqslant\frac{1}{\deg(u_m;V_i)}\,. \tag{38}$$

We do not necessarily know if u_m is (α,ε)-good itself, and that is why it is not necessarily the case that the two bounds on r_m are up to $(1+\mathsf{O}(\varepsilon))$ apart from each other. However, we will use the fact that v_j is (α,ε)-good and use property (c) of this definition.

By (37), the probability that all other edges emanating from v_j are closed, conditioned on this already occurring for $v_1, \dots, v_{j-1}$, is

$$\prod_{m=1}^{\deg(v_j)} \Big[1 - (1 + \mathsf{O}(\varepsilon))\big(r_m + \frac{1}{\sum_{q=j}^{\ell} \deg(v_q) - (m-1)}\big)\Big],$$

which equals

$$\exp\Big(-(1+\mathsf{O}(\varepsilon))\sum_m r_m\Big) \cdot \exp\Big(-(1+\mathsf{O}(\varepsilon))\sum_m \frac{1}{\sum_{q=j}^{\ell} \deg(v_q) - (m-1)}\Big). \tag{39}$$

Since v_j is (α,ε)-good, by property (c) of the definition we learn that

$$\sum_m r_m = (1+\mathsf{O}(\alpha^{1/2}))\sum_m \frac{1}{\deg(u_m)} = (1+\mathsf{O}(\alpha^{1/2}))b(v_j)\,.$$

Property (d) of the same definition asserts that $b(v_j) \leqslant \alpha^{-1/4}$, hence the first term in (39) is

$$\exp\Big(-(1+\mathsf{O}(\varepsilon))\sum_m r_m\Big) = (1+\mathsf{O}(\alpha^{1/4}))e^{-b(v_j)}\,.$$

To handle the second term of (39) we note that

$$\sum_m \frac{1}{\sum_{q=j}^{\ell} \deg(v_q) - (m-1)} = \log\Big(\frac{\sum_{q=j} \deg(v_q)}{\sum_{q=j+1} \deg(v_q)}\Big) + \mathsf{O}((\varepsilon n)^{-1})\,.$$

Since $\{v_1, \dots, v_\ell\}$ are (α,ε)-good we have that the ratio inside the logarithm is at most $1 + \ell\varepsilon^{-1}$ and so the second term of (39) equals

$$(1+\mathsf{O}(\varepsilon\log(\varepsilon^{-1})))\frac{\sum_{q=j+1}^{\ell} \deg(v_q)}{\sum_{q=j}^{\ell} \deg(v_q)}\,.$$

We multiply these two terms of (39) and get that (36) holds. □

Using Lemma 3.7 we may estimate the probability that $B_{\mathcal{T}}(X,r) \cong T$ "purely", that is, that $E(B_{\mathcal{T}}(X,r)) \cap E(G[V_i])$ is a rooted tree that is isomorphic to T. We denote this event by $B_{\mathcal{T}}(X,r) \overset{\text{pure}}{\cong} T$. Note that $E(B_{\mathcal{T}}(X,r)) \cap E(G[V_i])$ need not even be a tree, so the events $B_{\mathcal{T}}(X,r) \overset{\text{pure}}{\cong} T$ and $B_{\mathcal{T}}(X,r) \cong T$ can be very different.

Corollary 3.8. *Assume the setting as introduced at the beginning of Section 3.3, and let $\mathcal{T}$ be a UST of G. Fix $i \in [k]$ and let X be a uniformly chosen random vertex of V_i. Then*

$$\mathbb{P}(B_{\mathcal{T}}(X,r) \overset{\text{pure}}{\cong} T) = (1+\mathsf{O}(\ell\alpha^{1/4}))\mathrm{Freq}(T;G,i)\,,$$

where $\mathrm{Freq}(T;G,i)$ *is defined in Definition 3.6.*

Proof. This is immediate since the event $E(B_{\mathcal{T}}(X,r)) \cap E(G[V_i]) \cong T$ is the union over all i-pure tuples $(v_1,\dots,v_\ell)$ of the event $B_{\mathcal{T}}(v_1,r) \overset{\text{pure}}{\cong} T(v_1,\dots,v_\ell)$. These events are not disjoint, since for each automorphism of τ of T that fixes the root, we can permute $(v_1,\dots,v_\ell)$ according to τ and get the identical event. However, up to this invariance, the events are disjoint (which explains the factor of $|\mathrm{Stab}_T|^{-1}$ in the definition of $\mathrm{Freq}(T;G,i)$) and so Lemma 3.7 gives the proof. □

Corollary 3.8 is an annealed statement, that is, the probability space is the product of the UST probability measure and an independent uniform vertex X. A quenched statement follows by a second moment argument. We will first need the following assertion.

Lemma 3.9. *Suppose that G is a connected graph on n vertices and $S \subseteq V(G)$. Let $\mathcal{T}$ be a spanning tree of G and let X be a uniformly chosen vertex of some set $A \subseteq V(G)$. Then for any fixed rooted tree T of height r we have that*

$$\mathbb{P}\left(B_{\mathcal{T}}(X,r) \cong T \text{ and } V(B_{\mathcal{T}}(X,r)) \cap S \neq \varnothing\right) \leqslant |T|^2|A|^{-1}|S|\,.$$

Furthermore, if in addition G has some decomposition $V(G) = V_0 \sqcup V_1 \sqcup \ldots \sqcup V_k$, then for each $i \in [k]$,

$$\mathbb{P}\left(B_{\mathcal{T}}(X,r) \overset{\text{pure}}{\cong} T \text{ and } V(B_{\mathcal{T}}(X,r)) \cap V_i \cap S \neq \varnothing\right) \leqslant |T|^2|A|^{-1}|S|\,.$$

Proof. We prove only the first statement; the second follows by the same logic. Fix $v \in S$. For each vertex q of T, if $\mathcal{T}$ contains an induced copy of T such that v takes the place of q, then the probability that $B_{\mathcal{T}}(X,r) = T$ such that v takes the place of q is at most $\deg_T(q)/|A|$ since once we choose which edge of $\mathcal{T}$ that touches v corresponds to the edge of T that touches q towards the root of T, then the corresponding root in $\mathcal{T}$ is chosen and X has to be chosen precisely to be that root. We bound this probability by $|T|/|A|$ and use the union bound over the vertices q of T and v of S. □

We are now ready to prove the quenched version of Corollary 3.8. Observe that without the assumption that V_i is (α,ε)-big $\mathrm{Freq}(T;G,i)$ can be very small or event 0 (for instance, V_i could have size less than $\varepsilon^5 n$ and if all the vertices of V_0 are connected to each vertex of V_i, then V_i has no (α,ε)-good vertices at all). Thus, we cannot hope to have the have the concentration required for the following quenched statement without this assumption.

Lemma 3.10. *Assume the setting as introduced at the beginning of Section 3.3. Let $\mathcal{T}$ be a UST of G. Fix $i \in [k]$ and assume that V_i is (α,ε)-big. Let X be a uniformly chosen vertex of V_i. Then with probability at least $1 - \mathsf{O}(\ell\alpha^{1/8})$ the random tree $\mathcal{T}$ is such that*

$$\mathbb{P}\left(B_{\mathcal{T}}(X,r) \overset{\text{pure}}{\cong} T\right) = (1 + \mathsf{O}(\alpha^{1/16}))\mathrm{Freq}(T;G,i)\,.$$

Proof. We write by $Z = Z(\mathcal{T})$ the $\mathcal{T}$-measurable random variable counting the number of vertices v of V_i such that $B_{\mathcal{T}}(v,r) \overset{\text{pure}}{\cong} T$. Corollary 3.8 is equivalent to the assertion that

$$\mathbb{E}Z = (1 + \mathsf{O}(\ell\alpha^{1/4}))\text{Freq}(T;G,i)\cdot|V_i|\,. \tag{40}$$

The second moment of Z is

$$\mathbb{E}Z^2 = \sum_{v,v'\in V_i} \mathbb{P}\left(B_{\mathcal{T}}(v,r) \overset{\text{pure}}{\cong} T \text{ and } B_{\mathcal{T}}(v',r) \overset{\text{pure}}{\cong} T\right)\,. \tag{41}$$

We will show that for any $v \in V_i$,

$$\sum_{v'\in V_i} \mathbb{P}\left(B_{\mathcal{T}}(v',r) \overset{\text{pure}}{\cong} T \mid B_{\mathcal{T}}(v,r) \overset{\text{pure}}{\cong} T\right) \leqslant (1 + \mathsf{O}(\ell\alpha^{1/4}))\text{Freq}(T;G,i)|V_i|\,. \tag{42}$$

If we have this, then since Corollary 3.8 gives that

$$\sum_{v\in V_i} \mathbb{P}\left(B_{\mathcal{T}}(v,r) \overset{\text{pure}}{\cong} T\right) = (1 + \mathsf{O}(\ell\alpha^{1/4}))\text{Freq}(T;G,i)\cdot|V_i|\,,$$

by putting both in (41) we get that

$$\mathbb{E}Z^2 = (1 + \mathsf{O}(\ell\alpha^{1/4}))\big[\text{Freq}(T;G,i)\cdot|V_i|\big]^2\,.$$

Hence

$$\text{Var}(Z) = \mathsf{O}(\ell\alpha^{1/4})\big[\text{Freq}(T;G,i)\cdot|V_i|\big]^2\,,$$

and so by Chebychev's inequality we learn that

$$\mathbb{P}\left(|Z - \mathbb{E}Z| \geqslant \alpha^{1/16}\text{Freq}(T;G,i)\cdot|V_i|\right) \leqslant \mathsf{O}(\ell\alpha^{1/8})\,,$$

concluding the proof.

To prove (42) we fix $v \in V_i$ and condition on the event $B_{\mathcal{T}}(v,r) \overset{\text{pure}}{\cong} T$ and on the vertices $(v_1,\ldots,v_\ell)$ which form $B_{\mathcal{T}}(v,r)$. By Proposition 3.1 we contract the edges of $\mathcal{T}$ in $B_{\mathcal{T}}(v,r)$ and erase the edges we conditioned that are not in $\mathcal{T}$. Denote the resulting multigraph by G^* and by v^* the vertex to which the vertices $v_1,\ldots,v_\ell$ have been identified. We consider the loopless multigraph G^* with the partition

$$V(G^*) = V_0 \sqcup V_1 \sqcup \ldots \sqcup V_{i-1} \sqcup V_i^* \sqcup V_{i+1} \sqcup \ldots \sqcup V_k\,,$$

where V_i^* is V_i with the vertices $v_1,\ldots,v_\ell$ replaced by v^*. We now claim that this partition is a $(\beta,\alpha,\gamma/2,\varepsilon)$-good-decomposition (as in Definition 2.10). First, by Proposition 2.2 the graph $G^*[V_i^*]$ is still a $\gamma/2$-expander and so the partition $V_0 \sqcup V_1 \sqcup \ldots \sqcup V_{i-1} \sqcup V_i^* \sqcup V_{i+1} \sqcup \ldots \sqcup V_k$ is a $(\gamma/2,\varepsilon^5,\varepsilon^5)$-expander decomposition of G^*. Next we verify that the number of good vertices has not dropped by too much. Indeed, from each vertex $u \in V_i\backslash\{v_1,\ldots,v_\ell\}$ at most ℓ edges touching it were erased and hence it is immediate to check in Definition 2.9 that, as long as n is large enough (in terms of ε and α), if u was (α,ε)-good in G, then

it is (α,ε)-good in G^* — the constants may have changed, but they are swallowed in the $\mathsf{O}(\cdot)$ and $\Omega(\cdot)$ notation of Definition 2.9.

Thus we may apply Corollary 3.8 and obtain that

$$\sum_{v'\in V_i^*}\mathbb{P}\left(B_{\mathcal{T}}(v',r)\overset{\text{pure}}{\cong}T\mid G^*\right)=(1+\mathsf{O}(\ell\alpha^{1/4}))\mathrm{Freq}(T;G^*,i)\cdot|V_i^*|\,. \tag{43}$$

We wish to bound the sum in (42) by the above sum, however, there are two important differences between the sums. The first is the difference in the input to the Freq function.

Claim 3.10.1. *If V_i is (α,ε)-big, then*

$$c\gamma^{\ell+1}e^{-\ell/\gamma}\leqslant\mathrm{Freq}(T;G,i)\leqslant\ell\gamma^{-\ell-1}\,,$$

for some universal constant $c>0$.

Proof. Since V_i is (α,ε)-big we have at least $(1-\mathsf{O}(\alpha^{1/8}))|V_i|$ vertices which are (α,ε)-good and these must span at least $\Omega(\alpha^{1/9}|V_i|n)$ edges, since the number of edges of $G[V_i]$ touching non (α,ε)-good vertices of V_i is at most $\mathsf{O}(\alpha^{1/8}|V_i|n)$ and we also have $e(G[V_i])=\Omega(\alpha^{1/9}|V_i|n)$. Thus the average degree of this graph is $d=\Omega(\alpha^{1/9}n)$, and, by greedily removing vertices of degree $d/4$ we can obtain a subgraph of $G[V_i]$ of minimal degree $\Omega(\alpha^{1/9}n)=\Omega(\gamma n)$ such that all of its vertices are (α,ε)-good. Thus it is immediate to find $\Omega(\gamma^\ell n^\ell)$ copies of the tree T. We deduce that the number of ℓ-tuples counted in $\mathrm{Freq}(T;G,i)$ is at least $\Omega(\gamma^\ell n^\ell)$ and at most n^ℓ. Since each vertex degree in V_i is at most n and at least γn we learn that each ℓ-tuple contributes to $\mathrm{Freq}(T;G,i)$ at most $\ell\gamma^{-\ell-1}n^{-\ell}$ and at least $\gamma e^{-\ell/\gamma}n^{-\ell}$ and the claim follows. □

By this claim, since V_i and V_i^* differ by $\ell-1$ vertices, and ℓ-tuples of V_i which one of vertices is in $\{v_1,\dots,v_\ell\}$ can contribute at most $\mathsf{O}(\ell\gamma^{-\ell-1}n^{-1})$ to $\mathrm{Freq}(T;G,i)$ we learn that

$$\mathrm{Freq}(T;G,i)=(1+\mathsf{O}(\ell\gamma^{-\ell-1}n^{-1}))\mathrm{Freq}(T;G^*,i)\,, \tag{44}$$

and n can be taken large enough to that the error in the O-notation above is $\mathsf{O}(\ell\alpha^{1/4})$. This handles the first difference.

The second difference is that $B_{\mathcal{T}}(v',r)$ may be isomorphic to T in G by using some of the edges in $T(v_1,\dots,v_\ell)$ that were contracted to v^* in G^*. In this case, it does *not* necessarily hold that $B_{\mathcal{T}}(v',r)\cong T$ in G^*, so it is possible that this contribution to (42) is large and not counted for in (43). We will show that this is not the case.

We bound the LHS of (42) as follows. The terms corresponding to $v'\in\{v_1,\dots,v_\ell\}$ we bound by 1 and get a negligible contribution of ℓ. For any other v' we split the event $B_{\mathcal{T}}(v',r)\overset{\text{pure}}{\cong}T$ according to whether $V(B_{\mathcal{T}}(v',r))\cap V(B_{\mathcal{T}}(v,r))=\varnothing$. This intersection

is empty if and only if $v^* \notin V(B_{\mathcal{T}}(v',r))$ in the graph G^*, and if it is empty, then it holds that $B_{\mathcal{T}}(v',r) \overset{\text{pure}}{\cong} T$ in G^*. Thus the LHS of (42) is at most

$$\ell + \sum_{v' \in V_i^*} \mathbb{P}\left(B_{\mathcal{T}}(v',r) \overset{\text{pure}}{\cong} T \mid G^*\right) + \sum_{v' \in V_i^*} \mathbb{P}\left(B_{\mathcal{T}}(v',r) \overset{\text{pure}}{\cong} T \text{ and } v^* \in V(B_{\mathcal{T}}(v',r)) \mid G^*\right).$$

The first term in the above is negligible, the second we bound using (43) and (44) by the RHS of (42). The third term we bound using Lemma 3.9 applied on the graph G^* with $S = \{v^*\}$ and $A = V_i^*$, giving the bound

$$\sum_{v' \in V_i^*} \mathbb{P}\left(B_{\mathcal{T}}(v',r) \overset{\text{pure}}{\cong} T \text{ and } v^* \in V(B_{\mathcal{T}}(v',r)) \;\middle|\; G^*\right) \leqslant \ell^2,$$

which is also negligible. This completes the proof of (42) and concludes the proof of the lemma. □

We now turn to estimating the event $B_{\mathcal{T}}(X,r) \cong T$ rather than the "pure" version of this event. To that aim we define

$$\mathcal{O} = \left\{(x,y) \in \mathcal{T} : i(x) \neq i(y) \text{ or } x \in V_0 \text{ or } y \in V_0\right\}, \tag{45}$$

that is, $\mathcal{O}$ is the set of edges of $\mathcal{T}$ that are between components or contained in V_0. We first assert that this set cannot be too large.

Lemma 3.11. *Assume the setting as introduced at the beginning of Section 3.3. Let $\mathcal{T}$ be a UST of G and $\mathcal{O} \subseteq \mathcal{T}$ be defined in* (45)*. Then*

$$\mathbb{E}|\mathcal{O}| = \mathsf{O}(\alpha^{1/4} n).$$

Proof. Let us assume that n is large enough as in Lemma 3.4. By Lemma 3.4 and Kirchoff's formula (24) we have that

$$\mathbb{E}|\mathcal{T} \cap E(G[V_i])| \geqslant (1 - \mathsf{O}(\varepsilon)) \frac{1}{2} \sum_{u \in V_i} \sum_{v \in V_i : v \sim u} \left(\frac{1}{\deg(u)} + \frac{1}{\deg(v)}\right) = (1 - \mathsf{O}(\varepsilon)) \sum_{u \in V_i} \frac{\deg(u; V_i)}{\deg(u)}.$$

Hence

$$\mathbb{E}\left|\mathcal{T} \cap \bigcup_{i=1}^{k} E(G[V_i])\right| \geqslant (1 - \mathsf{O}(\varepsilon)) \sum_{u \in V \setminus V_0} \frac{\deg(u; V_{i(u)})}{\deg(u)}.$$

For any u that is (α, ε)-good we have by part (b) of Definition 2.9 that $\deg(u; V_{i(u)}) \geqslant (1 - \varepsilon^2)\deg(u)$. Since at least $(1 - \mathsf{O}(\alpha^{1/4}))n$ vertices are (α, ε)-good we deduce that

$$\mathbb{E}\left|\mathcal{T} \cap \bigcup_{i=1}^{k} E(G[V_i])\right| \geqslant (1 - \mathsf{O}(\alpha^{1/4}))n.$$

The proof is now completed since $|\mathcal{T}| = n - 1$ with probability 1. □

We now reach our final destination.

Lemma 3.12. *Assume the setting as introduced at the beginning of Section 3.3. Let $\mathcal{T}$ be a UST of G and let X be a uniformly chosen vertex of G. Then with probability at least $1-\mathsf{O}(\ell\alpha^{1/8})$ for the random tree $\mathcal{T}$ we have*

$$\mathbb{P}\left(B_{\mathcal{T}}(X,r)\cong T\right)=(1+\mathsf{O}(\ell^2\alpha^{1/16}))\mathrm{Freq}(T;G)+\mathsf{O}(\ell^2\beta^{1/8})\,.$$

Proof. Let $\mathcal{O}$ be defined in (45). We apply Lemma 3.10, together with Lemma 3.11 and Markov's inequality, and get that with probability at least $1-\mathsf{O}(\ell\alpha^{1/8})$ the random tree $\mathcal{T}$ satisfies

(1) $|\mathcal{O}|\leqslant\alpha^{1/8}n$, and
(2) For each $i\in[k]$ for which V_i is (α,ε)-big we have that

$$\mathbb{P}\left(B_{\mathcal{T}}(X,r)\overset{\text{pure}}{\cong}T\right)=(1+\mathsf{O}(\ell\alpha^{1/16}))\mathrm{Freq}(T;G,i)\,,$$

where X is a uniformly chosen vertex of V_i.

Let X be a uniformly chosen vertex of G. By Proposition 2.12 the probability that X is in either V_0 or in some V_i that is *not* (α,ε)-big is at most $\mathsf{O}(\beta^{1/8})$. Conditioned on $X\in V_i$ we have that X is uniform random vertex of V_i. Hence we may sum item (2) above over these i's and get that with probability at least $1-\mathsf{O}(\ell\alpha^{1/8})$ the random tree $\mathcal{T}$ satisfies

$$\mathbb{P}(B_{\mathcal{T}}(X,r)\overset{\text{pure}}{\cong}T)=(1+\mathsf{O}(\ell\alpha^{1/16}))\mathrm{Freq}(T;G)+\mathsf{O}(\beta^{1/8})\,,$$

by definition of $\mathrm{Freq}(T;G)$.

Assume $\mathcal{T}$ satisfies these. If $B_{\mathcal{T}}(X,r)\cong T$ but not $B_{\mathcal{T}}(X,r)\overset{\text{pure}}{\cong}T$, then we must have that $B_{\mathcal{T}}(X,r)\cong T$ and $B_{\mathcal{T}}(X,r)\cap V(\mathcal{O})\neq\varnothing$. Similarly, if $B_{\mathcal{T}}(X,r)\overset{\text{pure}}{\cong}T$ but not $B_{\mathcal{T}}(X,r)\cong T$, then we similarly must have that $B_{\mathcal{T}}(X,r)\overset{\text{pure}}{\cong}T$ and $B_{\mathcal{T}}(X,r)\cap V(\mathcal{O})\neq\varnothing$. Since $|\mathbb{P}(A)-\mathbb{P}(B)|\leqslant\mathbb{P}(A\backslash B)+\mathbb{P}(B\backslash A)$ we have that

$$\left|\mathbb{P}\left(B_{\mathcal{T}}(X,r)\cong T\right)-\mathbb{P}\left(B_{\mathcal{T}}(X,r)\overset{\text{pure}}{\cong}T\right)\right|=\mathsf{O}(\ell^2\beta^{1/8})\,,$$

where the last inequality is due to Lemma 3.9 with $A=V(G)$ and $S=V(\mathcal{O})$. □

4. Proof of Theorems 1.1 and 1.3

4.1. **Deriving Theorem 1.1 from Lemma 3.12 via anatomies.** In this section, we deduce Theorem 1.1 from Lemma 3.12. The main technical result of this section, Lemma 4.3, says that the quantity Freq is continuous in a certain robust sense. To prove Lemma 4.3, we need to group the elements of Ω for a given graphon $W:\Omega^2\to[0,1]$ to groups with similar degrees. Actually, we need a recursive refinement of this, as follows. We call the above partition of Ω according to the degrees *anatomy of depth 1*. Now, having defined an anatomy of depth d, *anatomy of depth* $d+1$ is a decomposition of Ω into groups in which elements have approximately similar degrees into every individual cell of the anatomy of depth d. Let us make this precise.

Suppose that $h \in \mathbb{N}$ and S is a finite set. Let $\mathfrak{C}(h,S)$ be the partition of $[0,1]^S$ into Voronoi cells generated by points $p \in \{0, \frac{1}{h}, \dots, \frac{h-1}{h}, 1\}^S$ (we assign each boundary point to one arbitrary neighboring cell to break ties). Note that these cells form $(h+1)^{|S|}$ cubes of the form $\prod_{i\in S}\left[\max\{0, \frac{2r_i-1}{2h}\}, \min\{1, \frac{2r_i+1}{2h}\}\right]$, for some $\left\{r_i \in \{0,\dots,h\}\right\}_{i\in S}$.[5] In the degenerate case $S = \varnothing$, we define $\mathfrak{C}(h,S) := \{\varnothing\}$. We call the points p the *centers of the cells*. When $C \in \mathfrak{C}(h,S)$ is a Voronoi cell with center $p = (p_i)_{i\in S}$, for $i \in S$ we write $center_i(C) := p_i$.

Let us write $\mathtt{b}$ for an (abstract) element. Below, the only purpose of $\mathtt{b}$ will be to refer to a coordinate that will have to do with the function $b_W(\cdot)$.

Now, suppose that $d \in \mathbb{N}$ and $\mathbf{h} \in \mathbb{N}^d$. For $t \in [d]$, we write $\mathbf{h}_t$ for the t-coordinate of $\mathbf{h}$, and $\mathbf{h}[\![t]\!]$ for the t-dimensional vector obtained from $\mathbf{h}$ by removing the last $d-t$ components. For $h \in \mathbb{N}$, let $\mathfrak{D}_{h,0} := \mathfrak{C}(h,\varnothing)$ and for $d \in \mathbb{N}$ and $\mathbf{h} \in \mathbb{N}^d$ let $\mathfrak{D}_{\mathbf{h}} := \mathfrak{C}\left(\mathbf{h}_d, \mathfrak{D}_{\mathbf{h}[\![d-1]\!]} \sqcup \{\mathtt{b}\}\right)$. We have

$$\sum_{t=0}^{d} \left|\mathfrak{D}_{\mathbf{h}[\![\mathbf{t}]\!]}\right| = \hbar(\mathbf{h}) \,, \tag{46}$$

for a suitable tower-function $\hbar(\cdot)$.

Suppose $W : \Omega^2 \to [0,1]$ is a graphon and $h \in \mathbb{N}$ is arbitrary. Let $\mathcal{A}_\varnothing := \Omega$. We call $\{\mathcal{A}_\varnothing\} = \{\mathcal{A}_C\}_{C\in\mathfrak{D}_{h,0}}$ *anatomy of W of depth 0*. Suppose that $d \in \mathbb{N}$, $\mathbf{h} \in \mathbb{N}^d$ and that we already know the anatomy $\{\mathcal{A}_C\}_{C\in\mathfrak{D}_{\mathbf{h}[\![d-1]\!]}}$ of W of depth $d-1$, which is a partition of Ω. Now, for each $\omega \in \Omega$ we consider the $|\mathfrak{D}_{\mathbf{h}[\![d-1]\!]}|$-tuple of degrees

$$\mathbf{deg}_d(\omega) := (\deg_W(\omega, \mathcal{A}_C))_{C\in\mathfrak{D}_{\mathbf{h}[\![d-1]\!]}} \,.$$

Then for each $F \in \mathfrak{D}_{\mathbf{h}}$ we define $\mathcal{A}_F := (\mathbf{deg}_d)^{(-1)}(F) \cap (\exp(-b_W(\cdot)))^{(-1)}(F)$. In words, each cell $\mathcal{A}_F \subseteq \Omega$ has the property that for each $C \in \mathfrak{D}_{\mathbf{h}[\![d-1]\!]}$ and for each $\omega \in \mathcal{A}_F$ we have that

$$\deg_W(\omega, \mathcal{A}_C) = center_C(F) \pm \tfrac{1}{2\mathbf{h}_d} \quad \text{and} \quad \exp(-b_W(\omega)) = center_{\mathtt{b}}(F) \pm \tfrac{1}{2\mathbf{h}_d} \,. \tag{47}$$

We call $\{\mathcal{A}_F\}_{F\in\mathfrak{D}_{\mathbf{h}}}$ the *anatomy of W of depth d*. Obviously, $\{\mathcal{A}_F\}_{F\in\mathfrak{D}_{\mathbf{h}}}$ is a partition of Ω. Consider an arbitrary $\omega \in \mathcal{A}_F$. Summing (47) over all $C \in \mathfrak{D}_{\mathbf{h}[\![d-1]\!]}$ (for which (46) tells us that there are $\hbar(\mathbf{h}[\![d-1]\!])$ summands), we get

$$\deg_W(\omega) = \sum_{C\in\mathfrak{D}_{\mathbf{h}[\![d-1]\!]}} center_C(F) \;\pm \tfrac{\hbar(\mathbf{h}[\![d-1]\!])}{2\mathbf{h}_d} \,. \tag{48}$$

For this reason, we shall call the number $\sum_{C\in\mathfrak{D}_{\mathbf{h}[\![d-1]\!]}} center_C(F)$ the *degree of $\mathcal{A}_F$* (in W), and denote it by $\deg_W^{\mathrm{anat}}(\mathcal{A}_F)$.

[5] Strictly speaking, when some but not all coordinates r_i are 0 or h, these are not cubes but rectangular prisms. This is however not important.

Suppose $U : \Omega^2 \to [0,1]$ and $X \subseteq \Omega$. Suppose that $d \in \mathbb{N}$, $\mathbf{h} \in \mathbb{N}^d$, and $\kappa \geqslant 0$ are given. Let $\mathcal{B}_\varnothing := \Omega$ and for each $t \in [d]$ let $\{\mathcal{B}_C\}_{C \in \mathfrak{D}_{\mathbf{h}[\![t]\!]}}$ be a partition of Ω. Suppose that for each $t \in [d]$, each $F \in \mathfrak{D}_{\mathbf{h}[\![t]\!]}$, each $C \in \mathfrak{D}_{\mathbf{h}[\![t-1]\!]}$ and each $\omega \in \mathcal{B}_F \backslash X$ we have

(49)
$$|\deg_U(\omega, \mathcal{B}_C) - center_C(F)| \leqslant \tfrac{1}{2\mathbf{h}_t} + \kappa \quad \text{and} \quad |\exp(-b_U(\omega)) - center_{\mathfrak{b}}(F)| \leqslant \tfrac{1}{2\mathbf{h}_d} + \kappa \,.$$

We then say that $\left\{\{\mathcal{B}_F\}_{F \in \mathfrak{D}_{\mathbf{h}[\![t]\!]}}\right\}_{t=0}^{d}$ are *κ-approximate anatomies of U up to depth d with exceptional set X*.

Note that when we take $\kappa = 0$ and $X = \varnothing$ then we recover the notion of anatomies.

Our next lemma says that two graphons that are close in the cut-distance have similar anatomies.

Lemma 4.1. *Suppose that $W : \Omega^2 \to [0,1]$ is a nondegerate graphon and $d \in \mathbb{N}$, $\mathbf{h} \in \mathbb{N}^d$ are arbitrary. Let $\left\{\{\mathcal{A}_F\}_{F \in \mathfrak{D}_{\mathbf{h}[\![t]\!]}}\right\}_{t=0}^{d}$ be the anatomies of W up to depth d. For every $\kappa > 0$ there exists a number $\delta > 0$ such that the following holds for every graphon $U : \Omega^2 \to [0,1]$ with $\|W - U\|_\square < \delta$. There exists a set $X \subseteq \Omega$ of measure at most κ so that $\left\{\{\mathcal{A}_F\}_{F \in \mathfrak{D}_{\mathbf{h}[\![t]\!]}}\right\}_{t=0}^{d}$ are κ-approximate anatomies of U up to depth d with exceptional set X.*

Proof. Suppose that we are given W, d, $\mathbf{h}$, and κ as above.

Since W is nondegenerate, there exists $\beta \in (0, 10^{-6})$ such that the measure of the set

$$S := \left\{\omega \in \Omega : \deg(\omega) < 16\sqrt[4]{\beta}\right\}$$

is less than $\frac{\kappa^2}{100}$. Let $\delta := \min\{\frac{\beta\kappa^4}{400}, \frac{\kappa^2}{4\hbar(\mathbf{h})^2}\}$. Suppose that U is given with $\|W - U\|_\square < \delta$. It is enough to prove that there exists a set $X_{\mathfrak{b}} \subseteq \Omega$ of measure at most $\frac{\kappa}{4}$ such that for each $t \in [d]$ and each $F \in \mathfrak{D}_{\mathbf{h}[\![t]\!]}$, we have for each $\omega \in F \backslash X_{\mathfrak{b}}$ that

$$|\exp(-b_U(\omega)) - center_{\mathfrak{b}}(F)| \leqslant \tfrac{1}{2\mathbf{h}_d} + \kappa \,, \tag{50}$$

and that (with t and F as above) for each $C \in \mathfrak{D}_{\mathbf{h}[\![t-1]\!]}$ we have that all but at most $\frac{\kappa}{4\hbar(\mathbf{h})^2}$ measure of elements $\omega \in \mathcal{A}_F$ satisfy

$$\deg_U(\omega, \mathcal{A}_C) \geqslant center_C(F) - \frac{1}{2\mathbf{h}_t} - \kappa \,, \tag{51}$$

and all but at most $\frac{\kappa}{4\hbar(\mathbf{h})^2}$ measure of elements $\omega \in \mathcal{A}_F$ satisfy

$$\deg_U(\omega, \mathcal{A}_C) \leqslant center_C(F) + \frac{1}{2\mathbf{h}_t} + \kappa \,. \tag{52}$$

The lemma will then follow from (46) by taking X to be the union of $X_{\mathfrak{b}}$ together with the exceptional sets from (51), (52) over all t, C, and F.

The following claim clearly implies (50).

Claim 4.1.1. *There exists a set $X_\mathfrak{b}$ of measure at most $\frac{\kappa}{4}$ such that for all $\omega \in \Omega \backslash X_\mathfrak{b}$ we have $|b_W(\omega) - b_U(\omega)| < \kappa$.*

For the proof of Claim 4.1.1, we need Claims 4.1.2–4.1.4 below.

Claim 4.1.2. *Suppose that $\Gamma : \Omega^2 \to [0,1]$ is a graphon and that $A \subseteq \Omega$. Then*

$$\int_{x\in\Omega}\int_{\omega\in A} \frac{\Gamma(x,\omega)}{\deg_\Gamma(\omega)} \leqslant \mu(A).$$

Proof. By Fubini's Theorem, we have

$$\int_{x\in\Omega}\int_{\omega\in A} \frac{\Gamma(x,\omega)}{\deg_\Gamma(\omega)} = \int_{\omega\in A} \frac{1}{\deg_\Gamma(\omega)} \int_{x\in\Omega} \Gamma(x,\omega) = \int_{\omega\in A} \frac{1}{\deg_\Gamma(\omega)} \deg_\Gamma(\omega, A) \leqslant \mu(A)\ .$$

□

Claim 4.1.3. *Suppose $\Gamma_1, \Gamma_2 : \Omega^2 \to [0,1]$ are two graphons and that $f : \Omega \to [0, C]$ is an arbitrary function. Then*

$$\int_{x\in\Omega} \left| \int_{\omega\in\Omega} f(\omega)(\Gamma_1(x,\omega) - \Gamma_2(x,\omega)\right| \leqslant 2C \cdot \|\Gamma_1 - \Gamma_2\|_\square\ .$$

Proof. By [15, (8.20)] we can alternatively express as $\|\Gamma_1 - \Gamma_2\|_\square$

$$\begin{aligned}\|\Gamma_1 - \Gamma_2\|_\square &= \sup_{F,G:\Omega\to[0,1]} \left|\int_x\int_y G(x)F(y)(\Gamma_1(x,y) - \Gamma_2(x,y))\right| = \\ &= \frac{1}{C} \cdot \sup_{F:\Omega\to[0,C],G:\Omega\to[0,1]} \left|\int_x\int_y G(x)F(y)(\Gamma_1(x,y) - \Gamma_2(x,y))\right|\ . \\ &\geqslant \frac{1}{2C} \cdot \sup_{F:\Omega\to[0,C],G:\Omega\to[-1,1]} \left|\int_x\int_y G(x)F(y)(\Gamma_1(x,y) - \Gamma_2(x,y))\right|\ .\end{aligned}$$

The claim follows by considering in the latter supremum functions

$$F(\cdot) := f(\cdot) \quad \text{and} \quad G(\cdot) := \operatorname{sgn}\left(\int_y f(y)(\Gamma_1(\cdot,y) - \Gamma_2(\cdot,y))\right)\ .$$

□

For the last auxiliary claim, we shall introduce an auxiliary notion. Below we shall work with not necessarily symmetric L^1-functions $K : \Omega^2 \to [0,+\infty)$. Note that the notions of cut-norm and cut-distance extend to this setting.[6] The following claim is then obvious.

Claim 4.1.4. *Suppose that $\varepsilon > 0$ and $\Gamma, \Gamma' : \Omega^2 \to [0,1]$ are two L^1-functions such that for each $x, y \in \Omega$, $\max(\Gamma(x,y), \Gamma'(x,y)) < (1+\varepsilon)\min(\Gamma(x,y), \Gamma'(x,y))$. Then $\|\Gamma - \Gamma'\|_\square \leqslant \varepsilon$.*

□

[6]Note that in this case, the cut-norm really has to be defined over all rectangles in (6), and not only over all squares.

Proof of Claim 4.1.1. Let

$$D := \left\{\omega \in \Omega\backslash S : \deg_W(\omega) \neq \left(1 \pm \tfrac{\beta^{0.3}\kappa^2}{4}\right)\deg_U(\omega)\right\} .$$

Since for each $\omega \in D$ we have $|\deg_W(\omega)-\deg_U(\omega)| \geqslant 3\beta^{0.55}\kappa^2$ we get from $\|W-U\|_\square < \frac{\beta\kappa^4}{400}$ that $\mu(D) \leqslant \frac{\kappa^2}{100}$.

Put $A := S \cup D$. We have $\mu(A) \leqslant \frac{\kappa^2}{50}$.

We have

$$\begin{aligned}&\int_{x\in\Omega} |b_W(x) - b_U(x)| \\ &\leqslant \int_{x\in\Omega}\left|\int_{\omega\in\Omega\backslash A}\frac{W(x,\omega)}{\deg_W(\omega)} - \frac{U(x,\omega)}{\deg_U(\omega)}\right| + \int_{x\in\Omega}\left|\int_{\omega\in A}\frac{W(x,\omega)}{\deg_W(\omega)} - \frac{U(x,\omega)}{\deg_U(\omega)}\right|\end{aligned} \tag{53}$$

The second term on the right-hand side of (53) is bounded by Claim 4.1.2 by at most $2\mu(A) \leqslant \kappa^2/50$. Now, let us consider the L^1-function $\tilde{U}$, defined by

$$\tilde{U}(x,y) := \begin{cases}\frac{\deg_W(y)}{\deg_U(y)}\cdot U(x,y) & \text{if } y \notin A \\ U(x,y) & \text{if } y \in A\end{cases} .$$

Observe that U and $\tilde{U}$ satisfy the assumptions of Claim 4.1.4 with error parameter $\left(\frac{\beta^{0.3}\kappa^2}{4}\right)$. Then the first term of the right-hand side of (53) can be rewritten as

$$\int_{x\in\Omega}\left|\int_{\omega\in\Omega}\frac{\mathbf{1}_{\Omega\backslash A}(\omega)}{\deg_W(\omega)}\cdot\left(W(x,\omega)-\tilde{U}(x,\omega)\right)\right| . \tag{54}$$

Observe that the function $\frac{\mathbf{1}_{\Omega\backslash A}(\cdot)}{\deg_W(\cdot)}$ is bounded by above by $\frac{1}{16\sqrt[4]{\beta}}$. Claim 4.1.3 tells us that (54) is at most

$$\frac{1}{8\sqrt[4]{\beta}}\cdot\left(\|W-U\|_\square + \|U-\tilde{U}\|_\square\right) \overset{\text{C4.1.4}}{\leqslant} \frac{1}{3200}\beta^{3/4}\kappa^2 + \frac{\beta^{0.3}\kappa^2}{4\cdot 8\beta^{1/4}} < \frac{\kappa^2}{25} .$$

Plugging this back into (53) we obtain that $\int_{x\in\Omega}|b_W(x)-b_U(x)| < \frac{\kappa^2}{4}$. The claim follows. □

It now remains to prove (51) and (52). We will only prove (51) since the proof of (52) is verbatim. So, suppose that (51) fails, i.e., the set $X := \{\omega \in \mathcal{A}_F : \deg_U(\omega,\mathcal{A}_C) > center_C(F) + \frac{1}{2\mathbf{h}_t} + \kappa\}$ satisfies $\mu(X) > \frac{\kappa}{2\hbar(\mathbf{h})}$. By the definition of $\mathcal{A}_F$ we have for every

$\omega \in X$ that $\deg_W(\omega, \mathcal{A}_C) \leqslant center_C(F) + \frac{1}{2\mathbf{h}_t}$. Therefore,

$$\int_{X\times\mathcal{A}_F} W \leqslant (center_C(F) + \frac{1}{2\mathbf{h}_t})\mu(X) \quad \text{but}$$
$$\int_{X\times\mathcal{A}_F} U > (center_C(F) + \frac{1}{2\mathbf{h}_t} + \kappa)\mu(X) \geqslant (center_C(F) + \frac{1}{2\mathbf{h}_t})\mu(X) + \frac{\kappa^2}{2\hbar(\mathbf{h})} \quad .$$

This is a contradiction to the fact that $\|U - W\|_\square < \delta$. □

For the proof of Lemma 4.3 we need the following result.

Proposition 4.2. *Suppose that $W : \Omega^2 \to [0,1]$ is a nondegenerate graphon. Then*

$$\mathbb{E}_{x\in\Omega}[b_W(x)] = 1.$$

Proof. By Fubini's Theorem,

$$\mathbb{E}_{x\in\Omega}[b_W(x)] = \int_x \left(\int_y \frac{W(x,y)}{\deg(y)}\,\mathrm{d}y\right)\mathrm{d}x = \int_y \left(\int_x \frac{W(x,y)}{\deg(y)}\mathrm{d}x\right)\mathrm{d}y = 1\,.$$

□

Lemma 4.3 puts a relation between quantities $\mathrm{Freq}(T;W)$ and $\mathrm{Freq}^-(T;G,V_0,E_0)$, where G is a graph, $V_0 \subseteq V(G)$ and $E_0 \subseteq E(G)$, and $\mathrm{Freq}^-(T;G,V_0,E_0)$ is defined as

$$\mathrm{Freq}^-(T;G,V_0,E_0)$$
$$:= |\mathrm{Stab}_T|^{-1} \sum_{\substack{v_1,\dots,v_\ell\in V(G)\setminus V_0 \\ \forall ij\in E(T): v_iv_j\in E(G)\setminus E_0}} \frac{1}{v(G)} \exp\left(-\sum_{j=1}^{p-1} b_G(v_j)\right) \frac{\sum_{j=p}^{\ell}\deg_G(v_j)}{\prod_{j=1}^{\ell}\deg_G(v_j)}\,.$$

We are now ready to state Lemma 4.3. It says that the parameter $\mathrm{Freq}(T;\cdot)$ is continuous in a certain sense. While it would be possible to prove continuity of $\mathrm{Freq}(T;\cdot)$ for nondegenerate graphons with respect to the cut-distance, here we need to put a relation between $\mathrm{Freq}(T;\cdot)$ of a graphon and the parameter $\mathrm{Freq}^-(T;\cdot)$ of a graph that is close to that graphon. Let us note that while our proof of continuity is long, the statement itself is natural. Indeed, the definition (1) is an integration involving products of values of the graphon over the edges of T (similar to the way that is used to define the density of T in the graphon), degrees in the graphon and the function $b(\cdot)$ from (2) which is also defined using degrees in the graphon. As subgraph densities are continuous with respect to the cut-metric, and so is the degree sequence (c.f. Lemma 5.1), it is actually plausible to have continuity of many graph(on) parameters obtained by combinations thereof.

Lemma 4.3. *Suppose that $W : \Omega^2 \to [0,1]$ is a nondegenerate graphon and T is a fixed tree. For every $\varepsilon > 0$ there exists $a > 0$ such that for every $\chi > 0$ there exists $n_0 \in \mathbb{N}$ and $\delta > 0$ such that if G is a graph and $V_0 \subseteq V(G)$, $E_0 \subseteq E(G)$ satisfy*

(a) $n > n_0$,
(b) $\delta_\square(W, G) < \delta$,
(c) $|V_0| \leqslant an$, $|E_0| \leqslant an^2$, and
(d) for each $v \in V(G) \backslash V_0$ we have $\deg_G(v) \geqslant \chi n$,

then we have

$$|\mathrm{Freq}(T; W) - \mathrm{Freq}^-(T; G, V_0, E_0)| < \varepsilon \,. \tag{55}$$

Proof. Suppose that W, T, and ε are given.

Let L be the height of T. Suppose that the vertices of T are $[\ell]$. Suppose that 1 is the root of T, suppose that the height of each vertex i is denoted g_i and that the vertices of T are enumerated so that we have $0 = g_1 < g_2 \leqslant g_3 \leqslant \ldots \leqslant g_{p-1} < g_p = g_{p+1} = \ldots = g_\ell = L$.

In the course of deriving Theorem 1.3 from Theorem 1.1 in Section 4.2 we prove that $\mathrm{Freq}(T; W) \leqslant 1$.[7] In particular, the function $f_W : \Omega^\ell \to [0, +\infty)$,

$$f_W(\omega_1, \ldots, \omega_\ell) := \exp\left(- \sum_{j=1}^{p-1} b_W(\omega_j) \right) \cdot \frac{\sum_{j=p}^{\ell} \deg_W(\omega_j)}{\prod_{j=1}^{\ell} \deg_W(\omega_j)} \cdot \prod_{(i,j) \in E(T)} W(\omega_i, \omega_j) \tag{56}$$

is integrable. Let $\delta_0 > 0$ be such that for any set $A \subseteq \Omega^\ell$, $\mu^{\otimes \ell}(A) < \delta_0$ we have

$$\int_A f_W \leqslant \frac{\varepsilon}{10} \,. \tag{57}$$

Since W is nondegenerate, there exists a number $\Delta \in (0, \frac{\delta_0}{4\ell})$ such that the set

$$\Omega_{\mathrm{small}} := \{\omega \in \Omega : \deg_W(\omega) < 2\Delta\}$$

has measure less than $\frac{\delta_0}{2\ell}$.

Let $a := \min\{\frac{\delta_0}{10\ell}, \frac{\varepsilon \Delta^\ell}{400\ell^\ell}\}$. Now, suppose that χ is given.

Let $\mathbf{h} \in \mathbb{N}^{L+1}$ and $d', d, \tau, \kappa > 0$ satisfy

$$\min(\Delta, \varepsilon, \chi, \delta_0) \gg d' \gg d \gg \tau \gg \frac{1}{\mathbf{h}_1} \gg \frac{1}{\mathbf{h}_2} \gg \ldots \gg \frac{1}{\mathbf{h}_L} \gg \frac{1}{\mathbf{h}_{L+1}} \gg \kappa > 0 \,. \tag{58}$$

We note that the above dependencies are tower-type, i.e.,

$$\frac{1}{\mathbf{h}_i} \gg \frac{\hbar(\mathbf{h}[\![i]\!])}{\mathbf{h}_{i+1}} \,. \tag{59}$$

[7] At this moment, we have not established Theorem 1.3 nor Theorem 1.1. However, the fact $\mathrm{Freq}(T; W) \leqslant 1$ did not rely on the validity of either of these theorems, but rather followed from making the connection to the branching process κ_W.

Let $\left\{\{\mathcal{A}_F\}_{F\in\mathfrak{D}_{\mathbf{h}[\![t]\!]}}\right\}_{t=0}^{L+1}$ be the anatomies of W up to depth $L+1$. Let δ be given by Lemma 4.1 for input graphon W and parameters $\mathbf{h}$ and κ.

Suppose now that the graph G is given. Take a graphon representation U of G such that $\|W-U\|_{\square}<\delta$. Lemma 4.1 tells us that there exists a set $X\subseteq\Omega$ of measure at most κ such that $\left\{\{\mathcal{A}_F\}_{F\in\mathfrak{D}_{\mathbf{h}[\![t]\!]}}\right\}_{t=0}^{L+1}$ are a κ-approximate anatomies of U up to depth $L+1$ with exceptional set X.

Given a vertex $v\in V(G)$, we write $\Omega_v\subseteq\Omega$ for the set representing v in U. We write $\Lambda_{\mathrm{vert}}:=\bigcup_{v\in V_0}\Omega_v$, and $\Lambda_{\mathrm{edge}}=\bigcup_{uv\in E_0}(\Omega_u\times\Omega_v\cup\Omega_v\times\Omega_u)$. By assumption (c) we have

$$\mu(\Lambda_{\mathrm{vert}})\leqslant a\quad\text{and}\quad\mu^{\otimes 2}(\Lambda_{\mathrm{edge}})\leqslant 2a\,. \tag{60}$$

To express $\mathrm{Freq}^{-}(T;G,V_0,E_0)$, we introduce a counterpart to (56) that reflects U,

$$f_U(\omega_1,\dots,\omega_\ell):=\exp\left(-\sum_{j=1}^{p-1}b_U(\omega_j)\right)\cdot\frac{\sum_{j=p}^{\ell}\deg_U(\omega_j)}{\prod_{j=1}^{\ell}\deg_U(\omega_j)}\cdot\prod_{(i,j)\in E(T)}U(\omega_i,\omega_j)$$

and a version f_U^{-}, which takes into the account the "deleted" vertices V_0 and edges E_0,

$$f_U^{-}(\omega):=\begin{cases}0 & \text{if }\exists\ i\in[\ell]\colon\ \omega_i\in\bigcup_{v\in V_0}\Omega_v\text{ or }\exists\ ij\in E(T)\colon\ (\omega_i,\omega_j)\in\bigcup_{vw\in E_0}\Omega_v\times\Omega_w,\\ f_U(\omega) & \text{otherwise.}\end{cases}$$

Note that the assumption (d) implies that

$$f_U^{-}\leqslant\frac{\ell}{\chi^\ell}\quad\text{on the whole domain }\Omega^\ell\,. \tag{61}$$

We say that a map $\pi:[\ell]\to\{\mathcal{A}_F\}_{t=1,\dots,L+1;F\in\mathfrak{D}_{\mathbf{h}[\![t]\!]}}$ is a *depth preserving anatomy assignment* if for each $i\in[\ell]$ we have $\pi(i)=\mathcal{A}_F$ for some $F\in\mathfrak{D}_{\mathbf{h}[\![L+1-g_i]\!]}$. This means that the root 1 is assigned an anatomy of depth $L+1$ while vertices further from the root are assigned anatomies of smaller depths. Note that depth preserving anatomy assignment partition the set Ω^ℓ it the sense that

$$\Omega^\ell=\bigsqcup_{\substack{\pi:[\ell]\to\{\mathcal{A}_F\}_{t=1,\dots,L+1;F\in\mathfrak{D}_{\mathbf{h}[\![t]\!]}}\\ \text{depth preserving anatomy assignment}}}\ \prod_{j=1}^{\ell}\pi(j)\,. \tag{62}$$

In particular,

$$1 = \sum_{\substack{\pi:[\ell]\to\{\mathcal{A}_F\}_{t=1,\dots,L+1;F\in\mathfrak{D}_{\mathbf{h}[\![t]\!]}} \\ \text{depth preserving anatomy assignment}}} \prod_{j=1}^{\ell} \mu\left(\pi(j)\right) \ , \text{ and} \tag{63}$$

$$1 = \sum_{\substack{\pi:[\ell]\to\{\mathcal{A}_F\}_{t=1,\dots,L+1;F\in\mathfrak{D}_{\mathbf{h}[\![t]\!]}} \\ \text{depth preserving anatomy assignment}}} \prod_{j=[\ell]\setminus\{j*\}}^{\ell} \mu\left(\pi(j)\right) \text{ for each } j^* \in [\ell]. \tag{64}$$

We need to define three additional classes of depth preserving anatomy assignments.

- We say that a depth preserving anatomy assignment π is *singular* if there exists $i \in [\ell]$ such that a positive measure of the elements $\omega \in \pi(i)$ satisfy $\deg_W(\omega) \leqslant \Delta$, or $b_W(\omega) > \Delta^{-1}$.
- We say that a depth preserving anatomy assignment π is *dense* if for every vertex $j \in [\ell]$ and every child j^* of j we have that $center_{\pi(j^*)}(\pi(j)) > d \cdot \mu(\pi(j^*))$. If this fails for some pair jj^*, then we call j^* a *witness*. (Of course, several witnesses may exist.)
- We say that a depth preserving anatomy assignment π is *non-problematic* if it is dense, and it is not singular.

For a depth preserving anatomy assignment π, we write $\mathbf{\Omega}_\pi = \prod_{i=1}^{\ell}(\pi(i)\backslash X)$.

The next claim establishes some basic properties of singular depth preserving anatomy assignment.

Claim 4.3.1. *Suppose that π is a singular depth preserving anatomy assignment, and that $i \in \ell$ is a witness of singularity. Then we have that $\pi(i) \subseteq \Omega_{\text{small}}$ or that $b_W(\omega') > \frac{1}{2\Delta}$ for all $\omega' \in \pi(i)$.*

Proof. The definition of singular depth preserving anatomy assignment says that there exists $\omega \in \pi(i)$ such that $\deg_W(\omega) \leqslant \Delta$, or $b_W(\omega) > \Delta^{-1}$.

Let us first deal with the former case. A double application of (48) gives that for each $\omega' \in \pi(i)$ we have

$$\deg_W(\omega') \leqslant \deg_W^{\text{anat}}(\pi(i)) + \frac{\hbar(\mathbf{h}[\![L-1-g_i]\!])}{2\mathbf{h}_{L-g_i}} \leqslant \deg_W(\omega) + \frac{2\hbar(\mathbf{h}[\![L-1-g_i]\!])}{2\mathbf{h}_{L-g_i}} \overset{(58),(59)}{<} 2\Delta \ .$$

Therefore, in this case, $\pi(i) \subseteq \Omega_{\text{small}}$.

We can deal with the latter case, but using (47) instead of (48). We have

$$b_W(\omega') > \Delta^{-1} - \frac{4}{2\mathbf{h}_{L-g_i}} \geqslant \frac{1}{2\Delta} \quad \text{for all } \omega' \in \pi(i) \ .$$

□

The next key claim says that the integrals of f_W and f_U^- over sets corresponding to non–problematic depth preserving anatomy assignment are almost the same.

Claim 4.3.2. *Suppose that $\pi : [\ell] \to \{\mathcal{A}_F\}_{t=1,\dots,L+1;F\in\mathfrak{D}_{\mathbf{h}[\![t]\!]}}$ is a non-problematic depth preserving anatomy assignment. Let*

$$e^{\pi}_{\mathrm{vert}} := \sum_{i=1}^{\ell} \mu\left(\pi(i) \cap \Lambda_{\mathrm{vert}}\right) \cdot \prod_{k\in V(T)\setminus\{i\}} \mu(\pi(k)) \quad , \text{ and}$$

$$e^{\pi}_{\mathrm{edge}} := \sum_{ij\in E(T)} \mu^{\otimes 2}\left(\Lambda_{\mathrm{edge}} \cap \bigcup_{ij\in E(T)} \pi(i)\times\pi(j)\right) \cdot \prod_{k\in V(T)\setminus\{i,j\}} \mu(\pi(k)) \ .$$

Then for the quantities

$$Q_1 := \int_{\boldsymbol{\omega}\in\boldsymbol{\Omega}_\pi} f_W(\boldsymbol{\omega}) \text{ and}$$

$$Q_2 := \int_{\boldsymbol{\omega}\in\boldsymbol{\Omega}_\pi} f_U^-(\boldsymbol{\omega})$$

we have $Q_1 = Q_2 \cdot (1 \pm d) \pm \frac{(e^{\pi}_{\mathrm{vert}}+e^{\pi}_{\mathrm{edge}})\ell}{\Delta^\ell}$.

Proof. Suppose that $j \in V(T)$. If $j \neq 1$, then we write $\mathrm{Par}(j)$ for the parent of j in T. Let us write F_j for the tree obtained from T by erasing the edge from j to $\mathrm{Par}(j)$ and taking the component containing j. When $j = 1$, we take $F_j := T$. Also, for $j = 1$ we define $\mathit{center}_{\pi(j)}\left(\pi(\mathrm{Par}(j))\right) := 1$.

For $j \in V(T)$, write

$$R_j := \prod_{k\in V(F_j)\setminus\{j\}} \mathit{center}_{\pi(k)}\left(\pi(\mathrm{Par}(k))\right) \ .$$

Here, recall the convention that a product over the empty set is 1. This in particular applies when j is a leaf and $j \neq 1$.

We first want to show that the quantities

$$\int_{(\omega_1,\dots,\omega_\ell)\in\boldsymbol{\Omega}_\pi} \prod_{(i,j)\in E(T)} W(\omega_i,\omega_j) \quad \text{and} \quad \int_{(\omega_1,\dots,\omega_\ell)\in\boldsymbol{\Omega}_\pi} \prod_{(i,j)\in E(T)} U(\omega_i,\omega_j)$$

are very close. To this end, for $j \in V(T)$ we define $A_j(\omega_j) = B_j(\omega_j) := 1$ if j is a leaf (different than the root) and otherwise we define

$$\begin{aligned} A_j(\omega_j) &:= \int\limits_{\{\omega_k \in \pi(k)\backslash X\}_{k\in V(F_j)\setminus\{j\}}} \prod_{(a,b)\in E(F_j)} W(\omega_a,\omega_b)\ , \\ B_j(\omega_j) &:= \int\limits_{\{\omega_k \in \pi(k)\backslash X\}_{k\in V(F_j)\setminus\{j\}}} \prod_{(a,b)\in E(F_j)} U(\omega_a,\omega_b)\ . \end{aligned} \tag{65}$$

Inductively for $t = L+1, L, \ldots, 0$, assume that for each vertex $j \in V(T)$ at height $g_j = t$,[8] for each $\omega_j \in \pi(j)$ we have

$$A_j(\omega_j) = R_j \cdot \exp(\pm \tfrac{v(F_j)}{\mathbf{h}_1}) \tag{66}$$

and for every $\omega_j \in \pi(j)\backslash X$ we have

$$B_j(\omega_j) = R_j \cdot \exp(\pm \tfrac{v(F_j)}{\mathbf{h}_1})\ . \tag{67}$$

Now, let suppose that the statement is true for all vertices at height $t+1$. Let j be an arbitrary vertex at height t, and let $\omega_j \in \pi(j)$ be arbitrary. When j is a leaf then (66) and (67) hold trivially. So, suppose that j has some children $j_1, j_2, \ldots, j_q$. Since these children are at height $t+1$, applying the induction hypothesis, we disintegrate (65) with respect to these children, and get

$$\begin{aligned} A_j(\omega_j) &= \prod_{c=1}^{q} \left(\int_{\omega_{j_c}\in\pi(j_c)\backslash X} W(\omega_j,\omega_{j_c}) \int_{\{\omega_k\in\pi(k)\backslash X\}_{k\in V(F_{j_c})\setminus\{j_c\}}} \prod_{(a,b)\in E(F_{j_c})} W(\omega_a,\omega_b) \right) \\ &= \prod_{c=1}^{q} \left(\int_{\omega_{j_c}\in\pi(j_c)\backslash X} W(\omega_j,\omega_{j_c}) \cdot A_{j_c}(\omega_{j_c}) \right) \\ &\overset{(66)}{=} \prod_{c=1}^{q} \left(\int_{\omega_{j_c}\in\pi(j_c)\backslash X} W(\omega_j,\omega_{j_c}) \cdot R_{j_c} \exp\left(\pm \tfrac{v(F_{j_c})}{\mathbf{h}_1}\right) \right) \\ &= \prod_{c=1}^{q} \deg_W(\omega_j,\pi(j_c)\backslash X) \cdot \prod_{c=1}^{q} \left(R_{j_c} \exp(\pm \tfrac{v(F_{j_c})}{\mathbf{h}_1}) \right) \\ &= \prod_{c=1}^{q} (\deg_W(\omega_j,\pi(j_c)) \pm \kappa) \cdot \prod_{c=1}^{q} \left(R_{j_c} \exp(\pm \tfrac{v(F_{j_c})}{\mathbf{h}_1}) \right)\ . \end{aligned}$$

Now, each of the factors in the first product equals to $center_{\pi(j_c)}(\pi(j)) \pm \frac{1}{\mathbf{h}_{L+1-g_j}}$ by using the property (47). Since π is a dense depth preserving anatomy assignment and since

[8] Note that the first step of the induction is satisfied trivially, as the height of T is L.

$\mu(\pi(j_c))$ is a positive number which depends only on $\mathbf{h}[\![L-g_j]\!]$, we have

$$\deg_W(\omega_j,\pi(j_c))\pm\kappa = center_{\pi(j_c)}(\pi(j))\cdot(1\pm\sqrt{\kappa})\ .$$

Since $(1\pm\sqrt{\kappa})^q=\exp(\pm\frac{1}{\mathbf{h}_1})$, we just verified (66) for j and ω_j.

We can get exactly the same calculations for $B_j(\omega)$, where $\omega\in\pi(j)\backslash X$ and get

$$B_j(\omega_j)=\prod_{c=1}^{q}(\deg_U(\omega_j,\pi(j_c))\pm\kappa)\cdot\prod_{c=1}^{q}\Big(R_{j_c}\exp(\pm\tfrac{v(F_{j_c})}{\mathbf{h}_1})\Big)$$

Now, the fact ω_j is a non-exceptional element of the cell $\pi(j)$ of the κ-approximate anatomy $\{\mathcal{A}_F\}_{F\in\mathfrak{D}_{\mathbf{h}[\![L+1-j]\!]}}$ (for U) implies (67).

Clearly, the term $\int_{\omega_1\in\pi(1)\backslash X}A_1(\omega_1)$ corresponds to the term $\int\prod_{(i,j)\in E(T)}W(\omega_i,\omega_j)$ in the definition of Q_1. We shall now control all the remaining terms. The idea is that these all are almost constant since we are working inside one anatomy cell. Let us first deal with the terms in the definition of Q_1.

As $\pi(j)$ is non-singular for each $j\in[\ell]$, the denominator of $\frac{\sum_{j=p}^{\ell}\deg_W(\omega_j)}{\prod_{j=1}^{\ell}\deg_W(\omega_j)}$ is at least Δ^ℓ. Further, the fluctuations in the denominator are at most $\sum_{j=1}^{\ell}\frac{\hbar(\mathbf{h}[\![L-g_j]\!])}{\mathbf{h}_{L+1-g_j}}\leqslant\tau$ around

$$C_1:=\prod_{j=1}^{\ell}\deg_W^{\mathrm{anat}}(\pi(j))$$

by (48). Similarly, the nominator is bounded from below by Δ, and the fluctuations in the nominator are at most $\sum_{j=p}^{\ell}\frac{\hbar(\mathbf{h}[\![L-g_j]\!])}{\mathbf{h}_{L+1-g_j}}\leqslant\tau$ around

$$C_2:=\sum_{j=p}^{\ell}\sum_{F\in\mathfrak{D}_{\mathbf{h}[\![L-g_j]\!]}}center_F(\pi(j))\ .$$

The fluctuations in the term $\exp\Big(-\sum_{j=1}^{p-1}b_W(\omega_j)\Big)$ are at most τ around $C_3:=\prod_{j=1}^{p-1}center_{\mathfrak{b}}(\pi(j))$. Also, since $\pi(j)$ is non-singular for each $j\in[p-1]$, we have $C_3>\exp(-\frac{\ell}{\Delta})$. The two upper-/lower- bounds we have derived using from the non-singularity will be used below to transform an additive error of the type *value* $\pm$ *error* to a multiplicative one,

$value \cdot (1 \pm \frac{error}{lower\ bound\ on\ the\ value})$. Therefore,

$$\begin{aligned}
Q_1 &= \int_{(\omega_1,\dots,\omega_\ell)\in\mathbf{\Omega}_\pi} \exp\left(-\sum_{j=1}^{p-1} b_W(\omega_j)\right)\cdot\frac{\sum_{j=p}^{\ell}\deg_W(\omega_j)}{\prod_{j=1}^{\ell}\deg_W(\omega_j)}\cdot\prod_{(i,j)\in E(T)} W(\omega_i,\omega_j)\\
&= \int_{(\omega_1,\dots,\omega_\ell)\in\mathbf{\Omega}_\pi} (C_3\pm\tau)\cdot\frac{C_2\pm\tau}{C_1\pm\tau}\cdot\prod_{(i,j)\in E(T)} W(\omega_i,\omega_j)\\
&= C_3\cdot\frac{C_2}{C_1}\cdot\left(1\pm 6\tau\cdot\exp(\tfrac{\ell}{\Delta})\right)\cdot\int_{(\omega_1,\dots,\omega_\ell)\in\mathbf{\Omega}_\pi}\prod_{(i,j)\in E(T)} W(\omega_i,\omega_j)\\
&= C_3\cdot\frac{C_2}{C_1}\cdot\left(1\pm 6\tau\cdot\exp(\tfrac{\ell}{\Delta})\right)\cdot\int_{\omega_1\in\pi(1)\setminus X} A_1(\omega_1)\\
&\overset{(66)}{=} C_3\cdot\frac{C_2}{C_1}\cdot\left(1\pm 6\tau\cdot\exp(\tfrac{\ell}{\Delta})\right)\cdot\mu(\pi(1)\setminus X)\cdot R_1\cdot\exp(\pm\tfrac{\ell}{\mathbf{h}_1})\\
&= C_3\cdot\frac{C_2}{C_1}\cdot\mu(\pi(1)\setminus X)\cdot R_1\cdot\left(1\pm\tfrac{d}{3}\right)\ . \qquad (68)
\end{aligned}$$

We now attempt to show that this quantity is close to Q_2. First, let us introduce a modified quantity in which even integration over V_0 and E_0 is considered, $Q_2^+ := \int_{\boldsymbol{\omega}\in\mathbf{\Omega}_\pi} f_U(\boldsymbol{\omega})$. Now, we claim that we get exactly the same bound as in (68) even for Q_2^+. Let us explain that there are no traps in doing so. First, we can use $\int_{\omega_1\in\pi(1)\setminus X} B_1(\omega_1)$ to control $\int\prod_{(i,j)\in E(T)} U(\omega_i,\omega_j)$, exactly in the same way as we used $\int_{\omega_1\in\pi(1)\setminus X} A_1(\omega_1)$ to control $\int\prod_{(i,j)\in E(T)} W(\omega_i,\omega_j)$, because (67) is a perfect counterpart to (66). The remaining quantities appearing in the definition of Q_2 are again centered around the constants C_1, C_2, and C_3 as above, except the fluctuations can be bigger by κ since $\left\{\{\mathcal{A}_F\}_{F\in\mathfrak{D}_{\mathbf{h}[\![t]\!]}}\right\}_{t=0}^{L+1}$ are only κ-approximate anatomies for U (as opposed to exact anatomies for W). However, κ is the smallest constant in (58) and thus this additional error can be easily swallowed by other error terms. Third, the upper- and lower- bounds based on non-nonsigularity that we used to transform an additive errors to multiplicative ones are still valid when the additional approximation term κ is taken into account. Therefore, we conclude that $Q_2^+ = C_3\cdot\frac{C_2}{C_1}\cdot\mu(\pi(1)\setminus X)\cdot R_1\cdot\left(1\pm\frac{d}{3}\right)$. This finishes the proof.

Now, let us show that we have $|Q_2^+ - Q_2| < \frac{(e^\pi_{\mathrm{vert}}+e^\pi_{\mathrm{edge}})\ell}{\Delta^\ell}$. Clearly, we have $Q_2^+ \geqslant Q_2$. On the other hand, if for some $(\omega_1,\dots,\omega_\ell)\in\mathbf{\Omega}_\pi$ we have that $f_U^-(\omega_1,\dots,\omega_\ell) = 0$ but $f_U(\omega_1,\dots,\omega_\ell) > 0$ then for some $i\in[\ell]$ we have $\omega_i\in\pi(i)\cap\Lambda_{\mathrm{vert}}$ or for some $ij\in E(T)$ we have $(\omega_i,\omega_i)\in\pi(i)\times\pi(j)\cap\Lambda_{\mathrm{edge}}$. □

Claim 4.3.3. *Suppose that $\pi:[\ell]\to\{\mathcal{A}_F\}_{t=1,\dots,L+1;F\in\mathfrak{D}_{\mathbf{h}[\![t]\!]}}$ is a depth preserving anatomy assignment that is not singular. Suppose that $j^*\in[\ell]$ is a witness that it is not dense*

either. Then

$$\int_{\prod_{j=1}^{\ell}\pi(j)} f_W < \frac{d\ell}{\Delta^\ell}\cdot\prod_{j\in[\ell]}\mu(\pi(j)) + \frac{\ell}{\mathbf{h}_{L+1-g_{j^*}}\cdot\Delta^\ell}\cdot\prod_{j\in[\ell]\setminus\{j^*\}}\mu(\pi(j))\,. \tag{69}$$

Proof. Let us bound all the terms in (56). More precisely, let $kj^* \in E(T)$ be any edge that witnesses that π is not a dense depth preserving anatomy assignment. To bound f_W, we shall use that

$$f_W(\omega_1,\dots,\omega_\ell) \leqslant \exp\left(-\sum_{j=1}^{p-1} b_W(\omega_j)\right)\cdot\frac{\sum_{j=p}^{\ell}\deg_W(\omega_j)}{\Pi_{j=1}^{\ell}\deg_W(\omega_j)}\cdot W(\omega_k,\omega_{j^*}).$$

The term $\exp\left(-\sum_{j=1}^{p-1} b_W(\omega_j)\right)$ is trivially at most 1. The nominator in the term $\frac{\sum_{j=p}^{\ell}\deg_W(\omega_j)}{\prod_{j=1}^{\ell}\deg_W(\omega_j)}$ is at most ℓ while the denominator is at least Δ^ℓ by non-singularity. Finally, because $kj^* \in E(T)$ is an edge that witnesses that π is not a dense depth preserving anatomy assignment. Then for every choice of $\omega_k \in \pi(k)$, we have that $\int_{\omega_{j^*}\in\pi(j^*)} W(\omega_k,\omega_{j^*}) = \deg_W(\omega_k,\pi(\omega_{j*})) \leqslant d\cdot\mu(\pi(j^*)) + \frac{1}{2\mathbf{h}_{L+1-g_{j^*}}}$. The claim now follows by integrating over the remaining dimensions. □

We are now ready to express $\mathrm{Freq}(T;W)$ and $\mathrm{Freq}^-(T;G,V_0,E_0)$ in a way that most terms will approximately cancel. To express $\mathrm{Freq}(T;W)$, we work with the function f_W defined in (56). Also, we partition the integration over Ω^ℓ in (1) into terms corresponding to individual depth preserving anatomy assignments as in (62).

$$\begin{aligned}|\mathrm{Stab}_T|\cdot\mathrm{Freq}(T;W) = &\sum_{\pi\ \text{non}-\text{problematic}}\int_{\mathbf{\Omega}_\pi} f_W\\ &+\sum_{\pi\ \text{non}-\text{problematic}}\int_{\prod_{j=1}^{\ell}\pi(j)\setminus\mathbf{\Omega}_\pi} f_W\\ &+\sum_{\pi\ \text{not dense and not singular}}\int_{\prod_{j=1}^{\ell}\pi(j)} f_W\\ &+\sum_{\pi\ \text{singular}}\int_{\prod_{j=1}^{\ell}\pi(j)} f_W\,.\end{aligned} \tag{70}$$

Now, we have

$$\begin{aligned}|\mathrm{Stab}_T| \cdot \mathrm{Freq}^-(T;G,V_0,E_0) = &\sum_{\pi \text{ non-problematic}} \int_{\boldsymbol{\Omega}_\pi} f_U^- \\ &+ \sum_{\pi \text{ non-problematic}} \int_{\prod_{j=1}^{\ell}\pi(j)\setminus\boldsymbol{\Omega}_\pi} f_U^- \\ &+ \sum_{\pi \text{ not dense and not singular}} \int_{\prod_{j=1}^{\ell}\pi(j)} f_U^- \\ &+ \sum_{\pi \text{ singular}} \int_{\prod_{j=1}^{\ell}\pi(j)} f_U^- \ .\end{aligned} \tag{71}$$

We will show that the first sums in (70) and (71) cancel almost perfectly, and that all the remaining terms are negligible.

Claim 4.3.2 tells us that for the first sums in (70) and (71) we have

$$\sum_{\pi \text{ non-probl.}} \int_{\boldsymbol{\Omega}_\pi} f_W = \sum_{\pi \text{ non-probl.}} \int_{\boldsymbol{\Omega}_\pi} f_U^- \cdot (1 \pm d) \pm \sum_{\pi \text{ non-probl.}} \frac{(e^\pi_{\mathrm{vert}} + e^\pi_{\mathrm{edge}})\ell}{\Delta^\ell} \ . \tag{72}$$

We have

$$\sum_{\pi \text{ non-probl.}} e^\pi_{\mathrm{vert}} = \sum_{i=1}^{\ell} \sum_{\pi \text{ non-probl.}} \mu\left(\pi(i) \cap \Lambda_{\mathrm{vert}}\right) \cdot \prod_{k \in V(T)\setminus\{i\}} \mu(\pi(k)) \overset{(64)}{\leqslant} \sum_{i=1}^{\ell} \mu\left(\Lambda_{\mathrm{vert}}\right) \overset{(60)}{\leqslant} \ell \cdot a \ , \text{ and}$$

$$\begin{aligned}\sum_{\pi \text{ non-probl.}} e^\pi_{\mathrm{edge}} &= \sum_{ij \in E(T)} \sum_{\pi \text{ non-probl.}} \mu^{\otimes 2}\left(\Lambda_{\mathrm{edge}} \cap \bigcup_{ij \in E(T)} \pi(i) \times \pi(j)\right) \cdot \prod_{k \in V(T)\setminus\{i,j\}} \mu(\pi(k)) \\ &\leqslant |E(T)| \mu^{\otimes 2}\left(\Lambda_{\mathrm{edge}}\right) \overset{(60)}{\leqslant} \ell \cdot 2a \ .\end{aligned}$$

Recall that in the course of proving Theorem 1.3 from Theorem 1.1 we showed that $\int_{\boldsymbol{\omega}\in\Omega} f_W(\boldsymbol{\omega}) \leqslant |\mathrm{Stab}_T| \leqslant \ell^\ell$. In particular, using the above bounds, we can turn the multiplicative error in (72) into an additive error,

$$\begin{aligned}\sum_{\pi \text{ non-probl.}} \int_{\boldsymbol{\Omega}_\pi} f_W &= \sum_{\pi \text{ non-probl.}} \int_{\boldsymbol{\Omega}_\pi} f_U^- \ \pm 2d\ell^\ell \pm \frac{3a\ell^2}{\Delta^\ell} \\ &= \sum_{\pi \text{ non-probl.}} \int_{\boldsymbol{\Omega}_\pi} f_U^- \ \pm \frac{\varepsilon}{100} \ .\end{aligned} \tag{73}$$

Let us now focus on the second sum in (70). Observe that the set $\bigcup_{\pi \text{ non-problematic}} \prod_{j=1}^{\ell}\pi(j)\setminus\boldsymbol{\Omega}_\pi$ has measure at most $\ell \cdot \mu(X) \leqslant \delta_0$. Therefore, (57) applies, and we get that

$$\sum_{\pi \text{ non-problematic}} \int_{\prod_{j=1}^{\ell}\pi(j)\setminus\boldsymbol{\Omega}_\pi} f_W \leqslant \frac{\varepsilon}{10} \ .$$

Let us now turn to the third sum in (70). We write

$$\sum_{\pi \text{ not dense and not singular}} \int_{\prod_{j=1}^{\ell} \pi(j)} f_W \leqslant \sum_{g=0}^{L} \sum_{\pi:\text{witness at height } g} \int_{\prod_{j=1}^{\ell} \pi(j)} f_W \ ,$$

where the last sum ranges over all non-singular depth preserving anatomy assignments π with a witness for not being dense at height g. By Claim 4.3.3 and by (63) and (64) for each $g \in \{0, \ldots, L\}$ we have

$$\sum_{\pi:\text{witness at height } g} \int_{\prod_{j=1}^{\ell} \pi(j)} f_W \leqslant \frac{d\ell}{\Delta^{\ell}} + \frac{\ell^2}{\mathbf{h}_{L+1-g} \cdot \Delta^{\ell}} \leqslant \frac{d'}{L+1} \ .$$

Therefore,

$$\sum_{\pi \text{ not dense and not singular}} \int_{\prod_{j=1}^{\ell} \pi(j)} f_W < d' \ . \tag{74}$$

The fourth sum of (70) can be bounded from above as follows. If π is singular then by Claim 4.3.1 there exists $i \in [\ell]$ such that we have $b_W(\omega) > 1/(2\Delta)$ for all $\omega \in \pi(i)$, or we have $\pi(i) \subseteq \Omega_{\text{small}}$. Since the average value of b_W is 1 by Proposition 4.2, the measure of the elements that satisfy the former condition is at most 2Δ. Thus, the measure of the set $\bigcup_{\pi} \prod_{j=1}^{\ell} \pi(j)$, where the union ranges over all depth preserving anatomy assignments π which are singular at coordinate i, is at most $2\Delta + \mu(\Omega_{\text{small}}) < \frac{\delta_0}{\ell}$. We conclude that the measure of $\bigcup_{\pi \text{ singular}} \prod_{j=1}^{\ell} \pi(j)$ is less than δ_0. Therefore, (57) applies, and $\sum_{\pi \text{ singular}} \int_{\prod_{j=1}^{\ell} \pi(j)} f_W \leqslant \frac{\varepsilon}{10}$. It now remains to bound the second, third, and the fourth term (71). Recall that when bounding the corresponding terms in (70) above, we argued that the domains of integration of the second and the fourth term have measures at most δ_0 each. Combining this with the upper bound (61), we get that the second and the fourth term are at most $\delta_0 \cdot \frac{\ell}{\chi^{\ell}} < \frac{\varepsilon}{100}$, using (58). So, the only term we have to control now is

$$\sum_{\pi \text{ not dense and not singular}} \int_{\prod_{j=1}^{\ell} \pi(j)} f_U^- =$$

$$\sum_{\pi \text{ not dense and not singular}} \int_{\mathbf{\Omega}_{\pi}} f_U^- + \sum_{\pi \text{ not dense and not singular}} \int_{\prod_{j=1}^{\ell} \pi(j) \setminus \mathbf{\Omega}_{\pi}} f_U^- \ . \tag{75}$$

The first summand can be bounded by $2d'$ in exactly the same way as we derived (74). To see this, recall that on $\mathbf{\Omega}_{\pi}$ the degrees into anatomies are similar for W and for U and thus we can make use of the "non-denseness" and "non-singularity" assumption even for f_U^- (with slightly worse constants). The second summand of (75) corresponds represents integration over a set of measure at δ_0. In other words, we can bound the second summand of (75) by $\frac{\varepsilon}{100}$ in the same way we bounded the second and the fourth term of (71).

The lemma now follows by expanding $\mathrm{Freq}(T;W) - \mathrm{Freq}^-(T;G,V_0,E_0)$ using (70) and (71) and using the bounds above. □

Theorem 1.1 now follows in a straightforward way by combining Lemma 3.12 and Lemma 4.3.

Proof of Theorem 1.1. Suppose that a graphon W, an ℓ-vertex tree T and $\varepsilon > 0$ are given. Let $a > 0$ by given by Lemma 4.3 for W, T, and $\frac{\varepsilon}{2}$. Lemma 2.11 with parameter $\beta := c_1 \cdot \min\{a^8, \varepsilon^{16}\}$ (for a sufficiently small constant c_1) yields positive constants $\alpha, \varepsilon', \gamma$ and ξ' with $\beta \gg \alpha \gg \varepsilon' \gg \gamma \gg \xi'$. Now, let n_0 and δ' be numbers given by Lemma 4.3 for input parameters as above together with $\chi := \gamma$. Set $\xi := \min(\xi', \delta, \frac{1}{n_0})$.

Suppose that G is a graph with $d_\square(G,W) \leqslant \xi$. Owing to Lemma 2.11, we know that G has a $(\beta,\alpha,\varepsilon',\gamma)$-good decomposition $V(G) = V_0 \sqcup V_1 \sqcup \ldots \sqcup V_k$. Lemma 3.12 now tells us that with probability at least $1 - \mathsf{O}(\ell\alpha^{1/8})$ the uniform spanning tree $\mathcal{T}$ of G satisfies

$$\mathbb{P}_X(B_{\mathcal{T}}(X,r) \cong T) = (1 + \mathsf{O}(\ell^2\alpha^{1/16}))\mathrm{Freq}(T;G) + \mathsf{O}(\ell^2\beta^{1/8})\,. \tag{76}$$

Recall that the quantity $\mathrm{Freq}(T;G)$ is defined in (34), and depends on the decomposition $V(G) = V_0 \sqcup V_1 \sqcup \ldots \sqcup V_k$. Observe that $\mathrm{Freq}(T;G) = \mathrm{Freq}^-(T;G,U_0,E_0)$, where U_0 is the union of V_0 and sets V_i ($i \geqslant 1$) that are not (α,ε')-big, and E_0 consists of all edges running across two distinct (α,ε')-big sets $\{V_i\}_i$. Proposition 2.12 tells us that $|U_0| = \mathsf{O}(\beta^{1/8}n) < an$. The fact that $V(G) = V_0 \sqcup V_1 \sqcup \ldots \sqcup V_k$ is a $(\gamma,\varepsilon'^5,\varepsilon'^5)$-expander decomposition (in particular, (G2) of Definition 2.6 applies) tells us that $|E_0| \leqslant \varepsilon'^5 n^2 < an^2$. Thus, Lemma 4.3 gives $\mathrm{Freq}(T;G) = \mathrm{Freq}(T;W) \pm \frac{\varepsilon}{2}$. Plugging this back to (76) and using that the term $(1 + \mathsf{O}(\ell^2\alpha^{1/16}))$ can be bounded by $(1 \pm \frac{\varepsilon}{4})$ and the term $\mathsf{O}(\ell^2\beta^{1/8})$ is smaller than $\frac{\varepsilon}{4}$, we get the theorem. □

4.2. **Proof of Theorem 1.3 from Theorem 1.1.** Let T be a fixed rooted tree with $\ell \geqslant 2$ vertices and height r. We denote the vertices of T, as before, by $\{1,\ldots,\ell\}$ so that 1 is the root and the vertices of distance r from the root are $\{p,\ldots,\ell\}$ for some $2 \leqslant p \leqslant \ell$. Our goal is to show that the probability that the first r generations of the branching process yield a tree which is root-isomorphic to T is $\mathrm{Freq}(T;W)$.

First, the factor of $|\mathrm{Stab}_T|^{-1}$ comes from the $|\mathrm{Stab}_T|$ different vertex labellings of T which yield the same event. Secondly, we may split this event to $\ell - p + 1$ disjoint events, indexed by the vertices $q \in \{p,\ldots,\ell\}$ at height r of T, and each of these is the event that the path from 1 to q is the first r steps of the ancestral path defined in κ_W. We will show that the probability of each of these events is precisely

$$\int_{\omega_1,\ldots,\omega_\ell} \exp\left(-\sum_{j=1}^{p-1} b_W(\omega_j)\right) \cdot \frac{\deg(\omega_q)}{\prod_{j=1}^{\ell}\deg(\omega_j)} \cdot \prod_{(i,j)\in E(T)} W(\omega_i,\omega_j)\mathrm{d}\omega_1\cdots\mathrm{d}\omega_\ell\,, \tag{77}$$

which is the q-th term in (1) when we expand the sum over j in the numerator. We denote the path from 1 to q in T by $x_1, \dots, x_r$ where $x_1 = 1$ and $x_r = q$. By definition of κ_W, the density function of the first r ancestral particles $\omega_{x_1}, \dots, \omega_{x_r}$ is

$$\prod_{j=1}^{q-1} \frac{W(\omega_{x_j}, \omega_{x_{j+1}})}{\deg(x_j)} .$$

Thirdly, by independence of the branching process, conditioned on this path, the rest of the branching process emanating from the particles on this path is independent and all new particles (if there are any) are "other" particles. Given a particle of type ω (either ancestral or other), the number of its other progeny is distributed as $\mathsf{Poisson}(b_W(w))$ random variable, so the probability that it equals $k \geqslant 0$ is $\frac{e^{-b_W(\omega)} b_W(\omega)^k}{k!}$. Given the the number of its "other" progeny is k, these k points of Ω are distributed as k i.i.d. points drawn from the probability density function $\frac{W(\omega,\omega')}{b_W(\omega)\deg(\omega')}$. Thus, conditioned on having k progeny, the density of these k-tuple $\{\omega_1, \dots, \omega_k\}$ of points is

$$\frac{k!}{b(\omega)^k} \prod_{i=1}^{k} \frac{W(\omega, \omega_i)}{\deg(\omega_i)} .$$

The $k!/b_W(\omega)^k$ cancels with the $b_W(\omega)^k/k!$ in the probability of having k points and we are left with last term times $e^{-b_W(\omega)}$. We now apply this throughout all the branching points of the tree T (even for $k = 0$) and this concludes the proof (note that we do not have a factor for $e^{-b_W(\omega)}$ for tree vertices at level r since we do not care if they have progeny or not). □

5. Bounds on the number of vertices of a given degree in the UST

In this section we prove Theorem 1.5. Let us make two remarks beforehand. Firstly, we cannot hope for converse bounds to the theorem. Indeed, let us take a small $\alpha > 0$ and let us consider the complete bipartite graph $K_{\alpha n,(1-\alpha n)}$ with color classes A and B, $|A| = \alpha n$, $|B| = (1-\alpha)n$. The handshaking lemma tells us that for *any* spanning tree T of $K_{\alpha n,(1-\alpha n)}$ we have

$$n - 1 = \sum_{b \in B} \deg_T(b) = (1-\alpha)n + \sum_{b \in B, \deg_T(b) > 1} (\deg_T(b) - 1) .$$

In particular, the last sum can have at most $\alpha n - 1$ summands. We conclude that T has more than $(1-2\alpha)n$ leaves, and less than $2\alpha n$ vertices of degrees more than 1.

Secondly, as we mentioned in the beginning of this paper, the degree distribution of a UST in a complete graph is approximately $\mathsf{Poisson}(1) + 1$. Considering Theorem 1.5, we see that the complete graph is an asymptotic minimizer for the number of leaves (density e^{-1}), the asymptotic maximizer for the number of vertices of degree 2 (density e^{-1}) and 3

(density $e^{-1}/2$). However, the density of vertices degree $k \geqslant 4$ in the UST of the complete graph is $e^{-(k-1)}/(k-1)!$ which is smaller than the $(k-2)^{k-2}e^{-(k-2)}/(k-1)!$ given by Theorem 1.5. Examining the proof shows that the bounds in Theorem 1.5 for $k \geqslant 4$ are optimal in the following sense. Let $\alpha > 0$ be arbitrarily small, and let $G_{n,k,\alpha}$ be the graph obtained by taking a complete graph on $n/(k-2)$ vertices and additional $n(k-3)/(k-2)$ vertices such that each of the $n^2(k-3)$ possible edges between the two parts is retained with probability α and erased otherwise, independently of all other edges. It is a straightforward computation using Theorem 1.3 (see also the proof of Theorem 1.5 below) to show that for n large, with high probability the proportion of vertices of degree k in a UST will be arbitrarily close, as $\alpha \to 0$, to $\frac{1}{(k-1)!}((k-2)/e)^{k-2}$.

We are almost ready to prove the theorem. We will need the following folklore statement that provides a relation between degrees in a graphon and finite graphs that converge to it. Lemma 5.1 is easy to prove directly, but let us note that is a straightforward consequence of Lemma 4.1 for anatomies of depth 1.

Lemma 5.1. *Suppose that $G_1, G_2, \ldots$ are finite graphs converging to a graphon $W : \Omega^2 \to [0,1]$.*

(1) Suppose that $0 \leqslant a < b \leqslant 1$ are given. and let d be the measure of points of Ω whose degree is in the interval (a,b). Then for an arbitrary $\varepsilon > 0$ there exists an n_0 such that for every $n > n_0$ we have that in G_n there are at least $(d-\varepsilon)v(G_n)$ many vertices with degrees in the interval $((a-\varepsilon)v(G_n), (b+\varepsilon)v(G_n))$ and there are at most $(d+\varepsilon)v(G_n)$ many vertices with degrees in the interval $((a+\varepsilon)v(G_n), (b-\varepsilon)v(G_n))$.

(2) Suppose that there exists $\alpha > 0$ such that in each G_n, all but at most $\mathsf{o}_n(1)v(G_n)$ vertices have degrees at least $\alpha v(G_i)$. Then W is nondegenerate.

Proof of Theorem 1.5. We start by proving (3). Assume by contradiction that there are $\varepsilon_0, \delta_0 > 0$ such that there exists a sequence of graphs G_n (which we assume have n vertices) such that

(A) at least $(1 - \mathsf{o}(1))n$ vertices of G_n are of degree at least $\delta_0 n$, and,
(B) $\mathbb{P}\big(L_1(n) \leqslant (e^{-1} - \varepsilon_0)n\big) > \varepsilon_0$,

where $L_1(n)$ is the number of leaves in a UST of G_n. By Theorem 1.6 we may assume without loss of generality that G_n is a converging sequence and let W be the corresponding limit graphon. By Lemma 5.1(2) and assumption (A) we deduce that W is nondegenerate. Let $\mathcal{T}_n$ be a UST of G_n. By Theorem 1.3 the sequence $\mathcal{T}_n$ almost surely satisfies that

$$\lim_{n\to\infty} \mathbb{P}(\deg_{\mathcal{T}_n}(X) = 1) = \mathbb{E}_{x\in\Omega}\left[\mathbb{P}[1 + \mathsf{Poisson}(b_W(x)) = 1]\right] = \mathbb{E}_{x\in\Omega}\left[e^{-b_W(x)}\right],$$

where X is a uniformly chosen vertex of G_n. By Proposition 4.2 we have that $\mathbb{E}_{x\in\Omega}[b_W(x)] = 1$. Hence Jensen's inequality implies that

$$\lim_{n\to\infty} \mathbb{P}(\deg_{\mathcal{T}_n}(X) = 1) \geqslant e^{-1}\,.$$

Since $L_1(n) = n\mathbb{P}(\deg_{\mathcal{T}_n}(X) = 1)$ we arrive at a contradiction to assumption (B), showing (3).

The proof of (4) and (5) proceeds similarly. We fix $k \geqslant 2$ and make the an analogous contradictory assumption as in (A) and (B). We have that

$$\lim_{n\to\infty} \mathbb{P}(\deg_{\mathcal{T}_n}(X) = k) = \mathbb{E}_{x\in\Omega}\left[\mathbb{P}[1 + \mathsf{Poisson}(b_W(x)) = k]\right] = \frac{1}{(k-1)!}\mathbb{E}_{x\in\Omega}\left[e^{-b_W(x)}b_W(x)^{k-1}\right].$$

When $k = 2$, since $ye^{-y} \leqslant e^{-1}$ for all $y \geqslant 0$ we learn that the last quantity is at most e^{-1}, proving (4) by contradiction. When $k \geqslant 3$, it follows by Lemma 5.2 (provided immediately below) that the last quantity is at most $\frac{1}{(k-1)!}((k-2)/e)^{k-2}$, similarly proving (5). □

The last missing piece is to state and prove the optimization result that was key for our proof of Theorem 1.5 above.

Lemma 5.2. *Let $b : [0,1] \to [0,\infty)$ satisfy $\int_{[0,1]} b(x)\mathrm{d}x = 1$. Then for any $k \geqslant 2$,*

$$\int_{[0,1]} e^{-b(x)}b(x)^k\mathrm{d}x \leqslant \left(\frac{k-1}{e}\right)^{k-1}.$$

Proof. For ease of notation, let $f(x) = e^{-x}x^k$. Fix $\kappa > 0$. Approximating $b(x)$ by simple functions, we can find $n \in \mathbb{N}$ and n non-negative numbers $b_1, \ldots, b_n$ such that $\frac{1}{n}\sum_{i=1}^n b_i = 1$, and $\int_{[0,1]} f(b(x))\mathrm{d}x \leqslant \frac{1}{n}\sum_{i=1}^n f(b_i) + \kappa$. Hence

$$\int_{[0,1]} e^{-b(x)}b(x)^k\mathrm{d}x \leqslant \max_{(b_1,\ldots,b_n)\in\mathcal{C}} \Phi(b_1,\ldots,b_n) + \kappa,$$

where $\Phi(b_1,\ldots,b_n) := \frac{1}{n}\sum_{i=1}^n f(b_i)$ and $\mathcal{C} := \{(b_1,\ldots,b_n) \in [0,+\infty)^n : \frac{1}{n}\sum_{i=1}^n b_i \leqslant 1\}$. To prove the lemma, it thus suffices to show that $\Phi|_{\mathcal{C}} \leqslant \left(\frac{k-1}{e}\right)^{k-1}$.

By abuse of notation, we use the same symbol $\vec{b} = (b_1,\ldots,b_n)$ to denote the maximizer of $\Phi|_{\mathcal{C}}$. As $f'(x) = e^{-x}x^{k-1}(k-x) < 0$ for every $x > k$, $f(x)$ is decreasing on (k,∞), and consequently

$$0 \leqslant b_i \leqslant k \quad \text{for all } i \in [n]. \tag{78}$$

There are two possibilities.

Case 1: There exists a non-negative number y so that $b_i \in \{0, y\}$ for every $i \in [n]$. Let $\lambda \in [0,1]$ denote the proportion of $i \in [n]$ with $b_i = y$. Then $\lambda y = \frac{1}{n}\sum_{i=1}^n b_i \leqslant 1$.

Hence

$$\Phi(\vec{b}) = \lambda e^{-y} y^k \leqslant e^{-y} y^{k-1} \leqslant \left(\frac{k-1}{e}\right)^{k-1},$$

as required.

Case 2: b_i receives at least two positive values.

Let i and j be two arbitrary indices such that $1 \leqslant i < j \leqslant n$ and $\min\{b_i, b_j\} > 0$. For every small $\varepsilon > 0$, two vectors $\vec{b}' := (b_1, \ldots, b_i + \varepsilon, \ldots, b_j - \varepsilon, \ldots, b_n)$ and $\vec{b}'' := (b_1, \ldots, b_i - \varepsilon, \ldots, b_j + \varepsilon, \ldots, b_n)$ clearly belong to the domain $\mathcal{C}$. By Taylor's formula and the optimality of $\vec{b}$, we thus obtain

$$0 \leqslant \Phi(\vec{b}) - \Phi(\vec{b}') = (f'(b_j) - f'(b_i))\varepsilon + O_{b_i,b_j}(\varepsilon^2),$$
$$0 \leqslant \Phi(\vec{b}) - \Phi(\vec{b}'') = -(f'(b_j) - f'(b_i))\varepsilon + O_{b_i,b_j}(\varepsilon^2).$$

This forces $f'(b_i) = f'(b_j)$.

As $f''(x) = e^{-x}x^{k-2}((x-k)^2 - k)$, we see that $f''(x) > 0$ for $x \in (0, k - \sqrt{k})$, and $f''(x) < 0$ for $x \in (k - \sqrt{k}, k)$. It follows that $f'(x)$ is increasing in $[0, k - \sqrt{k}]$, and decreasing in $[k - \sqrt{k}, k]$. Hence for each $\alpha \in \mathbb{R}$, the equation $f'(x) = \alpha$ has at most two solutions in $[0, k]$.

From the discussion above and (78), we learn that there are two numbers y and z satisfying the following properties:

(i) $b_i \in \{0, y, z\}$ for every $i \in [n]$,
(ii) $f'(y) = f'(z)$, that is, $e^{-y}y^{k-1}(k-y) = e^{-z}z^{k-1}(k-z)$,
(iii) $0 < y < z \leqslant k$.

Let λ and μ be the proportions of $i \in [n]$ such that $b_i = y$ and $b_i = z$, respectively. Then $\lambda + \mu \leqslant 1$, $\lambda y + \mu z \overset{(i)}{=} \frac{1}{n}\sum_{i=1}^n b_i \leqslant 1$, and $\Phi(\vec{b}) \overset{(i)}{=} \Psi(\lambda, \mu)$, where $\Psi(\lambda, \mu) := \lambda f(y) + \mu f(z)$.

We will view y and z as constants, and seek to maximize $\Psi(\lambda, \mu)$ under the constraints: $\lambda, \mu \geqslant 0$, $\lambda + \mu \leqslant 1$, and $\lambda y + \mu z \leqslant 1$. Let (λ_0, μ_0) be a maximizer of Ψ. We claim that $\min\{\lambda_0, \mu_0\} = 0$. Before proving the claim, let us show how it implies the lemma. Indeed, if one of λ_0 and μ_0 is zero, say μ_0, then

$$\Phi(\vec{b}) = \Psi(\lambda, \mu) \leqslant \Psi(\lambda_0, \mu_0) = \lambda_0 e^{-y} y^k \leqslant e^{-y} y^{k-1} \leqslant \left(\frac{k-1}{e}\right)^{k-1},$$

as desired.

It remains to prove that $\min\{\lambda_0, \mu_0\} = 0$. Suppose to the contrary that $\lambda_0, \mu_0 > 0$. Let $\lambda_1 = \lambda_0 - \varepsilon$ and $\mu_1 = \mu_1 + \varepsilon y/z$, where $\varepsilon > 0$ is a small constant. It is not difficult to verify that $\lambda_1, \mu_1 \geqslant 0$, $\lambda_1 + \mu_1 \leqslant 1$, and $\lambda_1 y + \mu_1 z \leqslant 1$. Thus, by the optimality of (λ_0, μ_0), we get $\Psi(\lambda_1, \mu_1) \leqslant \Psi(\lambda_0, \mu_0)$, giving $e^{-z}z^{k-1} \leqslant e^{-y}y^{k-1}$. Combined with (iii), we obtain $e^{-z}z^{k-1}(k-z) < e^{-y}y^{k-1}(k-y)$, contradicting (ii). This completes our proof of the lemma. □

Remark 5.3. *Our proof of Theorem 1.5 crucially relied on our running assumption that almost all the degrees in the graph G are linear. It seems natural to investigate similar extremal questions with a weakened form of this assumption. The minimal meaningful assumption seems to be that G contains no vertices of degree 2; this is to avoid the case of paths. This general setting seems much more complicated. For example, the proportion of leaves in a uniform spanning tree on an $n \times n$ torus is asymptotically almost surely $(1 - 2/\pi) \cdot \frac{8}{\pi^2} + \mathsf{o}(1) \approx 0.294$ (see for example* [18, p. 112]*), which is less than the lower bound of $e^{-1} \approx 0.368$ in Theorem 1.5.*

6. Acknowledgments

Hladký and Tran were supported by the Czech Science Foundation, grant number GJ16-07822Y and by institutional support RVO:67985807. Hladký is supported by the Alexander von Humboldt Foundation. Nachmias is supported by ISF grant 1207/15, and ERC starting grant 676970 RANDGEOM. Part of this work was done while Hladký and Tran were visiting Tel Aviv University. They thank the department of mathematical sciences at TAU for their hospitality. The trip was funded by the Czech Academy of Sciences.

We thank Omer Angel and Tom Hutchcroft for their assistance with the proof of Lemma 3.2 and to Michael Krivelevich and Wojciech Samotij for useful discussions.

References

[1] D. Aldous and R. Lyons. Processes on unimodular random networks. *Electron. J. Probab.*, 12:no. 54, 1454–1508, 2007.

[2] N. Alon. The number of spanning trees in regular graphs. *Random Structures Algorithms*, 1(2):175–181, 1990.

[3] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B*, 38(1):73–88, 1985.

[4] I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 1–13, 2001.

[5] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Percolation on dense graph sequences. *Ann. Probab.*, 38(1):150–183, 2010.

[6] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, 31(1):3–122, 2007.

[7] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[8] A. Frieze. On the number of perfect matchings and Hamilton cycles in ε-regular non-bipartite graphs. *Electron. J. Combin.*, 7:Research Paper 57, 11, 2000.

[9] G. R. Grimmett. Random labelled trees and their branching networks. *J. Austral. Math. Soc. Ser. A*, 30(2):229–237, 1980/81.

[10] G. Kirchhoff. Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird. *Ann. Phys. Chem.*, 72(12).

[11] V. F. Kolchin. Branching processes, random trees, and a generalized scheme of arrangements of particles. *Mathematical notes of the Academy of Sciences of the USSR*, 21(5):386–394, May 1977.

[12] A. V. Kostochka. The number of spanning trees in graphs with a given degree sequence. *Random Structures Algorithms*, 6(2-3):269–274, 1995.

[13] G. Lawler and A. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.*, 309:557–580, 1988.

[14] D. Levin and Y. Peres. with contributions by E. Wilmer. *Markov chains and mixing times*. American Mathematical Society, second edition, 2017. Available via http://darkwing.uoregon.edu/ dlevin/MARKOV/mcmt2e.pdf.

[15] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[16] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[17] R. Lyons. Asymptotic enumeration of spanning trees. *Combin. Probab. Comput.*, 14(4):491–522, 2005.

[18] R. Lyons and Y. Peres. *Probability on trees and networks*, volume 42 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2016.

[19] J. Matoušek and J. Nešetřil. *Invitation to discrete mathematics*. Oxford University Press, Oxford, second edition, 2009.

[20] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computing*, 82:93–133, 1989.

Institut für Geometrie, TU Dresden, 01062 Dresden, Germany.

Department of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel.

Department of Mathematics, ETH, 8092 Zurich, Switzerland.

FROM FLIP PROCESSES TO DYNAMICAL SYSTEMS ON GRAPHONS

FREDERIK GARBE, JAN HLADKÝ, MATAS ŠILEIKIS, AND FIONA SKERMAN

Abstract. We introduce a class of random graph processes, which we call *flip processes*. Each such process is given by a *rule* which is a function $\mathcal{R} : \mathcal{H}_k \to \mathcal{H}_k$ from all labelled k-vertex graphs into itself (k is fixed). The process starts with a given n-vertex graph G_0. In each step, the graph G_i is obtained by sampling k random vertices $v_1, \dots, v_k$ of G_{i-1} and replacing the induced graph $F := G_{i-1}[v_1, \dots, v_k]$ by $\mathcal{R}(F)$. This class contains several previously studied processes including the Erdős–Rényi random graph process and the triangle removal process. Actually, our definition of flip processes is more general, in that $\mathcal{R}(F)$ is a probability distribution on $\mathcal{H}_k$, thus allowing randomised replacements.

Given a flip process with a rule $\mathcal{R}$, we construct time-indexed trajectories $\Phi : \mathcal{W}_0 \times [0, \infty) \to \mathcal{W}_0$ in the space of graphons. We prove that for any $T > 0$ starting with a large finite graph G_0 which is close to a graphon W_0 in the cut norm, with high probability the flip process will stay in a thin sausage around the trajectory $(\Phi(W_0, t))_{t=0}^{T}$ (after rescaling the time by the square of the order of the graph).

These graphon trajectories are then studied from the perspective of dynamical systems. Among others topics, we study continuity properties of these trajectories with respect to time and initial graphon, existence and stability of fixed points and speed of convergence (whenever the infinite time limit exists). We give an example of a flip process with a periodic trajectory.

1. Introduction

1.1. **The Erdős–Rényi random graph process.** The uniform Erdős–Rényi random graph model $\mathbb{G}(n, M)$ is 63 years old and now the theory of random graphs is an active field at the border between graph theory and probability. While studying this model provides insights in many situations, sometimes one is in an evolving setting, of which $\mathbb{G}(n, M)$ is just one snapshot. More precisely, the *Erdős–Rényi random graph process*, for a given integer n, is a random sequence of graphs $G_0, G_1, \dots, G_{\binom{n}{2}}$, all on vertex set $[n]$, in which G_0 is edgeless and each $G_{\ell+1}$ is created from G_ℓ by flipping a uniformly random

FG: Research supported by GAČR project 18-01472Y and RVO: 67985840. Furthermore, this work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 648509). This publication reflects only its authors' view; the European Research Council Executive Agency is not responsible for any use that may be made of the information it contains. This author was also supported by the MUNI Award in Science and Humanities of the Grant Agency of Masaryk University; JH: Research supported by Czech Science Foundation Project 21-21762X. MŠ: Research supported by Czech Science Foundation Project GJ20-27757Y, with Institutional Support RVO:67985807. FS: Research supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP), the project AI4Research at Uppsala University, and a Simons-Berkeley Research Fellowship.

non-edge into an edge. Lastly, $G_{\binom{n}{2}}$ is a complete graph and so there is no meaningful way to continue the process.

It is easy to prove that as n goes to infinity, all the graphs $G_0, G_1, \dots, G_{\binom{n}{2}}$ are $o(1)$-quasirandom, asymptotically almost surely. Here and throughout the paper, we will use the concept of quasirandomness which was developed in the 1980s by Chung, Graham and Wilson [10] and by Thomason [25] and reflected the then emerging theory surrounding the Szemerédi regularity lemma [22]. Namely, we say that an n vertex graph is *ε-quasirandom of density d* if it has $d\binom{n}{2}$ edges and each set of vertices U induces $\binom{|U|}{2}d \pm \varepsilon n^2$ edges.

Let us consider a small modification of this process, which we call the *Erdős–Rényi flip process*. We again start with the edgeless graph on $[n]$. To create $G_{\ell+1}$ from G_ℓ, we pick a uniformly random pair xy of vertices and make that pair an edge. In particular, if xy was an edge, then $G_{\ell+1} = G_\ell$. Let us make some remarks about the Erdős–Rényi flip process. Firstly, this process is defined for any number of steps, and eventually stabilises at the complete graph (around time $\tau_{\mathrm{ER}} = n^2 \ln n$). Secondly, the probability that a fixed pair xy forms an edge in a graph G_ℓ is equal to

$$1 - \left(1 - \frac{1}{\binom{n}{2}}\right)^{\ell} \approx 1 - \exp\left(-\frac{2\ell}{n^2}\right) ,$$

and indeed it is easy to prove that throughout the evolution the graphs G_ℓ stay quasirandom, of density around

$$1 - \exp(-2t) \tag{1.1}$$

for each $t > 0$ and $\ell = tn^2$, asymptotically almost surely.

Note the number of edges in the Erdős–Rényi flip process after k steps is not concentrated on one number (unlike in the uniform Erdős–Rényi random graph) and is not binomial either (unlike in the binomial Erdős–Rényi random graph). However, both these models are good proxies since our view and results will be insensitive to changes of $o(n^2)$ edges, as usual in the theory of dense graph limits.

1.2. **The triangle removal process.** The *triangle removal process* was introduced by Bollobás and Erdős around 1990. We start with the complete graph G_0. Each $G_{\ell+1}$ is created from G_ℓ by firstly selecting a uniformly random triangle in G_ℓ and then deleting all its three edges. Thus, the number of edges in each step of the process decreases by three and the process cannot be continued when we arrive at a triangle-free graph. Observe this stopping time is random. A result of Bohman, Frieze, and Lubetzky [6] says that there are $n^{3/2+o(1)}$ edges left in this final graph, asymptotically almost surely. To show this, Bohman, Frieze, and Lubetzky set up complicated machinery which allows them to control quasirandomness of the sequence during the evolution. In particular, throughout the evolution, the graphs stay $o(1)$-quasirandom asymptotically almost surely.

As above, we can modify the triangle removal process into the *triangle removal flip process*. In this modification, in each step we sample a triple of vertices uniformly at random and if that triple induces a triangle, then we remove all the three edges. Note that the triangle removal flip process is exactly the triangle removal process, where between each pair of consecutive steps of the triangle removal process we insert a random number of copies of the first graph of the pair, the random number being geometrically distributed with a parameter corresponding to the density of triangles of this first graph. It can be proven that this process reaches a triangle-free graph in $\tau_{\mathrm{TR}} = n^{4+o(1)}$ steps asymptotically almost surely, at which point it stabilises (it is not trivial to prove this, however we will only need it for informal comments).

Let us now try to understand the evolution of the density in the triangle removal flip process. Suppose that the density of a graph G_ℓ is d. Then, because G_ℓ is quasirandom, the probability that a random triple of vertices induces a triangle is roughly d^3. In other words, if we fix $\varepsilon > 0$, then the number of triangle removals between rounds ℓ and $\ell + \varepsilon n^2$ is around $\varepsilon d^3 n^2$ and hence the density drops to around $d - 6\varepsilon d^3$. Note that one source of imprecision is the assumption that within this window, a triangle is still sampled with probability d^3 which is true only at the beginning of the window. This imprecision, however, is negligible if ε is small. So, if for $t \geq 0$ we denote by $d(t)$ the density of the graph $G_{\lfloor tn^2 \rfloor}$ and we treat $d(t)$ as a number (actually, it is a random variable, though highly concentrated), then the heuristics given above suggest a differential equation

$$\begin{aligned} d(0) &= 1 \,, \\ d'(t) &= -6d(t)^3 \,, \end{aligned} \tag{1.2}$$

which has a unique solution on $[0, \infty)$,

$$d(t) = (1 + 12t)^{-1/2} \,. \tag{1.3}$$

1.3. **Flip processes.** We can now generalise the idea of local replacements that defined the Erdős–Rényi flip process and the triangle removal flip process.

We first define a subclass of flip processes which we call *definite flip processes of order* k. Let $\mathcal{H}_k$ be the family of graphs on vertex set $[k]$. A *rule* is an arbitrary function $\mathcal{R} : \mathcal{H}_k \to \mathcal{H}_k$. We then consider a random discrete time process $(G_\ell : \ell = 0, 1, \dots)$ of graphs on the vertex set $[n]$, where $n \geq k$ is arbitrary. We are free to choose the initial graph G_0 (deterministically or randomly). In each step, the graph $G_{\ell+1}$ is obtained by sampling k random vertices $v_1, \dots, v_k$ of G_ℓ (conditioned to be distinct) and replacing the induced graph $G_\ell[v_1, \dots, v_k]$ — which we call the *drawn graph* — by $\mathcal{R}(G_\ell[v_1, \dots, v_k])$ — which we call the *replacement graph*.

Clearly, the Erdős–Rényi flip process and the triangle removal flip process are definite flip processes of order 2 and 3, respectively, see Figure 1.1. Also, note that there is no need

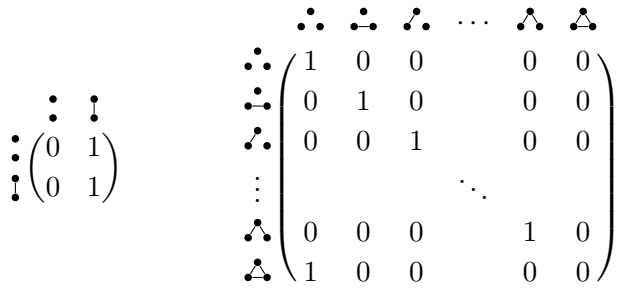


FIGURE 1.1. Rules of the Erdős–Rényi and triangle removal flip processes.

to start the Erdős–Rényi flip process with the edgeless graph and the triangle removal flip process with the complete graph. For example, if we start the triangle removal flip process with a triangle-free graph, then it stabilises immediately.

To generalise these definite flip processes we allow additional randomisation in the replacement procedure. A *rule of a flip process of order* k is a matrix $\mathcal{R} \in [0,1]^{\mathcal{H}_k \times \mathcal{H}_k}$ such that for every $H \in \mathcal{H}_k$ we have

$$\sum_{J \in \mathcal{H}_k} \mathcal{R}_{H,J} = 1. \tag{1.4}$$

So, given a rule $\mathcal{R}$ and integer $n \geq k$, the *flip process with rule* $\mathcal{R}$ is a discrete time process $(G_\ell : \ell = 0, \dots)$ of graphs on the vertex set $[n]$. Again, we start with a given graph G_0, and for $\ell \geq 0$ we obtain graph $G_{\ell+1}$ from graph G_ℓ as follows. Sample, uniformly at random, an ordered tuple $\mathbf{v} = (v_1, \cdots, v_k)$ of distinct vertices. Let $H \in \mathcal{H}_k$ be the unique graph, called the *drawn graph*, such that $v_i v_j$ is an edge in G_ℓ if and only if $ij \in E(H)$. Sample a graph J from the *replacement distribution (for the given drawn graph H)* $(\mathcal{R}_{H,J})_J$ and replace $G_\ell[\mathbf{v}]$ by J, that is, turn $v_i v_j$ into an edge if $ij \in J$ and $v_i v_j$ into a non-edge if $ij \notin J$. We call the resulting graph $G_{\ell+1}$. Note that the order of the elements in $\mathbf{v}$ matters.

For brevity we will sometimes say "flip process $\mathcal{R}$" meaning "flip process with rule $\mathcal{R}$", having in mind, precisely speaking, the class of processes (given by all possible initial graphs G_0).

Flip processes are defined for all times. However, we shall only be looking at typical behaviours at times $x \cdot n^2$, $x \in [0, +\infty)$, which we call *bounded horizon*. While there are certainly interesting features at later times, as we saw with τ_{ER} and τ_{TR} in Sections 1.1 and 1.2, these features cannot be captured by our theory.

Given a process of order k with rule $\mathcal{R}$, we call a graph $H \in \mathcal{H}_k$ *idle* if $\mathcal{R}_{H,H} = 1$. We say that the process is *trivial* if each graph is idle.

Observe that each flip process is a time-homogeneous Markov process.

Let us give a sample questions we address in this paper (of course, with respect to the asymptotics $n \to \infty$; all statements hold asymptotically almost surely; the number 99 below is given only for the sake of concreteness).

(Q1) Is there a flip process which would start with a quasirandom graph of order n and density 0.5 and in $99n^2$ steps produces a complete balanced bipartite graph? *No for two reasons. Firstly, flip processes cannot create additional macroscopic structure in a bounded horizon. Secondly, they cannot achieve densities 0 or 1 (unless the corresponding part of the graph had these densities initially) in a bounded horizon.*

(Q2) Is there a flip process which would start with a complete balanced bipartite graph and in $99n^2$ steps produces a quasirandom graph of density 0.5? *No, flip processes cannot entirely wipe out the macroscopic structure in a bounded horizon.*

(Q3) Is there a flip process and an initial graph G_0 so that the evolution is $99n^2$-periodic? That is, we want that G_ℓ is similar to $G_{\ell+99n^2}$ for all ℓ, while it is different from, say, $G_{\ell+40n^2}$. *Yes, we can construct an example of such a periodic process.*

(Q4) Suppose our partner runs a flip process with a known rule $\mathcal{R}$ but a hidden initial graph G_0 for $99n^2$ steps and shows us the final graph G_{99n^2}. Is there a way to approximately reconstruct G_0? *Yes.*

(Q5) Is there a flip process where a small change in the initial graph G_0 would result in a large change in G_{99n^2}? *No.*

1.4. **Passing to graphons.** In this section, we assume basic familiarity with the theory of graphons. We give more details in Section 3.2. In particular, $\mathcal{W}_0$ is the set of all graphons, that is, symmetric measurable functions defined on the square of a probability space Ω^2 and with values in the interval $[0,1]$. We write $d_\square(U,W) = \|U-W\|_\square$ for the cut norm distance between U and W, and regard $\mathcal{W}_0$ as a metric space with metric $d_\square$.

As we noted earlier, the Erdős–Rényi flip process, when started from the edgeless graph on a vertex set $[n]$, is typically quasirandom, with edge density around $1-e^{-2t}$ in step tn^2. We can thus say that after rescaling the time by n^2, the Erdős–Rényi flip process typically evolves like the function $t \mapsto \Phi^t_{\mathrm{ER}}$, where $\Phi^t_{\mathrm{ER}} \equiv 1-e^{-2t}$ is the constant graphon with density $(1-e^{-2t})$.

Similarly, we saw that in the triangle removal flip process, when started from the complete graph, the graphs are typically quasirandom with edge density around $(1+12t)^{-1/2}$ at step tn^2. We can thus say that after rescaling the time by n^2, the triangle removal flip process typically evolves like the function $t \mapsto \Phi^t_{\mathrm{TR}}$, where $\Phi^t_{\mathrm{TR}} \equiv U_{(1+12t)^{-1/2}}$ is the constant graphon with density $(1+12t)^{-1/2}$.

In general, given a fixed rule $\mathcal{R}$ of order k, we shall construct *trajectories*, that is, a we shall define the evolution function $\Phi : \mathcal{W}_0 \times [0,+\infty) \to \mathcal{W}_0$, where we treat the

first coordinate as the *initial graphon*, and the second coordinate as the *time*. Instead of writing $\Phi(W,t)$ we put the time in the superscript as $\Phi^t(W)$. It will turn out that the trajectories correspond to a dynamical system generalising the differential equation (1.2). However, (1.2) was a differential equation with a single numerical parameter (the edge density). In the general setting, the differential equation will involve a graphon-valued trajectory. So, the key objects here are Banach-space-valued differential equations (the main differential equation is given in (4.5) below). The main statement about the existence of trajectories is Theorem 4.5.

As a preview of our general results in Section 4, in Figure 1.2 we illustrate some trajectories that take values among non-constant graphons.

The *Transference Theorem* (Theorem 5.1) says that if we start with a large initial graph G_0 of order n, fix a representation of its vertices on Ω and hence obtain a graphon representation W_0 of G_0, then, with high probability, each graph G_{tn^2} (where $t \in [0,\infty)$) when represented as a graphon, is close to $\Phi^t(W_0)$ in the cut norm. Obviously, we then must have $\Phi^0(U) = U$ for each graphon U. Furthermore, the trajectories are continuous with respect to time, and with respect to the initial graphon. The last fact means that G_{tn^2} is also close to $\Phi^t(U)$, if W_0 is close to U in the cut norm. Let us state an informal version of this key result here.

Theorem (Informal and simplified consequence of Theorems 4.5, 4.16 and 5.1)**.** *Suppose that $\mathcal{R}$ is a rule. Then there exist $\Phi : \mathcal{W}_0 \times [0,+\infty) \to \mathcal{W}_0$ with the following property. Suppose that $T > 0$. Suppose that G_0 is an n-vertex graph and U is a graphon such that the graphon representation W_0 of G_0 satisfies $d_\square(W_0, U) = o(1)$.*

Consider the flip process $(G_i)_{i\geq 0}$ with rule $\mathcal{R}$ started from G_0. Write W_i for the graphon representation of G_i. Then with probability $1-o(1)$ we have $\max_{i\in\mathbb{N}\cap[0,Tn^2]} d_\square(W_i, \Phi^{i/n^2}U) = o(1)$. (The asymptotics $o(1)$ is for fixed $\mathcal{R}$, T and as $n \to \infty$.)

Note that we equipped the space of graphons with the cut norm distance $d_\square$ rather than the cut distance $\delta_\square$ (which factors out isomorphic graphons). This allows us not only to track how an initial graph evolves up to isomorphism, but also tells us how individual parts of the graph evolve. We give a specific example in Figure 1.3.

So, Theorem 5.1 transfers problems about random discrete trajectories on graphs to problems about deterministic continuous trajectories on graphons. With this latter view, combinatorial problems such as those we asked in Section 1.3 translate to problems studied in dynamical systems such as the existence and stability of fixed points or periodic trajectories.

This opens a new perspective and a new set of problems: we answer many of them, yet many more are left unexplored.

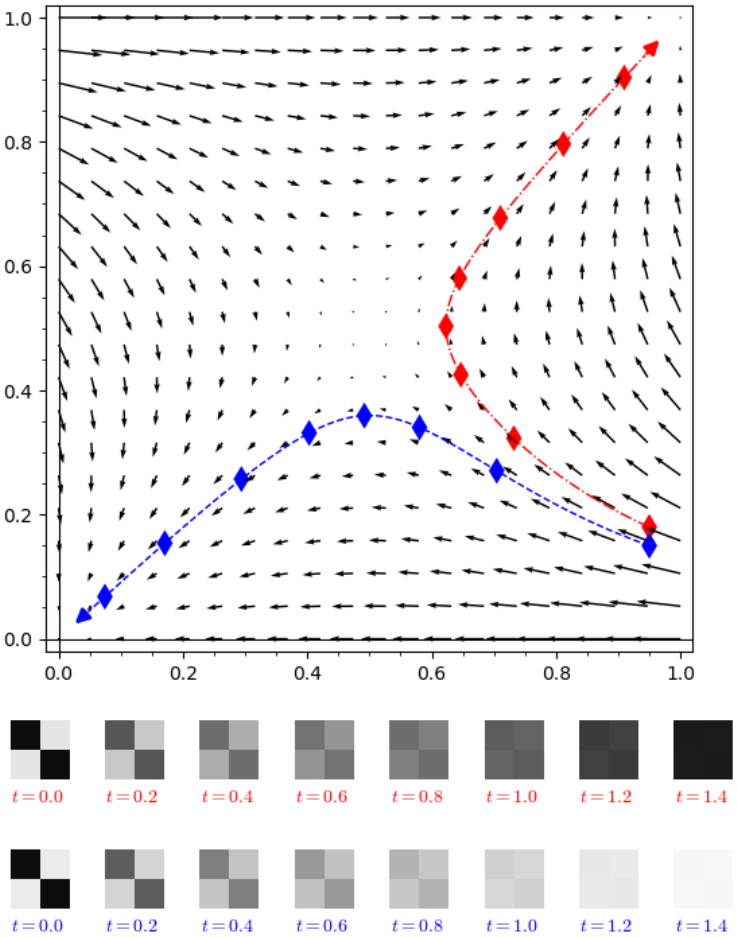


FIGURE 1.2. Extremist flip process of order three works as follows: if the drawn graph has two or three edges we replace it with the triangle and otherwise we replace it by the edgeless graph. Consider a class of step graphons $\{U_{x,y} : x, y \in [0,1]\}$ with steps Ω_1, Ω_2 of measure $1/2$, where $U_{x,y}$ takes value x on $\Omega_1^2 \cup \Omega_2^2$ and value y elsewhere. We represent $U_{x,y}$ by a point $(x, y) \in [0,1]^2$. It turns out that trajectories starting in this class do not leave it. The arrows in the diagram depict the tangent vectors of such trajectories. The red line is the trajectory from initial graphon $U_{0.95,0.18}$ with diamond markers showing the positions along the trajectory at times $t = 0$, $t = 0.2$, ..., $t = 1.4$; the graphon represented by each marker is plotted in grayscale underneath the diagram. The blue line is similar, from initial graphon $U_{0.95,0.15}$.

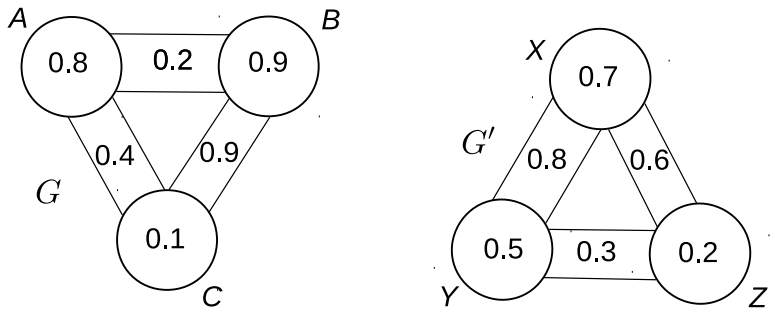


FIGURE 1.3. Two graphs G and G', consisting of three blobs A, B, C and X, Y, Z, respectively, each of order $n/3$. Inside the blobs and between the pairs of blobs we have quasirandom and quasirandom bipartite graphs, with densities as indicated. Suppose that we start a flip process with G. Then our results not only say whether after a certain number of steps we typically arrive to a graph with a structure as in G', but they also give, in the positive case, the correspondence of the sets (A, B, C) and of (X, Y, Z).

1.5. **Related work.** In addition to the Erdős–Rényi and triangle removal processes there are many other graph processes considered in the literature, the H-free process (see [7] and references therein), the Achlioptas process (see [20] and references therein), or the preferential attachment process (see [5]), to name a few. These processes do not fall into our framework, and in particular are typically studied in the sparse regime. An interesting open direction, suggested to us by Mihyun Kang, is to study modifications of flip processes which would incorporate some features of these other standard models.

For example one could model preferential attachment by introducing 'types' of vertices. The initial graph would consist of a single edge between two 'active' vertices (or loop at one 'active' vertex) and very many isolated 'unused' vertices. The rule would then be to pick two vertices at random and if one of them is 'active' and one 'unused' add an edge between them and change the type of the 'unused' vertex to 'active'; otherwise do nothing. Note the rate of growth would depend on the number of 'unused' vertices in the construction. Further generalisations with rules of higher order are also possible.

Another related work is that of Keliger [15]. In that paper, vertices have a finite number of types, and an update rule at each vertex takes into account the type of that vertex and distribution of the types among the neighborhood. The paper focuses on dense graphs and uses a somewhat similar theory of ODEs.

A different line of research looks at stochastic processes on graphs which satisfy abstract properties and studies their projection into the space of graphons. This direction was

started by Crane [11], who studied exchangeable[a] stochastic processes on graphs with countably many vertices. Roughly speaking, leaving out various regularity conditions, they proved that these correspond to Markov processes in the space of graphons. In the follow-up work of Athreya, den Hollander, and Röllin [3], more complicated processes on graphs are studied, to which the corresponding graphon processes exhibit diffusive behavior.

Last, let us draw a connection between our work and *Wormald's differential equation method* (DEM). This method was introduced in [26], unifying many previous ad-hoc analyses, mostly of randomized algorithms on random graphs. In such a setting one needs to keep track of random numerical parameters (say, the number of matched vertices in some sequential matching algorithm), and there is a heuristic suggesting that these parameters in expectation evolve according to a certain (deterministic) differential equation. DEM then gives general conditions which guarantee that with high probability the parameters indeed do follow the evolution predicted by the differential equation. In that sense there is a clear similarity to our Transference Theorem (Theorem 5.1). The main difference is that DEM links typical behavior of numerical parameters, whereas Theorem 5.1 links typical behaviour of the entire structure of the evolving graph (described by the language of graphons) to a graphon-valued differential equation.

1.6. **Organisation of the paper.** We start by giving examples of several interesting classes of flip processes in Section 2. It uses only a moderate amount of references to later parts of the paper and is therefore a good starting point for reading the paper. Section 3 gives necessary preliminaries, mostly on graphons. In Section 4 we construct trajectories $\Phi : \mathcal{W}_0 \times [0, +\infty) \to \mathcal{W}_0$ for an arbitrary flip process as advertised in Section 1.4. In fact, the construction extends also to negative times, as was already hinted by (Q4) in Section 1.3. In Section 4, we also state basic properties of trajectories, such as continuity mentioned in (Q5). In Section 5 we state and prove the Transference Theorem, which tells us that a typical evolution of a flip process on a finite graph stays close to the graphon trajectory. In Section 6 we establish many general features of trajectories, including those that settle questions (Q1)–(Q3).

2. Examples

We believe that our framework provides countless interesting specific examples. Below we give seven classes of flip processes which we found interesting. See [2] for details. In this section, k is the order of the flip process in question.

[a] A processes is *exchangeable* if it is invariant under arbitrary relabeling of vertices by finite permutations. The concept of exchangeability is connected to graph limits as explained in [4, 12].

2.1. **Ignorant flip processes.** Ignorant flip processes are processes where the replacement graph does not depend on the drawn graph. More formally, we require that the rule $\mathcal{R}$ of an ignorant flip process satisfies for each $H, H', J \in \mathcal{H}_k$ that $\mathcal{R}_{H,J} = \mathcal{R}_{H',J}$. We then have the *ignorant replacement distribution* $\mathcal{D} = \{\mathcal{R}_{H,J}\}_J$ (here, H is arbitrary), and its *average density* $d := \frac{1}{\binom{k}{2}} \cdot \mathbb{E}_{J\in\mathcal{D}}[e(J)]$.

Observe that the Erdős–Rényi flip process is an ignorant flip process of order $k = 2$ with ignorant replacement distribution which is the Dirac measure on the edge K_2.

A crucial (and easy to establish) property is that each ignorant process of order $k \geq 2$ has all trajectories converging to the constant graphon corresponding to the average density of its replacement distribution.

2.2. **Monotone flip processes.** Roughly speaking, a monotone flip process should be one, in which the replacement graph is 'monotone' with respect to the drawn graph. Besides an obvious choice whether 'monotone' should mean non-decreasing or non-increasing, and whether we require non-strict monotonicity or strict monotonicity,[b] there are three further natural definitions: (we now stick to the choice of non-decreasing flip processes)

- *Pointwise non-decreasing flip processes* in which we require that $\mathcal{R}_{H,J} = 0$ unless J is a supergraph of H (as a labelled graph).
- *Edge non-decreasing flip processes* in which we require that $\mathcal{R}_{H,J} = 0$ unless $e(J) \geq e(H)$.
- *Averaged-edge non-decreasing flip processes* in which we require that $\sum_J \mathcal{R}_{H,J} \cdot e(J) \geq e(H)$ for every H.

Obviously, the 'averaged-edge' version is weaker than the 'edge' version which is weaker than the pointwise version. Edge non-decreasing flip processes (and thus pointwise non-decreasing flip processes) are always convergent whereas averaged-edge non-decreasing flip processes need not be.

2.3. **Removal flip processes.** These flip processes generalise the triangle removal flip process. That is, we have a graph $H \in \mathcal{H}_k$. When the drawn graph H is a supergraph of F, the replacement graph is $H \setminus F$. Other drawn graphs are idle.

Obviously, each removal flip process is a pointwise non-increasing flip process. The set of destinations of all trajectories is the set $\mathcal{W}_F$ of F-free graphons, that is, the graphons W with $t(F, W) = 0$.

The question of speed of convergence in removal flip processes is a non-trivial one with a connection to the removal lemma.

[b] Of course, this strictness has to be exempted when the drawn graph is complete (for increasing flip processes) or edgeless (for decreasing flip processes).

2.4. **The complementing flip process.** The complementing flip process is a process in which the drawn graph is replaced by its complement.

Trajectories of this flip process converge very quickly to the constant-$\frac{1}{2}$ graphon.

2.5. **The component completion flip process.** The component completion flip process is a process in which the drawn graph is replaced by its *component closure*, that is the graph is a disjoint union of cliques which has the same component structure.

For $k \geq 3$, the trajectories of a component completion process converge to a component closure of the original graphon.

2.6. **The stirring flip process.** The stirring process is defined in two variants, *firm* or *loose*, as follows. Suppose that the drawn graph is $H \in \mathcal{H}_k$. Set $M := e(H)$ and $p := M/\binom{k}{2}$. The replacement graph is the uniform Erdős–Rényi random graph $\mathbb{G}(k, M)$ (the firm variant) or the binomial Erdős–Rényi random graph $\mathbb{G}(k, p)$ (the loose variant).

So, in the firm stirring process we always preserve the number of edges of the underlying graph, and in the loose stirring process we preserve it in expectation.

Each trajectory converges to a constant graphon of the same density as the initial graph. For initial graphons that are regular (i.e., all the degrees are equal to a constant d) these trajectories are the same as for an ignorant process of replacement distribution with average density d. However, for non-regular initial graphons, the trajectories are different than trajectories of ignorant processes.

2.7. **The extremist process.** The extremist process is defined as follows. Suppose $H \in \mathcal{H}_k$ is the sampled graph. Set $\ell := e(H)$. If $\ell > \binom{k}{2}/2$, then the replacement graph is the complete graph. If $\ell < \binom{k}{2}/2$, then the replacement graph is the edgeless graph. If $\ell = \binom{k}{2}/2$, then $\mathcal{R}$ is idle.

If W is a constant-d graphon, then the trajectory converges to constant-0 for $d < \frac{1}{2}$ and to constant-1 for $d > \frac{1}{2}$. Trajectories starting at non-constant graphons are more subtle. For $k \geq 5$ there exist a *density threshold* $d_k > 0$ such that the trajectory of each graphon of edge density less than d_k converges to constant-0, and symmetrically, the trajectory of each graphon of edge density more than $1 - d_k$ converges to constant-1. Further, $d_k \to \frac{1}{2}$ as $k \to \infty$. No positive density thresholds exist for $k = 3, 4$.

3. Preliminaries

3.1. Basic notation and simple facts.

3.1.1. *Basic notation.* We write $(n)_k$ for the *falling factorial*, $(n)_k := n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$. Given a set A, we write $\mathbb{I}_A$ for the indicator function of A. The symmetric difference is denoted by $\ominus$.

3.1.2. *Graphs.* All graphs in this paper are simple, loopless and undirected. Given a graph $G = (V, E) = (V(G), E(G))$, we write $v(G) = |V|$ and $e(G) = |E|$. Given two sets $A, B \subset V$, we write $e_G(A, B)$ for the number of pairs $(a, b) \in A \times B$ such that $ab \in E$. In particular, edges in $A \cap B$ are counted twice.

Let $k \in \mathbb{N}$. We denote the family of all labelled graphs on the vertex set $[k]$ by $\mathcal{H}_k$. Given a labeled graph $F = (V(F), E(F))$ and an *ordered* subset $R \subseteq V(F)$ of vertices, we call $(R, V(F), E(F))$ a *rooted graph*. Whenever $R = (a, b)$ is an ordered pair, we simply denote $(R, V(F), E(F))$ by $F^{a,b}$.

3.2. **Graphons.** Our notation concerning graphons mostly follows [16], which is recommended for further reading.

All our graphons will be defined on an atomless measure space (Ω, π). The sigma-algebra is given implicitly and is assumed to be *separable*, that is, there is a countable set $\mathcal{C}$ of measurable sets such that for every measurable set S and every $\delta > 0$ there is $T \in \mathcal{C}$ such that $\pi(S \ominus T) < \delta$. We will often also make the measure implicit and, in particular, for $k \in \mathbb{N}$ we write Ω^k for the product probability space. By $\mathcal{W}$ we denote the space of measurable functions $W : \Omega^2 \to \mathbb{R}$ which are symmetric (i.e. $W(x, y) = W(y, x)$ almost everywhere[c]) and bounded (i.e., $\operatorname{ess\,sup} |W| < \infty$). We identify two functions that are equal a.e. and tacitly assume $\mathcal{W}$ consists of equivalence classes of such functions. Hence $\mathcal{W}$ is a closed linear subspace of the space $L^\infty(\Omega^2)$ and thus Banach when endowed with the norm $\|\cdot\|_\infty$. We, however, will also consider other norms on $\mathcal{W}$. We call elements of $\mathcal{W}$ *kernels* and elements of $\mathcal{W}_0 := \{W \in \mathcal{W} : W \in [0, 1]\}$ *graphons*. Here and elsewhere, we write $W \in S$ to denote that up to a null set the W takes values in S, that is, the set $\{(x, y) \in \Omega^2 : W(x, y) \notin S\}$ is null. We will also write, for two measurable functions defined on the same measure space, $f = g$ and $f \leq g$ meaning that equality or inequality holds a.e. With this notation, we recall that the *essential supremum* $\operatorname{ess\,sup} f$ is the least constant c such that $f \leq c$ and the *essential infimum* $\operatorname{ess\,inf} f$ is defined analogously.

The set $\mathcal{W}_0$ is a closed $\|\cdot\|_\infty$-ball of radius $1/2$ centered at the constant-$\frac{1}{2}$ graphon. For $\varepsilon \geq -1/2$, we define a ball with the same center but radius $\frac{1}{2} + \varepsilon$,

$$\mathcal{W}_0^\varepsilon := \{W \in \mathcal{W} : W \in [-\varepsilon, 1 + \varepsilon]\}. \tag{3.1}$$

3.2.1. *Graphon representation.* Suppose that $G = (V, E)$ is a graph. Suppose that we have a partition $(\Omega_v : v \in V)$ of Ω into sets of measure $1/|V|$ each. Then we define the *graphon representation of* G, which we will denote by W_G, as the graphon $W : \Omega^2 \to \{0, 1\}$ defined to be 1 on each rectangle $\Omega_u \times \Omega_v$ for which $uv \in E$, and 0 otherwise. Note that this graphon representation may depend on the choice of the partition $(\Omega_v : v \in V)$. However,

[c]Here and below we say *almost everywhere* or *a.e.* when we mean μ-*almost everywhere*, whenever the measure μ is clear from the context; say, in this particular instance $\mu = \pi^2$.

we will frequently keep the partition implicit. The important thing is that we will assume that graphs on the same vertex set are represented using the same partition of Ω.

3.2.2. *Cut norm, cut distance and densities.* The *cut norm* on $\mathcal{W}$ is defined to be

$$\|W\|_\square := \sup_{A,B\subseteq\Omega} \left| \int_{A\times B} W \,\mathrm{d}\pi^2 \right|, \tag{3.2}$$

where the supremum is taken over measurable sets.

We will occasionally use the obvious inequalities

$$\|W\|_\square \le \|W\|_1 \le \|W\|_\infty, \quad W \in \mathcal{W}. \tag{3.3}$$

The metric induced by the cut norm is called the *cut norm distance*,

$$d_\square(U,W) := \|U-W\|_\square .$$

The *cut distance* is defined for $U, W \in \mathcal{W}$ as

$$\delta_\square(U,W) := \inf_{\varphi\in\mathbb{S}_\Omega} d_\square\left(U, W^\varphi\right), \tag{3.4}$$

where $W^\varphi(x,y) = W(\varphi(x),\varphi(y))$ and $\mathbb{S}_\Omega$ is the set of all measure preserving bijections from Ω to Ω.

Remark 3.1. *It is known that $\delta_\square(U,W) = 0$ if and only if U and W are* weakly isomorphic*, that is, when for every finite graph F, we have $t(F,U) = t(F,W)$ (using the notion of densities defined in Definition 3.3 below). Since weak isomorphism is an equivalence relation, we define $\widetilde{\mathcal{W}}$ to be $\mathcal{W}$ with equivalent kernels identified, which, together with $\delta_\square$ turns $\widetilde{\mathcal{W}}$ into a metric space.*

The following fact will also be useful.

Lemma 3.2. *The metric space $(\mathcal{W}_0, d_\square)$ is complete.*

A proof of a substantially more general result can be found in [19, Proposition 1]. We give a self-contained proof of Lemma 3.2 in the Appendix.

For a function $f : \Omega^{[k]} \to \mathbb{R}$ and a subset $S \subset [k]$ we write

$$\int f \,\mathrm{d}\pi^S = \int_{\Omega^S} f(x_i : i \in [k]) \prod_{i\in S} \pi(\,\mathrm{d}x_i).$$

We use this notation to introduce the key concept of density of a finite graph in a kernel.

Definition 3.3 (Densities in kernels)**.** *Given a kernel $W \in \mathcal{W}$ and a graph $F = (V,E)$, we define the* density *of F in W as*

$$t(F,W) := \int \prod_{ij\in E} W(x_i,x_j) \,\mathrm{d}\pi^V.$$

and the induced density *as*

$$t_{\mathrm{ind}}(F,W) \coloneqq \int \prod_{ij\in E} W(x_i,x_j) \prod_{ij\notin E} (1-W(x_i,x_j))\, \mathrm{d}\pi^V.$$

Furthermore, we need to work with a notion of a rooted density.

Definition 3.4 (Rooted densities in kernels)**.** *Given a kernel $W \in \mathcal{W}$, a rooted graph $F=(R,V,E)$, and $\mathbf{x}\in\Omega^R$ we define*

$$t^{\mathbf{x}}(F,W) \coloneqq \int \prod_{ij\in E} W(x_i,x_j)\, \mathrm{d}\pi^{V\setminus R}\,, \text{ and} \tag{3.5}$$

$$t^{\mathbf{x}}_{\mathrm{ind}}(F,W) \coloneqq \int \prod_{ij\in E} W(x_i,x_j) \prod_{ij\notin E} [1-W(x_i,x_j)]\, \mathrm{d}\pi^{V\setminus R}\,. \tag{3.6}$$

Note that $\mathbf{x}\mapsto t^{\mathbf{x}}(F,W)$ and $\mathbf{x}\mapsto t^{\mathbf{x}}_{\mathrm{ind}}(F,W)$ are functions in $L^\infty(\Omega^R)$ and are not necessarily symmetric.

For convenience we define the operator $T_F^{a,b}:\mathcal{W}\to L^\infty(\Omega^2)$, by setting

$$(T_F^{a,b}W)(x,y) = t^{(x,y)}_{\mathrm{ind}}(F^{a,b},W)\,. \tag{3.7}$$

For future reference we note that for any rooted graph F and any $k\geq v(F)$,

$$t^{\mathbf{x}}(F,W) = \sum_{H\in\mathcal{H}_k : E(F)\subseteq E(H)} t^{\mathbf{x}}_{\mathrm{ind}}(H,W)\,, \tag{3.8}$$

where in the summation F is treated as a fixed rooted labeled graph, say, with root $R=[|R|]$ and vertex set $V(F)=[v(F)]$, and each H is treated as the rooted graph $(R,[k],E(H))$. Moreover, for any two graphs F,G with $v(F)\leq v(G)$ and for a graphon representation W_G of G we have (see [16, Exercise 7.7]) that

$$|t_{\mathrm{ind}}(F,G) - t_{\mathrm{ind}}(F,W_G)| \leq \frac{\binom{v(F)}{2}}{v(G)}\,. \tag{3.9}$$

Similarly, if W_G is a graphon representation of G with respect to a partition $(\Omega_v : v\in V(G))$, then for every $u,v\in V(G)$ and $x\in\Omega_u$, $y\in\Omega_v$, we have

$$\left|t^{(u,v)}_{\mathrm{ind}}(F,G) - t^{(x,y)}_{\mathrm{ind}}(F,W_G)\right| \leq \frac{\binom{v(F)}{2}}{v(G)}\,. \tag{3.10}$$

(The proof is also an easy exercise.)

Next, we need a bound on expansion of the operator $T_F^{a,b}$ with respect to the L^∞-metric.

Lemma 3.5. *Let $F\in\mathcal{H}_k$ and $a,b\in[k]$ be two distinct vertices. Let $\varepsilon\geq 0$ be a constant. If $U,W\in\mathcal{W}_0^\varepsilon$, then*

$$\left\|T_F^{a,b}U - T_F^{a,b}W\right\|_\infty \leq (1+\varepsilon)^{\binom{k}{2}-1}\binom{k}{2}\|U-W\|_\infty\,.$$

We postpone the proof of the lemma to Appendix B.

3.2.3. *Decorated homomorphisms.* We will also need to consider homomorphism densities of graphs in m-tuples of graphons which can be viewed as a density of edge-coloured subgraphs, where each graphon describes a colour class.

Definition 3.6 (Densities in vectors of kernels)**.** *Given a graph $F = (V, E)$ and a tuple $\mathbf{W} = (W_e : e \in E)$ of kernels, we set*

$$t(F, \mathbf{W}) \coloneqq \int \prod_{ij \in E} W_{ij}(x_i, x_j) \, \mathrm{d}\pi^V \, .$$

We can define similarly the induced density. This time, suppose that we have a graph F, a complete graph K on $V(F)$, and a tuple $\mathbf{W} = (W_e : e \in E(K))$ of kernels. Set

$$t_{\mathrm{ind}}(F, \mathbf{W}) \coloneqq \int \prod_{ij \in E(K) \setminus E(F)} (1 - W_{ij}(x_i, x_j)) \prod_{ij \in E(F)} W_{ij}(x_i, x_j) \, \mathrm{d}\pi^{V(F)} \, .$$

We will make use of a counting lemma similar to [16, Lemma 10.24]. It is convenient to extend this to kernels. A proof can be found in Appendix B.

Lemma 3.7. *Given $\varepsilon \geq 0$, a simple graph $F = (V, E)$, and two tuples $\mathbf{U} = (U_e \in \mathcal{W}_0^{\varepsilon} : e \in E)$ and $\mathbf{W} = (W_e \in \mathcal{W}_0^{\varepsilon} : e \in E)$ we have*

$$|t(F, \mathbf{U}) - t(F, \mathbf{W})| \leq (1 + 2\varepsilon)^{|E|-1} \sum_{e \in E} \|U_e - W_e\|_{\square} \, .$$

3.2.4. *Step functions and twins.* We say that $f \in L^{\infty}(\Omega^2)$ is a *step function* if there is a partition (finite or countable) $S_1 \cup S_2 \cup \ldots$ of Ω into sets (*steps*) of positive measure so that f is constant almost everywhere on each set $S_i \times S_j$. We say that a partition into steps is *minimal* for f if f is not a step function with respect to any coarser partition of Ω. A *step graphon* is a graphon which is a step function.

We say that $A \subset \Omega$ is a *twin-set* of a kernel W if there exists a conull set $A' \subset A$ such that for all $x, x' \in A'$, the (single variable) functions $W(x, \cdot)$ and $W(x', \cdot)$ are equal a.e. A variant of this concept is studied in detail in [16, Section 13.1.1]. Later we shall need the following easy result, a proof of which can be found in Appendix B.

Lemma 3.8. *Suppose that $U, U_1, U_2, \ldots$ are kernels and $\lim_{n \to \infty} \|U_n - U\|_{\square} = 0$. If a set $A \subset \Omega$ has positive measure and is a twin-set for each U_n, then A is a twin-set for U.*

3.2.5. *Sampling from a graphon.* Each graphon W defines a probability distribution on $\mathcal{H}_k$ as follows. Sample k independent π-random elements $\mathbf{U}_v, v \in [k]$ of Ω and then, for each pair of distinct vertices $u, v \in [k]$, make uv an edge independently with probability $W(\mathbf{U}_u, \mathbf{U}_v)$. For a random graph G obtained this way we write $G \sim \mathbb{G}(k, W)$ or even use $\mathbb{G}(k, W)$ to denote the random graph itself.

Sometimes it will be convenient to write the rooted density defined in (3.6) in a probabilistic form as

$$t^{\mathbf{x}}_{\mathrm{ind}}(F,W) = \mathbb{P}\left(\mathbb{G}(k,W) = F \mid \mathbf{U}_v = x_v : v \in R\right). \tag{3.11}$$

3.3. **Differential equations.** To develop our framework we will need some theory of differential equations, both for real-valued functions and for Banach-space-valued functions. This is because our trajectories on graphons will be implicitly defined as solutions to differential equations with the derivative which is a function of the flip rule and current position/graphon.

3.3.1. *Grönwall's inequality.* Among real-valued functions, the differential equation

$$f'(t) = Cf(t) \tag{3.12}$$

with the initial condition $f(0) = z$ has a unique solution $f(t) = ze^{Ct}$. We will often be able to assume only a differential *inequality*, which implies that a function satisfying it is bounded by the solution of a corresponding equation. A statement of this kind are known as Grönwall's inequality. There are many versions of it with different assumptions on signs of constants and continuity/differentiability properties of the function in question. We shall need two versions; the first is borrowed from the literature and the second one (which we could not find a reference for) is derived using standard tools.

Lemma 3.9 (Integral Grönwall's inequality, [23, Theorem 1.10])**.** *Let $f : [t_0, t_1] \to [0, \infty)$ be a continuous non-negative function. Suppose that for some constants $A \geq 0$, $C > 0$ we have*

$$f(t) \leq A + C\int_{t_0}^{t} f(x)\,\mathrm{d}x \tag{3.13}$$

for every $t \in [t_0, t_1]$. Then we have

$$f(t) \leq Ae^{C(t-t_0)} \quad \text{for all } t \in [t_0, t_1]. \tag{3.14}$$

Lemma 3.10 (Differential Grönwall's inequality)**.** *Let $F : [t_0, t_1] \to \mathbb{R}$ be an absolutely continuous function, that is, there exists an integrable function $f : [t_0, t_1] \to \mathbb{R}$ such that*

$$F(t) = F(t_0) + \int_{t_0}^{t} f(x)\,\mathrm{d}x$$

for all $t \in [t_0, t_1]$. Suppose that for some constant $C \in \mathbb{R}$

$$f(t) \leq CF(t) \tag{3.15}$$

for almost every $t \in [t_0, t_1]$. Then we have

$$F(t) \leq F(t_0)e^{C(t-t_0)} \quad \text{for all } t \in [t_0, t_1]. \tag{3.16}$$

Moreover if (3.15) *holds with $\geq$ instead of $\leq$, then* (3.16) *holds with $\geq$ instead of $\leq$.*

Proof. We revisit the proof of [23, Theorem 1.12]. Define a function $G : [t_0, t_1] \to \mathbb{R}$ by

$$G(t) = F(t)e^{-C(t-t_0)} .$$

Since $G(t)$ is a product of two absolutely continuous functions, it is absolutely continuous (see for example [21, §6.4, Exercise 42]). Hence, we can use the product rule (see for example [21, §6.5, Exercise 52]),

$$G(t) = G(t_0) + \int_{t_0}^{t} \left(f(x)e^{-C(x-t_0)} + F(x) \left(-Ce^{-C(x-t_0)} \right) \right) \mathrm{d}x.$$

By (3.15) the integrand is non-positive almost everywhere, whence $G(t) \leq G(t_0)$, which is equivalent to (3.16). If we assume the opposite inequality in (3.15), then we get the opposite inequality in (3.16). □

3.3.2. *Banach-space-valued differential equations.* A key step in our construction involves a solution to a first-order autonomous differential equation. This solution is however not real-valued but rather function-valued. This falls within the more general setting of differential equations in Banach spaces. Recall Definition C.1 of the *derivative*.

The following theorem is a basic statement about local existence and uniqueness of solutions to ordinary differential equations in Banach spaces. For a reference, see, e.g., [1], Lemma 4.1.6 (existence and uniqueness) and Lemma 4.1.8 (Lipschitz continuity on the initial condition[d]).

Theorem 3.11. *Consider a Banach space $\mathcal{E}$ with norm $\|\cdot\|$. Let $\mathcal{U} \subset \mathcal{E}$ be an open set, and $K, M > 0$ be constants. Let $X : \mathcal{U} \to \mathcal{E}$ be a K-Lipschitz function, that is*

$$\|X(x) - X(y)\| \leq K \|x - y\| \text{ for all } x, y \in \mathcal{U}.$$

Suppose that

$$\sup\{\|X(x)\| : x \in \mathcal{U}\} \leq M. \tag{3.17}$$

Let $x_0 \in \mathcal{U}$. If b is a positive number such that the closed ball $\mathcal{B}(x_0, b) := \{x \in \mathcal{E} : \|x - x_0\| \leq b\}$ is contained in $\mathcal{U}$, then for $\alpha := \min\{1/K, b/M\}$ there is a unique continuously differentiable function $f = f_{x_0} : [-\alpha, \alpha] \to \mathcal{U}$ satisfying the equation

$$f'(t) = X(f(t))$$

and the initial value condition $f(0) = x_0$.

Furthermore, if $x_0, y_0 \in \mathcal{U}$ are such that $\mathcal{B}(x_0, b), \mathcal{B}(y_0, b)$ are contained in $\mathcal{U}$ and f_{x_0} and f_{y_0} correspond, respectively, to the initial values x_0 and y_0, then for any $t \in [-\alpha, \alpha]$ we have

$$\|f_{x_0}(t) - f_{y_0}(t)\| \leq e^{K|t|} \|x_0 - y_0\| . \tag{3.18}$$

[d] In [1] the claim is actually weaker, but by examining the proof one easily shows (3.18) .

4. Trajectories

4.1. **Heuristics for the trajectories.** Let us explain the main idea behind constructing trajectories $\Phi : \mathcal{W}_0 \times [0, +\infty) \to \mathcal{W}_0$ for a given flip process. As we explained in Section 1.4, the main motivation for trajectories is that we want to have a Transference Theorem, which says that if G_0 is an n-vertex graph which is close in cut norm to a graphon W_0, then performing tn^2 steps of the flip process we get a graph G_{tn^2}, which is, with high probability, close to $\Phi^t(W_0)$. The Markov property of the flip process can be informally rephrased as "my nearest future depends only on where I am now". Vaguely transferring this to the graphon realm, "the nearest future" corresponds to the time derivative $\frac{\mathrm{d}}{\mathrm{d}t}\Phi^t(W_0)$, and "depends only on where I am now" corresponds to the equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi^t(W_0) = \mathfrak{V}\left(\Phi^t(W_0)\right),$$

where $\mathfrak{V} : \mathcal{W} \to \mathcal{W}$ is some operator mapping kernels (and thus graphons) to kernels. In this section we will intuitively describe this (so far) hypothetical operator in the case of triangle removal flip process.

To interpret the time derivative, we want to see how a typical graph $G_{\delta n^2}$ evolved in δn^2 steps from an n-vertex graph G_0 which is close to U. Consider any points $x, y \in \Omega$. The key idea behind the cut norm is that the points x and y correspond to certain sets $X \subset V(G_0)$ and $Y \subset V(G_0)$, where $|X| = |Y| = \gamma n$ (and γ should be thought of as infinitesimally small[e]), and that

$$U(x, y) \approx \frac{e_{G_0}(X, Y)}{\gamma^2 n^2} \,. \tag{4.1}$$

Since in the triangle removal process edges are only removed, $U - \Phi^\delta U$ at (x, y) corresponds to the number of edges removed from $G_0[X, Y]$ during the first δn^2 steps. How can we remove edges in one step, say the first one? Firstly, we need one of the three sampled vertices to go in X and one in Y, which happens with probability $\approx 6\gamma^2$ (the approximation justified by the smallness of γ). Conditioning on this, we need all three vertices to form a triangle, in which case we remove *one* edge from $G_0[X, Y]$. This conditional event happens with a probability approximately reflected in the 2-rooted triangle density $t^{(x,y)}(\triangle^{1,2}, U)$ (cf. (3.11)). Hence,

$$\mathbb{E}\left[e_{G_0}(X, Y) - e_{G_1}(X, Y)\right] \approx 6\gamma^2 t^{(x,y)}(\triangle^{1,2}, U) \,.$$

Since the overall changes of the host graph are fairly minor for the first δn^2 steps, the calculations are almost unchanged for $e_{G_1}(X, Y) - e_{G_2}(X, Y), \ldots, e_{G_{\delta n^2}}(X, Y) - e_{G_{\delta n^2 - 1}}(X, Y)$.

[e]Instead of points x, y one can also approximate U with a step graphon of step size γ and take sets I_x, I_y containing x and y of measure γ.

Hence,

$$\mathbb{E}[e_{G_0}(X,Y) - e_{G_{\delta n^2}}(X,Y)] \approx 6\gamma^2 t^{(x,y)}(\triangle^{1,2}, U) \cdot \delta n^2.$$

Recalling the scaling (4.1), we therefore have $(U - \Phi^\delta U)(x,y) \approx 6\gamma^2 t^{(x,y)}(\triangle^{1,2}, U) \cdot \delta$, and hence we expect that the time derivative $\lim_{\delta\to 0}[\Phi^\delta U(x,y) - U(x,y)]/\delta$ equals

$$\mathfrak{V}\, U(x,y) := -6t^{(x,y)}(\triangle^{1,2}, U)\ . \tag{4.2}$$

To generalise to more complicated flip processes, we just need to consider edge erasures and edge additions coming from different subgraph replacements of the host graph.

4.2. **The velocity operator and definition of trajectories.** In this section, we first define the velocity operator $\mathfrak{V}_\mathcal{R} : \mathcal{W} \to \mathcal{W}$ for a flip process with a rule $\mathcal{R}$ of order k. As we saw in the motivating Section 4.1, 2-rooted densities of graphs in $\mathcal{H}_k$ will be key parameters in that definition. Having defined the velocity, we then in Theorem 4.5 construct trajectories $\Phi : \mathcal{W}_0 \times [0,+\infty) \to \mathcal{W}_0$ using Theorem 3.11. In fact it will be more convenient to define trajectories starting at arbitrary kernels, justifying the domain of the operator $\mathfrak{V}_\mathcal{R}$.

Definition 4.1. *Recall the operator $T_F^{a,b} : \mathcal{W} \to L^\infty(\Omega^2)$, defined in* (3.7)*, which, when applied to a graphon, gives the rooted induced density. Given a rule $\mathcal{R}$ of order k, we define an operator $\mathfrak{V}_\mathcal{R} : \mathcal{W} \to \mathcal{W}$ by*

$$\mathfrak{V}_\mathcal{R} W := \sum_{F,H\in\mathcal{H}_k} \mathcal{R}_{F,H} \sum_{1\le a\ne b\le k} T_F^{a,b}W \cdot (\mathbb{I}_{\{ab\in H\setminus F\}} - \mathbb{I}_{\{ab\in F\setminus H\}})\ , \tag{4.3}$$

where ab is a shortening of $\{a,b\}$. We call $\mathfrak{V}_\mathcal{R}W$ the velocity at W for rule $\mathcal{R}$*. We will often omit the rule that is clear from the context, denoting $\mathfrak{V} = \mathfrak{V}_\mathcal{R}$.*

Note that $T_F^{a,b}W$ is not necessarily symmetric in (x,y), but $T_F^{a,b}W$ and $T_F^{b,a}W$ appear with the same coefficient in $\mathfrak{V}_\mathcal{R}W$, making it a symmetric function.

With this definition, we may formally calculate the velocity at W for the rule $\mathcal{R}^{\mathrm{TR}}$ which defines the triangle removal flip process (see Figure 1.1, right). We have $k = 3$, $\mathcal{R}^{\mathrm{TR}}_{\triangle,\bullet\bullet\bullet} = 1$ and $\mathcal{R}^{\mathrm{TR}}_{F,F} = 1$ for all $F \in \mathcal{H}_3\setminus\{\triangle\}$. This gives

$$\mathfrak{V}_{\mathcal{R}^{\mathrm{TR}}}W = -\sum_{1\le a\ne b\le 3} T^{a,b}_{\triangle} W = -6T^{1,2}_{\triangle} W\ ,$$

which matches our heuristic in (4.2).

Calculation of the velocity will often be simplified by the following reformulation using notation from Section 3.2.5.

Lemma 4.2. *Suppose that a graphon W and a rule $\mathcal{R}$ of order k are given. Let $(\mathbf{F}, \mathbf{H})$ be a random pair of graphs in $\mathcal{H}_k$ such that $\mathbf{F} \sim \mathbb{G}(k, W)$ and $\mathbb{P}(\mathbf{H} = H \mid \mathbf{F} = F) = \mathcal{R}_{F,H}$ for every $F, H \in \mathcal{H}_k$. We have*

$$(\mathfrak{V}_{\mathcal{R}} W)(x,y) = \sum_{a \neq b} \left[\mathbb{P}(ab \in \mathbf{H} | \mathbf{U}_a = x, \mathbf{U}_b = y) - W(x,y)\right] . \tag{4.4}$$

Proof. By (3.11),

$$\begin{aligned}
\mathfrak{V}_{\mathcal{R}} W(x,y) &= \sum_{a \neq b} \left[\mathbb{P}\left(ab \in \mathbf{H} \setminus \mathbf{F} \,|\, \mathbf{U}_a = x, \mathbf{U}_b = y\right) - \mathbb{P}\left(ab \in \mathbf{F} \setminus \mathbf{H} \,|\, \mathbf{U}_a = x, \mathbf{U}_b = y\right)\right] \\
&= \sum_{a \neq b} \left[\mathbb{P}\left(ab \in \mathbf{H} \,|\, \mathbf{U}_a = x, \mathbf{U}_b = y\right) - \mathbb{P}\left(ab \in \mathbf{F} \,|\, \mathbf{U}_a = x, \mathbf{U}_b = y\right)\right] \\
&= \sum_{a \neq b} \left[\mathbb{P}\left(ab \in \mathbf{H} \,|\, \mathbf{U}_a = x, \mathbf{U}_b = y\right) - W(x,y)\right],
\end{aligned}$$

where the second equality follows by adding and subtracting, in each term, the number $\mathbb{P}(ab \in \mathbf{F} \cap \mathbf{H} \,|\, \mathbf{U}_a = x, \mathbf{U}_b = y)$. □

Let us explain how Definition 4.1 relates to a flip process with rule $\mathcal{R}$ of order k. Suppose that $(\mathbf{G}_i)_{i \geq 0}$ is the flip process on n vertices. Fix distinct $u, v \in [n]$. The probability that the ath sampled vertex is u and the bth is v is $1/(n)_2$, whence

$$\begin{aligned}
&\mathbb{E}\left(\mathbb{I}_{\{uv \in \mathbf{G}_{i+1}\}} - \mathbb{I}_{\{uv \in \mathbf{G}_i\}} \,|\, \mathbf{G}_i\right) \\
=&\tfrac{1}{(n)_2} \sum_{1 \leq a \neq b \leq k} \sum_{F,H \in \mathcal{H}_k} \mathcal{R}_{F,H} \cdot t_{\mathrm{ind}}^{(u,v)}(F^{a,b}, \mathbf{G}_i)(\mathbb{I}_{\{ab \in H \setminus F\}} - \mathbb{I}_{\{ab \in F \setminus H\}}) .
\end{aligned}$$

Furthermore let W_i be the graphon representation of $\mathbf{G}_i$ with respect to the same partition. Since, by (3.10),

$$|t_{\mathrm{ind}}^{(x,y)}(F^{a,b}, \mathbf{G}_i) - t_{\mathrm{ind}}^{(x,y)}(F^{a,b}, W_i)| \leq \binom{k-2}{2} \frac{1}{n},$$

we get the following formula linking the flip process with the velocity operator.

Remark 4.3. *Consider the flip process $(\mathbf{G}_i)_{i \geq 0}$ with rule $\mathcal{R}$ of order k. If W_i is the graphon representation of $\mathbf{G}_i$, then for a.e. $x, y \in \Omega$,*

$$\mathbb{E}[W_{i+1}(x,y)] = W_i(x,y) + \frac{1}{(n)_2} \mathfrak{V}_{\mathcal{R}} W_i(x,y)(1 + O(1/n)) .$$

We can now define the key object of our paper.

Definition 4.4. *Given a rule $\mathcal{R}$ and a kernel $W \in \mathcal{W}$, a* trajectory starting at W *is a differentiable function $\Phi^{\cdot} W : I \to (\mathcal{W}, \|\cdot\|_\infty)$ defined on an open interval $I \subseteq \mathbb{R}$ containing zero, that satisfies the autonomous differential equation*

$$\frac{\mathrm{d}}{\mathrm{d}t} \Phi^t W = \mathfrak{V}\, \Phi^t W \tag{4.5}$$

with the initial condition

$$\Phi^0 W = W\ . \tag{4.6}$$

The ultimate goal of this section is the following theorem (proved in Section 4.4), which asserts that trajectories exist, are unique, and whenever they start at a graphon (rather than at a general kernel), they stay among graphons for every positive time.

Theorem 4.5. *Suppose that $\mathcal{R}$ is a rule.*

(i) *For each kernel $W \in \mathcal{W}$, there is an open interval $\mathcal{D}_W \subseteq \mathbb{R}$ containing 0 and a trajectory $\Phi^\cdot W : \mathcal{D}_W \to \mathcal{W}$ starting at W such that any other trajectory starting at W is a restriction of $\Phi^\cdot W$ to a subinterval of $\mathcal{D}_W$.*

(ii) *For any $u \in \mathcal{D}_W$ we have $\mathcal{D}_{\Phi^u W} = \{t \in \mathbb{R} : t + u \in \mathcal{D}_W\}$ and for every $t \in \mathcal{D}_{\Phi^u W}$ we have*

$$\Phi^t \Phi^u W = \Phi^{t+u} W. \tag{4.7}$$

(iii) *Whenever $W \in \mathcal{W}_0$, the set $I_W := \{t \in \mathcal{D}_W : \Phi^t W \in \mathcal{W}_0\}$ is a closed interval containing $[0, \infty)$.*

Definition 4.6. *For $W \in \mathcal{W}_0$ we write* $\mathsf{age}(W) := -\inf I_W \in [0, \infty]$.

We will sometimes (as in the following remark) abuse the notation by saying *trajectory*, when we actually mean the orbit (image) of a trajectory, that is the set $\{\Phi^t W : t \in I\}$.

Remark 4.7. *Note that part* (ii) *also implies that trajectories from two different kernels W, U are disjoint, unless $W = \Phi^t U$ for some t. In other words, trajectories partition $\mathcal{W}$.*

4.3. **Bounds on the velocity.** The first claim of the next lemma, (4.9), gives bounds on the velocity in terms of the corresponding graphon. Roughly speaking, it says that the $\mathfrak{V}_{\mathcal{R}} W$ cannot be too positive for those (x, y) for which $W(x, y)$ is large, and cannot be too negative for those (x, y) for which $W(x, y)$ is small. Combinatorially, this is quite intuitive: if, for example, $W(x, y)$ is close to 0, then this corresponds to a graph with a sparse bipartite spot (X, Y), where the vertex sets X and Y correspond to the points x and y. The lower bound in (4.9) corresponds to the fact that deletions of (X, Y)-edges are indeed rare, since they may only happen when we actually sample an (X, Y)-edge.

The following lemma also claims two more facts about the velocity operator on the Banach space $(\mathcal{W}, \|\cdot\|_\infty)$ that are crucial in application of Theorem 3.11: (4.10) says it is locally Lipschitz, while (4.11) gives a bound on the norm on a ball around the constant-1/2 graphon.

Lemma 4.8. *Suppose that $\mathcal{R}$ is a rule of order k. Let $\varepsilon \geq 0$ be a constant. Define*

$$C_k := (k)_2^2 2^{\binom{k}{2}-1}, \quad C_\infty(k, \varepsilon) := C_k (1 + \varepsilon)^{\binom{k}{2}-1}. \tag{4.8}$$

For every graphon $W \in \mathcal{W}_0$ we have

$$-(k)_2 \cdot W \le \mathfrak{V}_{\mathcal{R}} W \le (k)_2 \cdot (1 - W)\,. \tag{4.9}$$

For every $U, W \in \mathcal{W}_0^\varepsilon$ we have

$$\|\mathfrak{V}_{\mathcal{R}} U - \mathfrak{V}_{\mathcal{R}} W\|_\infty \le C_\infty(k,\varepsilon)\,\|U - W\|_\infty\,. \tag{4.10}$$

For every $W \in \mathcal{W}_0^\varepsilon$ we have

$$\|\mathfrak{V}_{\mathcal{R}} W\|_\infty \le (k)_2 + C_\infty(k,\varepsilon)\varepsilon. \tag{4.11}$$

Proof. Using the probabilistic form of the velocity (4.4), and noting that the conditional probability takes values in $[0,1]$, inequalities (4.9) follow.

To obtain (4.10), using Definition 4.1 we have

$$\begin{aligned}
\|\mathfrak{V}_{\mathcal{R}} U - \mathfrak{V}_{\mathcal{R}} W\|_\infty &= \left\| \sum_{a\neq b} \sum_{F\in\mathcal{H}_k} \left(T_F^{a,b} U - T_F^{a,b} W\right) \sum_{H\in\mathcal{H}_k} \mathcal{R}_{F,H} \left(\mathbb{I}_{\{ab\in H\setminus F\}} - \mathbb{I}_{\{ab\in F\setminus H\}}\right) \right\|_\infty \\
&\le \sum_{a\neq b} \sum_{F\in\mathcal{H}_k} \left\| T_F^{a,b} U - T_F^{a,b} W \right\|_\infty \sum_{H\in\mathcal{H}_k} \mathcal{R}_{F,H} \\
\boxed{\text{by (1.4)}} \quad &= \sum_{a\neq b} \sum_{F\in\mathcal{H}_k} \left\| T_F^{a,b} U - T_F^{a,b} W \right\|_\infty \\
\boxed{\text{Lemma 3.5}} \quad &\le (k)_2 |\mathcal{H}_k| \cdot (1+\varepsilon)^{\binom{k}{2}-1} \binom{k}{2} \|U - W\|_\infty = C_\infty(k,\varepsilon)\,\|U - W\|_\infty\,.
\end{aligned}$$

Finally, to see (4.11), given W, pick the nearest $U \in \mathcal{W}_0$ (so that $\|W - U\|_\infty \le \varepsilon$). Using the triangle inequality, (4.9) and (4.10), we infer

$$\|\mathfrak{V}\, W\|_\infty \le \|\mathfrak{V}\, U\|_\infty + C_\infty(k,\varepsilon)\,\|W - U\|_\infty \le (k)_2 + C_\infty(k,\varepsilon)\varepsilon.$$

□

The continuity property (4.10) allows us to rewrite the differential equation (4.5) as an integral equation. This is formulated in the remark below.

Remark 4.9. *Since $\mathfrak{V}$ is continuous on $(\mathcal{W}, \|\cdot\|_\infty)$ by (4.10), Theorem C.2 tells us that the trajectory $\Phi^\cdot W : I \to \mathcal{W}$ satisfies*

$$\Phi^t W = W + \int_0^t \mathfrak{V}_{\mathcal{R}} \Phi^\tau W \,\mathrm{d}\tau, \quad t \in I,$$

where the integral is defined with respect to the norm $\|\cdot\|_\infty$.

We will also need that the velocity operator is locally Lipschitz with respect to the cut norm.

Lemma 4.10. *For any rule of order k and every $\varepsilon \geq 0$ and $U, W \in \mathcal{W}_0^\varepsilon$ we have*

$$\|\mathfrak{V}_{\mathcal{R}} U - \mathfrak{V}_{\mathcal{R}} W\|_\square \leq C_\square \|U - W\|_\square ,$$

where $C_\square = C_\square(k, \varepsilon) = (k)_2^2 \, (2 + 4\varepsilon)^{\binom{k}{2}}$.

Proof. First, we want to show that for every $a, b \in [k]$, $a \neq b$, and $G \in \mathcal{H}_k$ the following holds.

Claim 4.11. *If $U, W \in \mathcal{W}_0^\varepsilon$, then*

$$\left\| T_G^{a,b} U - T_G^{a,b} W \right\|_\square \leq (k)_2 (1 + 2\varepsilon)^{\binom{k}{2}} \|U - W\|_\square .$$

Proof of Claim 4.11. Fix an arbitrary set $A \subseteq \Omega$ and let $\mathbb{I}_{A\times A} : \Omega^2 \to \{0, 1\}$ be the indicator function of the set $A \times A$. Given W, G, a and b, we define a tuple $\mathbf{W} = (W_e : e \in E(K_k))$ by

$$W_e = \begin{cases} W, & \text{if } e \neq ab, e \in E(G) \\ 1 - W, & \text{if } e \neq ab, e \notin E(G) \\ W \cdot \mathbb{I}_{A\times A}, & \text{if } e = ab \in E(G) \\ (1 - W) \cdot \mathbb{I}_{A\times A}, & \text{if } e = ab \notin E(G) . \end{cases}$$

and similarly $\mathbf{U} = (U_e : e \in E(K_k))$, with W replaced by U. We claim that

$$\|U_e - W_e\|_\square \leq \|U - W\|_\square \tag{4.12}$$

for every $e \in K_k$. When $e \neq ab$, regardless whether $e \in E(G)$ or not, we have $U_e - W_e = U - W$ and therefore $\|U_e - W_e\|_\square = \|U - W\|_\square$.

When $e = ab$, we have $U_e - W_e = (U - W)\mathbb{I}_{A\times A}$ and therefore

$$\begin{aligned} \|U_e - W_e\|_\square = \|(U - W)\mathbb{I}_{A\times A})\|_\square &= \sup_{S,T\subseteq\Omega} \left| \int_{S\times T} (U - W) \cdot \mathbb{I}_{A\times A} \, \mathrm{d}\pi^2 \right| \\ &= \sup_{S,T\subseteq\Omega} \left| \int_{(S\cap A)\times(T\cap A)} (U - W) \, \mathrm{d}\pi^2 \right| \leq \|U - W\|_\square . \end{aligned}$$

Note that

$$\begin{aligned} \int_{A\times A} T_G^{a,b} U \, \mathrm{d}\pi^2 &= \int \mathbb{I}_{A\times A}(x_a, x_b) \int \prod_{ij\in E(G)} U(x_i, x_j) \prod_{ij\notin E(G)} (1 - U(x_i, x_j)) \, \mathrm{d}\pi^{[k]\setminus\{a,b\}} \, \mathrm{d}\pi^{\{a,b\}} \\ &= t(K_k, \mathbf{U}). \end{aligned}$$

And similarly $\int_{A\times A} T_G^{a,b}W \,\mathrm{d}\pi^2 = t(K_k, \mathbf{W})$. Hence

$$\begin{aligned}\left|\int_{A\times A}(T_G^{a,b}U - T_G^{a,b}W)\,\mathrm{d}\pi^2\right| &= |t(K_k,\mathbf{U}) - t(K_k,\mathbf{W})| \\ \boxed{\text{Lemma 3.7}} \quad &\le (1+2\varepsilon)^{\binom{k}{2}} \sum_{e\in K_k} \|U_e - W_e\|_\square \\ \boxed{\text{by (4.12)}} \quad &\le (1+2\varepsilon)^{\binom{k}{2}} \binom{k}{2} \|U-W\|_\square \ .\end{aligned}$$

Since this was verified for each set A, the result follows from Fact B.1. □

Now, using the triangle inequality and Claim 4.11, we get

$$\begin{aligned}\|\mathfrak{V}_\mathcal{R}U - \mathfrak{V}_\mathcal{R}W\|_\square &\le \sum_{1\le a\ne b\le k}\ \sum_{(G,F)\in\mathcal{H}_k\times\mathcal{H}_k} \mathcal{R}_{G,F}\cdot\left\|T_G^{a,b}U - T_G^{a,b}W\right\|_\square \\ \boxed{\text{by (1.4)}} \quad &\le \sum_{1\le a\ne b\le k}\ \sum_{G\in\mathcal{H}_k} \left\|T_G^{a,b}U - T_G^{a,b}W\right\|_\square \\ &\le (k)_2 2^{\binom{k}{2}}\cdot(k)_2(1+2\varepsilon)^{\binom{k}{2}}\|U-W\|_\square < C_\square\|U-W\|_\square \ ,\end{aligned}$$

which proves the lemma. □

4.4. **Proof of Theorem 4.5.** Fix a rule $\mathcal{R}$ of order k and further write $\mathfrak{V} = \mathfrak{V}_\mathcal{R}$.

We start with existence of a time-local solution for every $W \in \mathcal{W}$.

Claim 4.12. *For every $\rho \ge 0$ there is $T_\rho > 0$, so that for every $W \in \mathcal{W}_0^\rho$ there is a unique trajectory $\Phi^\cdot W : J_\rho = (-T_\rho, T_\rho) \to \mathcal{W}_0^{1+\rho}$. Moreover, for every $t \in J_\rho$ the function $W \mapsto \Phi^t W$ is continuous on $(\mathcal{W}_0^\rho, \|\cdot\|_\infty)$.*

Proof. We apply Theorem 3.11 for the Banach space $\mathcal{E} = (\mathcal{W}, \|\cdot\|_\infty)$ with function $X = \mathfrak{V}$. Using the notation of Theorem 3.11, we set $\mathcal{U}$ to be the interior of $\mathcal{W}_0^{1+\rho}$ (equivalently, an open ball around constant-$\frac{1}{2}$ of radius $\frac{3}{2}+\rho$), that is,

$$\mathcal{U} = \bigcup_{\beta>0} \mathcal{W}_0^{1+\rho-\beta} \ .$$

Let function C_∞ be as in (4.8). By (4.10), operator $\mathfrak{V}$ is K-Lipschitz on $\mathcal{U}$ with $K = C_\infty(k, 1+\rho)$. Setting $b := 0.9$, say, note that for every $x_0 = W \in \mathcal{W}_0^\rho$, we have that the closed ball $\mathcal{B}(x_0, b)$ lies in $\mathcal{U}$. By (4.11), condition (3.17) is satisfied by $M := (k)_2 + (1+\rho)C_\infty(k, 1+\rho)$. Theorem 3.11 implies the claim with $T_\rho = \alpha = \min\{b/M, 1/K\}$, with the continuity of $W \mapsto \Phi^t W$ following from (3.18). □

For a given rule, consider all trajectories starting at W that are defined on an open interval containing 0 and let $\mathcal{D}_W$ be the union of their domains. Using uniqueness given by Claim 4.12 it is easy to see (cf. [24, Section 2.6]) that any two such trajectories

agree on the intersection of their domains and therefore determine a unique trajectory $\Phi^{\cdot}W : \mathcal{D}_W \to \mathcal{W}$ with a maximal domain. This proves (i).

Uniqueness also implies (ii), by noting that $f_1(t) = \Phi^{t+u}W$ and $f_2(t) = \Phi^t\Phi^u W$ are two solutions for the same initial condition $f(0) = \Phi^u W$.

It remains to prove (iii). Despite being "heuristically obvious", its analytic proof is somewhat tedious. The first step is to show this for graphons that lie in the interior of $\mathcal{W}_0$ with respect to $\|\cdot\|_\infty$.

Claim 4.13. *For every $\varepsilon > 0$ and every $W \in \mathcal{W}_0$ such that $\varepsilon \le W \le 1-\varepsilon$ and every $t \in \mathcal{D}_W \cap [0,\infty)$ we have $\Phi^t W \in \mathcal{W}_0$.*

Proof. Let $T := \sup\{t \in \mathcal{D}_W : \Phi^u W \in \mathcal{W}_0 \forall u \in [0,t]\}$ and suppose for contradiction that $T < \infty$.

The continuity[f] of the function $t \mapsto \Phi^t W$ implies the following three facts: $T > 0$ (since W belongs to the interior of $\mathcal{W}_0$ with respect to $\|\cdot\|_\infty$); $\Phi^T W \in \mathcal{W}_0$ (since $\mathcal{W}_0$ is a closed set); by maximality of T, $\Phi^T W$ lies on the boundary of $\mathcal{W}_0$, that is,

$$\operatorname{ess\,inf} \Phi^T W = 0 \quad \text{or} \quad \operatorname{ess\,sup} \Phi^T W = 1\,. \tag{4.13}$$

We will obtain a contradiction by showing that $\Phi^T W$ lies in the interior of $\mathcal{W}_0$.

Define continuous kernel-valued functions $f_\cdot, g_\cdot$ with domain $[0,T]$ by $f_t := \mathfrak{V}\,\Phi^t W$ and $g_t = -(k)_2 \Phi^t W$. By (4.9) we have that $f_t \ge g_t$ for $t \in [0,T]$ and by Remark 4.9 $\Phi^t W = W + \int_0^t f_\tau \,\mathrm{d}\tau$. In view of Proposition C.4, we can assume that for every $(x,y) \in \Omega^2$,

- the function $\tau \mapsto \mathfrak{V}\,\Phi^\tau W(x,y)$ is continuous on $[0,T]$;
- $\mathfrak{V}\,\Phi^\tau W(x,y) \ge -(k)_2 \Phi^\tau W(x,y)$, $\tau \in [0,T]$;
- and
$$\Phi^t W(x,y) = W(x,y) + \int_0^t \mathfrak{V}\,\Phi^\tau W(x,y)\,\mathrm{d}\tau, \quad t \in [0,T].$$

Let $h(t) = \mathfrak{V}\,\Phi^t W(x,y)$ so that $H(t) := \Phi^t W(x,y)$ satisfies $H(t) = H(0) + \int_0^t h(x)\,\mathrm{d}x$ and $h(t) \ge -(k)_2 H(t)$. By Lemma 3.10, $H(T) \ge H(0) e^{-(k)_2 T}$. Since $W \ge \varepsilon$, we conclude that $\Phi^T W \ge \varepsilon e^{-(k)_2 T} =: \varepsilon' > 0$. By an analogous argument applied to graphon $1 - W$, we get that $1 - W \ge \varepsilon'$. This is a contradiction to (4.13). □

We now extend Claim 4.13 to all graphons.

Claim 4.14. *For every $W \in \mathcal{W}_0$ and every $t \in \mathcal{D}_W \cap [0,\infty)$ we have $\Phi^t W \in \mathcal{W}_0$.*

Proof. Let T_0 be given by Claim 4.12 for $\rho = 0$. Observe that it is enough to prove $\Phi^t W \in \mathcal{W}_0$ only for times $t \in \mathcal{D}_W \cap [0, T_0)$. Indeed, with such a weaker statement we can write any time t as $t = t_1 + \ldots + t_\ell$ where each $t_s \in (0, T_0)$ and then inductively for

[f] The continuity with respect to $\|\cdot\|_\infty$ is implied by the fact that Theorem 3.11 outputs a function which is even differentiable with respect to $\|\cdot\|_\infty$.

$s = 1, \dots, \ell$ deduce that $\Phi^{\sum_{i=1}^{s} t_i} W \in \mathcal{W}_0$. The advantage for working on this smaller time interval is that Claim 4.12 tells us that for each $\tau \in (-T_0, T_0)$, the map $U \mapsto \Phi^\tau U$ is $\|\cdot\|_\infty$-continuous on $\mathcal{W}_0^0 = \mathcal{W}_0$.

Let $W \in \mathcal{W}_0$ and $t \in \mathcal{D}_W \cap [0, T_0)$ be arbitrary. Then for every $\varepsilon > 0$ let $U_\varepsilon := \varepsilon + (1 - 2\varepsilon)W$. Noting that $\varepsilon \le U_\varepsilon \le 1 - \varepsilon$ we conclude by Claim 4.13 that $\Phi^t U_\varepsilon \in \mathcal{W}_0$. Since $\|W - U_\varepsilon\|_\infty \to 0$ as $\varepsilon \to 0$, continuity of the map $U \mapsto \Phi^t U$ implies that $\Phi^t W \in \mathcal{W}_0$, as was needed. □

Claim 4.14 and (ii) imply that $I_W \subseteq \mathbb{R}$ is an (open or closed) interval containing $[0, \infty)$. Since $\Phi^\cdot W : \mathcal{D}_W \to (\mathcal{W}, \|\cdot\|_\infty)$ is continuous and $\mathcal{W}_0$ is closed, interval I_W is closed with respect to $\mathcal{D}_W$. Hence, recalling the notation $\mathsf{age}(W) = -\inf I_W$, we have that either $I_W = \mathcal{D}_W$ or $I_W = [-\mathsf{age}(W), \infty)$. It remains to check that in the former case $\mathsf{age}(W) = \infty$, so that $I_W = \mathcal{D}_W = \mathbb{R}$. Suppose for contradiction that we have $\mathsf{age}(W) < \infty$, so that $\mathcal{D}_W = (-\mathsf{age}(W), \infty)$. Let T_0 be given by Claim 4.12 for $\rho = 0$. Let $t = -\mathsf{age}(W) + T_0/4 \in I_W$. Since $\Phi^t W \in \mathcal{W}_0$, we can use Claim 4.12 to extend the trajectory from the graphon $\Phi^t W$ to the interval $(-T_0, T_0)$. In particular, we have extended the trajectory of the graphon W by at least $3T_0/4$ before the $-\mathsf{age}(W)$, contradicting the maximality of $\mathcal{D}_W$.

4.5. **Smoothness of trajectories in time.** The following lemma says that for any graphon the trajectory is Lipschitz in time. Inequality (4.15) will be used in the proof of concentration of flip processes (Theorem 5.1).

Lemma 4.15. *For every rule $\mathcal{R}$, every graphon $W \in \mathcal{W}_0$ and every $t, u \in I_W$,*

$$\left\|\Phi^t W - \Phi^u W\right\|_\infty \le (k)_2 |t - u|, \tag{4.14}$$

and for every $\delta > 0$,

$$\left\| \frac{\Phi^\delta W - W}{\delta} - \mathfrak{V}\, W \right\|_\infty \le C_k (k)_2 \delta, \tag{4.15}$$

where C_k is defined in (4.8).

Proof. By Remark 4.9 we have, for every $t, u \ge -\mathsf{age}(W)$,

$$\Phi^t W - \Phi^u W = \int_u^t \mathfrak{V}\, \Phi^\tau W \, \mathrm{d}\tau \,. \tag{4.16}$$

Combining (C.1) and the inequality $\|\mathfrak{V}\, U\|_\infty \le (k)_2$ for $U \in \mathcal{W}_0$ (see (4.11)), we obtain

$$\left\|\Phi^t W - \Phi^u W\right\|_\infty \le |t - u| \sup_{\tau \in [u,t]} \|\mathfrak{V}\, \Phi^\tau W\|_\infty \le (k)_2 |t - u|,$$

thus proving (4.14).

To prove (4.15), setting $t = \delta, u = 0$ in (4.16) becomes

$$\Phi^\delta W - W = \int_0^\delta \mathfrak{V}\, \Phi^\tau W \, \mathrm{d}\tau \ . \tag{4.17}$$

By (4.14) we have, for every $\tau \geq 0$,

$$\|\Phi^\tau W - W\|_\infty = \left\|\Phi^\tau W - \Phi^0 W\right\|_\infty \leq (k)_2 \tau \ ,$$

and since $\Phi^\tau W \in \mathcal{W}_0$, by (4.10),

$$\|\mathfrak{V}\, \Phi^\tau W - \mathfrak{V}\, W\|_\infty \leq C_\infty(k,0)(k)_2\tau = C_k(k)_2\tau \ .$$

Hence, using (4.17), and elementary properties of the integral (see Section C),

$$\begin{aligned}\left\|\Phi^\delta W - W - \delta\mathfrak{V}\, W\right\|_\infty &= \left\|\int_0^\delta (\mathfrak{V}\, \Phi^\tau W - \mathfrak{V}\, W)\, \mathrm{d}\tau\right\|_\infty \\ &= (\delta - 0) \sup_{\tau\in[0,\delta]} \|\mathfrak{V}\, \Phi^\tau W - \mathfrak{V}\, W\|_\infty \leq C_k(k)_2\delta^2.\end{aligned}$$

Dividing by δ we obtain (4.15). □

4.6. **Dependence on the initial conditions.** The following theorem implies that on any bounded time interval the trajectories depend continuously on the initial conditions. We have already used such a fact in the proof of Claim 4.14, where it relied on the well-known machinery of Banach-space-valued differential equations and on (3.18) in particular. However, there the metric was induced by the L^∞-norm. For many combinatorial problems the cut norm is more relevant, and to deduce Lipschitz continuity we need to be more careful. Even though the velocity is also Lipschitz with respect to the cut norm (see Lemma 4.10), we are not in the setting of Banach spaces.[g]

Theorem 4.16. *Given $k \in \mathbb{N}$, let $C_\square = C_\square(k,0)$ as in Lemma 4.10. Let $\mathcal{R}$ be a rule of order k, and U, W be any graphons. If $t \in I_U \cap I_W$, then*

$$\left\|\Phi^t U - \Phi^t W\right\|_\square \leq e^{C_\square |t|} \, \|U - W\|_\square \ .$$

Proof. For $t \in I_U \cap I_W = [-\min(\mathsf{age}(U), \mathsf{age}(W)), +\infty)$, let $W_t := \Phi^t W$ and $U_t := \Phi^t U$. Let us first consider the case $t \geq 0$ and set $F(t) := \|U_t - W_t\|_\square$. By Remark 4.9, for any s and t such that $-\min(\mathsf{age}(U), \mathsf{age}(W)) \leq s \leq t < \infty$ we have

$$W_t - W_s = \int_s^t \mathfrak{V}\, W_x \, \mathrm{d}x \ , \tag{4.18}$$

where the integral is understood in terms of the Banach space $(\mathcal{W}, \|\cdot\|_\infty)$. Since $x \mapsto \mathfrak{V}\, W_x$ is also continuous with respect to the cut norm, integral can be treated with respect of

[g] That is, we do not know if $(\mathcal{W}, \|\cdot\|_\square)$ is complete. Note that the subset of uniformly bounded functions is complete by Lemma 3.2.

the completion $\mathcal{E}$ of the normed space $(\mathcal{W}, \|\cdot\|_\square)$. Note that $\mathcal{W}_0$ is a closed subset of $\mathcal{E}$ by Lemma 3.2. Hence, using the triangle inequality, property of the integral (C.1), and Lemma 4.10, we obtain, for $t \geq s \geq 0$,

$$\begin{aligned}
&\boxed{\text{triangle inequality}} && F(t) - F(s) \leq \|U_t - U_s - (W_t - W_s)\|_\square \\
&\boxed{\text{by (4.18)}} && \leq \left\| \int_s^t (\mathfrak{V}\, U_x - \mathfrak{V}\, W_x)\, \mathrm{d}x \right\|_\square \\
&\boxed{\text{by (C.1)}} && \leq \int_s^t \|\mathfrak{V}\, U_x - \mathfrak{V}\, W_x\|_\square \, \mathrm{d}x \\
&\boxed{\text{Lemma 4.10}} && \leq C_\square \int_s^t \|U_x - W_x\|_\square = C_\square \int_s^t F(x)\, \mathrm{d}x\,.
\end{aligned}$$

In particular this implies that F is $C_\square$-Lipschitz and so continuous. Moreover, it implies that F satisfies (3.13) with $t_0 = 0$, $A = F(0)$ and $C = C_\square$. Hence Lemma 3.9 implies Theorem for $t \geq 0$.

Now considering $t < 0$, we define

$$F(x) = \|U_{-x} - W_{-x}\|_\square\,, \quad x \in [0, \min\{\mathsf{age}(U), \mathsf{age}(W)\}]$$

and similarly as above obtain that, for $0 \leq s \leq t \leq \min\{\mathsf{age}(U), \mathsf{age}(W)\}$

$$F(t) - F(s) \leq C_\square \int_s^t \|U_{-x} - W_{-x}\|_\square \, \mathrm{d}x = C_\square \int_s^t F(x)\, \mathrm{d}x\,,$$

whence by Lemma 3.9,

$$F(t) \leq F(0) e^{C_\square t}, \quad t \in [0, \min\{\mathsf{age}(U), \mathsf{age}(W)\}].$$

□

5. The Transference Theorem

In this section we prove that for a quadratic number of steps the flip process starting with some large graph G, viewed as a graphon-valued process, is likely to stay close (with respect to the cut norm) to the deterministic trajectory $t \mapsto \Phi^t W_G$.

Let W_G be the graphon representation of a graph G with respect to some partition $(\Omega_v : v \in [n])$ of Ω (see Section 3.2.1).

Theorem 5.1. *For every $k \in \mathbb{N}$ there is a constant $C > 0$ so that the following holds. Given a rule $\mathcal{R}$ of order k and a graph G on the vertex set $[n]$, let $(G_i)_{i \geq 0}$ be the flip process starting with $G_0 = G$.*

For any $T > 0$ and $\varepsilon > 0$ have

$$\max_{i \in (0, Tn^2] \cap \mathbb{Z}} \left\| W_{G_i} - \Phi^{i/n^2} W_G \right\|_\square < \varepsilon \tag{5.1}$$

with probability at least

$$1 - \frac{CTe^{CT}}{\varepsilon} \exp\left((2\ln 2)n - \frac{C\varepsilon^3 n^2}{e^{CT}}\right). \tag{5.2}$$

Remark 5.2. *A typical application of Theorem 5.1 will also invoke Theorem 4.16. That is, instead of relating W_{G_i} to a trajectory started at W_G (at time i/n^2), it seems more useful to relate it to a cleaned version of that trajectory started at some (simple) graphon W which is close to W_G. For example, if G is a quasirandom graph of density $\frac{1}{2}$, then taking $W \equiv \frac{1}{2}$ seems to be a good choice.*

Remark 5.3. *While throughout the paper we are motivated by the idea that T is an arbitrarily large constant (corresponding to 'bounded horizons'), the probability bound* (5.2) *is useful even for $T = O(\log n)$. We are not sure if this bound can be pushed further.*

This exact limit of applicability does not seem that important because in practice we will also be limited by Theorem 4.16 as we explained in Remark 5.2. For example, suppose that we have a rule of order 3 *in which the edgeless graph stays edgeless and other graphs are replaced by a triangle. Let G be a graph on n vertices and $o(n^2)$ edges. This is a sparse graph and so $W \equiv 0$ seems like a good choice of a starting point for the trajectory. Obviously, we have $\Phi^t W \equiv 0$ but since $\|W_G - 0\|_\square \geq 1/n^2$, we cannot conclude anything useful about G_i for $i \gg n^2 \log n$ due to the Lipschitz constant in Theorem 4.16.*

For convenience, we also include an easy corollary of Theorem 5.1 phrased in terms of the cut distance.

Corollary 5.4. *For every $k \in \mathbb{N}$ the following holds with the same constant C as in Theorem 5.1. Given a rule $\mathcal{R}$ of order k and an n-vertex graph G, let $(G_i)_{i\geq 0}$ be the flip process starting with $G_0 = G$. Let $W_0, W_1, \dots$ be arbitrary graphon representations of the graphs $G_0, G_1, \dots$. For any $T > 0$ and $\varepsilon > 0$ have $\max\{\delta_\square(W_i, \Phi^{i/n^2} W_0) : i \in (0, Tn^2] \cap \mathbb{Z}\} < \varepsilon$ with probability at least* (5.2).

5.1. **Proof of Theorem 5.1.** The idea is to subdivide the time domain into intervals of size δn^2 for some small enough δ (depending on k, ε and T) and approximate, in each of the intervals, the expected change of the graphon W_{G_i} linearly in terms of the velocity at the start of the interval. The following lemma quantifies the error of this approximation.

Lemma 5.5. *Let $k, t, n \in \mathbb{N}$ be such that $n \geq 2$ and $\delta := t/n^2 \geq 1/n$. Furthermore, let $\mathcal{R}$ be a rule of order k and G be a graph with vertex set $[n]$. Let $(G_i)_{i\geq 0}$ be the flip process with initial graph $G_0 = G$. With probability at least*

$$1 - 2^{2n+1} \exp\left(-\frac{\delta^3 n^2}{2(k)_2^2}\right) \tag{5.3}$$

we have

$$\|W_{G_t} - (W_G + \delta \mathfrak{V}_{\mathcal{R}} W_G)\|_\square \le C''' \delta^2 \,, \tag{5.4}$$

with a constant

$$C''' := \tbinom{k}{2}\left(2\tbinom{k}{2}|\mathcal{H}_k| + 6 + C_\square\right) + 1, \tag{5.5}$$

where $C_\square = C_\square(k,0) = (k)_2^2 2^{\binom{k}{2}}$*, as in Lemma 4.10.*

Proof. In addition to $C_\square$ and C''' we need to define two further constants,

$$C' := 2\left(\binom{k}{2}^2 |\mathcal{H}_k| + (k)_2\right) \quad \text{and} \quad C'' := C_\square \binom{k}{2}.$$

For simplicity write $W_i = W_{G_i}$. For $S \subseteq [n]$ write $\Omega_S = \cup\{\Omega_u : u \in S\}$. Since all graphons and kernels involved in the proof are constant on each of the sets $\Omega_u \times \Omega_v$, in view of Fact B.2 we will bound the cut norm only on sets of the form $\Omega_S \times \Omega_T$.

We plan to use the Hoeffding–Azuma inequality (see Section A) to show concentration of

$$e_{G_t}(S,T) - e_{G_0}(S,T)$$

for fixed $S, T \subseteq [n]$ and then apply the union bound over the 2^{2n} choices of S, T. For $i \in [t]$, let $\mathbf{v}_i$ be the k-tuple of vertices chosen at step i from graph G_{i-1} and $\mathbf{H}_i \in \mathcal{H}_k$ be the graph by which we replace $G_{i-1}[\mathbf{v}_i]$. We consider a filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_t$, where $\mathcal{F}_i$ is generated by $(\mathbf{v}_j, \mathbf{H}_j)_{1 \le j \le i}$.

For each $i \in [t]$, define a random variable $X_i := e_{G_i}(S,T) - e_{G_{i-1}}(S,T)$. We have $e_{G_t}(S,T) - e_{G_0}(S,T) = \sum_{i=1}^t X_i$. Note that X_i is determined by $(\mathbf{v}_j, \mathbf{H}_j)_{j \in [i]}$, and so is $\mathcal{F}_i$-measurable. Defining $Y_i := X_i - \mathbb{E}\,(X_i \,|\, \mathcal{F}_{i-1})$, we have that $Y_1, \ldots, Y_t$ is a sequence of martingale differences.

We will show that, with probability one,

$$\left|\sum_{i=1}^t \mathbb{E}\,(X_i \,|\, \mathcal{F}_{i-1}) - t \int_{\Omega_S \times \Omega_T} \mathfrak{V}_{\mathcal{R}} W_{G_0}\right| \le (C''' - 1)\delta^2 n^2, \tag{5.6}$$

and

$$\mathbb{P}\left(\left|\sum_{i=1}^t Y_i\right| \ge \delta^2 n^2\right) \le 2\exp\left(-\frac{\delta^4 n^4}{2t(k)_2^2}\right) = 2\exp\left(-\frac{\delta^3 n^2}{2(k)_2^2}\right). \tag{5.7}$$

The triangle inequality and inequalities (5.6) and (5.7) imply that, with probability at least $1 - 2\exp\left(-\delta^3 n^2 / \left(2(k)_2^2\right)\right)$,

$$\left|e_{G_t}(S,T) - e_{G_0}(S,T) - t\int_{\Omega_S \times \Omega_T} \mathfrak{V}_{\mathcal{R}} W_{G_0}\right| \le C''' \delta^2 n^2.$$

Taking the union bound over 2^{2n} choices of S, T, we obtain that

$$\|W_{G_t} - W_{G_0} - \delta \mathfrak{V}_{\mathcal{R}} W_{G_0}\|_{\square} \le C'''\delta^2$$

holds with probability (5.3).

To prove (5.6), we fix $i \in [t]$. Then for any two distinct vertices u and v we define random variables

$$Y_{uv} = Y_{vu} := \mathbb{E}\left(\mathbb{I}_{uv\in G_i} - \mathbb{I}_{uv\in G_{i-1}} \,|\, \mathcal{F}_{i-1}\right) ,$$

and write

$$\mathbb{E}\left(X_i \,|\, \mathcal{F}_{i-1}\right) = \sum_{u\in S, v\in T: u\neq v} Y_{uv} .$$

Note that Y_{uv} is a function of G_{i-1}. Hence we can fix ("condition on") G_{i-1} and calculate the expectation of $\mathbb{I}_{uv\in G_i} - \mathbb{I}_{uv\in G_{i-1}}$ with respect to the random pair $(\mathbf{v}_i, \mathbf{H}_i) \in [n]^k \times \mathcal{H}_k$.

Writing $\mathbf{v}_i = (v_1, \dots, v_k)$, note that events $E_{a,b} := \{v_a = u, v_b = v\}$, $1 \le a < b \le k$, are mutually exclusive and equally likely, each of probability $1/(n)_2$. If we condition on $E_{a,b}$, then for any $F, H \in \mathcal{H}_k$ the probability that the function $j \mapsto v_j, j \in [k]$ defines an isomorphism between F and $G_{i-1}[\mathbf{v}_i]$ is exactly $t_{\mathrm{ind}}^{(u,v)}(F^{a,b}, G_{i-1})$. The probability that then $G_{i-1}[\mathbf{v}_i]$ is replaced by H is $\mathcal{R}_{F,H}$. Hence

$$Y_{uv} = \sum_{1\le a\neq b\le k} \frac{1}{(n)_2} \sum_{F,H\in\mathcal{H}_k} t_{\mathrm{ind}}^{(u,v)}(F^{a,b}, G_{i-1}) \cdot \mathcal{R}_{F,H} \cdot \left(\mathbb{I}_{\{ab\in H\setminus F\}} - \mathbb{I}_{\{ab\in F\setminus H\}}\right) . \tag{5.8}$$

Now we want to relate (5.8) to the function $\mathfrak{V}_{\mathcal{R}} W_{i-1}$, which is constant on the set $\Omega_u \times \Omega_v$. By (3.10), for $x \in \Omega_u, y \in \Omega_v$ we have

$$\left| t_{\mathrm{ind}}^{(u,v)}(F^{a,b}, G_{i-1}) - t_{\mathrm{ind}}^{(x,y)}(F^{a,b}, W_{i-1}) \right| \le \frac{\binom{k}{2}}{n},$$

which allows us to rewrite (5.8) as

$$Y_{uv} = \sum_{1\le a\neq b\le k} \frac{1}{(n)_2} \sum_{F,H\in\mathcal{H}_k} \left(t_{\mathrm{ind}}^{(x,y)}(F^{a,b}, W_{i-1}) \pm \frac{\binom{k}{2}}{n} \right) \cdot \mathcal{R}_{F,H} \cdot \left(\mathbb{I}_{\{ab\in H\setminus F\}} - \mathbb{I}_{\{ab\in F\setminus H\}}\right) .$$

To bound the aggregate contribution of the error term $\pm\frac{\binom{k}{2}}{n}$, we need include a factor of $\binom{k}{2}$ because of the summation over $1 \le a \ne b \le k$, a factor of $|\mathcal{H}_k|$ because of the summation of $H \in \mathcal{H}_k$, and a factor of 1 because of the summation $F \in \mathcal{H}_k$ (here, we make use of the term $\mathcal{R}_{F,H}$ and (1.4)). Hence,

$$Y_{uv} = \sum_{1\le a\neq b\le k} \frac{1}{(n)_2} \sum_{F,H\in\mathcal{H}_k} t_{\mathrm{ind}}^{(x,y)}(F^{a,b}, W_{i-1}) \cdot \mathcal{R}_{F,H} \cdot \left(\mathbb{I}_{\{ab\in H\setminus F\}} - \mathbb{I}_{\{ab\in F\setminus H\}}\right) \pm \frac{\binom{k}{2}^2 |\mathcal{H}_k|}{n} ,$$

and we recover (4.3). Thus, for any $x \in \Omega_u, y \in \Omega_v$,

$$\left|Y_{uv} - \tfrac{1}{(n)_2}\mathfrak{V}_{\mathcal{R}} W_{i-1}(x,y)\right| \le \frac{\binom{k}{2}^2 |\mathcal{H}_k|}{n^2(n-1)} \,. \tag{5.9}$$

Since $\mathfrak{V}_{\mathcal{R}} W_{i-1}(x,y) = n^2 \int_{\Omega_u\times\Omega_v} \mathfrak{V}_{\mathcal{R}} W_{i-1}$ for any $x \in \Omega_u, y \in \Omega_v$, and (4.11) implies

$$\left|\int_{\Omega_u\times\Omega_v} \mathfrak{V}_{\mathcal{R}} W_{i-1}\right| \le \frac{(k)_2}{n^2},$$

we obtain

$$\left|Y_{uv} - \int_{\Omega_u\times\Omega_v} \mathfrak{V}_{\mathcal{R}} W_{i-1}\right| \le \frac{\binom{k}{2}^2 |\mathcal{H}_k| + (k)_2}{n^2(n-1)} \le \frac{C'}{n^3}, \tag{5.10}$$

where the extra term $(k)_2$ is due to replacement of $(n)_2$ by n^2 in (5.9). Again using the bound $\|\mathfrak{V}\, W_{i-1}\|_\infty \le (k)_2$ from (4.11),

$$\begin{aligned} &\left|\int_{\Omega_S\times\Omega_T} \mathfrak{V}_{\mathcal{R}} W_{i-1} - \sum_{u\in S, v\in T: u\neq v} \int_{\Omega_u\times\Omega_v} \mathfrak{V}_{\mathcal{R}} W_{i-1}\right| \\ &= \left|\sum_{u\in S\cap T} \int_{\Omega_u\times\Omega_v} \mathfrak{V}_{\mathcal{R}} W_{i-1}\right| \le \frac{(k)_2 |S\cap T|}{n^2} \,. \end{aligned} \tag{5.11}$$

Summing (5.10) over $u \neq v$ and using (5.11), we obtain

$$\left|\mathbb{E}\,(X_i \,|\, \mathcal{F}_{i-1}) - \int_{\Omega_S\times\Omega_T} \mathfrak{V}_{\mathcal{R}} W_{i-1}\right| \le \frac{C'|S||T|}{n^3} + \frac{(k)_2|S\cap T|}{n^2} \le \frac{C'+(k)_2}{n}. \tag{5.12}$$

Now Lemma 4.10 implies that, for $i = 1, \dots, t$,

$$\begin{aligned} &\Big|\int_{\Omega_S\times\Omega_T} \mathfrak{V}_{\mathcal{R}} W_i - \int_{\Omega_S\times\Omega_T} \mathfrak{V}_{\mathcal{R}} W_0\Big| \le C_\square \,\|W_i - W_0\|_\square \\ &\le C_\square \int_{\Omega^2} |W_i - W_0| = \frac{C_\square \,|E(G_i) \ominus E(G_0)|}{n^2} \\ &\le C_\square \binom{k}{2} \frac{i}{n^2} \le C_\square \binom{k}{2} \delta = C''\delta \,. \end{aligned} \tag{5.13}$$

Recalling the assumption $\delta \ge 1/n$, from (5.12) and (5.13) we infer that the left-hand side of (5.6) is at most

$$(C' + (k)_2)\delta n + C''\delta^2 n^2 \le \big(C' + (k)_2 + C''\big)\,\delta^2 n^2 = (C''' - 1)\delta^2 n^2.$$

For every i we have $X_i \in [-\binom{k}{2}, \binom{k}{2}]$ and hence $\mathbb{E}\,(X_i \,|\, \mathcal{F}_{i-1}) \in [-\binom{k}{2}, \binom{k}{2}]$. Consequently $|Y_i| \le (k)_2$. Hence, (5.7) follows by applying Lemma A.1 to both $\sum_{i=1}^t Y_i$ and $\sum_{i=1}^t (-Y_i)$. □

We can now prove Theorem 5.1.

Proof of Theorem 5.1. For simplicity we write $W_i = W_{G_i}$ and $\mathfrak{V}\, W_i = \mathfrak{V}_{\mathcal{R}} W_{G_i}$.

We first show that it is enough to fix a small number $\delta > 0$ such that $t := \delta n^2$ is an integer and to prove (5.1) only for i being multiples of t that fall in the interval $(0, Tn^2]$. For any integer $i \in [Tn^2]$ let i' be the largest multiple of t not larger than i, namely $i' = t\lfloor i/t \rfloor$. By the triangle inequality,

$$\Big|\, \Big\|W_i - \Phi^{i/n^2} W_0\Big\|_\square - \Big\|W_{i'} - \Phi^{i'/n^2} W_0\Big\|_\square \Big| \le \|W_i - W_{i'}\|_\square + \Big\|\Phi^{i'/n^2} W_0 - \Phi^{i/n^2} W_0\Big\|_\square .$$

We have $\|W_i - W_{i'}\|_\square \overset{(3.3)}{\le} \|W_i - W_{i'}\|_1 = 2|E(G_i) \ominus E(G_i')|/n^2 \le 2(i-i')\binom{k}{2}/n^2$. Moreover, by (3.3) and (4.14), $\Big\|\Phi^{i'/n^2} W_0 - \Phi^{i/n^2} W_0\Big\|_\square \le (k)_2 (i-i')/n^2$. Hence

$$\begin{aligned} & \Big|\, \Big\|W_i - \Phi^{i/n^2} W_0\Big\|_\square - \Big\|W_{i'} - \Phi^{i'/n^2} W_0\Big\|_\square \Big| \\ \boxed{\text{triangle inequality}} \quad & \le \|W_i - W_{i'}\|_\square + \Big\|\Phi^{i'/n^2} W_0 - \Phi^{i/n^2} W_0\Big\|_\square \\ & \le \frac{2(i-i')}{n^2}\binom{k}{2} + (k)_2 \frac{i-i'}{n^2} \le 4\binom{k}{2}\delta. \end{aligned}$$

Hence if we choose δ small enough to make sure that, say, $4\binom{k}{2}\delta \le \varepsilon/2$, it suffices to prove that with high probability

$$\max_{i \in (0, Tn^2] \cap t\mathbb{Z}} \Big\|W_i - \Phi^{i/n^2} W_0\Big\|_\square \le \frac{\varepsilon}{2}. \tag{5.14}$$

With foresight we define a constant (depending on k)

$$C^* := \frac{C_\square}{2(C''' + C_k (k)_2 + 1)},$$

where $C_\square = C_\square(k, 0)$ is the constant defined in Lemma 4.10, C_k is defined in (4.8), and C''' is defined in (5.5). Now we set

$$\delta = \frac{C^* \varepsilon}{e^{C_\square T}}\ . \tag{5.15}$$

We further set

$$r := |\,\{i \in (0, Tn^2] \cap t\mathbb{Z}\}\,| = \left\lfloor \frac{Tn^2}{t} \right\rfloor = \left\lfloor \frac{T}{\delta} \right\rfloor\ .$$

By Lemma 5.5 and union bound we have

$$\max_{i \in [r]} \big\|W_{it} - W_{(i-1)t} - \delta \mathfrak{V}\, W_{(i-1)t}\big\|_\square \le \big(C''' + 1\big)\, \delta^2 \tag{5.16}$$

with probability at least

$$1 - r2^{2n+1} \exp\left(-\frac{\delta^3 n^2}{2(k)_2^2}\right)$$
$$\geq 1 - \frac{Te^{C_\square T}}{C^*\varepsilon} \exp\left((2\ln 2)n - \frac{C^*\varepsilon^3 n^2}{2(k)_2^2 e^{3C_\square T}}\right) ,$$

For the following fix an outcome of the flip process such that (5.16) holds. For every $i \in [0, r]$ we have that

$$\left\|\Phi^\delta W_i - W_i - \delta\mathfrak{V}\, W_i\right\|_\square \overset{(3.3)}{\leq} \left\|\Phi^\delta W_i - W_i - \delta\mathfrak{V}\, W_i\right\|_\infty \overset{(4.15)}{\leq} C_k(k)_2\delta^2 . \tag{5.17}$$

From (5.16) and (5.17) we get that for every $i \in [r]$ that

$$\begin{aligned}\left\|W_{it} - \Phi^\delta W_{(i-1)t}\right\|_\square &\leq \left\|W_{it} - W_{(i-1)t} - \delta\mathfrak{V}\, W_{(i-1)t}\right\|_\square \\ &\quad + \left\|W_{(i-1)t} + \delta\mathfrak{V}\, W_{(i-1)t} - \Phi^\delta W_{(i-1)t}\right\|_\square \\ &\leq (C''' + 1 + C_k(k)_2)\delta^2 =: \varepsilon_2 . \end{aligned} \tag{5.18}$$

Theorem 4.16 implies that for every $i \in [r]$

$$\left\|\Phi^\delta W_{(i-1)t} - \Phi^\delta\Phi^{(i-1)\delta} W_0\right\|_\square \leq e^{C_\square\delta}\left\|W_{(i-1)t} - \Phi^{(i-1)\delta} W_0\right\|_\square . \tag{5.19}$$

We further prove by induction that for $i = 0, 1, \ldots, r$ we have

$$\left\|W_{it} - \Phi^{i\delta} W_0\right\|_\square \leq \frac{\left(e^{C_\square\delta i} - 1\right)\varepsilon_2}{e^{C_\square\delta} - 1} . \tag{5.20}$$

The base $i = 0$ is trivial and for the induction step assuming $i \geq 1$ we have

$$\begin{aligned}\left\|W_{it} - \Phi^{i\delta} W_0\right\|_\square &\overset{(4.7)}{\leq} \left\|W_{it} - \Phi^\delta W_{(i-1)t}\right\|_\square + \left\|\Phi^\delta W_{(i-1)t} - \Phi^\delta\Phi^{(i-1)\delta} W_0\right\|_\square \\ &\boxed{\text{by (5.18), (5.19)}} \quad \leq \varepsilon_2 + e^{C\delta}\left\|W_{(i-1)t} - \Phi^{(i-1)\delta} W_0\right\|_\square \\ &\boxed{\text{induction hypothesis}} \quad \leq \varepsilon_2 + \frac{e^{C_\square\delta}\left(e^{C_\square\delta(i-1)} - 1\right)\varepsilon_2}{e^{C_\square\delta} - 1} = \frac{\left(e^{C_\square\delta i} - 1\right)\varepsilon_2}{e^{C_\square\delta} - 1} , \end{aligned}$$

as was needed.

This puts us in a position to prove (5.14), which was needed to finish the proof. Indeed, for every $i \in [r]$ we have

$$\begin{aligned}\left\|W_{it} - \Phi^{i\delta} W_0\right\|_\square &\overset{(5.20)}{\leq} \frac{e^{C_\square\delta r} - 1}{e^{C_\square\delta} - 1}\varepsilon_2 \\ &\boxed{r \leq T/\delta,\ e^x \geq 1 + x} \quad \leq e^{C_\square T}\varepsilon_2/C_\square\delta \\ &\boxed{\text{by (5.18)}} \quad = e^{C_\square T}\cdot(C''' + C_k(k)_2 + 1)\delta/C_\square \\ &\boxed{\text{by (5.15)}} \quad = \varepsilon/2 . \end{aligned}$$

□

5.2. **Updates along a Poisson point process.** In the model of flip processes we discussed so far, the flips occur at discrete times. The following modification may be more suitable for applications. Starting with a graph $G = G_0^*$ we consider a continuous time Markov process $(G_t^*)_{t\geq 0}$ in which we update the graph along a Poisson point process with intensity 1 according to the same procedure as in the discrete case. We then have a counterpart of Theorem 5.1. Indeed, at any given time Tn^2, by the law of large numbers, the number of arrivals is $(1 + o(1))Tn^2$, with exponentially high probability.

6. Properties of trajectories

In this part we collect various useful properties of trajectories with respect to a fixed flip process. Some of these properties are interesting per se, other are more of an auxiliary nature and used later on. Let us quickly summarise these results. In Section 6.1 we study backwards evolution of trajectories. Recall that in Definition 4.6, we defined the age of each graphon W as the maximum time by which the backward trajectory stays in the space of graphons. We define the *origin of* W as $\Phi^{-\mathsf{age}(W)}W$. We also prove that age is upper semicontinuous. For many natural flip processes and many initial graphons W, the trajectory $\Phi^t W$ converges as $t \to \infty$. In Section 6.2 we define these limits as *destinations*. Section 6.4 treats graphons which have a step structure, and in particular proves that this step structure is preserved throughout the evolution. In Section 6.5 we prove that values 0 and 1 cannot emerge in a graphon throughout its evolution. Section 6.6 proves the plausible fact that a zero velocity on a part $\Omega_1 \times \Omega_2$ of Ω^2 is a sufficient condition for a graphon to stay constant on $\Omega_1 \times \Omega_2$ throughout its evolution. Section 6.7 treats evolution of constant graphons. In particular, we get that each flip process has at least one *fixed point*, that is, a graphon where the trajectory stays constant. In Section 6.8 we prove that any nontrivial rule has at least one nontrivial trajectory. In Section 6.9 we introduce stable destinations as counterparts to asymptotically stable fixed points in the theory of dynamical systems.

The most interesting results about trajectories are given in Sections 6.10 and 6.11. We try to understand the situation in which a given trajectory does not converge, thus complementing Section 6.2. In Section 6.10 we prove the existence (in a non-constructive way) of a periodic trajectory, using the Poincaré–Bendixson theorem as the main tool. In Section 6.11 we study how complicated a single trajectory can be. For example, we prove that a single trajectory cannot be dense in the space of graphons.

Section 6.12 studies the speed of convergence, that is, how fast convergent trajectories approach their destination. The last Section 6.13 is short, but contains an important uniqueness question: what can we say about rules which yield the same trajectories?

6.1. **Going back in time.** In this section, we look at origins of each graphon with respect to our trajectories. Recall that, given a graphon W, $\mathsf{age}(W)$ is the time for which the trajectory can be extended from W backwards within the space $\mathcal{W}_0$ of graphons. For many flip processes, the age of many graphons is finite. For example, in the Erdős–Rényi flip process, when going backwards, the values at each point of the graphon decrease faster and faster, eventually becoming negative.

If $\mathsf{age}(W) < \infty$, we define *the origin of* W

$$\mathsf{orig}_{\mathcal{R}}(W) := \Phi^{-\mathsf{age}(W)}W.$$

which, by Theorem 4.5(iii), is well-defined and is indeed a graphon.

Remark 6.1. *Given a rule $\mathcal{R}$, let $W \in \mathcal{W}_0$ be a graphon such that $\mathsf{age}(W) < \infty$. Writing $U := \mathsf{orig}\,(W)$, we have $\operatorname{ess\,inf} U = 0$ or $\operatorname{ess\,sup} U = 1$, since otherwise U belongs to the interior of $\mathcal{W}_0$ with respect to the norm $\|\cdot\|_\infty$, and the trajectory starting at U stays, by continuity (cf. Lemma 4.15), inside $\mathcal{W}_0$ for some negative time contradicting the fact that $t = -\mathsf{age}(W)$ was the earliest time $\Phi^t W$ was a graphon.*

Let us now look at continuity properties of the map $W \mapsto \mathsf{age}(W)$ in the space of graphons. As usual, the topology of interest is that of the cut norm. Unfortunately, the map age need not be continuous. Consider the triangle removal flip process. The age of the constant-0 graphon is ∞. However, for an arbitrarily small $\alpha > 0$ we can take a set $A \subset \Omega$ of measure α and a graphon that is zero everywhere except on a set $A \times A$ where it is 1. The age of that graphon is 0. To be more exact, this example shows that the map age need not be lower semicontinuous. However, upper semicontinuity is guaranteed by the following theorem.

Theorem 6.2. *For any rule $\mathcal{R}$, if $(W_n)_{n=1}^\infty$ is a sequence of graphons that converges to a graphon W in the cut norm, then $\limsup_n \mathsf{age}(W_n) \le \mathsf{age}(W)$.*

Proof. The statement is vacuous when $\limsup_n \mathsf{age}(W_n) = 0$, so assume the contrary. We need to prove that for any (finite) $\gamma \in (0, \limsup_n \mathsf{age}(W_n))$ the trajectory of W is well-defined in backward time until $-\gamma$, that is, there exists a graphon U so that $\Phi^\gamma U = W$. By passing to a subsequence, we can assume that for every n we have $\mathsf{age}(W_n) > \gamma$, and hence graphons $U_n := \Phi^{-\gamma}W_n$ exist. We apply Theorem 4.16 with $t = -\gamma$ and graphons $(W_n)_n$, and see that the sequence $(U_n)_n$ is Cauchy with respect to the cut norm distance. Hence, by Lemma 3.2 the sequence $(U_n)_n$ converges to some graphon U in the cut norm. Applying Theorem 4.16, this time with $t = \gamma$ and to graphons $U, U_1, U_2, \dots$, we see that $(\Phi^\gamma U_n = W_n)_n$ converges to $\Phi^\gamma U$, implying $\Phi^\gamma U = W$. □

6.1.1. *Going back in time even more.*

- If $\mathsf{age}(W) < \infty$, by Theorem 4.5 the trajectory $\Phi^t W$ is defined on an *open* interval $\mathcal{D}_W$ containing $[-\mathsf{age}(W), \infty)$. In particular it is defined for some $t < -\mathsf{age}(W)$. We may have $\mathcal{D}_W \neq (-\infty, \infty)$. For example, we saw that the trajectory of the triangle removal flip process along constant graphons given by (1.3) satisfies $\lim_{t \searrow -\frac{1}{12}} \Phi^t(1) = +\infty$, so $\mathcal{D}_1 = (-\frac{1}{12}, \infty)$.
- Theorem 4.5(iii) implies that if we follow the trajectory from a graphon backwards, once we leave the graphon space, we never come back.
- We do not have a combinatorially meaningful interpretation of the trajectories outside the space of graphons. For example the velocity of the constant-(-1) kernel with respect to the triangle removal flip process is positive, contrary to the intuition that the triangle removal flip process decreases values.

6.2. **Destinations.** Suppose that we have a flip process with a rule $\mathcal{R}$. Given a graphon U, we say that *the trajectory of U is convergent* if there exists a *destination of U*, denoted $\mathsf{dest}_{\mathcal{R}}(U)$ such that $\Phi^t U \xrightarrow{\square} \mathsf{dest}_{\mathcal{R}}(U)$ as $t \to \infty$. We say that a the flip process with rule $\mathcal{R}$ is *convergent* if all trajectories converge.

Trivially each fixed point W is a destination (of the trajectory $\Phi^{\cdot} W \equiv W$). We observe that a destination is necessarily a fixed point.

Proposition 6.3. *Suppose that $\mathcal{R}$ is a flip process. If U is a graphon with a destination W, then $\mathfrak{V}_{\mathcal{R}} W = 0$.*

Proof. For every $t \geq 0$, by the cut norm continuity of $W \mapsto \Phi^t W$ (see Lemma 4.10) we have that (will all limits with respect to the cut norm)

$$W = \lim_{u \to \infty} \Phi^{t+u} U = \lim_{u \to \infty} \Phi^t \Phi^u U = \Phi^t \big(\lim_{u \to \infty} \Phi^u U \big) = \Phi^t W$$

whence $\mathfrak{V} W = \lim_{t \searrow 0} (\Phi^t W - W)/t \equiv 0$. □

The next conjecture asserts that destinations are not only cut norm limits but also L^1-norm limits. It could be that they are even L^∞-norm limits.

Conjecture 6.4. *Suppose that $\mathcal{R}$ is a flip process. If U is a graphon with a destination W, then $(\Phi^t U)_{t \to \infty}$ converges to W also in L^1.*

6.3. **Permuting the ground space.** Recall that $\mathbb{S}_\Omega$ is the set of all measure preserving bijections from Ω to Ω. Given a rule $\mathcal{R}$ and $\varphi \in \mathbb{S}_\Omega$, it is easy to check that for every graphon W we have $\mathfrak{V}\,(W^\varphi) = (\mathfrak{V}\, W)^\varphi$. Since the operator $W \mapsto W^\varphi$ is continuous on $L^\infty(\Omega^2)$, it easily follows from the definition that for every graphon W the function $t \mapsto (\Phi^t W)^\varphi$ is a trajectory starting at $U := W^\varphi$. By uniqueness, $I_U = I_W$ and $\Phi^{\cdot} U = (\Phi^{\cdot} W)^\varphi$. Continuity also implies that $\mathsf{dest}\, U = (\mathsf{dest}\, W)^\varphi$, should at least one of the destinations exist.

6.4. **Step structure.** One of the structurally simplest classes of graphons are step graphons. Related random graph models called *stochastic block models* were actually studied prior to the notion of graphons (recall the terminology from Section 3.2.4). One can check that the velocity of a step graphon W (for any fixed flip process) is a kernel with the same step structure, which in turn yields that $\Phi^t(W)$ still has the same step structure (cf. Remark 4.9). We derive this claim from the following slightly stronger proposition.

Proposition 6.5. *Consider an arbitrary flip process and a graphon W. If $A \subset \Omega$ is a twin-set of W, then for any $t \in I_W$, the set A is also a twin-set for $\Phi^t W$.*

Proof. The set $\mathcal{W}_A := \{W \in \mathcal{W} : A \text{ is a set of twins in } W\}$ is clearly a linear subspace of the space of kernels $\mathcal{W}$. Lemma 3.8 tells us that $\mathcal{W}_A$ is a closed subspace with respect to the $\|\cdot\|_\infty$-norm. Crucially, observe that whenever $U \in \mathcal{W}_A$, then it follows from Definition 4.1 that $\mathfrak{V}\, U \in \mathcal{W}_A$. Applying Theorem 3.11, along the same lines as in Theorem 4.5, to the Banach space $(\mathcal{W}_A, \|\cdot\|_\infty)$ (rather than to the whole $\mathcal{W}$) it follows that $\Phi^t W \in \mathcal{W}_A$. □

The following corollary says that the step structure of a graphon stays the same along the trajectory. A special case of the corollary is that if W is not a constant graphon, then $\Phi^t(W)$ is not a constant graphon for every $t \in I_W$.

Corollary 6.6. *If W is a step graphon with a minimal step partition $S_1 \cup S_2 \cup \dots$, then for every $t \in I_W$ the graphon $\Phi^t W$ is a step graphon with a minimal step partition $S_1 \cup S_2 \cup \dots$.*

Proof. If follows immediately from Proposition 6.5 that $S_1 \cup S_2 \cup \dots$ is a step partition for any $\Phi^t W$. It remains to argue that it stays minimal.

Suppose $\Phi^t W =: U$ is a step graphon with respect to a coarser step partition $T_1 \cup T_2 \cup \dots$. By Proposition 6.5, applied to graphon U, graphon $W = \Phi^{-t} U \in \mathcal{W}_0$ is a step graphon with partition $T_1 \cup T_2 \cup \dots$, contradicting the minimality of partition $S_1 \cup S_2 \cup \dots$ for W. □

6.5. **Values 0 and 1.** Here, we give a quick proof of the fact that we cannot arrive to values 0 and 1. This result will be useful in Section 6.11.

Proposition 6.7. *For any flip process, any graphon W, and every $t > 0$ the sets*

$$A_t := \left\{(x,y) \in \Omega^2 : W(x,y) > 0, \Phi^t W(x,y) = 0\right\},$$
$$B_t := \left\{(x,y) \in \Omega^2 : W(x,y) < 1, \Phi^t W(x,y) = 1\right\}$$

have measure zero.

Proof. The proof is similar to the proof of Claim 4.13, but things are now simpler since we know $\Phi^t W \in \mathcal{W}_0$ for every $t \geq 0$. By (4.9), we have $\mathfrak{V}\, \Phi^t W \geq -(k)_2 \Phi^t W$, so using

Remark 4.9 and Proposition C.4 for justification, we can fix $x, y \in \Omega$ and assume that $F(t) = \Phi^t W(x,y)$ and $f(t) = \mathfrak{V}\,\Phi^t W(x,y)$ satisfy $F(t) - F(0) = \int_0^t f(\tau)\,\mathrm{d}\tau$ and $f(t) \geq -(k)_2 F(t)$ for $t \geq 0$. Lemma 3.10 implies that

$$\Phi^t W(x,y) = F(t) \geq F(0) e^{-(k)_2 t} = e^{-(k)_2 t} W(x,y).$$

It follows that $\Phi^t W > 0$ wherever $W > 0$. Since, when applying Proposition C.4, we modified W and $\Phi^t W$ only on sets of measure zero, the proof is complete for the set A_t. The proof for B_t is analogous. □

6.6. **Zero velocity on a section.** Consider a subspace of graphons which agree on a fixed subset S. Our next proposition states an intuitive and sometimes "obvious" thing: if the velocity of every graphon in the subspace equals zero on S, then any trajectory started in the subspace stays inside it.

Proposition 6.8. *Consider a flip process of order k. Given two sets $\Omega_1, \Omega_2 \subset \Omega$, let $S := \Omega_1 \times \Omega_2 \cup \Omega_2 \times \Omega_1$ and let $\alpha : S \to [0,1]$ be a symmetric function. Let $\mathcal{A} = \{W \in \mathcal{W}_0 : W_{\restriction S} = \alpha\}$. Suppose that for each $W \in \mathcal{A}$ we have $(\mathfrak{V}\,W)_{\restriction S} = 0$.*

If $U \in \mathcal{A}$ and $t \geq 0$, then $\Phi^t U \in \mathcal{A}$.

Proof. We denote the L^∞-norm on the space of functions with domain S as $\|\cdot\|_*$. That is to make a distinction from the L^∞-space on Ω^2 for which we use the symbol $\|\cdot\|_\infty$. Given an arbitrary graphon $W \in \mathcal{W}_0$, let its *α-surgery* be a graphon $W^{[\alpha]}$ which is equal to α on S and to W on $\Omega^2 \setminus S$. Observe that we always have

$$\|W - W^{[\alpha]}\|_\infty = \|W_{\restriction S} - \alpha\|_* \,, \tag{6.1}$$

and also

$$\|W - U\|_\infty \geq \|W_{\restriction S} - U_{\restriction S}\|_* \,. \tag{6.2}$$

Suppose that $U \in \mathcal{A}$ is given. For each $t \geq 0$ we define $U_t := \Phi^t U$, and write $f(t) := \|(U_t)_{\restriction S} - \alpha\|_*$. In particular, $f(0) = 0$. Using that $U_{\restriction S} = \alpha$ and $\left(\mathfrak{V}\,\left(U_\tau^{[\alpha]}\right)\right)_{\restriction S} \equiv 0$, we can write

$$\alpha = \left(U + \int_{\tau=0}^{t} \mathfrak{V}\,\left((U_\tau)^{[\alpha]}\right)\right)_{\restriction S} .$$

We have

$$\begin{aligned}
f(t) &= \left\| \left(U + \int_{\tau=0}^{t} \mathfrak{V}\, U_\tau \right)_{\restriction S} - \left(U + \int_{\tau=0}^{t} \mathfrak{V}\, \left((U_\tau)^{[\alpha]} \right) \right)_{\restriction S} \right\|_* \\
\boxed{\text{by (6.2)}} \quad &\le \left\| \left(U + \int_{\tau=0}^{t} \mathfrak{V}\, U_\tau \right) - \left(U + \int_{\tau=0}^{t} \mathfrak{V}\, \left((U_\tau)^{[\alpha]} \right) \right) \right\|_\infty \\
\boxed{\text{triangle inequality}} \quad &\le \int_{\tau=0}^{t} \left\| \mathfrak{V}\, U_\tau - \mathfrak{V}\, \left((U_\tau)^{[\alpha]} \right) \right\|_\infty \\
\boxed{\text{by (4.10)}} \quad &\le C_k \int_{\tau=0}^{t} \left\| U_\tau - \left((U_\tau)^{[\alpha]} \right) \right\|_\infty \\
\boxed{\text{by (6.1)}} \quad &= C_k \int_{\tau=0}^{t} \left\| (U_\tau)_{\restriction S} - \alpha \right\|_* = C_k \int_{\tau=0}^{t} f(\tau)\, .
\end{aligned}$$

Applying Lemma 3.9 (with $A = 0$), we obtain, for every $t \ge 0$, that $f(t) \le A \exp(C_k t) = 0$, which implies that $f(t) = 0$ and therefore $\Phi^t U = U_t \in \mathcal{A}$, as was needed. □

6.7. **Constant graphons.** Suppose that $\mathcal{R}$ is a flip process of order k. We shall study the trajectories when restricted on the space of constant graphons (recall that by Corollary 6.6 trajectories started from constants stay constant and that no trajectory started at a non-constant reaches a constant). Moreover, the velocity of a constant graphon is also constant. Given a constant-d graphon W, let $\mathbf{F}$ be the drawn graph and $\mathbf{H}$ be the replacement graph. By integrating (4.4) over Ω^2, it follows that

$$\mathfrak{V}\, d = \int_{\Omega^2} \mathfrak{V}\, d = 2\mathbb{E}\left(e(\mathbf{H}) - e(\mathbf{F}) \right). \tag{6.3}$$

For each $\ell \in \{0, \dots, \binom{k}{2}\}$, write

$$\Delta_\ell := \mathbb{E}\left(e(\mathbf{H}) - e(\mathbf{F}) \,|\, e(\mathbf{F}) = \ell \right) = \binom{\binom{k}{2}}{\ell}^{-1} \sum_{H \in \mathcal{H}_k : e(H) = \ell} \sum_{H' \in \mathcal{H}_k} \mathcal{R}_{H,H'} \cdot (e(H') - \ell)\, . \tag{6.4}$$

That is, Δ_ℓ is the expected change of the number of edges in the replacement graph in one step of the flip process, if the drawn graph is a uniformly random graph with ℓ edges. Note that

$$\Delta_0 \ge 0 \quad \text{and} \quad \Delta_{\binom{k}{2}} \le 0\, . \tag{6.5}$$

Since $\mathbf{F}$ is a binomial random graph $\mathbb{G}(k, d)$, for each constant $d \in [0, 1]$ the velocity satisfies

$$\mathfrak{V}\, d = 2 \sum_{\ell=0}^{\binom{k}{2}} \binom{\binom{k}{2}}{\ell} d^\ell (1-d)^{\binom{k}{2} - \ell} \Delta_\ell\, . \tag{6.6}$$

Viewed as a function in d, this is a polynomial of degree at most $\binom{k}{2}$. Firstly, let us characterise the case when this polynomial is trivial.

Fact 6.9. *$\mathfrak{V}\, d \equiv 0$ if and only if $\Delta_\ell = 0$ for all $\ell \in \{0, \dots, \binom{k}{2}\}$.*

Proof. If $\Delta_\ell = 0$ for all $\ell \in \{0, \dots, \binom{k}{2}\}$, then obviously $\mathfrak{V}\, d \equiv 0$. Suppose, on the other hand, that there exists $\Delta_h \neq 0$, and let h be minimum such. Then looking at (6.6), we have $\lim_{d\to 0} \frac{\mathfrak{V}\, d}{d^h} = \binom{\binom{k}{2}}{h}\Delta_h$ which implies that $\mathfrak{V}\, d$ is indeed non-zero. □

So, let us proceed under the assumption that $\mathfrak{V}\, d$ is nontrivial. The roots of $\mathfrak{V}\, d$ on the interval $[0, 1]$ are those constant graphons which are fixed points of $\mathcal{R}$. By (6.5) and the intermediate value theorem, there is at least one such fixed point. Hence we get the following.

Proposition 6.10. *Any flip process has at least one fixed point.*

A trajectory started at a value $d \in (r, r')$ between two consecutive roots r and r' converges either to r or r' depending on the signum of $\mathfrak{V}$ on that interval. Further, polynomial speed of convergence (see Section 6.12 for general treatment of non-constant graphons) can be easily deduced.

So, we have reduced the problem of studying trajectories on constants to questions about roots of polynomials of the form

$$f(x) = \sum_{\ell=0}^{\binom{k}{2}} c_\ell x^\ell (1-x)^{\binom{k}{2}-\ell}\,, \tag{6.7}$$

where $c_\ell \in 2\binom{\binom{k}{2}}{\ell} \cdot [-\ell, \binom{k}{2} - \ell]$. (It is clear from the second form of Δ_ℓ in (6.4) that one can choose a rule $\mathcal{R}$ so that $\mathfrak{V}_{\mathcal{R}} d = f(d)$.) This coefficient restriction can be relaxed by multiplying by a small constant. After this the only restrictions that remain are $c_0 \geq 0$ and $c_{\binom{k}{2}} \leq 0$, which translates as

$$f(0) \geq 0 \text{ and } f(1) \leq 0\,. \tag{6.8}$$

The space of polynomials of the form (6.7) (with arbitrary real coefficients) is a linear space of dimension $\binom{k}{2} + 1$. This can be seen from the fact that each of the $\binom{k}{2} + 1$ coefficients c_ℓ can be chosen, and the only constant-0 polynomial is the one with all coefficients zero. Hence this space is the space of all polynomials of degree at most $\binom{k}{2}$.

Recall that f was the velocity. From (6.8) we conclude, for a given k and arbitrary points $0 \leq x_1 \leq \ldots \leq x_m \leq 1$, where $m \leq \binom{k}{2}$, there exists a rule of order $\mathcal{R}$, which, when restricted to constant graphons, has fixed points $\{x_1, \dots, x_m\}$ and non-fixed trajectories are monotone and take values in the intervals between the roots.

6.8. **Examples of graphons not fixed for any (nontrivial) rule.** In this section, we shall find graphons W for which we cannot have $\mathfrak{V}_{\mathcal{R}}W \equiv 0$ for any nontrivial rule $\mathcal{R}$.

We shall work with step graphons with a small number of blocks. Hence, throughout this section Ω is finite and π is a measure such that $\pi(\{\omega\}) > 0$ for every $\omega \in \Omega$. Suppose there are steps $x, y, z \in \Omega$ (possibly $x = y$) such that $W(x, y) = 0$ and $W(x, z), W(y, z), W(z, z) \in (0, 1)$, see Figure 6.1, top two examples. The value $\mathfrak{V}_{\mathcal{R}}W(x, y)$ is a linear combination of numbers $\mathcal{R}_{G,H}$, where the coefficient of each $\mathcal{R}_{G,H}$ is proportional to

$$\sum_{ab\in H\setminus G} t_{\mathrm{ind}}^{(x,y)}(G^{ab}, W) - \sum_{ab\in G\setminus H} t_{\mathrm{ind}}^{(x,y)}(G^{ab}, W) = \sum_{ab\in H\setminus G} t_{\mathrm{ind}}^{(x,y)}(G^{ab}, W)$$

because of $W(x, y) = 0$. Note that the first sum is positive whenever $H \setminus G \neq \emptyset$, since if we fix $ab \in H \setminus G$, then by mapping a to x, b to y and the remaining vertices of G to z, we see that $t_{\mathrm{ind}}^{x,y}(G^{ab}, W) > 0$. So in order to have $\mathfrak{V}\, W(x, y) = 0$ we need $\mathcal{R}_{G,H} = 0$ whenever $|H \setminus G| \geq 1$. In other words, we can have $\mathcal{R}_{H,G} > 0$ only if $H \subseteq G$. Suppose that $\mathcal{R}_{G,H} > 0$ for some G, H such that $H \subsetneq G$. Since in $\mathfrak{V}\, W(z, z)$ the coefficient of such $\mathcal{R}_{G,H}$ is, because of $W(z, z) \in (0, 1)$,

$$-\sum_{ab\in G\setminus H} t_{\mathrm{ind}}^{(z,z)}(G^{ab}, W) < 0,$$

implying that $\mathcal{R}_{G,H} = 0$ for $H \subsetneq G$ as well. This leaves only the possibility $\mathcal{R}_{G,H} > 0$ when $G = H$, hence the rule $\mathcal{R}$ is trivial.

By a symmetric argument $\mathfrak{V}_{\mathcal{R}}W \equiv 0$ implies $\mathcal{R}$ is trivial if value 0 for $W(x, y)$ is replaced by value 1. There are a few more configurations which prevent a step graphon from being a fixed point in a nontrivial rule, see Figure 6.1.

The most important corollary is the following.

Corollary 6.11. *If a rule $\mathcal{R}$ is such that every graphon is a fixed point, then $\mathcal{R}$ is trivial.*

6.9. **Stable destinations.** We say that a graphon $S \in \mathcal{W}_0$ is a *stable destination* for a given flip process if there exists $\delta > 0$ such that for each $W \in \mathcal{W}_0$ with $d_{\square}(W, S) < \delta$ we have $\mathsf{dest}\,(W) = S$. We call the supremum of such δ the *radius of the stable destination*. In the dynamical systems terminology a stable destination is called an *asymptotically stable fixed point*.

Many nontrivial flip processes have at least one stable destination. However, this probably is not always the case. In [2] it is shown that the "extremist flip process" (see Section 2.7) for $k = 3, 4$ does not have stable destinations at its most likely locations (though it does not rule out the existence of a stable destination at some other locations).

It is plausible that Theorem 5.1 (see also Remarks 5.2 and 5.3) can be strengthened for trajectories converging to a stable destination. That is, it is likely that a flip process started

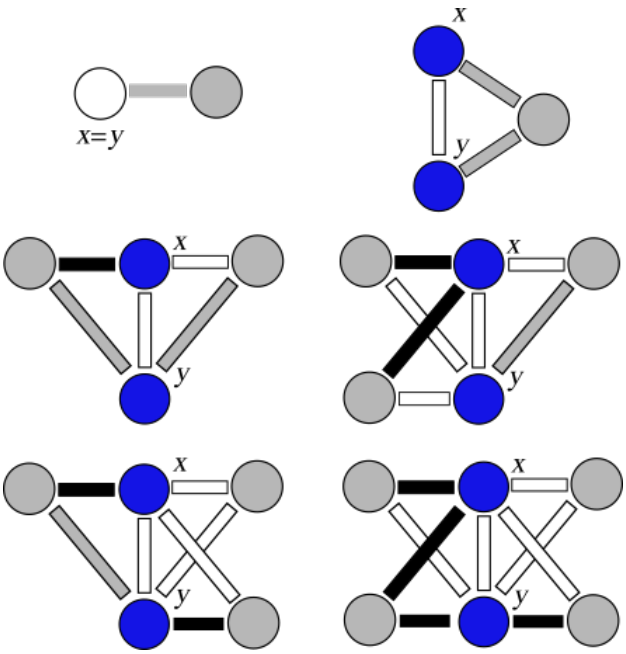


FIGURE 6.1. Examples of configurations in a step graphon preventing $\mathfrak{V}_{\mathcal{R}}W \equiv 0$ for any nontrivial rule $\mathcal{R}$. White has density 0, black has density 1. Different grey objects allow different densities (but necessarily in $(0,1)$). Blue clusters have unrestricted densities. Changing density of $W(x,y)$ from 0 to 1 gives six more configurations.

with a finite n-vertex graph close to such a trajectory typically stays in the neighbourhood of that destination for $\exp(\Theta(n^2))$ steps, until a really rare event kicks in.

6.10. **A periodic trajectory.** In this section we prove existence of a flip process which has a periodic trajectory in $\mathcal{W}_0$ (which is not a fixed point).

Let us first present the main idea, introducing necessary notation on the way. Fix a partition $\Omega = \Omega_1 \cup \Omega_2$ with $\pi(\Omega_1) = \pi(\Omega_2) = 1/2$. We define a map $\Gamma : \mathbb{R}^2 \to \mathcal{W}$ which associates to each $(p_1, p_2) \in \mathbb{R}^2$ a step function

$$\Gamma(p_1, p_2) = p_1 \mathbb{I}_{\Omega_1^2} + p_2 \mathbb{I}_{\Omega_2^2}, \tag{6.9}$$

that is, $\Gamma(p_1, p_2)$ is constant p_i on Ω_i^2 for $i = 1, 2$, and zero elsewhere. Define a class of graphons $\mathcal{U}_0 := \Gamma([0,1]^2)$ and a class of kernels $\mathcal{U} := \Gamma(\mathbb{R}^2)$. We will show that for some $\mathcal{R}$ there is a periodic trajectory taking values in $\mathcal{U}_0$. We will sometimes omit the symbol Γ and make no distinction between points in $\mathbb{R}^2$ and elements of $\mathcal{U}$. In particular, we write $\mathfrak{V}_{\mathcal{R}}(p_1, p_2) := \mathfrak{V}_{\mathcal{R}}(\Gamma(p_1, p_2))$. It is immediate from the definition that $\mathfrak{V}_{\mathcal{R}}(p_1, p_2)$ is a step function with respect to the partition $\Omega = \Omega_1 \cup \Omega_2$. Furthermore, the trajectory will be

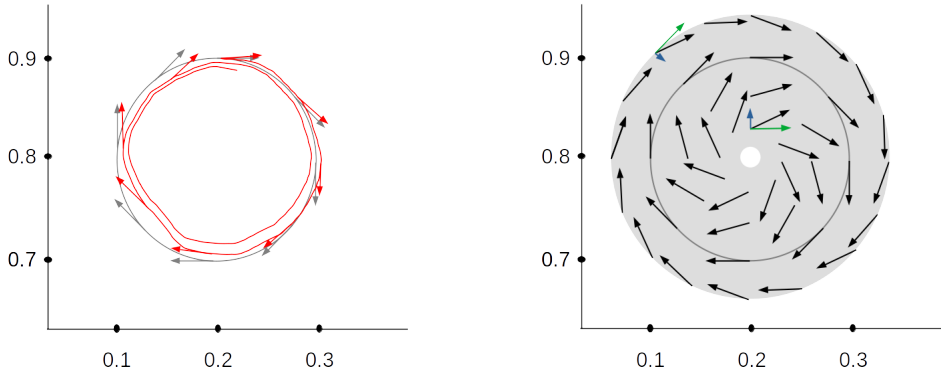


FIGURE 6.2. *Left:* Ideal periodic trajectory given by the field g shown in grey arrows. Red is the actual trajectory (if the rule comes only from g) which is perturbed by small inaccuracies and rare events. *Right:* Vector field $g + h$ in black defines $\mathcal{R}$ on an annulus containing $\mathsf{circle}(r, q)$. Vectors of g and h appear in green and blue, respectively.

confined to $\mathcal{U}_0$, which is equivalent (with the help of Proposition 6.8) to

$$\mathfrak{V}_{\mathcal{R}}(p_1, p_2) \text{ is zero on } \Omega_1 \times \Omega_2. \tag{6.10}$$

Hence, we can treat $\mathfrak{V}_{\mathcal{R}}(p_1, p_2)$ as a point in $\mathbb{R}^2$ via Γ^{-1}.

For $r > 0$ and $p \in \mathbb{R}^2$, we denote by $B(r, p)$ the open ball of radius r around a point p, by $\overline{B(r, p)}$ the corresponding closed ball and we set $\mathsf{circle}(r, p) = \overline{B(r, p)} \setminus B(r, p)$.

(The following description uses tacitly the map Γ, as discussed above.) Our first attempt is to have the trajectory constrained to $\mathsf{circle}(r, q)$ with $r = 0.1$ and $q = (a, b) := (0.2, 0.8)$. If we want to achieve that, then for each point $(x, y) \in \mathsf{circle}(r, q)$, the velocity $\mathfrak{V}(x, y)$ has to be a constant multiple of the tangent field $g : \mathbb{R}^2 \setminus \{q\} \to \mathbb{R}^2$,

$$g(x, y) = (-y + b, x - a)/\sqrt{(-y + b)^2 + (x - a)^2}\,. \tag{6.11}$$

See Figure 6.2(left). So let us aim at $\mathfrak{V}_{\mathcal{R}}(x, y) = g(x, y)$. Assume for a while an idealised setting where we also have that

(A) in the drawn graph F, each vertex carries the information whether it was sampled from Ω_1 or from Ω_2, and also that it carries the information about the density x on $\Omega_1 \times \Omega_1$ and the density y on $\Omega_2 \times \Omega_2$, and

(B) the replacement distribution of the rule $\mathcal{R}$ is allowed to depend on these additional input parameters.

In such a setting, we could easily add or remove edges on the vertices sampled from Ω_1 according to the first coordinate of $g(x, y)$ (even cases where, for example, the first coordinate of $g(x, y)$ is irrational could be handled by randomised outputs), and similarly

for the vertices sampled from Ω_2. We shall leave the pairs of vertices corresponding to $\Omega_1 \times \Omega_2$ intact because of (6.10).

In reality we do not have (A). But when the order k of the flip process is large, then we can, with high probability, get a fairly precise estimate of all this data. To this end we use basic large deviation tools and further standard calculations to see that with probability $1 - o(1)$ for the drawn graph F for the rule, $F \sim \mathbb{G}(n, \Gamma(x, y))$ satisfies the following properties:

(♣) F consists of two components, each of size $(1 + o(1))\frac{k}{2}$ (henceforth until the end of section asymptotics are for $k \to \infty$). Further, one component forms a graph of density $d_1 = x + o(1)$ and all its vertices were sampled from Ω_1. The other component forms a graph of density $d_2 = y + o(1)$ and all its vertices were sampled from Ω_2.

Recall that for each $(x, y) \in \mathsf{circle}(r, q)$ we have $x \leq 0.3$ and $y \geq 0.7$. This means that with high probability the vertices in the sparser component were sampled from Ω_1 and the vertices from the denser component from Ω_2. So, to circumvent the unavailability of (B) we might define the rule $\mathcal{R}$ to be nontrivial on graphs of the form (♣) and for each such drawn graph the rule would change the density in the first and the second component according to the first and the second coordinate of $g(d_1, d_2)$, respectively.

Unfortunately, such a rule alone cannot guarantee that $\mathsf{circle}(r, q)$ will be the trajectory, since we do not have any control over the $o(1)$-errors. To give a specific example, even when the current graphon is (x, y), with a small but positive probability it may happen that the drawn graph F will consist of two components of respective densities $d_1 = \hat{x} \pm o(1)$ and $d_2 = \hat{y} \pm o(1)$ for some very different $(\hat{x}, \hat{y}) \in \mathsf{circle}(r, q)$. Hence, unwanted changes to the densities of the components of F will be applied.

To amend the situation, rather than focusing solely on the circle, we work in an annulus C containing $\mathsf{circle}(r, q)$ in its interior, and $\mathcal{R}$ will involve all possible graphs F as in (♣) where $(d_1, d_2) \in C$. On $\mathsf{circle}(r, q)$, the rule $\mathcal{R}$ is as before. On the rest of the annulus, the rule is given by the tangent vector field g composed with a vector field $h : \mathbb{R}^2 \setminus \{q\} \to \mathbb{R}^2$ oriented towards $\mathsf{circle}(r, q)$,

$$h(x, y) = \left(\frac{r}{\sqrt{(x-a)^2 + (y-b)^2}} - 1 \right) (x - a, y - b) \,. \tag{6.12}$$

See Figure 6.2(right). For drawn graphs F which do not consist of exactly two components whose joint densities lie in C, we define the replacement graph H arbitrarily, subject to the condition that

(6.13) no edges are introduced between any two components of F.

The point is that even though we still know that $\mathfrak{V}_{\mathcal{R}}(x,y)$ is only approximately equal to $(h+g)(x,y)$ on C, we can prove that the undesired perturbation from rare events cannot cause that the actual velocity vector escapes from C. The classical Poincaré–Bendixson theorem then implies that there must exist a periodic trajectory. Note that (6.13) is needed to preserve (6.10); here even rare violations of (6.13) would kill (6.10).

Lemma 6.12 below allows us to produce a rule whose velocities approximate a given vector field; in fact such a rule is produced in the way we described above. To motivate some of the conditions of the lemma, we recall that we relied on the fact that the two components of the graph F are distinguishable by their densities (see (ii)), and also that in the planned change, either component cannot have density less than 0 or more than 1 (see (iv)).

Lemma 6.12. *Let Γ be defined by (6.9). For every $\varepsilon,\delta>0$ exists $k\in\mathbb{N}$ such that the following holds. If $S_1,S_2\subseteq[0,1]^2$ and $f:S_1\to\mathbb{R}^2$ are such that*

(i) *$B(\delta,p)\subseteq S_1$ for every $p\in S_2$,*
(ii) *$0<\delta\le p_1\le p_2-\delta\le 1-2\delta$ for every $(p_1,p_2)\in S_1$,*
(iii) *f is uniformly continuous on S_1, and*
(iv) *$0\le p_i+(k)_2^{-1}f_i(p_1,p_2)\le 1$ for every $i=1,2$ and $(p_1,p_2)\in S_1$,*

then there exists a rule $\mathcal{R}$ of order k such that (6.10) holds for every $p=(p_1,p_2)\in S_2$ and such that

$$\sup_{p\in S_2}|f(p)-\Gamma^{-1}(\mathfrak{V}_{\mathcal{R}}\Gamma(p))|\le\varepsilon. \tag{6.14}$$

Proof. Within the proof the asymptotic notation is with respect to $k\to\infty$ and whenever we say "for k large enough", we mean k larger than some constant depending on ε,δ. We fix some sequences $\alpha_1=\alpha_1(k),\alpha_2=\alpha_2(k)$ such that

$$\tfrac{1}{\sqrt{k}}\ll\alpha_1\ll\alpha_2\ll 1.$$

We define a rule $\mathcal{R}$ as follows. Assume that $F\in\mathcal{H}_k$ is the drawn graph. We define the probabilities $\mathcal{R}_{F,\cdot}$ by describing a random replacement graph $\mathbf{H}$. If F has an isolated vertex or the number of connected components in F is not two, we do nothing, that is $\mathbf{H}:=F$. If F has two components and each of them consists of at least 2 vertices, let F_1 be the component of F with smaller edge density and let F_2 be the other one (the ties can be broken arbitrarily). For $i=1,2$, we write $t_i:=e(F_i)/\binom{v(F_i)}{2}$ for the edge density of F_i. If $t_i+f_i(t_1,t_2)<0$ or $t_i+f_i(t_1,t_2)>1$ for some $i\in[2]$, then we again set $\mathbf{H}:=F$. Otherwise we replace F_i by a random binomial graph with edge probability $t_i+(k)_2^{-1}f_i(t_1,t_2)$. Hence $\mathbf{H}$ is the vertex-disjoint union of these two (independently drawn) random graphs.

Write $p_{1,1}:=p_1$, $p_{2,2}:=p_2$ and $p_{1,2}=p_{2,1}:=0$. Let $\mathbf{F}\sim\mathbb{G}(k,\Gamma(p))$ be a random k-vertex graph sampled from $\Gamma(p_1,p_2)$. We get the same random graph by sampling

independent uniform elements $\mathbf{U}_v \in [2]$ for each $v \in [k]$ and then connecting each pair u, v independently with probability $p_{\mathbf{U}_u,\mathbf{U}_v}$. By Lemma 4.2,

$$(\mathfrak{V}_{\mathcal{R}}\Gamma(p))_{\restriction \Omega_i \times \Omega_j} = \sum_{a\neq b, a,b\in[k]} (\mathbb{P}\left(ab \in \mathbf{H} \,|\, \mathbf{U}_a = i, \mathbf{U}_b = j\right) - p_{i,j}), \quad i,j \in [2]\,. \tag{6.15}$$

Whenever $i \neq j$, the condition $\mathbf{U}_a = i, \mathbf{U}_b = j$ implies that pair ab is a non-edge in $\mathbf{H}$ and thus each term in (6.15) is zero. This implies that (6.10) holds for every $p \in [0,1]^2$.

We further assume that $p = (p_1, p_2) \in S_2$ and prove (6.14). Let D_i be the edge density of the subgraph $\mathbf{F}[i]$ of $\mathbf{F}$ induced by $\{v : \mathbf{U}_v = i\}$. A standard Chernoff bound implies that with probability at least $1 - 2\exp(-\alpha_1^2 k/6)$ we have $v(\mathbf{F}[i]) = (1 \pm \alpha_1)k/2$ for $i \in [2]$. We denote this event by $\mathcal{A}$. Furthermore, $\mathbb{E}\left(e(\mathbf{F}[i]) \,|\, \mathcal{A}\right) = \binom{(1\pm\alpha_1)k/2}{2} p_i \sim k^2 p_i/8$ and, assuming that $\mathcal{A}$ holds, again a standard Chernoff bound implies that with probability at least $1 - 2\exp(-\alpha_1^2(1-\alpha_1)^2 k^2 p_i/24)$ we have $e(\mathbf{F}[i]) = (1 \pm \alpha_1)\binom{(1\pm\alpha_1)k/2}{2} p_i$ and therefore $D_i = (1 \pm \alpha_2)p_i$. Lastly, assuming $\mathcal{A}$ holds, with probability at least

$$1 - \binom{v(\mathbf{F}[i])}{2}\left(1 - p_i^2\right)^{v(\mathbf{F}[i])-2} \geq 1 - \binom{k}{2}(1-p_i^2)^{(1-\alpha_1)k/2-2}\,.$$

we have that in $\mathbf{F}[i]$ every two vertices have a common neighbour and hence $\mathbf{F}[i]$ is connected. We denote the event such that all of the three events above hold by $\mathcal{C}$ and note that $\mathbb{P}(\mathcal{C}) = 1 - \exp(-\Omega(\alpha_1^2 k)) = 1 - o(1)$. In particular $\mathcal{C}$ together with (ii) implies that $D_2 \geq p_2 - \alpha_2 > p_2 - \delta/2 \geq p_1 + \delta/2 > p_1 + \alpha_2 \geq D_1$ and also that $(D_1, D_2) \in B(\delta, (p_1, p_2))$ and therefore by (i) and (iv) we have that $0 \leq D_i + (k)_2^{-1} f_i(D_1, D_2) \leq 1$. In view of (i), for large enough k we have $p(1 \pm \alpha_2) \in S_1$, so that condition (iii) implies $|f_i(1 \pm \alpha_2)p) - f_i(p)| = o(1)$, uniformly over all choices of $p \in S_2$.

Hence for $i \in [2]$ and $a, b \in [k]$ (note that by symmetry of the rule the following does not depend on the actual choice of ab as long as $\mathbf{U}_a = \mathbf{U}_b = i$),

$$\begin{aligned}
\mathbb{P}(ab \in \mathbf{H} \mid \mathbf{U}_a = \mathbf{U}_b = i) &= \mathbb{P}\left(ab \in \mathbf{H} \,|\, \mathbf{U}_a = \mathbf{U}_b = i, \mathcal{C}\right) \cdot \mathbb{P}(\mathcal{C}) \pm \mathbb{P}(\overline{\mathcal{C}}) \\
&= \left((1 \pm \alpha_2)p_i + (k)_2^{-1} f_i\left((1 \pm \alpha_2)p_1, (1 \pm \alpha_2)p_2\right)\right) \cdot \mathbb{P}(\mathcal{C}) \pm \mathbb{P}(\overline{\mathcal{C}}) \\
\boxed{\text{by (iii)}} \quad &= \left((1 \pm \alpha_2)p_i + (k)_2^{-1} f_i(p_1, p_2) + o(1)\right) \cdot (1 - o(1)) + o(1) \\
\boxed{\text{for large enough } k} \quad &= p_i + (k)_2^{-1} f_i(p_1, p_2) \pm \varepsilon/(2(k)_2)\,.
\end{aligned}$$

Therefore, recalling (6.15), we finally get that

$$(\mathfrak{V}_{\mathcal{R}}\Gamma(p))_{\restriction \Omega_i \times \Omega_i} = f_i(p_1, p_2) \pm \varepsilon/2, \quad i \in [2]\,,$$

and in particular

$$\sup_{p \in S_2} |f(p) - \Gamma^{-1}(\mathfrak{V}_{\mathcal{R}}\Gamma(p))| \leq \varepsilon.$$

□

We now construct the set where we will find a periodic trajectory. Fix values $r = 0.1$, $r_s = 0.09$, $r_w = 0.11$, $\theta < 0.0001$, $a = 0.2$, $b = 0.8$ and set $q = (a, b)$. Recall that functions $g, h : B(0.15, q) \setminus q \to \mathbb{R}^2$ were defined by (6.11) and (6.12). Define a function $f : B(0.15, q) \setminus q \to \mathbb{R}^2$ by

$$f(x, y) = \theta \cdot (g(x, y) + h(x, y)).$$

Note that the dynamical system over $B(0.15, q) \setminus q$ defined by f has a periodic trajectory $\{(x, y) \in \mathbb{R}^2 \mid (x - a)^2 + (y - b)^2 = r^2\}$. To finish our construction we apply Lemma 6.12 to the function f, with the constants

$$\varepsilon < \min\{|r - r_s|, |r - r_w|, \theta\} \quad \text{and} \quad \delta < 0.005,$$

and the sets $S_1 = B(0.15, q) \setminus \overline{B(0.005, q)}$ and $S_2 = B(0.145, q) \setminus \overline{B(0.01, q)}$. Note that f is continuous on the closure of S_1, which is compact, implying that condition (iii) holds. Verifying the other conditions is straightforward. Hence by Lemma 6.12 there is a rule $\mathcal{R}$ of order k so that for the vector field $f_{\mathcal{R}} : S_2 \to \mathbb{R}^2$ defined by $f_{\mathcal{R}}(p) = \Gamma^{-1}(\mathfrak{V}_{\mathcal{R}}\Gamma(p))$ we have

$$|f(p) - f_{\mathcal{R}}(p)| \leq \varepsilon \quad \text{for all} \quad p \in S_2. \tag{6.16}$$

Using the continuity and linearity of Γ and Γ^{-1} we see that the trajectory $\Phi^{\cdot}_{\mathcal{R}}\Gamma(p)$ (bijectively) corresponds to the solution $x(t) = \Gamma^{-1}\left(\Phi^t\Gamma(p)\right)$ of the dynamical system defined by $F = f_{\mathcal{R}}$ with initial condition $F(0) = p$. Of course, one is periodic if and only if the other is periodic. We will show that the dynamical system defined by F has a periodic trajectory in a closed annulus using the Poincaré–Bendixson theorem, see [18, Theorem 1, p. 245].

Theorem 6.13. *[Poincaré–Bendixson] Suppose $S \subseteq \mathbb{R}^2$ is an open set and a continuously differentiable function $F : S \to \mathbb{R}^2$ defines the dynamical system*

$$x'(t) = F(x(t)) \,. \tag{$*$}$$

If $C \subseteq S$ is a compact set and

(i) *there exists $p \in C$ and a trajectory with $x(0) = p$ such that $x(t) \in C$ for every $t \geq 0$, and*

(ii) *there are no points $p \in C$ with $F(p) = 0$,*

then the system $(*)$ has at least one periodic trajectory inside C.

We apply Theorem 6.13 to the function $F = f_{\mathcal{R}}$ and the sets $S = S_2$ and $C = \overline{B(r_w, q)} \setminus B(r_s, q)$ The following claim shows that our choices fulfil the conditions of Theorem 6.13, implying the existence of a periodic trajectory.

Claim 6.14. *The following is true for $F = f_{\mathcal{R}}$.*

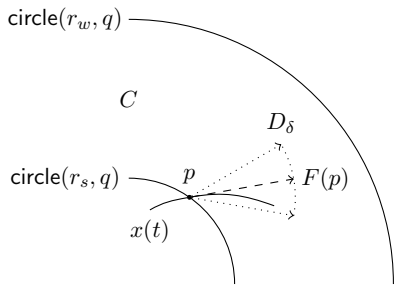


FIGURE 6.3. The trajectory $x(t)$ entering the set C at the point p, is approximated by the tangent $F(p)$ and thus stays in the wedge-shaped set D_δ.

(i) F points inwards *at every point p on the boundary of C, that is, $F(p)$ is not collinear with the tangent of the boundary at p and $p + \delta F(p) \in C$ for small enough $\delta > 0$.*
(ii) *If p lies on the boundary of C, that is, $\mathsf{circle}(r_w, q) \cup \mathsf{circle}(r_s, q)$, and $x(T) = p$ for some trajectory defined by $(*)$ and some $T > 0$, then for some $\tau > 0$ we have $x([T, T + \tau]) \in C$.*
(iii) *Condition* (i) *of Theorem 6.13 holds (in fact, for every $p \in C$).*
(iv) *Condition* (ii) *of Theorem 6.13 holds.*

Proof. For (i), assume first that $(x, y) = p \in \mathsf{circle}(r_s, q)$. The component $g(p)$ of $f(p)$ is tangent to $\mathsf{circle}(r_s, q)$ and the component $h(p)$ points away from the point q. Since $p \in \mathsf{circle}(r_s, q)$, we have $\sqrt{(x-a)^2 + (y-b)^2} = r_s$, therefore $|h(p)| = |r - r_s| > \varepsilon$ and hence, in view of (6.16), vector $F(p)$ points inwards. An analogous argument shows that $F(p)$ also points inwards for every $p \in \mathsf{circle}(r_w, q)$.

We now prove (ii). By (i), vector $F(p)$ forms a non-zero angle $\gamma \le \pi/2$ with the tangent at p to a boundary circle and by adding to p a small multiple of $F(p)$ we end up in C. Let D_δ be a wedge-shaped set consisting of vectors of length at most δ at angle at most $\gamma/2 (\le \pi/4)$ with $F(p)$ (see Figure 6.3). For sufficiently small δ we have that $p + y \in C$ for every $y \in D_\delta$. Since the trajectory x has the derivative $x'(t) = F(p)$, we have

$$|x(t + \tau) - (p + \tau F(p))| = o(\tau) .$$

Note that for small enough τ (depending on δ) the vector $\tau F(p)$ is $\Omega(\tau)$-far from the edge of D_δ. This implies that $x(t + \tau) - p \in D_\delta$, for small enough τ, and hence $x(t + \tau) \in C$.

Now we prove (iii). Pick an arbitrary $p \in C$ and let $x(t)$ be the solution of $(*)$ with the initial condition $x(0) = p$. Writing $T = \sup\{t \ge 0 : x([0, t]) \subset C\}$, we need to show $T = \infty$. Suppose the contrary, that is, $T < \infty$. Clearly, $x([0, T)) \subset C$, and $x(T)$ lies on

the boundary of C (as otherwise the trajectory would stay inside C a bit longer). By (ii) there exists $\tau > 0$ such that $x([0, T + \tau]) \subseteq C$ contradicting the maximality of T.

Finally, to prove (iv), fix $p \in C$. Since $C \subset S_2$, $g(p)$ is perpendicular to $h(p)$, and $|g(p)| = 1$, we get that

$$|F(p)| \overset{(6.16)}{\geq} |f(p)| - \varepsilon = \theta|g(p) + h(p)| - \varepsilon \geq \theta|g(p)| - \varepsilon = \theta - \varepsilon > 0 \,,$$

which proves (iv). □

Remark 6.15. *We stated the Poincaré–Bendixson theorem in an abridged form, sufficient for our purposes. Using the original version of the Poincaré–Bendixson theorem (see* [18, Theorem 1, p. 245]*) and Claim 6.14.*(iii)*, it actually follows that every trajectory starting in C is periodic or spirals towards a periodic orbit.*

We note that our proof of the existence of a periodic trajectory is not constructive, since the Poincaré–Bendixson theorem is not. We leave it as an open problem to find an explicit construction.

6.11. **Complicated trajectories.** Given a rule, a natural question about a rule and an initial graphon W is to understand which graphons a trajectory approaches infinitely often, namely to determine the set

$$\omega_+(W) := \{U \in \mathcal{W}_0 : \left\|\Phi^{t_n} W - U\right\|_\square \to 0 \text{ for some sequence } t_n \to \infty\}.$$

Convergent trajectories are simplest in the sense that for them $\omega_+(W)$ contains only $\mathsf{dest}\,(W)$. Even though the compactness of $\widetilde{\mathcal{W}_0}$ implies that every trajectory approximates some graphon U in the cut *distance* $\delta_\square$, we still do not have an answer to the following question.

Question 6.16. *Is it true that, for every rule $\mathcal{R}$, and every graphon W, set $\omega_+(W)$ is nonempty?*

We have seen examples where $\omega_+(W)$ is a continuum set, namely a periodic trajectory from Section 6.10 and a spiraling trajectory in Remark 6.15.

The following question has been suggested to us by Joel Spencer.

Question 6.17. *Does there exist a flip process and an initial graphon W so that the edge density oscillates between constant graphons 0 and 1, that is $\{0, 1\} \in \omega_+(W)$?*

What we find the most intriguing open question in this project is to find flip processes and graphons with really complicated trajectories. There is a topological way to express this (and it would be interesting to see alternative notions of complexity of trajectories). Recall that the metric $d_\square(U, W) = \|U - W\|_\square$ induces a non-compact topology on $\mathcal{W}_0$. We say that the trajectory starting at W is *complicated* if $\{\Phi^t W : t \geq 0\}$ is not totally

bounded, or equivalently, the closure of its orbit, that is, $\overline{\{\Phi^t W : t \geq 0\}}$ is not compact with respect to the metric $d_\square$. Then the main question is this.

Question 6.18. *Is there a flip process and a graphon W such that $\overline{\{\Phi^t W : t \geq 0\}}$ is not compact?*

A negative answer to Question 6.16 (which would be rather surprising) would give a positive answer to Question 6.18.

The proposition below excludes convergent trajectories from the search.

Proposition 6.19. *If a rule and a graphon W are such that $\mathsf{dest}(W)$ exists, then $\overline{\{\Phi^t W : t \geq 0\}}$ is compact.*

Proof. Let $\mathcal{R}$ be a rule of order k and W be the initial graphon with $U := \mathsf{dest}(W)$. We need to prove total boundedness, that is, for any given $\varepsilon > 0$ there exist finitely many $d_\square$-balls $B_1, \ldots, B_h$ in $\mathcal{W}_0$ of radius ε so that $\bigcup B_i \supset \{\Phi^t W : t \geq 0\}$. Since $U = \mathsf{dest}(W)$, there exists $t_0 \geq 0$ such that for each $t \geq t_0$, $d_\square(\Phi^t W, U) < \varepsilon$. Now, the appropriate balls have centers U and $\Phi^\tau W$ where τ ranges over multiples of ε/k^2 from 0 until (it exceeds) t_0. To prove the statement of the proposition, we need to argue that, for any $t \geq 0$, the graphon $\Phi^t W$ is covered by one of the balls above. If $t \geq t_0$, then $\Phi^t W$ is covered by the ball around U. If $t \in [0, t_0)$, then $\Phi^t W$ is covered by the ball around $\Phi^\tau W$ for the largest τ with $\tau \leq t$. To see this, recall that by (4.14), $\Phi^\tau W$ and $\Phi^t W$ at most $(k)_2(t - \tau) \leq \varepsilon$ apart in the L^∞-norm (and hence also in $d_\square$). □

The strongest possible (positive) answer to Question 6.18 (and also to Question 6.17) would be a trajectory whose closure covers the entire graphon space. However, there is no such trajectory.

Proposition 6.20. *There is no flip process and a graphon W such that $\overline{\{\Phi^t W : t \geq 0\}} = \mathcal{W}_0$.*

Proof. Suppose that such a flip process exists, say with a rule $\mathcal{R}$ of order k, and let W be a graphon whose closure of the trajectory covers $\mathcal{W}_0$.

Graphon W cannot have a set of twins of positive measure. Indeed, if it had, Proposition 6.5 and Lemma 3.8 would tell us that we can approximate only graphons with that set of twins, making it impossible to approximate graphons without a set of twins of positive measure. In particular,

(6.17) W is not a graphon representation of any k-vertex graph.

The following claim is the core of the proof.

Claim. For every $G \in \mathcal{H}_k$ we have $\mathcal{R}_{G,G} = 1$.

Proof of Claim. Suppose that this is not the case, that is, there exists a graph $H \in \mathcal{H}_k \setminus \{G\}$ such that $p := \mathcal{R}_{G,H} > 0$. Fix a pair $ij \in E(G) \ominus E(H)$. Let us assume that (*) $ij \notin E(G)$ and $ij \in E(H)$. The argument in the case (**) $ij \in E(G)$ and $ij \notin E(H)$ is symmetric and we will not include it. However, to clarify the adjustments needed and also for its own sake, let us give an overview of the proof of (*). We shall consider a graphon representation W_G of G with respect to a fixed partition $(\Omega_v)_{v\in[k]}$. Our focus will be on the square $\Omega_i \times \Omega_j$ where W_G is constant 0. The core argument (expressed in (6.19)) is that if the trajectory $U = \Phi^t W$ gets somewhat close to W_G then necessarily $\int_{\Omega_i\times\Omega_j} \mathfrak{V}\, U$ is positive (with an explicit positive lower bound) thus repelling from W_G on $\Omega_i \times \Omega_j$. Then inapproximability of W_G follows. Obviously, if we have (**) instead of (*) then the counterpart of (6.19) is a negative upper bound on $\int_{\Omega_i\times\Omega_j} \mathfrak{V}\, U$ showing that an ambition of having a trajectory approaching W_G must fail due to $(W_G)_{\restriction \Omega_i\times\Omega_j} = 1$.

Let W_G be the graphon representation of G with respect to an arbitrary fixed partition $(\Omega_v)_{v\in[k]}$. By (6.17) we have that $\Phi^0 W \neq W_G$. Also, as W_G is $\{0,1\}$-valued, we cannot have $\Phi^t W = W_G$ for any $t > 0$ by Proposition 6.7. In particular this means that there are times $t'_1 < t_1 < t'_2 < t_2 < \dots$ such that

$$d_\square(W_G, \Phi^{t'_\ell} W) > \frac{1}{4}, \quad d_\square(W_G, \Phi^{t_\ell} W) < \frac{1}{\ell}. \tag{6.18}$$

To see this, note that there are two $\{0,1\}$-valued graphons with distance more than $1/2$ between them (say, representation of complete and empty graphs); hence at least one of them, F, is such that $d_\square(W_F, W_G) > 1/4$; since W_F also does not lie on the trajectory, the trajectory has to alternate between small neighbourhoods of W_G and W_F.

Let $C_\square = C_\square(k, 0)$ be given by Lemma 4.10. Set $\delta := pk^{-k}/(2C_\square)$. We claim that for every graphon U with $d_\square(U, W_G) \le \delta$ we have

$$\int_{\Omega_i\times\Omega_j} \mathfrak{V}\, U \ge \frac{pk^{-k}}{2}\,. \tag{6.19}$$

To this end, first we compute $\mathfrak{V}\, W_G$, focusing on the value on the square $\Omega_i \times \Omega_j$ (on which $\mathfrak{V}\, W_G$ is constant 0). So, let us fix $x \in \Omega_i$, $y \in \Omega_j$ and look at the summands in the representation (4.4). Since $W_G(x, y) = 0$, each term in (4.4) is nonnegative. Keeping the term for $ab = ij$, we get (recall that $\mathbf{U}_v$s are independent π-random variables)

$$\begin{aligned}\mathfrak{V}\, W_G(x, y) &\ge \mathbb{P}\left(\mathbb{G}(k, W_G) = G \,|\, \mathbf{U}_i = x, \mathbf{U}_j = y\right) \mathcal{R}_{G,H} \\ &\ge \mathbb{P}\left(\mathbf{U}_v \in \Omega_v \forall v \in [k]\right) \mathcal{R}_{G,H} = k^{-k} p.\end{aligned}$$

By Lemma 4.10, $\int_{\Omega_i\times\Omega_j} U \ge \int_{\Omega_i\times\Omega_j} W_G - C_\square d_\square(U, W_G) \ge pk^{-k}/2$. This proves (6.19).

Let us show how (6.19) leads to a contradiction, thus proving the Claim. Letting τ_ℓ be the last time (see (6.18)) smaller than t_ℓ such that $d_\square(W_G, \Phi^\tau W) \geq \delta$, we have that

$$d_\square(\Phi^{\tau_\ell} W, \Phi^{t_\ell} W) \geq \delta - \frac{1}{\ell}\,,$$

and hence (4.14) (taking into account that $\|\cdot\|_\square \leq \|\cdot\|_\infty$) implies that $t_\ell - \tau_\ell \geq (\delta - 1/\ell)/(k)_2$. Starting with the integral form for the trajectory (see Remark 4.9), $\Phi^{t_\ell} W = \Phi^{\tau_\ell} W + \int_{\tau_\ell}^{t_\ell} \mathfrak{V}\, \Phi^t W \,\mathrm{d}t$ and making use of the fact that for each $t \in [\tau_\ell, t_\ell]$ we can apply (6.19) with $U = \Phi^t W$, we get

$$\begin{aligned} \int_{\Omega_i \times \Omega_j} \Phi^{t_\ell} W &= \int_{\Omega_i \times \Omega_j} (\Phi^{\tau_\ell} W) + \int_{\Omega_i \times \Omega_j} \left(\int_{\tau_\ell}^{t_\ell} \mathfrak{V}\, \Phi^t W \,\mathrm{d}t \right) \\ &\geq 0 + \frac{\delta - \frac{1}{\ell}}{(k)_2} \cdot \frac{pk^{-k}}{2}\,. \end{aligned} \tag{6.20}$$

To justify the last inequality, note that by definition of the integral (see Appendix C) we can approximate (in L^∞-norm and thus in the cut norm) $\int_{\tau_\ell}^{t_\ell} \mathfrak{V}\, \Phi^t W \,\mathrm{d}t$ by a linear combination of kernels taking values in $\{\mathfrak{V}\, \Phi^t W : t \in [\tau_\ell, t_\ell]\}$ and applying (6.19) to each of them.

By choosing ℓ large enough we get that $\int_{\Omega_i \times \Omega_j} \Phi^{t_\ell} W > \frac{1}{\ell}$. On the other hand, $ij \notin E(G)$ implies $\int_{\Omega_i \times \Omega_j} W_G = 0$. We conclude that

$$d_\square(\Phi^{t_\ell} W, W_G) \geq \int_{\Omega_i \times \Omega_j} \left(\Phi^{t_\ell} W - W_G\right) > 1/\ell,$$

contradicting (6.18). □

The Claim indeed proves the main statement, since by applying it to all graphs $G \in \mathcal{H}_k$ we deduce that the flip process in question is trivial. This is clearly a contradiction. □

Remark 6.21. *A strengthening of Proposition 6.20 can be proved for the factor space $\widetilde{\mathcal{W}}$ (recall Remark 3.1). Namely, there is no trajectory whose closure in the cut distance covers $\widetilde{\mathcal{W}}$. This follows by making minor changes to the proof of Proposition 6.20: in* (6.18) *we allow the graphon representation W_G to depend on ℓ.*

Remark 6.22. *In this section, we looked at the closure of the* forward *orbit $\overline{\{\Phi^t W : t \geq 0\}}$. As an alternative, we could look at the closure of the portion of the "whole" orbit $\overline{\{\Phi^t W : t \in I_W\}}$. Obviously, the qualitative difference between these two images can be only in the case $\mathsf{age}(W) = \infty$. Question 6.18 is also open in this alternative setting as is the question about non-emptiness of the set*

$$\omega_-(W) := \{U \in \mathcal{W}_0 : \left\|\Phi^{t_n} W - U\right\|_\square \to 0 \text{ for some sequence } t_n \to -\infty\}.$$

Nevertheless, a counterpart to Proposition 6.19 holds if we assume, in addition, that $\Phi^t W$ converges as $t \to -\infty$. Also, Proposition 6.20 holds even in this stronger setting. Indeed, the repulsion argument used to show the inapproximability of W_G holds uniformly even at negative times.

6.12. **Speed of convergence.** For now let us focus on convergent flip processes. For each such corresponding rule $\mathcal{R}$, we can define for each $\delta > 0$ *lower convergence time*

$$\tau^-_{\mathcal{R}}(\delta) = \sup\{\inf\{T : d_\square(\Phi^T W, \mathsf{dest}_{\mathcal{R}}(W)) < \delta\} : W \in \mathcal{W}_0\} ,$$

and the *upper convergence time*,

$$\tau^+_{\mathcal{R}}(\delta) = \sup\{\sup\{T : d_\square(\Phi^T W, \mathsf{dest}_{\mathcal{R}}(W)) \geq \delta\} : W \in \mathcal{W}_0\} .$$

Remark 6.23. *The definition of the cut norm as*

$$\|W\|_\square = \sup_{A,B\subseteq\Omega} \left| \int_{A\times B} W \,\mathrm{d}\pi^2 \right|$$

is not the only natural one, as sometimes it makes more sense to work only with

$$\|W\|_\square = \sup_{A\subseteq\Omega} \left| \int_{A\times A} W \,\mathrm{d}\pi^2 \right| \quad \text{or} \quad \|W\|_\square = \sup_{A,B\subseteq\Omega : A\cap B=\emptyset} \left| \int_{A\times B} W \,\mathrm{d}\pi^2 \right| .$$

As is well known (see, e.g., [8]*), these alternative definitions are off from the original one by a factor of at most 4. Thus, when pinning down the speed of convergence precisely, it matters which of these definitions is used.*

Remark 6.24. *In view of Conjecture 6.4, we might also define the speed using the L^1-distance.*

Below, we show that $\tau^-_{\mathcal{R}}(\delta)$ is finite for flip processes with a single destination.

Proposition 6.25. *If $\mathcal{R}$ is a convergent flip processes with a single destination, then for every $\delta \in (0, 1]$ we have $\tau^-_{\mathcal{R}}(\delta) < \infty$.*

Proof. Let U be the unique destination (and thus the unique fixed point). Hence by Proposition 6.10 graphon U is constant.

Suppose for contradiction that the statement does not hold. Then for each $\ell \in \mathbb{N}$ we have a graphon W_ℓ such that $\inf\{T : d_\square(\Phi^T W_\ell, U) < \delta\} > \ell$. By the Lovász–Szegedy compactness theorem, there exists a graphon Z which is an accumulation point of $W_1, W_2, \ldots$ in the cut distance. Without loss of generality, let us assume that Z is actually a cut distance limit of $W_1, W_2, \ldots$. Since the trajectory of Z converges to U, there exists a time T such that $d_\square(\Phi^T Z, U) < \delta/2$. Combining this with Theorem 4.16, for any W sufficiently close to Z in the cut norm, we have $d_\square(\Phi^T W, U) < \delta/2$. Since U is a constant graphon, by Section 6.3, the same conclusion holds for any W sufficiently close to Z in the cut distance. Taking $W = W_\ell$ for sufficiently large ℓ gives a contradiction. □

The same question is open for the quantity $\tau^+_{\mathcal{R}}(\delta)$.

Sometimes the convergence time is infinite for trivial reasons. For example if, given a rule $\mathcal{R}$, there is a graphon W with $\mathsf{age}(W) = \infty$ such that $\lim_{t\to-\infty} \Phi^t W$ exists and $\mathfrak{V}\, W \neq 0$, then one can make the convergence time arbitrarily long by choosing the initial graphon as $\Phi^{-t}W$ with t large enough.

Question 6.26. *Suppose that $\mathcal{R}$ is a convergent flip processes in which every graphon is either a fixed point or has finite age. Is it then true that for each $\delta \in (0, 1]$ we have $\tau^-_{\mathcal{R}}(\delta) < \infty$? Is the same conclusion true if, in addition, $\mathcal{R}$ has only finitely many destinations?*

We can go back to our motivating examples. Indeed, (1.1) tells us that for the Erdős–Rényi flip process (whose only destination is the constant-1 graphon) we have $\tau^+(\delta) = \ln(-\delta)/2$. The triangle removal flip process (generalised in Section 2.3) is much more interesting. In particular, in [2] a uniform bound on convergence time is derived using the Szemerédi's regularity lemma, thus giving extremely poor bounds.

It can be shown that for any flip process $\mathcal{R}$ with a single destination we have a lower bound $\tau^+(\delta) = \Omega(\ln(-\delta))$. Let us discuss the possibility of any general upper bounds. Note that there is no 'slowest process'. Indeed, we can slow down any rule $\mathcal{R}$ by any factor $C > 1$. That is, we define a new rule $\mathcal{R}'$ which, for any drawn graph H, replaces it with H with probability $1 - 1/C$ and replaces it with a graph sampled according to $\mathcal{R}$ with probability $1/C$. Still, this slowing down does not change the order of magnitude of the function $\tau^-_{\mathcal{R}}(\delta)$ as $\delta \to 0$. Hence, we could ask what is the slowest order of magnitude of convergence.

6.12.1. *Variants.* Many variants of these questions exist. For example, one might consider local versions of convergence times, such as follows. Suppose that we have a convergent flip process with a stable destination U of radius γ^*. Then define

$$\tau^-_{\mathcal{R}}(\delta, \gamma) = \sup\{\inf\{T : d_\square(\Phi^T W, U) < \delta\} : W \in \mathcal{W}_0, d_\square(W, U) < \gamma\}\,,$$

for $\gamma^* > \gamma > \delta > 0$.

6.13. **Uniqueness of the rule.** In an upcoming preprint, Hng [14] provides a complete characterisation of equivalence classes of flip process rules (of the same order) with the same trajectories. As an application, Hng demonstrates that the symmetric deterministic rules[h] and the rules of order 2 are precisely the rules which are unique in their equivalence

[h] A rule $\mathcal{R}$ of order k is *symmetric* if we have for each permutation ψ on $[k]$ (which naturally acts on $\mathcal{H}_k$) and each $H, J \in \mathcal{H}_k$ that $\mathcal{R}_{H,J} = \mathcal{R}_{\psi(H),\psi(J)}$. A rule $\mathcal{R}$ is *deterministic* if for each $H \in \mathcal{H}_k$ we have that $\mathcal{R}_{H,J} = 0$ for all except one $J \in \mathcal{H}_k$.

classes. The class of symmetric deterministic rules includes several examples from Section 2, namely the complementing rules, the component completion rules, the extremist rules, and the clique removal rules.

6.14. **Large deviations.** Our Theorem 5.1 provides a description of the typical evolution for flip processes. It would be interesting to work out a large deviation theory for flip processes as well. For flip processes of order 2 such a large deviation theory was worked out in [9], but in a variant with updates along a Poisson point process (cf. Section 5.2).[i]

Appendix A. Probability

We recall the Hoeffding–Azuma martingale concentration inequality including necessary preliminaries. To simplify some definitions we assume that the sample space Ω of the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is finite. Consider a sequence of sigma-algebras $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}_t$, such that $\mathcal{F}_i \subseteq \mathcal{F}$, which we call a *filtration*. We say that a sequence of random variables $Y_1, \ldots, Y_t$ is a *martingale difference sequence* if Y_i is $\mathcal{F}_i$-measurable and $\mathbb{E}(Y_i \,|\, \mathcal{F}_{i-1}) = 0$ for $i = 1, \ldots, t$.

The following lemma is a special case of Theorem 3.10 in [17].

Lemma A.1. *Let c be a constant. If $Y_1, \ldots, Y_t$ is a martingale difference sequence with $|Y_i| \leq c$ for each i, then for any $x \geq 0$,*

$$\mathbb{P}(|Y_1 + \cdots + Y_t| \geq x) \leq \exp\left(-\frac{x^2}{2tc^2}\right) .$$

Appendix B. Graphons

Here we include some more facts about graphons and the proofs for Section 3.2.

The next two facts give alternative ways of bounding the cut norm. Variants appear in the literature, but we provide proofs for completeness.

Fact B.1 (Inequality (7.1) in [8])**.** *For each $W \in \mathcal{W}$, we have*

$$\|W\|_\square \leq 2 \sup_{A \subseteq \Omega} \left| \int_{A \times A} W \, \mathrm{d}\pi^2 \right| .$$

[i] As we explain in Section 5.2 there is essentially no difference between updates at integer times and along a Poisson point process for the typical behaviour. However, there usually is a difference between these two regimes from the point of large deviations.

Proof. To bound $\|W\|_\square$, let A and B be arbitrary as in (3.2). Using symmetry of W,

$$\begin{aligned}\left|\int_{A\times B} W\right| &= \frac{1}{2}\left|\int_{A\times B} W + \int_{B\times A} W\right| \\ &= \frac{1}{2}\left|\int_{(A\cup B)^2} W + \int_{(A\cap B)^2} W - \int_{(A\setminus B)^2} W - \int_{(B\setminus A)^2} W\right| \\ &\le \frac{1}{2}\cdot 4 \sup_{C\subseteq\Omega}\left|\int_{C\times C} W\right| .\end{aligned}$$

□

Fact B.2. *Suppose that $\mathcal{A}$ is a partition of Ω into finitely many sets. For each $W \in \mathcal{W}$ which is constant on each cell of $\mathcal{A}\times\mathcal{A}$ we have*

$$\|W\|_\square \le \sup_{S,T\subseteq\mathcal{A}}\left|\int_{\bigcup S\times\bigcup T} W\,\mathrm{d}\pi^2\right| . \tag{B.1}$$

Proof. To bound $\|W\|_\square$, let A and B be arbitrary as in (3.2). We shall prove that for an arbitrary $D\in\mathcal{A}$ we have

$$\left|\int_{A\times B} W\right| \le \max\left\{\left|\int_{(A\cup D)\times B} W\right|, \left|\int_{(A\setminus D)\times B} W\right|\right\} . \tag{B.2}$$

A similar 'rounding' statement could be obtained for B as well. Once this is done, (B.1) follows by rounding each of A and B on each cell of $\mathcal{A}$.

Let $\alpha := \pi(A\cap D)/\pi(D)$. Since $x\mapsto\int_{y\in B} W(x,y)$ is constant on D, we have

$$\int_{A\times B} W = \alpha\cdot\int_{(A\cup D)\times B} W \;+\; (1-\alpha)\cdot\int_{(A\setminus D)\times B} W .$$

Since $\alpha\in[0,1]$, inequality (B.2) follows from the fact that $x\mapsto|x|$ is a convex function. □

Here, we give a self-contained proof of Lemma 3.2. This proof was suggested to us by Jan Grebík and Guus Regts.

Proof of Lemma 3.2. The proof shall use the concept of weak* convergence of graphons (which we treat as elements of $L^\infty(\Omega^2)$ with a predual space $L^1(\Omega^2)$), which we quickly recall. We refer the reader to [13] for details. A sequence of graphons $W_1, W_2, \ldots$ *converges weak** to a graphon W if for every $S,T\subset\Omega$ and every $\varepsilon>0$ there exists a number $N_{S,T,\varepsilon}\in\mathbb{N}$ such that

$$\left|\int_{S\times T} W_n - \int_{S\times T} W\right| < \varepsilon \quad \text{for every } n\ge N_{S,T,\varepsilon} . \tag{B.3}$$

So, the cut norm topology can be regarded as a certain uniformisation of the weak* topology, and indeed convergence in the former topology implies convergence in the latter.

It follows from the Banach–Alaoglu theorem that the closed unit ball in $L^\infty(\Omega^2)$ is compact with respect to the weak* topology. Since the set of graphons $\mathcal{W}_0$ is contained in this unit ball and is closed with respect to the L^∞ norm (and therefore closed in the weak* topology), $\mathcal{W}_0$ is compact in the weak* topology, too.

We are now ready to prove the lemma. Suppose that $W_1, W_2, \dots$ is a Cauchy sequence in $d_\square$. That is, for every $\varepsilon > 0$ there exists an $M_\varepsilon \in \mathbb{N}$ such that

$$\sup_{S,T\subset\Omega} \left| \int_{S\times T} W_n - \int_{S\times T} W_m \right| < \varepsilon \quad \text{for every } n, m \geq M_\varepsilon \,. \tag{B.4}$$

From compactness of $\mathcal{W}_0$ and (B.4) it easily follows that (W_n) converges weak* to some graphon W. We claim that W is the sought cut norm limit. Suppose that $\varepsilon > 0$ is arbitrary. We then have that for $n \geq M_\varepsilon$

$$\begin{aligned} d_\square(W_n, W) &= \sup_{S,T\subset\Omega} \left| \int_{S\times T} W_n - \int_{S\times T} W \right| \\ &\leq \sup_{S,T\subset\Omega} \left\{ \left| \int_{S\times T} W_n - \int_{S\times T} W_{\max\{M_\varepsilon, N_{S,T,\varepsilon}\}} \right| \right. \\ &\quad \left. + \left| \int_{S\times T} W_{\max\{M_\varepsilon, N_{S,T,\varepsilon}\}} - \int_{S\times T} W \right| \right\}, \end{aligned}$$

where the terms $N_{S,T,\varepsilon}$ are defined as in (B.3). From (B.3) and (B.4) it follows that $d_\square(W_n, W) \leq \varepsilon + \varepsilon$, which implies $\lim_{n\to\infty} d_\square(W_n, W) = 0$, as was needed. □

Next, we turn to proving Lemma 3.5. We will make use of the following fact.

Fact B.3. *Given a real number $M > 0$, any numbers $a_1, \dots, a_\ell, b_1, \dots, b_\ell$ of modulus at most M satisfy the inequality*

$$\left| \prod_i a_i - \prod_i b_i \right| \leq M^{\ell-1} \sum_i |a_i - b_i| \,.$$

Proof. To apply induction on ℓ, write $A = \prod_{i\leq \ell-1} a_i$, $B = \prod_{i\leq \ell-1} b_i$, so that the left-hand-side is $|Aa_\ell - Bb_\ell|$. We have $Aa_\ell - Bb_\ell = A(a_\ell - b_\ell) + (A - B)b_\ell$. Hence, by the triangle inequality and the induction hypothesis

$$\begin{aligned} |Aa_\ell - Bb_\ell| &\leq |A||a_\ell - b_\ell| + |A - B||b_\ell| \\ &\leq M^{\ell-1}|a_\ell - b_\ell| + |A - B|M \\ &\leq M^{\ell-1} \sum_{i=1}^{\ell} |a_i - b_i|, \end{aligned}$$

as was needed. □

Proof of Lemma 3.5. This follows by fixing (x, y), writing $(T_F^{a,b}U)(x, y)$ and $(T_F^{a,b}W)(x, y)$ as integrals (see (3.6)) with the same variables of integration, applying the inequality $|\int f| \le \int |f|$ and then applying Fact B.3 to the integrand. (Note that $W(x_i, x_j)$ and $1 - W(x_i, x_j)$ have modulus at most $1 + \varepsilon$ because of $W \in \mathcal{W}_0^\varepsilon$.) □

We want to move on to prove Lemma 3.7. For this we need to cite a counting lemma from [16, Lemma 10.24].

Lemma B.4 (Counting lemma for vectors of graphons)**.** *Given a simple graph $F = (V, E)$ and two tuples $\mathbf{U} = (U_e : e \in E)$ and $\mathbf{W} = (W_e : e \in E)$ of graphons, we have*

$$|t(F, \mathbf{U}) - t(F, \mathbf{W})| \le \sum_{e \in E} \|U_e - W_e\|_\square .$$

Proof of Lemma 3.7. Every $W \in \mathcal{W}_0^\varepsilon$ can be written as $W = -\varepsilon + (1+2\varepsilon)\hat{W} = (1+\varepsilon)\hat{W} - \varepsilon(1 - \hat{W})$ for some graphon $\hat{W} \in \mathcal{W}_0$. We have, always assuming that the arguments of W_{ij} and $\hat{W}_{ij}$ are (x_i, x_j),

$$\begin{aligned}
t(F, \mathbf{W}) &= \int \prod_{e \in E} W_e \, \mathrm{d}\pi^V \\
&= \int \prod_{e \in E} \Big((1+\varepsilon)\hat{W}_e - \varepsilon(1 - \hat{W}_e) \Big) \, \mathrm{d}\pi^V \\
&= \sum_{S \subseteq E} (1+\varepsilon)^{|S|} (-\varepsilon)^{|E|-|S|} \int \prod_{e \in S} \hat{W}_e \prod_{e \in E \setminus S} (1 - \hat{W}_e) \, \mathrm{d}\pi^V \\
&= \sum_{S \subseteq E} (1+\varepsilon)^{|S|} (-\varepsilon)^{|E|-|S|} t(F, \hat{\mathbf{W}}_S) \, ,
\end{aligned}$$

where we defined $\hat{\mathbf{W}}_S = (\hat{W}_{S,e} : e \in E)$ by setting $\hat{W}_{S,e} = \hat{W}_e$ if $e \in S$ and $\hat{W}_{S,e} = 1 - \hat{W}_e$ otherwise.

Analogously defining $\hat{\mathbf{U}}_S$ and expressing $t(F, \mathbf{U})$, it follows that

$$\begin{aligned}
|t(F, \mathbf{U}) - t(F, \mathbf{W})| &\le \sum_{S \subseteq E} (1+\varepsilon)^{|S|} \varepsilon^{|E|-|S|} \Big| t(F, \hat{\mathbf{U}}_S) - t(F, \hat{\mathbf{W}}_S) \Big| \\
\boxed{\text{Lemma B.4}} \quad &\le \sum_{S \subseteq E} (1+\varepsilon)^{|S|} \varepsilon^{|E|-|S|} \sum_{e \in E} \Big\| \hat{U}_{S,e} - \hat{W}_{S,e} \Big\|_\square \\
\boxed{\hat{U}_{S,e} - \hat{W}_{S,e} = (U_e - W_e)/(1+2\varepsilon)} \quad &= \sum_{S \subseteq E} (1+\varepsilon)^{|S|} \varepsilon^{|E|-|S|} \cdot \frac{\sum_{e \in E} \|U_e - W_e\|_\square}{1+2\varepsilon} \\
\boxed{\text{binomial theorem}} \quad &= (1+2\varepsilon)^{|E|-1} \sum_{e \in E} \|U_e - W_e\|_\square .
\end{aligned}$$

□

Lastly, we add the proof of Lemma 3.8.

Proof of Lemma 3.8. Fix a measurable $T \subset \Omega$ and let $f_n(x) = \int_T U_n(x,\cdot)\,\mathrm{d}\pi$ and $f(x) = \int_T U(x,\cdot)\,\mathrm{d}\pi$. Given a set of positive measure B and $g \in L^\infty(\Omega)$, let us write $\mathbb{E}_B g = \frac{1}{\pi(B)}\int_B g\,\mathrm{d}\pi$. Since A is a twin-set for U_n, for every $B \subset A$ of positive measure we have

$$\mathbb{E}_B f_n = \frac{1}{\pi(B)}\int_{B\times T} U_n = \frac{1}{\pi(A)}\int_{A\times T} U_n = \mathbb{E}_A f_n.$$

Since $U_n \to U$ in the cut norm, this implies that

$$\mathbb{E}_B f = \frac{1}{\pi(B)}\int_{B\times T} U = \frac{1}{\pi(A)}\int_{A\times T} U = \mathbb{E}_A f. \tag{B.5}$$

We claim that f is constant a.e. on A. Let $\varepsilon > 0$ be arbitrary. Let

$$B = \{x \in A : f(x) \geq \varepsilon + \mathbb{E}_A f\}\ .$$

If $\pi(B) > 0$, then $\mathbb{E}_B f \geq \varepsilon + \mathbb{E}_A f$, contradicting (B.5). Since ε was arbitrary, $f \leq \mathbb{E}_A f$ a.e., and by a similar argument $f \geq \mathbb{E}_A f$.

Recalling that the sigma-algebra of Ω is separable, let $\mathcal{C}$ be a countable set of measurable sets which approximate any measurable $S \subset \Omega$. Writing $f_T(x) = \int_T U(x,\cdot)\,\mathrm{d}\pi$, by what we showed above, for every $T \in \mathcal{C}$ we have $f_T = c_T := \mathbb{E}_A f_T$ a.e. Hence the set $A' := \{x \in A : \forall T \in \mathcal{C}\ f_T(x) = c_T\}$ is a conull subset of A.

Fix $x, y \in A'$ and $\varepsilon > 0$ and let $B_\varepsilon = \{z \in \Omega : U(x,z) - U(y,z) \geq \varepsilon\}$. For any $T \in \mathcal{C}$ we have

$$\varepsilon\pi(B_\varepsilon) \leq \int_{B_\varepsilon} (U(x,\cdot) - U(y,\cdot))\,\mathrm{d}\pi \leq \pi(B_\varepsilon \ominus T)\cdot 2\,\|U\|_\infty + \int_T (U(x,\cdot) - U(y,\cdot))\,\mathrm{d}\pi\ .$$

Since $\int_T (U(x,\cdot) - U(y,\cdot)) = f_n(x) - f_n(y) = 0$ and $\pi(B_\varepsilon \ominus T)$ can be made arbitrarily small, we get that $\pi(B_\varepsilon) = 0$, that is, $U(x,\cdot) \leq U(y,\cdot) + \varepsilon$ a.e. Since x, y, ε were arbitrary, this implies that A is a twin-set for U. □

Appendix C. Calculus of normed-space-valued functions

Let $\mathcal{E}$ be a Banach space with norm $\|\cdot\|$.

Definition C.1. *Suppose that a function $f : I \to \mathcal{E}$ is defined on some subset $I \subset \mathbb{R}$. If there is an element $z \in \mathcal{E}$ such that*

$$\lim_{I \ni \varepsilon \to 0} \left\| \frac{f(t+\varepsilon) - f(t)}{\varepsilon} - z \right\| = 0,$$

then we say that f is differentiable at t *and call $\frac{\mathrm{d}f}{\mathrm{d}t}(t) = f'(t) = z$ the* derivative of f at t.

For any $a, b \in \mathbb{R}$, let $\mathcal{S}([a,b],\mathcal{E})$ be the set of *step functions* $f : [a,b] \to \mathcal{E}$, that is, functions such that f is fixed on each set of some finite partition of $[a,b]$ into intervals $(P_1, \dots, P_n)$. For a step function f, the integral $\int_a^b f$ is an element of $\mathcal{E}$ defined by

$$\int_a^b f = \sum_i |P_i| f_i, \quad \text{where } f(t) = f_i \text{ on } P_i.$$

Considering the space of bounded functions from $[a,b]$ to $\mathcal{E}$ with uniform norm $\|f\|_\infty := \sup_{t\in[a,b]} \|f(t)\|$, it is easy to show that any continuous function $f : [a,b] \to \mathcal{E}$ can be approximated by step functions (see [1, p. 54]), that is, there is a sequence $f_n \in \mathcal{S}([a,b],\mathcal{E})$ such that $\lim_{n\to\infty} \|f - f_n\|_\infty = 0$. Moreover, for any such approximating sequence the elements $\int_a^b f_n$ converge to the same element of $\mathcal{E}$ which we denote by $\int_a^b f$ and call the *Cauchy–Bochner integral of f over* $[a,b]$. This integral satisfies the following properties

$$\left\| \int_a^b f \right\| \le \int_a^b \|f\| \le |b-a| \, \|f\|_\infty , \tag{C.1}$$

$$\int_a^b f = \int_a^c f + \int_c^b f \quad \text{with a convention that} \quad \int_a^b f := -\int_b^a f. \tag{C.2}$$

For proofs below, whenever we approximate a continuous function f by a step function $f^{(n)}$, $n \in \mathbb{N}$, we let $[a,b] = P_1 \cup \dots \cup P_n$ be a partition into intervals of equal lengths and assume that $f^{(n)}$ on P_i takes some (constant) value which f takes at some point in P_i. We may occasionally choose this value not arbitrarily; say if $f(t)$ has some property for almost every t, then we may want $f^{(n)}$ to inherit this property and to achieve this we will choose the value in P_i avoiding the exceptional measure-zero set for t.

Since f is continuous on a compact interval, it is also uniformly continuous. Hence the sequence $(f^{(n)})_n$ converges to f in the uniform norm implying that $\left(\int f^{(n)}\right)_n$ converges to $\int f$.

We now state the fundamental theorem of calculus for normed space-valued functions.

Theorem C.2 (Theorem 2.3.11 in [1])**.** *If $f : [a,b] \to \mathcal{E}$ is continuous, differentiable on (a,b) and if f' extends to a continuous function on $[a,b]$, then*

$$f(b) - f(a) = \int_a^b f' \, .$$

Further, given $f, g \in L^\infty(\Omega^2)$ we write $f \ge g$ meaning $f \ge g$ a.e.

Proposition C.3. *Given a continuous function $f : [a,b] \to L^\infty(\Omega^2)$ such that for every t we have $f(t) \ge 0$, we have that*

$$\int_a^b f(t) \, \mathrm{d}t \ge 0 \, .$$

Proof. Let f_n be the explicit step functions defined above so that

$$I_n := \int_a^b f_n(t)\,\mathrm{d}t \xrightarrow{\infty} \int_a^b f(t)\,\mathrm{d}t =: I.$$

Since f_n takes values in the range of f, for every t we have that $f_n(t) \geq 0$ and therefore $I_n \geq 0$. Assume for contradiction that $I \leq -\varepsilon$ on a set A of measure $\varepsilon > 0$. Then

$$-\varepsilon^2 \geq \int_A I = \int_A I_n + \int_A (I_n - I) \geq 0 - \|I_n - I\|_1 \geq -\|I_n - I\|_\infty \to 0,$$

giving a contradiction. □

The following fact says that for L^∞-valued integrals one can calculate the integral pointwise.

Proposition C.4. *Let $\mathcal{E}$ be the normed space $L^\infty(\Omega^2)$. Consider a continuous function $[a,b] \ni t \mapsto f_t \in \mathcal{E}$. One can modify, for every $t \in [a,b]$, f_t on a set of measure zero (that possibly depends on t) so that for every $(x,y) \in \Omega^2$ function $t \mapsto f_t(x,y)$ is continuous (and thus integrable) and the function $J \in \mathcal{E}$ defined by $J(x,y) = \int_a^b [f_t(x,y)]\,\mathrm{d}t$ represents the same element of $\mathcal{E}$ as $I := \int_a^b f_t\,\mathrm{d}t$.*

Furthermore, if we have continuous functions $t \mapsto f_t$ and $t \mapsto g_t$ such that $f_t \geq g_t$ for every $t \in [a,b]$, then each of them can be modified to satisfy the conclusion above and also $f_t(x,y) \geq g_t(x,y)$ for every $(x,y) \in \Omega^2$ and $t \in [a,b]$.

Proof. Since $[a,b]$ is compact, $f.$ is uniformly continuous, and so there is a function ω such that $\omega(x) \to 0$, $x \to 0$, for which

$$\|f_{t_1} - f_{t_2}\|_\infty \leq \omega(|t_1 - t_2|) \quad \text{for every } t_1, t_2 \in [a,b]. \tag{C.3}$$

Let us choose $D_1 \subset \Omega^2$ of full measure such that

$$\sup_{(x,y)\in D_1} |f_{t_1}(x,y) - f_{t_2}(x,y)| \leq \omega(|t_1 - t_2|) \text{ for every rational } t_1, t_2 \in [a,b]. \tag{C.4}$$

For $(x,y) \in D_1$, the function $t \mapsto f_t(x,y)$ is continuous on $[a,b] \cap \mathbb{Q}$, so we can extend it to a continuous function $t \mapsto \tilde{f}_t(x,y)$ defined on the whole interval $[a,b]$. For $(x,y) \in \Omega^2 \setminus D_1$ and $t \in [a,b]$, we define $\tilde{f}_t(x,y) = 0$, say. We claim that $\left\|f_t - \tilde{f}_t\right\|_\infty = 0$ for every $t \in [a,b]$. By the triangle inequality, for any rational $\tau \in [a,b]$,

$$\begin{aligned}\left\|f_t - \tilde{f}_t\right\|_\infty &\leq \|f_t - f_\tau\|_\infty + \left\|f_\tau - \tilde{f}_\tau\right\|_\infty + \left\|\tilde{f}_\tau - \tilde{f}_t\right\|_\infty \\ &\leq |t-\tau| + 0 + |\tau - t| = 2|t-\tau|.\end{aligned}$$

By sending τ to t the claim follows, implying that for every $t \in [a,b]$, $f_t = \tilde{f}_t$ almost everywhere.

We have $\int f_t \,\mathrm{d}t = \int \tilde{f}_t \,\mathrm{d}t$, since f_t and $\tilde{f}_t$ are the same element of $\mathcal{E}$. We declare $\tilde{f}_t$ to be the claimed modification of f_t, for every t. For convenience let us further drop the tilde and write f_t instead of $\tilde{f}_t$. By construction, we have that

$$\forall x, y \in \Omega \; \forall t_1, t_2 \in [a,b] \quad |f_{t_1}(x,y) - f_{t_2}(x,y)| \le \omega(|t_1 - t_2|). \tag{C.5}$$

In particular, $t \mapsto f_t(x,y)$ is continuous, so $J(x,y) := \int_a^b f_t(x,y)\,\mathrm{d}t$ is well defined for *every* $(x,y) \in \Omega^2$ (on $\Omega^2 \setminus D_1$ we have $J(x,y) = 0$). It remains to show that $\|I - J\|_\infty = 0$.

Let I be some bounded measurable function $I : \Omega^2 \to \mathbb{R}$ that represents $\int_a^b f_t \,\mathrm{d}t$. Let $f_t^{(n)}$ be a n-step function that approximates f_t (recalling the convention we made above), let $I_n(x,y) := (\int f_t^{(n)} \,\mathrm{d}t)(x,y) = \int f_t^{(n)}(x,y)\,\mathrm{d}t$, where the equality follows from the definition of the integral for the step function $f_t^{(n)}$.

By (C.1) and (C.3) we get

$$\begin{aligned} \|I_n - I\|_\infty &\le \sum_{i\in[n]} \left\| \int_{P_i} (f_t^{(n)} - f_t)\,\mathrm{d}t \right\|_\infty \\ &\le \sum_{i\in[n]} |P_i| \sup_{t\in P_i} \left\| f_t^{(n)} - f_t \right\|_\infty \\ &\le \sum_{i\in[n]} |P_i| \omega(|P_i|) = (b-a)\omega\left(\frac{b-a}{n}\right) =: d_n. \end{aligned} \tag{C.6}$$

We define a set

$$D_2 := \left\{ (x,y) \in \Omega^2 : \forall n \in \mathbb{N} \;\; |I_n(x,y) - I(x,y)| \le d_n \right\},$$

which by (C.6) has full measure.

Using (C.5), we have, for every $x, y \in \Omega$,

$$\begin{aligned} |I_n(x,y) - J(x,y)| &= \left| \int_a^b [f_t^{(n)}(x,y) - f_t(x,y)]\,\mathrm{d}t \right| \\ &\le \sum_{i\in[n]} |P_i| \sup_{t\in P_i} |f_t^{(n)}(x,y) - f_t(x,y)| \\ &\le (b-a)\omega\left(\frac{b-a}{n}\right) = d_n. \end{aligned}$$

We get that for $(x,y) \in D_2$

$$\begin{aligned} &|I(x,y) - J(x,y)| \\ &\le |I(x,y) - I_n(x,y)| + |I_n(x,y) - J(x,y)| \\ &\le 2d_n. \end{aligned}$$

Since $d_n \to 0$ as $n \to \infty$, we obtain $I = J$ on D_2 and thus $I = J$ as elements of $\mathcal{E}$.

To prove the "furthermore" part, we modify the set D_1 we chose at the beginning of the proof (still maintaining that it has full measure) so that (C.4) is satisfied for both $f.$ and $g.$, while, in addition, for every rational t we have $f_t(x,y) \geq g_t(x,y)$ for $(x,y) \in D_1$. Then we notice that the continuous extensions of $t \mapsto f_t(x,y)$ and $t \mapsto g_t(x,y)$ satisfy $\tilde{f}_t(x,y) \geq \tilde{g}_t(x,y)$ for $(x,y) \in D_1$ and *every* t. Since we defined $\tilde{f}_t = \tilde{g}_t = 0$ outside of D_1, the inequality everywhere follows. □

4. Acknowledgements

We thank Hanka Řada, Joel Spencer and Lutz Warnke for their comments. We also thank an anonymous referee for their feedback.

References

[1] R. Abraham, J. E. Marsden, and T. Ratiu. *Manifolds, tensor analysis, and applications*, volume 75 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1988.

[2] P. Araújo, J. Hladký, E. K. Hng, and M. Šileikis. Prominent examples of flip processes. arXiv:2206.03884.

[3] S. Athreya, F. den Hollander, and A. Röllin. Graphon-valued stochastic processes from population genetics. *Ann. Appl. Probab.*, 31(4):1724–1745, 2021.

[4] T. Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.*, 5:80–145, 2008.

[5] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[6] T. Bohman, A. Frieze, and E. Lubetzky. Random triangle removal. *Adv. Math.*, 280:379–438, 2015.

[7] T. Bohman and P. Keevash. The early evolution of the H-free process. *Invent. Math.*, 181(2):291–336, 2010.

[8] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[9] P. Braunsteins, F. den Hollander, and M. Mandjes. A sample-path large deviation principle for dynamic erdős-rényi random graphs. arXiv:2009.12848.

[10] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[11] H. Crane. Dynamic random networks and their graph limits. *Ann. Appl. Probab.*, 26(2):691–721, 2016.

[12] P. Diaconis and S. Janson. Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)*, 28(1):33–61, 2008.

[13] M. Doležal, J. Grebík, J. Hladký, I. Rocha, and V. Rozhoň. Relating the cut distance and the weak* topology for graphons. *J. Combin. Theory Ser. B*, 147:252–298, 2021.

[14] E. K. Hng. Characterisation of flip process rules with the same trajectories. manuscript.

[15] D. Keliger. Markov processes on quasi-random graphs. arXiv:2105.13446.

[16] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[17] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.

[18] L. Perko. *Differential Equations and Dynamical Systems*, volume 7 of *Texts in Applied Mathematics*. Springer-Verlag, New York, third edition, 2001.

[19] G. Regts and A. Schrijver. Compact orbit spaces in Hilbert spaces and limits of edge-colouring models. *European J. Combin.*, 52(part B):389–395, 2016.

[20] O. Riordan and L. Warnke. Achlioptas process phase transitions are continuous. *Ann. Appl. Probab.*, 22(4):1450–1464, 2012.

[21] M. Royden and P. M. Fitzpatrick. *Real analysis*. New York, NY: Prentice Hall, 2010.

[22] Endre Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

[23] T. Tao. *Nonlinear dispersive equations*, volume 106 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2006. Local and global analysis.

[24] G. Teschl. *Ordinary differential equations and dynamical systems*, volume 140 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[25] Andrew Thomason. Pseudorandom graphs. In *Random graphs '85 (Poznań, 1985)*, volume 144 of *North-Holland Math. Stud.*, pages 307–331. North-Holland, Amsterdam, 1987.

[26] Nicholas C Wormald. Differential equations for random processes and random graphs. *Ann. Appl. Probab.*, pages 1217–1235, 1995.

Faculty of Informatics, Masaryk University, Brno, Czechia

Institute of Computer Science of the Czech Academy of Sciences, Prague, Czechia

Institute of Computer Science of the Czech Academy of Sciences, Prague, Czechia

Department of Mathematics, Uppsala University, Uppsala, Sweden

PROMINENT EXAMPLES OF FLIP PROCESSES

PEDRO ARAÚJO, JAN HLADKÝ, ENG KEAT HNG, AND MATAS ŠILEIKIS

Abstract. Flip processes, introduced in [Garbe, Hladký, Šileikis, Skerman: From flip processes to dynamical systems on graphons], are a class of random graph processes defined using a rule which is just a function $\mathcal{R} : \mathcal{H}_k \to \mathcal{H}_k$ from all labelled graphs of a fixed order k into itself. The process starts with an arbitrary given n-vertex graph G_0. In each step, the graph G_i is obtained by sampling k random vertices $v_1, \dots, v_k$ of G_{i-1} and replacing the induced graph $G_{i-1}[v_1, \dots, v_k]$ by $\mathcal{R}(G_{i-1}[v_1, \dots, v_k])$.

Using the formalism of dynamical systems on graphons associated to each such flip process from ibid. we study several specific flip processes, including the triangle removal flip process and its generalizations, 'extremist flip processes' (in which $\mathcal{R}(H)$ is either a clique or an independent set, depending on whether $e(H)$ has less or more than half of all potential edges), and 'ignorant flip processes' in which the output $\mathcal{R}(H)$ does not depend on H.

1. Introduction

1.1. **Random graph processes.** Graphs are among the most universal structures, and they have therefore served as mathematical models in many settings. It is often the case that these settings evolve in time according to a stochastic rule. A variety of random graph processes have been introduced and studied extensively to this end. Most of these processes are defined so that they start with a specific initial graph on vertex set $[n]$ (for each $n \in \mathbb{N}$), and this vertex set remains unchanged. Further, it is usually asymptotic properties as n tends to infinity which are investigated. Let us give some examples. The simplest such process is the *Erdős–Rényi graph process*, introduced in the seminal paper [8] of Erdős and Rényi in 1960. The process starts with the edgeless graph G_0 on $[n]$. In each step $t = 1, \dots, \binom{n}{2}$, the graph G_t is obtained by taking G_{t-1} and turning a uniformly random non-edge into an edge. Obviously, the graph $G_{\binom{n}{2}}$ is complete. Another process introduced by Bollobás and Erdős at a "conference in Fairbanks, Alaska, in 1990, while admiring the playful bears" [6], is the *triangle removal process*. The process starts with the complete graph G_0 on $[n]$. In each step $t = 1, 2, \dots$, the graph G_t is obtained by taking G_{t-1}, choosing a random triangle (uniformly among the set of all triangles in G_{t-1}), and erasing its three edges. The process stops at the (random) time when G_t is triangle free. A major question raised already by Bollobás and Erdős in 1990 was what the number of edges of the final graph G_t is. This has been a subject of a long line of research, culminating in [3] which establishes that $e(G_t) = n^{3/2+o(1)}$. General H-removal processes were studied much

With institutional support RVO:67985807. JH and EKH were supported by the Czech Science Foundation, grant number 21-21762X; PA and MŠ were supported by the Czech Science Foundation, grant number 20-27757Y.

less, mostly motivated by extending the setting of the Erdős–Hanani conjecture resolved in [25], see e.g. [22, 26]. A process which looks like a dual process in many aspects is the following one, called the *H-free process*. Let H be a fixed graph. We start with the edgeless graph G_0. In each step t we select, uniformly at random, a non-edge whose addition to G_{t-1} does not create a copy of H, and turn it into an edge. Hence, when the process stops, G_t is a maximal H-free graph. The case $H = K_3$ is now well understood due to analyses by Bohman and Keevash [5] and Fiz Pontiveros, Griffiths, and Morris [10]. In particular, these results imply the currently best known lower bound for the off-diagonal Ramsey number $R(3, k)$. Among the numerous further papers studying the triangle-free process and other H-free processes, let us highlight [21, 4, 28]. The last class of processes we mention before turning to the main subject of this paper are the *Achlioptas processes*, introduced in [1]. We start with the edgeless graph on $[n]$. In each step we select $k = 2$ random non-edges (natural generalizations for larger k exist) $f_1 = x_1y_1$ and $f_2 = x_2y_2$. Then, we choose one of them to add into the graph according to a predetermined rule. For example, the so-called 'product rule' selects the non-edge f_i which minimizes the product of the sizes of the connected components to which x_i and y_i belong. Indeed, the motivation behind many Achlioptas rules was to alter the time of the emergence of the giant component. See, e.g. [23, 24], for more on Achlioptas processes.

Let us note that an obvious generalization of the processes above would be to start them from graphs other than the complete or the edgeless one. It seems that the reason such generalizations were not studied was that it was not clear what the observed features should be.

1.2. **Flip processes, basics.** In [12], the following general class of random graph processes was introduced. Let $k \in \mathbb{N}$ be given and let $\mathcal{H}_k$ be the set of all graphs on the vertex set $[k]$. Let $\mathcal{R} \in [0,1]^{\mathcal{H}_k \times \mathcal{H}_k}$ be a matrix which is stochastic, that is, for every $F \in \mathcal{H}_k$ we have $\sum_{H \in \mathcal{H}_k} \mathcal{R}_{F,H} = 1$. We call k the *order* and $\mathcal{R}$ the *rule* of the flip process. Given $\mathcal{R}$, an integer $n \geq k$ and an *initial graph* G_0 with vertex set $[n]$, the corresponding *flip process* proceeds as follows. For each $t \in \mathbb{N}$, we sample uniformly at random an ordered tuple $\mathbf{v} = (v_1, \cdots, v_k)$ of distinct vertices in $[n]$. Let $F \in \mathcal{H}_k$ be the *drawn graph*, that is, the graph on $[k]$ such that $ij \in E(F)$ if and only if $v_iv_j \in G_{t-1}$. Next, we sample a graph H, which we term the *replacement graph*, according to the probability distribution $(\mathcal{R}_{F,H})_H$ and replace $G_{t-1}[\mathbf{v}]$ by H in the graph G_{t-1} to obtain G_t. To be precise, we place a copy of H on $\mathbf{v}$ so that $ij \in E(H)$ if and only if $v_iv_j \in G_t$. A flip process is a discrete time Markov process. It is (unlike all the previously mentioned processes) defined for all times.

We observe that the Erdős–Rényi process and the triangle removal process have flip process variants. The *Erdős–Rényi flip process* starts with an arbitrary graph on $n \geq 2$ vertices. In each step, a random pair of vertices is chosen and turned into an edge. If the

pair is already an edge then the step is idle. The *triangle removal flip process* starts with an arbitrary graph on $n \geq 3$ vertices. In each step, a random triple of vertices is chosen. If the triple forms a triangle then its three edges are removed. Otherwise, the step is idle. Note that the difference between the original processes and their flip process variants is the idle steps in the latter. While H-removal processes can be viewed as flip processes, other classes mentioned in Section 1.1, like the H-free process and Achlioptas processes, cannot.

In [12], flip processes were studied from the perspective of dense graph limits. Recall that the main objects of this theory, introduced in [19, 7], are two-variable symmetric measurable functions defined on the square of a probability space (Ω, π) (which we assume to have a separable sigma-algebra and atomless measure) with values in $[0, 1]$. These functions are called *graphons*. We denote the set of all graphons by $\mathcal{W}_0$. Recall that one of the most important metrics on $\mathcal{W}_0$ is the *cut norm distance*, which is given by

$$d_\square(U, W) = \|U - W\|_\square = \sup_{S,T \subseteq \Omega} \left| \int_{S \times T} (U - W) \, \mathrm{d}\pi^2 \right| . \tag{1}$$

Recall also that each finite graph G can be represented as a graphon. To this end, take an arbitrary partition $(\Omega_v : v \in V(G))$ of Ω into sets of equal measure. Let the graphon W_G be defined to be 1 on each rectangle $\Omega_u \times \Omega_v$ for which $uv \in E(G)$, and 0 elsewhere. Note that this graphon representation depends on the choice of the partition $(\Omega_v : v \in V(G))$. However, we will assume that graphs on the same vertex set are represented using the same partition of Ω. Since each step of a flip process of order k flips at most $\binom{k}{2}$ pairs of vertices (from a non-edge to an edge or vice versa), it follows that we need to perform $\Omega(n^2)$ steps to see a change in the evolution of an n-vertex graph through the lens of dense graph limits. Indeed, one of the main results of [12] is the following correspondence.

Theorem 1.1 (Informal, with loose quantification; see Theorem 5.1 in [12])**.** *For every rule $\mathcal{R}$ of order $k \in \mathbb{N}$, there exists an* evolution function $\Phi : \mathcal{W}_0 \times [0, \infty) \to \mathcal{W}_0$ *so that, writing $\Phi^t W = \Phi(W, t)$ and treating $\mathcal{W}_0$ as the metric space with metric $d_\square$, we have the following.*

- (i) *For any fixed graphon $W \in \mathcal{W}_0$, we have $\Phi^0 W = W$ (W is the 'initial graphon') and the map $t \mapsto \Phi^t W : [0, \infty) \to \mathcal{W}_0$, which we call the* trajectory starting at W*, is continuous.*
- (ii) *The trajectory depends continuously on the initial conditions, i.e., for any fixed time $t \geq 0$, the map $\Phi^t(\cdot) : \mathcal{W}_0 \to \mathcal{W}_0$ is continuous. The variable t is 'time'.*
- (iii) *Let $G_0, G_1, \ldots$ be the flip process starting with an n-vertex graph G_0. Denote by W_ℓ, $\ell = 0, 1, \ldots$, the (random) graphon that represents the graph G_ℓ. For every $\varepsilon > 0$ and $T \geq 0$, with probability $1 - \exp(-\Omega(n^2))$ for every $\ell \in \mathbb{N} \cap [0, Tn^2]$ we have $d_\square(W_\ell, \Phi^{\ell/n^2} W_0) < \varepsilon$.*

The evolution function $\Phi : \mathcal{W}_0 \times [0,\infty) \to \mathcal{W}_0$ is constructed in a way which is only partly explicit. More specifically, given a rule $\mathcal{R}$, a *velocity field* $\mathfrak{V} : \mathcal{W}_0 \to \mathcal{W}$ is explicitly constructed (here, $\mathcal{W}$ is the space of 2-variable symmetric functions in $L^\infty(\Omega^2)$), and the trajectory starting at any initial graphon $W \in \mathcal{W}_0$ is then defined as the unique solution of the (Banach-space-valued) differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi^t W = \mathfrak{V}\left(\Phi^t W\right) \text{ with initial condition } \Phi^0 W = W\ ,$$

which can be equivalently formulated as the integral equation

$$\Phi^t W = W + \int_0^t \mathfrak{V}\left(\Phi^\tau W\right)\,\mathrm{d}\tau\ . \tag{2}$$

As we shall see in this paper, obtaining explicit trajectories is a difficult task even for fairly simple rules.

In the rest of this introduction we motivate further key notions in the theory of flip processes.

1.3. **Flip processes, going back in time.** Theorem 1.1(iii) says that for any time $t \geq 0$ the graphon $\Phi^t W$ approximately describes the structure of a typical graph after tn^2 steps of the flip process started at an n-vertex graph G_0 which is close to W. We may ask a complementary question: does there exist a graphon U such that, starting at an n-vertex graph H_0 that is close to U, after tn^2 steps we typically arrive to a graph which is similar to W? In such a case, it is natural to write $\Phi^{-t}W := U$, and the above property can be written as $\Phi^t(\Phi^{-t}W) = W$. Note that $\Phi^{-t}W$ need not always exist. To give a specific example, in the Erdős–Rényi flip process, there is no way we could arrive at an edgeless graph after performing any number of steps starting from any graph. This means that $\Phi^{-t}0$ does not exist for any $t > 0$. As a generalization of this example, for any constant graphon $C \in (0,1)$, the backwards trajectory $\Phi^{-t}C$ is a constant graphon which decreases as t increases until it reaches 0 (at time $t = -\frac{1}{2}\ln(1-C)$; this quantity can be obtained by considering the inverse of the trajectory starting at zero, see (1) in [12], and for larger t, $\Phi^{-t}C$ is not defined. Switching back to the general setting, we call the supremum of times t for which $\Phi^{-t}W$ exists the *age of a graphon* W. Furthermore, if $\mathrm{age}(W) < \infty$ then it can be shown that $\mathrm{orig}(W) := \Phi^{-\mathrm{age}(W)}W$ is well-defined (see Theorem 4.5(iii)Item.8 in [12]). We call $\mathrm{orig}(W)$ the *origin* of W. To return to the example of the Erdős–Rényi flip process, the origin of every constant graphon $C \in [0,1)$ is the constant-0. The constant-1 does not have an origin, but it has an infinite age and $\Phi^t 1 = 1$ for every $t \in (-\infty,\infty)$.

1.4. **Flip processes, destinations.** For a general rule and a fixed initial graphon W, there are several qualitatively different possible evolutions of the trajectory $(\Phi^t W)_{t\geq 0}$. The simplest one is when W is a *fixed point*, that is, $\Phi^t W = W$ for all $t \in \mathbb{R}$. The next

simplest is when $\Phi^t W$ converges as $t \to \infty$. Here we choose the weakest sensible notion of convergence, that is, cut norm distance convergence.[a] In such a setting we denote the limit by $\mathrm{dest}(W)$ and call it the *destination of* W. There are more complicated types of trajectories. For example, in [12] it was shown that a periodic trajectory exists for a certain rule. Further, when the initial graphon is taken in the proximity of such a periodic trajectory, its trajectory spirals towards the periodic one. However, at this moment we do not see a classification of trajectories beyond convergent/nonconvergent (where the former also contains fixed points).

For a convergent trajectory, it is of interest to investigate the speed of convergence. Bounds which are independent of the starting graphon are particularly interesting. More formally, suppose that $\mathcal{R}$ is a rule for which every trajectory converges. We can define for each $\delta > 0$ the *lower convergence time*

$$\tau^-(\delta) = \sup\{\inf\{T : d_\square(\Phi^T W, \mathrm{dest}(W)) < \delta\} : W \in \mathcal{W}_0\}$$

and the *upper convergence time*

$$\tau^+(\delta) = \sup\{\sup\{T : d_\square(\Phi^T W, \mathrm{dest}(W)) \geq \delta\} : W \in \mathcal{W}_0\} .$$

1.5. **Our results.** This paper investigates properties of some natural flip processes, with a particular emphasis on the description of their trajectories. Due to the key Theorem 1.1, this provides information about the typical evolution of an n-vertex graph in the next $\Theta(n^2)$ steps.

Let us briefly and informally introduce these flip processes and highlight the main results we obtain about them. The main results that concern either explicit description of trajectories, destinations, speed of convergence or speed of these flip processes are listed in Table 1. (In addition to those, some flip processes have specific features.)

1.5.1. *Ignorant flip processes (Section 4).* Ignorant flip processes are flip processes where the distribution for the replacement graph does not depend on the drawn graph. That is, the rule $\mathcal{R}$ satisfies that for any F, F', H, we have $\mathcal{R}_{F,H} = \mathcal{R}_{F',H}$. Ignorant flip processes are one of the few examples where we can describe the trajectories explicitly; this gives a complete description of their behaviour. In particular, all the trajectories of an ignorant flip process converge to the same constant graphon whose value is the expected edge density[b] according to the output distribution $\{\mathcal{R}_{F,H}\}_H$ (here F is arbitrary). The Erdős–Rényi flip process is a prominent example of an ignorant flip process.

[a]However, in [12] it is conjectured that this setting actually implies convergence in the L^1-norm and possibly even in the L^∞-norm.

[b]The *density* of a graph H is defined as $e(H)/\binom{v(H)}{2}$.

	trajectories	destination	age	speed
ignorant	Proposition 4.2			
balanced stirring	Proposition 5.2		Proposition 5.4	
complementing	Proposition 6.1			
extremist		Theorem 7.4 Theorem 7.6		
monotone		Proposition 8.2 Proposition 8.3		
component completion		Theorem 9.6		
removal		Fact 10.5 Theorem 10.7		Theorem 10.9 Proposition 10.10

TABLE 1. Summary of our main (and easily categorizable) results

1.5.2. *Stirring flip processes (Section 5).* While the output distribution in an ignorant flip process does not depend on the drawn graph at all, in a stirring flip process it depends only on a very basic graph parameter — its number of edges. In Section 5 we investigate balanced stirring flip processes, in which the expected number of edges in the replacement graph is equal to the number of edges in the drawn graph. We explicitly describe the trajectories of any balanced stirring flip process. Not surprisingly, the edge density is preserved along each trajectory. The description does not depend on the actual rule but only on its order. That is, any balanced stirring flip process of a given order yields the same trajectories.

1.5.3. *Complementing flip processes (Section 6).* The complementing flip process is a flip process where we replace the drawn graph by its graph complement. Informally, we thus sparsify the host graph at places where its density lies above $\frac{1}{2}$ and densify it at places where its density lies below $\frac{1}{2}$. It follows that each trajectory should converge to the constant-$\frac{1}{2}$ graphon. We work out the trajectories explicitly, confirming this intuition.

1.5.4. *Extremist flip processes (Section 7).* The extremist flip process is defined as follows. If the drawn graph has edge density less than 1/2, we replace it by the edgeless graph. If the edge density is greater than 1/2, we replace it by the clique. If the density is exactly 1/2, then there is some freedom in defining the replacement graph (our default choice is do nothing). Intuitively, dense graph(on)s converge to the complete graph (constant-1 graphon) while sparse graph(on)s converge to the edgeless graph (constant-0 graphon). Our main focus regarding extremist flip processes is to make this precise. Further, we study the limit behaviour as k grows.

1.5.5. *Monotone flip processes (Section 8).* We introduce several variants, but, roughly speaking, a monotone flip process is a flip process where in each step we add edges, or

at least add edges in expectation. In such a situation we expect the trajectories to be increasing in time and in particular to converge. Indeed, most of our treatment of monotone flip processes focuses on these features (including one negative example).

1.5.6. *Component completion flip processes (Section 9).* In a component completion flip process, we replace the drawn graph (which may have several connected components) by its so-called component completion graph, that is, we turn each connected component into a clique. Hence, the component structure of the graph does not change during the evolution. Note also that this procedure is void for orders $k = 1, 2$. For $k \geq 3$ we prove that the trajectories converge to a graphon with the same component structure, where each component is constant-1.

1.5.7. *Removal flip processes (Section 10).* Removal flip processes generalize the triangle removal flip process. We fix a graph $F \in \mathcal{H}_k$. In each step of the process, we delete a copy of F, should the drawn graph contain it. More formally, for $H \in \mathcal{H}_k$ which contains F as a subgraph (preserving the labels) we have $\mathcal{R}_{H,H-F} = 1$ while for H which does not contain F as a labelled subgraph we have $\mathcal{R}_{H,H} = 1$. The theory of removal flip processes is particularly elegant. There are two results which we could obtain only with the help of Szemerédi's regularity lemma: that convergence times for removal processes are finite and uniformly bounded and that the map which maps each graphon to its destination is continuous in the cut norm distance.

1.5.8. *Quadratically many steps and beyond.* Most of the results we obtain concern trajectories. Due to the correspondence in Theorem 1.1, this tells us about the typical evolution of a finite n-vertex graph in a flip process for $\Theta(n^2)$ steps. However, for the extremist and removal flip processes we found interesting questions which go beyond this time horizon.

2. Definitions and basics

This section recalls some basic facts, which we sometimes reprove in variants that are tailored to our applications.

For $k \in \mathbb{N}$, we write $(k)_2 = k(k-1)$. When using little-o notation, we sometimes use a subscript to emphasize the parameter of asymptotics. For example, $\lim_{k\to\infty} \frac{o_k(k^2)}{k^2} = 0$. In expressions of the form $a \pm \varepsilon$ we imply the closed interval $[a - \varepsilon, a + \varepsilon]$.

2.1. **Differential equations.**

2.1.1. *Grönwall's inequality.* Here we state two variants of Grönwall's inequality. The following lemma is a variant of Grönwall's inequality in integral form.

Lemma 2.1 ([27, Theorem 1.10])**.** *Let $f : [t_0, t_1] \to [0, \infty)$ be a continuous non-negative function. Suppose that for some constants $A \geq 0$, $C > 0$ we have*

$$f(t) \leq A + C \int_{t_0}^{t} f(x)\,\mathrm{d}x \tag{3}$$

for every $t \in [t_0, t_1]$. Then we have

$$f(t) \leq Ae^{C(t-t_0)} \quad \text{for all } t \in [t_0, t_1]\,. \tag{4}$$

The following lemma is a variant of Grönwall's inequality for differentiable functions where the differential inequality holds subject to a condition on the value of the function.

Lemma 2.2. *Let $D, K > 0$ and $f\colon [0, \infty) \to [0, \infty)$ be a differentiable function satisfying $f(0) \leq D$ and $f'(t) \leq -Kf(t)$ for all $t \in [0, \infty)$ such that $f(t) \leq D$. Then for all $t \in [0, \infty)$ we have $f(t) \leq f(0)e^{-Kt}$.*

Proof. First we show that we have $f(t) \leq D$ for all $t \in [0, \infty)$. Indeed, suppose that we have $f(t') > D$ for some $t' > 0$. Let $T := \inf\{t \geq 0 : f(t) > D\}$. By the continuity of f we have $f(T) = D$, so we have $f'(T) \leq -Kf(T) = -KD < 0$. By the definition of $f'(T)$, there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ we have

$$f(T + \varepsilon) \leq f(T) + \varepsilon(KD/2 + f'(T)) \leq f(T) - KD\varepsilon/2 < D.$$

This implies that $f(t) < D$ in the interval $(T, T + \varepsilon_0)$, which contradicts the definition of T as the infimum t such that $f(t) > D$. Hence, we have $f(t) \leq D$ for all $t \in [0, \infty)$.

The function $g\colon [0, \infty) \to [0, \infty)$ given by $g(s) := f(s)e^{Ks}$ is differentiable with $g'(s) = f'(s)e^{Ks} + Kf(s)e^{Ks} \leq 0$, so for every $t \geq 0$ we have $g(t) \leq g(0) = f(0)$ and hence $f(t) = g(t)e^{-Kt} \leq f(0)e^{-Kt}$, as desired. □

2.1.2. *On notation of integrals.* For the sake of brevity and clarity we sometimes use simplified notation to denote an integral, namely we often write $\int_{x=a}^{b} f(x)$ (or $\int_{x=a}^{b} f$) instead of $\int_a^b f(x)\,\mathrm{d}x$ for functions of a single real argument (that is, the underlying measure is the Lebesgue measure) and $\int_{x\in A} f(x)$ (or $\int_{x\in A} f$) instead of $\int_A f(x)\,\mathrm{d}\mu(x)$ for functions on some abstract measure space when the measure μ is clear from the context.

2.1.3. *Differential equations in Banach spaces.* We need to recall concepts of the derivative and the integral in Banach spaces. However, we do not need almost any results from the abstract theory of differential equations in Banach spaces since we only need to recall results obtained already in [12]. Let $\mathcal{E}$ be a Banach space with norm $\|\cdot\|$. Suppose that a function $f : I \to \mathcal{E}$ is defined for an open set $I \subset \mathbb{R}$. If $t \in I$ is such that there is an element $z \in \mathcal{E}$ for which

$$\lim_{\varepsilon\to 0} \left\| \frac{f(t+\varepsilon) - f(t)}{\varepsilon} - z \right\| = 0\,,$$

then we say that f *is differentiable at* t and call $\frac{\mathrm{d}}{\mathrm{d}t} f(t) = z$ the *derivative of* f *at* t. At the heart of the theory of flip processes are first-order autonomous differential equations of the type

$$\frac{\mathrm{d}}{\mathrm{d}t} f(t) = \mathfrak{O}(f(t)) \text{ with initial condition } f(0) = z \,, \tag{5}$$

where $z \in \mathcal{E}$, f takes values in $\mathcal{E}$, and $\mathfrak{O} : \mathcal{E} \to \mathcal{E}$ is (locally) Lipschitz. Such a differential equation can be rewritten as the integral equation

$$f(t) = z + \int_0^t \mathfrak{O}(f(s)) \, \mathrm{d}s \,, \tag{6}$$

where the integral is the so-called Cauchy–Bochner integral in the Banach space $\mathcal{E}$ (it is easy to show that this integral is well-defined, since $s \mapsto \mathfrak{O}(f(s))$ is continuous). The integral has all usual properties of the Riemann integral of continuous real-valued functions, with the norm playing the role of the absolute value (see Section C in [12]).

In this paper, we specifically consider differential equations in the Banach space $(\mathcal{W}, \|\cdot\|_\infty)$, which is a closed subspace of $L^\infty(\Omega^2)$ (where Ω is the probability space on which our graphons are defined). Recall that the elements of $L^\infty(\Omega^2)$ are equivalence classes (where equivalent means equal almost everywhere) of measurable functions, which means that it is not formally meaningful to evaluate an element $f \in L^\infty(\Omega^2)$ at a given point $(x, y) \in \Omega^2$. However, it is much more intuitive to pretend we can do this and, for example (i) evaluate $f(t)$ (see (2) for a prominent example) at point (x, y) by integrating a real-valued curve $s \mapsto \mathfrak{V}\,(f(s))\,(x, y)$; (ii) deduce that $f(t) \leq g(t)$ for every $t \in [a, b]$ implies $\int_a^b f(t) \leq \int_a^b g(t)$ (where the inequalities are understood to hold almost everywhere). In [12] it was shown that this apparently non-rigorous approach can be perfectly justified. We believe that this approach of dealing with trajectories "pointwise" has an advantage on the intuitive level, and we stick to it, referring a meticulous reader to the required technical tools for justification in Appendix C in [12].

What follows is the only abstract fact we need that is not available in [12]. It says that a solution of a differential equation of the type in (5) changes only little if we perturb the velocity field. This statement is very likely standard, but we found it easier to give a four-line proof than to find a source in the literature.

Lemma 2.3. *Consider a Banach space* $\mathcal{E}$ *with norm* $\|\cdot\|$. *Let* $\mathcal{U} \subset \mathcal{E}$ *be a set, and* $A, K > 0$ *be constants. Let* $X, Y : \mathcal{U} \to \mathcal{E}$ *be functions. Suppose that* X *is* K*-Lipschitz, that is, we have*

$$\|X(x) - X(y)\| \leq K \, \|x - y\| \qquad \text{for all } x, y \in \mathcal{U} \,. \tag{7}$$

Suppose further that for each $x \in \mathcal{U}$ *we have*

$$\|X(x) - Y(x)\| \leq A \,. \tag{8}$$

Given $x_0 \in \mathcal{U}$, let $\alpha, \beta : [0,T] \to \mathcal{U}$ satisfy the integral equations

$$\alpha(t) = x_0 + \int_{s=0}^{t} X(\alpha(s)) \quad \beta(t) = x_0 + \int_{s=0}^{t} Y(\beta(s)) \quad \text{for each } t \in [0,T]$$

Then for each $t \in [0,T]$ we have $\|\alpha(t) - \beta(t)\| \le AT \exp(Kt)$.

Proof. Set $f(t) := \|\alpha(t) - \beta(t)\|$. Then

$$\begin{aligned} f(t) &\le \int_{s=0}^{t} \|X(\alpha(s)) - Y(\beta(s))\| \overset{(8)}{\le} \int_{s=0}^{t} \big(A + \|X(\alpha(s)) - X(\beta(s))\|\big) \\ &\overset{(7)}{\le} \int_{s=0}^{t} \big(A + Kf(s)\big) \le AT + K\int_{s=0}^{t} f(t)\,. \end{aligned}$$

Hence, the claim follows from Lemma 2.1. □

2.2. **Graphons.** Our notation regarding graphons follows [18]. Recall that (Ω, π) is an atomless probability space with an implicit separable sigma-algebra. We write $\mathcal{W} \subset L^\infty(\Omega^2)$ for the set of bounded symmetric measurable functions; these are called *kernels*. *Graphons*, which we denote by $\mathcal{W}_0$, are then those kernels whose range is a subset of $[0,1]$.

2.2.1. *Densities and degrees.* The notion of homomorphism densities of a finite graph is one of the cornerstones of the theory of graph limits.

For a function $f : \Omega^V \to \mathbb{R}$ and a subset $S \subset V$ we write

$$\int f \,\mathrm{d}\pi^S = \int_{\Omega^S} f(x_i : i \in V) \prod_{i \in S} \pi(\,\mathrm{d}x_i).$$

Definition 2.4. *Given a kernel W and a graph $F = (V,E)$, we define the* density of F *in* W *as*

$$t(F,W) \coloneqq \int \prod_{ij \in E} W(x_i, x_j)\,\mathrm{d}\pi^V\,.$$

and the induced density of F *in* W *as*

$$t_{\mathrm{ind}}(F,W) \coloneqq \int \prod_{ij \in E} W(x_i,x_j) \prod_{ij \notin E} (1 - W(x_i,x_j))\,\mathrm{d}\pi^V\,.$$

In particular, the edge density of W *is the number $t(K_2,W) = \int W \,\mathrm{d}\pi^2$.*

A *rooted graph* is a triple (R,V,E), where (V,E) is a graph and R is a subset of V; we call the elements of R the *roots*.

Definition 2.5 (Rooted densities)**.** *Given a graphon W, a rooted graph $F = (R, V, E)$, and $\mathbf{x} = (x_i : i \in R) \in \Omega^R$ we define the functions*

$$t^{\mathbf{x}}(F, W) \coloneqq \int \prod_{ij \in E} W(x_i, x_j) \, \mathrm{d}\pi^{V \setminus R} \quad \textit{and}$$

$$t^{\mathbf{x}}_{\mathrm{ind}}(F, W) \coloneqq \int \prod_{ij \in E} W(x_i, x_j) \prod_{ij \notin E} (1 - W(x_i, x_j)) \, \mathrm{d}\pi^{V \setminus R} \, .$$

Further, if $\mathbf{A} = (A_i)_{i \in R}$ is a collection of sets $A_i \subset \Omega$ then we define the quantities

$$t^{\mathbf{A}}(F, W) \coloneqq \int_{\mathbf{x} \in \prod_i A_i} t^{\mathbf{x}}(F, W) \quad \textit{and} \quad t^{\mathbf{A}}_{\mathrm{ind}}(F, W) \coloneqq \int_{\mathbf{x} \in \prod_i A_i} t^{\mathbf{x}}_{\mathrm{ind}}(F, W) \, .$$

One of the key properties of densities is that they, as functions of a graphon, are continuous with respect to the cut norm distance.

Lemma 2.6 (Counting lemma, [18, Lemma 10.23])**.** *Suppose that we have a graph F of order k and two graphons U and W. Then we have*

$$|t(F, U) - t(F, W)| \leq \tbinom{k}{2} d_{\square}(U, W).$$

Lastly, we introduce the notions of degree and degree-regularity.

Definition 2.7. *The* degree function *of a kernel W is the function $\deg_W : \Omega \to \mathbb{R}$ defined as $\deg_W(x) = \int_{y \in \Omega} W(x, y)$. We say that a kernel W of edge density d is* degree-regular *if for π-almost every $x \in \Omega$ we have $\deg_W(x) = d$.*

2.2.2. *Sampling.* A graphon W defines, for every $k \in \mathbb{N}$, a probability distribution on $\mathcal{H}_k$ as follows. Sample independent π-random elements U_v, $v \in [k]$ of Ω and then connect each pair of distinct vertices $u, v \in [k]$ independently with probability $W(U_u, U_v)$. We write $F \sim \mathbb{G}(k, W)$ for the random graph obtained this way and sometimes even use $\mathbb{G}(k, W)$ to denote the random graph itself. For many properties of sampling we refer to Section 10 of [18]. The following lemma asserts that for large k this random graph is close to W with high probability. To this end we use the notion of *cut distance* between a finite graph G and a graphon W, which is defined as $\delta_{\square}(G, W) = \inf\{d_{\square}(U, W)\}$, where U ranges over all graphon representations of G.

Lemma 2.8 (Sampling lemma [18, Lemma 10.16])**.** *For $k = 2, 3, \ldots$ and every $W \in \mathcal{W}_0$ we have*

$$\mathbb{P}\left(\delta_{\square}(\mathbb{G}(k, W), W) \leq 22/\sqrt{\log_2 k}\right) \geq 1 - \exp(-k/(2 \log_2 k)) \, .$$

2.2.3. *Regularity lemma.* In Section 10 we study removal flip processes. To study their speed of convergence, we need to recall a regularity lemma. While for many purposes in the area of graph limits the Frieze–Kannan regularity lemma is sufficient, here we need

Szemerédi's regularity lemma. The notions of density and regularity are standard, but typically appear in the setting of finite graphs. The graphon setting is used for example in [20].

Definition 2.9. *Suppose that W is a graphon and $C, D \subset \Omega$ are sets of positive measure. The* density of (C, D) *in W is defined as*

$$d(C,D) := \frac{1}{\pi(C)\pi(D)} \int_{C\times D} W \,\mathrm{d}\pi^2 \,.$$

Given $\alpha > 0$, we say that (C, D) is α-regular *in W if for every $C' \subset C$ and $D' \subset D$ we have*

$$\left| \pi(C')\pi(D')d(C,D) - \int_{C'\times D'} W \,\mathrm{d}\pi^2 \right| \leq \alpha\pi(C)\pi(D) \,. \tag{9}$$

Definition 2.10. *Suppose that W is a graphon. For $\alpha > 0$, we say that a partition $\{P_1, \dots, P_M\}$ of Ω is an* α-regular partition of W *if*

- $\alpha \geq 1/M$,
- *for each i we have $\pi(P_i) = 1/M$,*
- *the number of pairs $(i, j) \in [M]^2$ for which (P_i, P_j) is not an α-regular pair in W is at most αM^2.*

We call the sets $P_1, \dots, P_M$ clusters.

With this terminology, we can state a version of Szemerédi's regularity lemma tailored to our purposes.

Theorem 2.11. *For each $\alpha > 0$ there exists $M_0 \in \mathbb{N}$ so that for each graphon W there exists an α-regular partition of W with at most M_0 sets.*

In the setting of regular pairs, homomorphism densities can be controlled. This is the subject of the counting lemma, which we state in a version tailored to our purposes.

Lemma 2.12. *Let $\alpha > 0$ and W be a graphon. Let $F = (R, R, E)$ be a graph which satisfies $|R| \geq 3$ and whose vertices are all roots. Suppose that $\mathbf{A} = (A_i)_{i\in R}$ is a tuple of subsets of Ω such that for every $i, j \in R$ we have $\pi(A_i) = \pi(A_j)$, and for every $\{i, j\} \in E$ the pair (A_i, A_j) is an α-regular pair in W. Then*

$$\left| t^{\mathbf{A}}(F, W) - \prod_{ij\in E} d(A_i, A_j) \prod_{i\in R} \pi(A_i) \right| \leq \alpha |R|^{|R|} \tag{10}$$

Proof. We derive this from the version of the counting lemma stated as Lemma 10.24 in [18], where we refer to the notion of decorated graphs and the corresponding homomorphism density. That counting lemma is tailored to the notion of cut norm distance, which is

linked to the weak regularity lemma. On the other hand, Lemma 2.12 is phrased in the setting of Szemerédi's regularity lemma, which is used much less in the theory of graph limits.[c] In particular, the error bound (10) is relative to $\prod_{i\in R}\pi(A_i)$. So, the routine work below is to transfer to that setting, which we do so by zooming in on the sets A_i.

Further, assume that $R=[r]$ for some $r\in\mathbb{N}$. For $i\in[r]$, write $I_i:=[\frac{i-1}{r},\frac{i}{r})$. Further, write $I=\bigcup_{i=1}^r I_i=[0,1)$. For every i we define two probability measures: the r-times rescaled Lebesgue measure $\mu_i(\cdot)=r\mathrm{Leb}(\cdot)$ on I_i and $\pi_i(\cdot)=\frac{\pi(\cdot)}{\pi(A_i)}$ on A_i. Since (Ω,π) is atomless and separable, the Isomorphism theorem for measure spaces tells us that there exists a measure-preserving map $f_i:(I_i,\mu_i)\to(A_i,\pi_i)$. For example, if $\Omega=([0,1],\mathrm{Leb})$, we can define $f_i:I_i\to A_i$ as $f_i(z):=\inf\{t\in A_i:\pi(A_i\cap[0,t])\geq(rz-(i-1))\cdot\pi(A_i)\}$.

Define $f:I\to\Omega$ by setting $f(z)=\sum_{i\in R}f_i(z)\mathbb{I}_{z\in I_i}$. Define $W^f:I^2\to[0,1]$ by setting

$$W^f(x,y):=W\left(f(x),f(y)\right)\ .$$

Consider the complete graph K on the vertex set R. We construct two decorated graphs $u=(U_e:e\in E(K))$ and $u'=(U'_e:e\in E(K))$ on the ground space I as follows. For every $e=\{i,j\}\in E(K)$, writing $d_e=d(A_i,A_j)$, we define

$$U_e=\mathbb{I}_{I_i\times I_j\cup I_j\times I_i}\cdot\begin{cases}W^f, & \text{if } e\in E,\\ 1, & \text{if } e\in E(K)\setminus E,\end{cases}$$

$$U'_e=\mathbb{I}_{I_i\times I_j\cup I_j\times I_i}\cdot\begin{cases}d_e, & \text{if } e\in E,\\ 1, & \text{if } e\in E(K)\setminus E.\end{cases}$$

It is obvious that for every $x_1,\dots,x_r\in I$ we have

$$\prod_{ij\in E(K)}U_{ij}(x_i,x_j)=\prod_{i,j\in E}W^f(x_i,x_j)\prod_{i\in[r]}\mathbb{I}_{I_i}(x_i). \tag{11}$$

Similarly,

$$\prod_{ij\in E(K)}U'_{ij}(x_i,x_j)=\prod_{e\in E}d_e\prod_{i\in[r]}\mathbb{I}_{I_i}(x_i). \tag{12}$$

Note that $(x_1,\dots,x_r)\mapsto(f(x_1),\dots f(x_r))$ is a measure-preserving map between the product of probability spaces (I_i,μ_i) and the product of probability spaces (A_i,π_i). Therefore by (11),

$$r^r t(K,u)=\int_{\prod_i I_i}\prod_{ij\in E}W^f(x_i,x_j)\prod_i\mathrm{d}\mu_i=\int_{\prod_i A_i}\prod_{ij\in E}W(y_i,y_j)\prod_i\mathrm{d}\pi_i=\frac{t^{\mathbf{A}}(F,W)}{\prod_i\pi(A_i)}$$

[c]Actually, we are not aware of a similar application of Szemerédi's regularity lemma in the setting of graphons.

and, by (12),

$$r^r t(K,u') = \int_{\prod_i I_i} \prod_{e\in E} d_e \prod_i \mathrm{d}\mu_i = \prod_{e\in E} d_e.$$

To sum up,

$$\begin{aligned} t(K,u) &= t^{\mathbf{A}}(F,W)\cdot \frac{r^{-r}}{\prod_{i\in R}\pi(A_i)} \quad \text{and} \\ t(K,u') &= \prod_{e\in E} d_e \cdot r^{-r}\ . \end{aligned} \tag{13}$$

We claim that for each $e \in E(K)$ we have $\|U_e - U'_e\|_\square \le 2\alpha r^{-2}$. Postponing the proof of this, we first show how it implies the asserted statement. By Lemma 10.24 in [18], we have

$$|t(K,u) - t(K,u')| \le \sum_{e\in E(K)} \left\|U_e - U'_e\right\|_\square \le \binom{r}{2} 2\alpha r^{-2} < \alpha\ .$$

Substituting the expressions from (13) and multiplying by $r^r \cdot \prod_{i\in R}\pi(A_i)$, we obtain

$$\left| t^{\mathbf{A}}(F,W) - \prod_{e\in E} d_e \cdot \prod_{i\in R}\pi(A_i) \right| \le \alpha \cdot r^r \cdot \prod_{i\in R}\pi(A_i)\ ,$$

as needed.

We now return to the postponed claim that for each $e \in E(K)$ we have $\|U_e - U'_e\|_\square \le 2\alpha r^{-2}$. We write ab as shorthand for $\{a,b\}$. Note that for $e = ab \notin E$ we have $U_{ab} = U'_{ab} = \mathbb{I}_{I_a\times I_b \cup I_b \times I_a}$ hence trivially $\|U_e - U'_e\|_\square = 0$. It remains to consider the case $e \in E$. Suppose that $S, T \subset I$ are arbitrary as in (1). Let $C_a := S\cap I_a$, $C_b := S\cap I_b$, $D_a := T\cap I_a$, and $D_b := T\cap I_b$. Since I_a, I_b are disjoint and U_e, U'_e are zero outside of $I_a\times I_b \cup I_b \times I_a$, we have

$$\left|\int_{S\times T} (U_e - U'_e)\,\mathrm{d}x_a\,\mathrm{d}x_b\right| \le \left|\int_{C_a\times D_b} (U_e - U'_e)\,\mathrm{d}x_a\,\mathrm{d}x_b\right| + \left|\int_{C_b\times D_a} (U_e - U'_e)\,\mathrm{d}x_a\,\mathrm{d}x_b\right|\ .$$

Since $(x_a, x_b) \mapsto (f(x_a), f(x_b))$ is a measure preserving map between $(I_a,\mu_a)\times(I_b,\mu_b)$ and $(A_a,\pi_a)\times(A_b,\pi_b)$, we have

$$\begin{aligned} r^2 \left|\int_{C_a\times D_b} (U_e - U'_e)\,\mathrm{d}x_a\,\mathrm{d}x_b\right| &= \left|\int_{C_a\times D_b} (W^f - d_e)\,\mathrm{d}\mu_a\,\mathrm{d}\mu_b\right| \\ &= \left|\int_{f(C_a)\times f(D_b)} (W - d_e)\,\mathrm{d}\pi_a\,\mathrm{d}\pi_b\right| \\ &= \frac{1}{\pi(A_a)\pi(A_b)} \left|\int_{f(C_a)\times f(D_b)} W\,\mathrm{d}\pi^2 - \pi(f(C_a))\pi(f(D_b))d_e\right| \overset{(9)}{\le} \alpha. \end{aligned}$$

To conclude,

$$\left|\int_{C_a\times D_b}(U_e-U'_e)\,\mathrm{d}\pi^2\right|\le\alpha r^{-2}\,.$$

Together with the same bound for $\left|\int_{C_b\times D_a}(U_e-U'_e)\,\mathrm{d}\pi^2\right|$ we get that

$$\left|\int_{S\times T}(U_e-U'_e)\,\mathrm{d}\pi^2\right|\le 2\alpha r^{-2}\,,$$

as was needed. □

2.2.4. *Inequality between clique densities.* Here we shall prove an inequality which relates clique densities in graphons. This is a consequence of the Kruskal–Katona theorem [16, 15]. In fact, our proof uses a slightly weaker version due to Lovász [17, Exercise 13.31(b)]. While we need this inequality only for the case $r=2$ (which can be derived from [9]), we shall state and prove the general case for possible future use. For $m\in\mathbb{N}$ and a real number $x\ge m$ we write $\binom{x}{m}:=\frac{x(x-1)\dots(x-m+1)}{m!}$.

Proposition 2.13. *Let $W\in\mathcal{W}_0$ be a graphon and $r,k\in\mathbb{N}$ satisfy $r<k$. Then we have*

$$t(K_r,W)\ge t(K_k,W)^{r/k}\,.$$

In particular, for all $k\ge 3$ we have $\int W\,\mathrm{d}\pi^2=t(K_2,W)\ge t(K_k,W)^{2/k}$.

Proof. Define $\alpha:=t(K_k,W)^{1/k}$. Pick a sequence $(G_n)_{n\in\mathbb{N}}$ of graphs with $v(G_n)=n$ which converges to the graphon W. Let $(u_n)_{n\in\mathbb{N}}$ be a sequence of real numbers so that the number of k-cliques in G_n is $\binom{u_n}{k}$, that is, $t(K_k,G_n)=\binom{u_n}{k}/\binom{n}{k}$; this implies $u_n\sim\alpha n$. Exercise 13.31 in Lovász [17] (for H consisting of the vertex sets of k-cliques) implies that the number of r-cliques in G_n is at least $\binom{u_n}{r}$, whence

$$t(K_r,W)=\lim_{n\to\infty}t(K_r,G_n)\ge\lim_{n\to\infty}\frac{\binom{u_n}{r}}{\binom{n}{r}}=\alpha^r=t(K_k,W)^{r/k}\,,$$

as required. □

2.3. **Probability.** The classical Berry–Esseen theorem establishes quantitative bounds on the accuracy of the approximation of the probability distribution of a sum of independent and identically distributed random variables by a normal distribution with the same mean and variance. We will apply this to prove Theorem 7.7.

Theorem 2.14 (Berry–Esseen theorem)**.** *There is a positive constant C such that the following holds. Suppose that $X_1,\dots,X_k$ are i.i.d. random variables with $\mathbb{E}X_i=0$, $\operatorname{Var}X_i=\sigma^2<\infty$ and $\mathbb{E}|X_i|^3=\beta<\infty$ for all $i\in[k]$. Let $X=\sum_{i=1}^k X_i$. Then for all $x\in\mathbb{R}$ we have*

$$\left|\mathbb{P}\left(X\le x\sigma\sqrt{k}\right)-\mathbb{P}\left(Z\le x\right)\right|\le\frac{C\beta}{\sigma^3\sqrt{k}},$$

where $Z \sim N(0,1)$ is a standard Gaussian random variable.

3. Flip processes as dynamical systems on graphons: a recap

In this section, we review the theory from [12]. With the kind permission of the authors of [12], we include the relevant parts into this section with varying degrees of additional edits.

3.1. **The velocity operator.** Let W be a graphon and let $\mathcal{R}$ be a rule of order k. Given a graph $F \in \mathcal{H}_k$ and an ordered pair $1 \le a \ne b \le k$ define an operator $T_F^{a,b} : \mathcal{W} \to L^\infty(\Omega^2)$ by setting

$$(T_F^{a,b}W)(x,y) := t_{\mathrm{ind}}^{(x,y)}(F^{a,b}, W)\,.$$

The *velocity operator for* $\mathcal{R}$ is the operator $\mathfrak{V}_\mathcal{R} : \mathcal{W} \to \mathcal{W}$ given by

$$\mathfrak{V}_\mathcal{R}W := \sum_{F,H\in\mathcal{H}_k} \sum_{1\le a\ne b\le k} \mathcal{R}_{F,H} \cdot T_F^{a,b}W \cdot \left(\mathbb{I}_{\{ab\in H\setminus F\}} - \mathbb{I}_{\{ab\in F\setminus H\}}\right)\,, \tag{14}$$

where ab is shorthand for $\{a,b\}$.

It is convenient to have alternative reformulations of the velocity operator. Let W be a graphon and $\mathcal{R}$ be a rule of order k. Let the random variables $(U_v : v \in [k])$ and the random graph $\mathbb{G}(k,W)$ be as defined in Section 2.2.2. Let $(\mathbb{F},\mathbb{H})$ be a pair of random graphs in $\mathcal{H}_k$ such that $\mathbb{F} \sim \mathbb{G}(k,W)$ and $\mathbb{P}(\mathbb{H} = H \,|\, \mathbb{F} = F) = \mathcal{R}_{F,H}$ for all $F, H \in \mathcal{H}_k$. Then we have (see equation (26) in [12])

$$\mathfrak{V}_\mathcal{R}W(x,y) = \sum_{a\ne b} \left[\mathbb{P}\left(ab \in \mathbb{H} \,|\, U_a = x, U_b = y\right) - W(x,y)\right]. \tag{15}$$

It follows that

$$\int \mathfrak{V}_\mathcal{R}W \,\mathrm{d}\pi^2 = 2\mathbb{E}\left(e(\mathbb{H}) - e(\mathbb{F})\right) \tag{16}$$

$$= 2\mathbb{E}e(\mathbb{H}) - (k)_2 t(K_2, W). \tag{17}$$

3.2. **Trajectories.** In this subsection, we recap the notion of trajectories which we mentioned informally in Theorem 1.1. We make several adjustments, following the way trajectories were constructed in [12]. Firstly, we define trajectories not only for graphons but for general kernels. Secondly, we define not only forward evolution in time but also backward evolution. It turns out that in order to allow for these generalizations, we cannot prescribe the time domain for a trajectory in advance.

Definition 3.1. *Given a rule $\mathcal{R}$ and a kernel $W \in \mathcal{W}$, a* trajectory starting at W *is a differentiable function $\Phi^\cdot W : I \to (\mathcal{W}, \|\cdot\|_\infty)$ defined on an open interval $I \subseteq \mathbb{R}$ containing*

zero, that satisfies the autonomous differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi^t W = \mathfrak{V}\,\Phi^t W \tag{18}$$

with the initial condition

$$\Phi^0 W = W\,. \tag{19}$$

The following theorem implies that if a trajectory starts at a graphon (rather than a general kernel), it is unique (provided we extended it maximally) and is defined for all positive times. Such a trajectory is always defined for some negative times (as $\mathcal{D}_W$ is open), but it might leave the graphon space immediately (i.e., we might have $I_W = [0,\infty)$).

Theorem 3.2 (Theorem 4.5 in [12])**.** *Suppose that $\mathcal{R}$ is a rule and $W \in \mathcal{W}$ is a kernel. Then the following hold.*

(i) *There is an open interval $\mathcal{D}_W \subseteq \mathbb{R}$ containing 0 and a trajectory $\Phi^\cdot W : \mathcal{D}_W \to \mathcal{W}$ starting at W such that any other trajectory starting at W is a restriction of $\Phi^\cdot W$ to a subinterval of $\mathcal{D}_W$.*

(ii) *For any $u \in \mathcal{D}_W$ we have $\mathcal{D}_{\Phi^u W} = \{t \in \mathbb{R} : t + u \in \mathcal{D}_W\}$ and for every $t \in \mathcal{D}_{\Phi^u W}$ we have*

$$\Phi^t \Phi^u W = \Phi^{t+u} W.$$

(iii) *If $W \in \mathcal{W}_0$ is a graphon, the set $I_W := \{t \in \mathcal{D}_W : \Phi^t W \in \mathcal{W}_0\}$ is a closed interval containing $[0,\infty)$.*

Definition 3.3. *For $W \in \mathcal{W}_0$ we write* $\operatorname{age}(W) := -\inf I_W \in [0,\infty]$.

Remark 3.4 (Remark 4.9 in [12])**.** *A trajectory $\Phi^\cdot W : I \to \mathcal{W}$ satisfies*

$$\Phi^t W = W + \int_0^t \mathfrak{V}_{\mathcal{R}} \Phi^\tau W \,\mathrm{d}\tau, \quad t \in I,$$

where the integral is defined with respect to the norm $\|\cdot\|_\infty$.

The last preliminary result we shall need about trajectories says the following. Let $\mathcal{R}$ be a rule and $\mathcal{A}$ be a set of graphons which have a certain profile on a fixed rectangle $\Omega_1 \times \Omega_2$. Suppose that for each graphon in $\mathcal{A}$, the velocity restricted to the rectangle $\Omega_1 \times \Omega_2$ is zero. Then each trajectory which starts in $\mathcal{A}$ stays confined to $\mathcal{A}$.

Proposition 3.5 (Proposition 6.8 in [12])**.** *Let $\mathcal{R}$ be a rule and $\alpha : S \to [0,1]$ be a symmetric function where $S := \Omega_1 \times \Omega_2 \cup \Omega_2 \times \Omega_1 \subset \Omega^2$. Let $\mathcal{A} = \{W \in \mathcal{W}_0 : W\restriction_S = \alpha\}$. Suppose that for each $W \in \mathcal{A}$ we have $(\mathfrak{V}\,W)\restriction_S = 0$. Then for each $U \in \mathcal{A}$ and each $t \geq 0$ we have $\Phi^t U \in \mathcal{A}$.*

3.3. **Continuity properties of trajectories.** In this subsection we assume that $\mathcal{R}$ is a rule of order k.

Lemma 3.6 (Lemma 4.8 in [12]). *For every $W \in \mathcal{W}_0$ we have*

$$-(k)_2 \cdot W \le \mathfrak{V}_{\mathcal{R}} W \le (k)_2 \cdot (1 - W)\ . \tag{20}$$

Lemma 3.7 (Lemma 4.15 in [12]). *For every $W \in \mathcal{W}_0$ and every $t, u \in [-\mathrm{age}(W), +\infty)$ we have*

$$\left\|\Phi^t W - \Phi^u W\right\|_\infty \le (k)_2 |t-u|,$$

and for every $\delta > 0$ we have

$$\left\|\frac{\Phi^\delta W - W}{\delta} - \mathfrak{V}_{\mathcal{R}} W\right\|_\infty \le (k)_2^3 2^{\binom{k}{2}-1}\delta. \tag{21}$$

Theorem 3.8 (Theorem 4.16 in [12]). *Write $C_{\square,k} = 4\binom{k}{2}^2 2^{\binom{k}{2}}$. Then for each $U, W \in \mathcal{W}_0$ and $t \in [-\min(\mathrm{age}(U), \mathrm{age}(W)), +\infty)$, we have*

$$\left\|\Phi^t U - \Phi^t W\right\|_\square \le \exp(C_{\square,k}|t|) \cdot \|U - W\|_\square\ .$$

3.4. **Destinations and speed of convergence.** Let $\mathcal{R}$ be a rule. We say that the trajectory of a graphon U is *convergent* if there exists a *destination* of U, that is, a graphon denoted by $\mathrm{dest}_{\mathcal{R}}(U)$ such that $\Phi^t U \xrightarrow{\square} \mathrm{dest}_{\mathcal{R}}(U)$ as $t \to \infty$. We say that the flip process with rule $\mathcal{R}$ is *convergent* if all its trajectories converge.

Proposition 3.9 ([12, Proposition 6.3]). *If $W = \mathrm{dest}_{\mathcal{R}}(U)$ for some $U \in \mathcal{W}_0$, then $\mathfrak{V}_{\mathcal{R}} W = 0$.*

For a rule $\mathcal{R}$, and $\delta > 0$ we define the *lower convergence time* $\tau^-_{\mathcal{R}}(\delta)$ and the *upper convergence time* $\tau^+_{\mathcal{R}}(\delta)$ by

$$\tau^-_{\mathcal{R}}(\delta) := \sup\{\inf\{t : d_\square(\Phi^t W, \mathrm{dest}_{\mathcal{R}}(W)) < \delta\} : W \in \mathcal{W}_0\}\ ,$$
$$\tau^+_{\mathcal{R}}(\delta) := \sup\{\sup\{t : d_\square(\Phi^t W, \mathrm{dest}_{\mathcal{R}}(W)) \ge \delta\} : W \in \mathcal{W}_0\}\ .$$

4. Ignorant flip processes

The first and in a sense the most trivial class of flip processes we introduce are the ignorant flip processes. These are flip processes in which the replacement graph is selected independently of the drawn graph.

Definition 4.1. *Let $k \in \mathbb{N}$. We say that a rule $\mathcal{R}$ of order k is* ignorant *if for any $F, F', H \in \mathcal{H}_k$ we have $\mathcal{R}_{F,H} = \mathcal{R}_{F',H}$. The* ignorant output density *is*

$$d := \frac{1}{\binom{k}{2}} \sum_{H \in \mathcal{H}_k} \mathcal{R}_{F,H} e(H)$$

(here F is arbitrary).

Observe that the Erdős–Rényi flip process is an ignorant flip process of order 2 with output distribution $\{\mathcal{R}_{F,H}\}_{H\in\mathcal{H}_2} = \boldsymbol{\delta}_{K_2}$ for each $F \in \mathcal{H}_2$.

Our main result about ignorant rules is Proposition 4.2, which explicitly describes their trajectories, destinations, speed of convergence and age.

Proposition 4.2. *Let $k \in \mathbb{N}$. Suppose that $\mathcal{R}$ is an ignorant rule of order k with ignorant output density d. Then for each graphon $W \in \mathcal{W}_0$ and $t \in I_W$ we have*

$$\Phi^t W = d + (W - d) \cdot e^{-(k)_2 t}. \tag{22}$$

Hence, we have $\mathrm{dest}(W) \equiv d$ *and*

$$\tau^-_{\mathcal{R}}(\delta) = \tau^+_{\mathcal{R}}(\delta) = \frac{1}{(k)_2} \ln\left(\frac{\max\{d, 1-d\}}{\delta}\right) \tag{23}$$

for each $0 < \delta < \frac{1}{2}$. Furthermore, writing $i_W := \operatorname{ess\,inf}(W)$ and $s_W := \operatorname{ess\,sup}(W)$, we have

$$\mathrm{age}(W) = \begin{cases} \frac{1}{(k)_2} \min\left\{\ln\frac{d}{d-i_W}\ ,\ \ln\frac{1-d}{s_W-d}\right\} & \text{if } i_W < d < s_W\ , \\ \frac{1}{(k)_2} \ln\frac{d}{d-i_W} & \text{if } i_W < d \text{ and } s_W \le d\ , \\ \frac{1}{(k)_2} \ln\frac{1-d}{s_W-d} & \text{if } i_W \ge d \text{ and } s_W > d\ , \\ \infty & \text{if } W \equiv d. \end{cases} \tag{24}$$

In particular, $\mathrm{age}(W)$ *is finite unless $W \equiv d$.*

Proof. Let $W \in \mathcal{W}_0$. We first compute the velocity $\mathfrak{V}_{\mathcal{R}}$ using the notation preceding equation (15). Since the distribution of $\mathbb{H}$ is independent of $(U_v, v \in [k])$, by (15) we have

$$\begin{aligned} \mathfrak{V}_{\mathcal{R}} W(x,y) &= \sum_{a\neq b} \left[\mathbb{P}\left(ab \in \mathbb{H} \,|\, U_a = x, U_b = y\right) - W(x,y)\right] \\ &= 2\mathbb{E}e(\mathbb{H}) - (k)_2 W(x,y) = (k)_2 \left(d - W(x,y)\right). \end{aligned}$$

It is clear that the initial condition $\Phi^0 W = W$ is satisfied. By the uniqueness of trajectories, it is enough to verify that the function on the right hand side of (22) satisfies the differential equation (18). For $x, y \in \Omega$ we set $f(t) = d + (W(x,y) - d) \cdot e^{-(k)_2 t}$. Observe that for each $t \ge 0$ we have

$$\frac{f(t+\varepsilon) - f(t)}{\varepsilon} \to f'(t) = (k)_2 \left(d - W(x,y)\right) e^{-(k)_2 t} = (k)_2 \left(d - f(t)\right)$$

as $\varepsilon \to 0$, where the convergence is uniform over all choices of $(x,y) \in \Omega^2$ by Taylor's theorem because W is bounded. This implies that (22) indeed describes the trajectory starting at W under the ignorant rule $\mathcal{R}$.

Clearly, all trajectories converge to the same destination, the constant-d graphon. By (22) we have $d_\square(\Phi^t W, d) \le \left\|\Phi^t W - d\right\|_\infty \le \max\{d, 1-d\} e^{-(k)_2 t}$, with both equalities attained

at the constant-0 graphon if $d \geq 1/2$ and at the constant-1 graphon if $d \leq 1/2$. Hence, we have $\tau^-_{\mathcal{R}}(\delta) = \tau^+_{\mathcal{R}}(\delta) = \frac{1}{(k)_2} \ln\left(\frac{\max\{d,1-d\}}{\delta}\right)$ for each $\delta \in (0, \frac{1}{2})$. Now observe that (22) tells us that for a point (x, y) where $W(x, y) < d$, the value of $\Phi^{-t}W$ crosses 0 at $t = \frac{1}{(k)_2} \ln \frac{d}{d-W(x,y)}$, and for a point (x, y) where $W(x, y) > d$, the value of $\Phi^{-t}W$ crosses 1 at $t = \frac{1}{(k)_2} \ln \frac{1-d}{W(x,y)-d}$. This implies (24). □

5. Stirring flip processes

All trajectories of ignorant flip processes converge to a constant given by the ignorant output density. Being a bit less ignorant this time, we introduce *stirring flip processes*, which take into account the edge count (but nothing else) of the drawn graph. In this section we focus on *balanced* stirring flip processes, in which the expected number of edges in the replacement graph is equal to the number of edges in the drawn graph. In Section 7 we will study *extremist flip processes*, which are "very unbalanced" stirring flip processes.

Definition 5.1. *Let $k \in \mathbb{N}$. We say that a rule $\mathcal{R}$ of order k is* stirring *if for each F and F' with $e(F) = e(F')$ and each H, we have $\mathcal{R}_{F,H} = \mathcal{R}_{F',H}$. We say that the rule $\mathcal{R}$ is* balanced *if for each $F \in \mathcal{H}_k$ we have $e(F) = \sum_{H \in \mathcal{H}_k} \mathcal{R}_{F,H} e(H)$.*

Two natural examples of balanced stirring rules are the rule where a drawn graph with ℓ edges is replaced by the Erdős–Rényi uniform random graph with ℓ edges and the rule where a drawn graph with ℓ edges is replaced by the Erdős–Rényi binomial random graph with edge probability $\ell/\binom{k}{2}$.

Our main result describes the trajectories of stirring flip processes in the most important balanced case.

Proposition 5.2. *Suppose that $\mathcal{R}$ is a balanced stirring rule of order $k \geq 3$ and let $W \in \mathcal{W}_0$. Then for every $t \in I_W$ we have*

$$\begin{aligned}\Phi^t W(x,y) = \|W\|_1 &+ (\deg_W(x) + \deg_W(y) - 2\|W\|_1) \cdot e^{-t(k-1)_2} \\ &+ (W(x,y) - \deg_W(x) - \deg_W(y) + \|W\|_1) \cdot e^{-t((k)_2-2)}\,.\end{aligned} \tag{25}$$

In particular, the edge density stays constant along the trajectory and $\mathrm{dest}(W) = \|W\|_1$.

Proof. Let $W \in \mathcal{W}_0$ and set $d := \|W\|_1$. The following is our key claim.

Claim 5.3. *Given a graphon $W \in \mathcal{W}_0$, for all $x, y \in \Omega$ we have*

$$\mathfrak{V}_{\mathcal{R}} W(x,y) = (2k-4)(\deg_W(x) + \deg_W(y)) + (k-2)_2 d - ((k)_2 - 2)\, W(x,y). \tag{26}$$

Before we prove this claim, let us first demonstrate how it implies our theorem. By the uniqueness of trajectories, it is enough to verify that (25) satisfies the differential equation

(18). Write $f_{(x,y)}(t)$ for the right-hand side of (25), set U_t to be the graphon given by $U_t(x,y) := f_{(x,y)}(t)$ and set $g_x(t) := \deg_{U_t}(x) = d + (\deg_W(x) - d)e^{-t(k-1)_2}$. Observe that

$$\begin{aligned} f'_{(x,y)}(t) &= -(k-1)_2(\deg_W(x) + \deg_W(y) - 2d)e^{-t(k-1)_2} \\ &\quad - ((k)_2 - 2)(W(x,y) - \deg_W(x) - \deg_W(y) + d)e^{-t((k)_2-2)} \\ &= ((k)_2 - 2)\,(d - f_{(x,y)}(t)) + 2(k-2)(\deg_W(x) + \deg_W(y) - 2d)e^{-t(k-1)_2} \\ &= (2k-4)(g_x(t) + g_y(t)) + (k-2)_2 d - ((k)_2 - 2)\,f_{(x,y)}(t)\,. \end{aligned}$$

Furthermore, for each $t \geq 0$ the convergence in $(f_{(x,y)}(t+\varepsilon) - f_{(x,y)}(t))/\varepsilon \to f'_{(x,y)}(t)$ is uniform over all choices of $(x,y) \in \Omega^2$ by Taylor's theorem because W is bounded. This implies that $\frac{\mathrm{d}}{\mathrm{d}t}U_t = \mathfrak{V}_{\mathcal{R}}U_t$ in the norm $\|\cdot\|_\infty$. Moreover, we have that $U_0 = W$. Hence, by uniqueness (25) indeed describes the trajectory starting at W under a balanced stirring rule.

It remains to prove Claim 5.3.

Proof of Claim 5.3. Let $W \in \mathcal{W}_0$ and $x, y \in \Omega$. Using the notation preceding equation (15), observe that

$$\begin{aligned} \mathbb{E}\,(e(\mathbb{F}) \,|\, U_1 = x, U_2 = y) &= \frac{1}{2}\sum_{a \neq b} \mathbb{P}\,(ab \in \mathbb{F} \,|\, U_1 = x, U_2 = y) \\ &= W(x,y) + (k-2)\,(\deg_W(x) + \deg_W(y)) + \binom{k-2}{2} d. \end{aligned} \tag{27}$$

Setting $q_m(ab) := \mathbb{P}\,(ab \in \mathbb{H} \,|\, e(\mathbb{F}) = m)$, we have

$$\sum_{a \neq b} q_m(ab) = 2\mathbb{E}\,(e(\mathbb{H}) \,|\, e(\mathbb{F}) = m) = 2m. \tag{28}$$

Since the distribution of $\mathbb{F}$ is invariant under permutation of vertices, for any distinct vertices a and b the quantity $\mathbb{P}\,(e(\mathbb{F}) = m \,|\, U_a = x, U_b = y)$ is the same; we denote it by $p_{x,y}(m)$. We have

$$\mathbb{P}\,(ab \in \mathbb{H} \,|\, U_a = x, U_b = y) = \sum_{m=0}^{\binom{k}{2}} p_{x,y}(m) q_m(ab). \tag{29}$$

By summing (29) over all pairs a, b of distinct vertices, interchanging the order of summation, and applying (28), we obtain

$$\begin{aligned} \sum_{a \neq b} \mathbb{P}\,(ab \in \mathbb{H} \,|\, U_a = x, U_b = y) &= \sum_{m=0}^{\binom{k}{2}} p_{x,y}(m) \sum_{a \neq b} q_m(ab) = \sum_{m=0}^{\binom{k}{2}} [p_{x,y}(m) \cdot 2m] \\ &= 2\mathbb{E}\,(e(\mathbb{F}) \,|\, U_1 = x, U_2 = y)\,. \end{aligned}$$

Now by recalling (15) and applying (27), we obtain

$$\begin{aligned}\mathfrak{V}_{\mathcal{R}}W(x,y) &= \sum_{a\neq b}\left[\mathbb{P}\left(ab\in\mathbb{H}\,|\,U_a=x,U_b=y\right)-W(x,y)\right],\\ &= 2\mathbb{E}\left(e(\mathbb{F})\,|\,U_1=x,U_2=y\right)-(k)_2W(x,y)\\ &= (2k-4)(\deg_W(x)+\deg_W(y))+(k-2)_2d-W(x,y)\left((k)_2-2\right)\end{aligned}$$

as required. ∎

This completes the proof of the proposition. □

5.1. **Age of a balanced stirring process.** Recall that the age of a graphon W is the infimum of times $t\geq 0$ for which the range of $\Phi^{-t}W$ is not wholly contained in $[0,1]$. That is, given a balanced stirring rule as in Proposition 5.2, we want to know the infimum of $t\geq 0$ where we have for some $x,y\in\Omega$ that

$$\begin{aligned}\|W\|_1+(\deg_W(x)+\deg_W(y)-2\|W\|_1)\cdot e^{t(k-1)_2}+\\ +(W(x,y)-\deg_W(x)-\deg_W(y)+\|W\|_1)\cdot e^{t((k)_2-2)}\end{aligned}\tag{30}$$

is more than 1 or less than 0. While there is no simple formula to deduce the age from (30) in general because we have two exponential terms with different coefficients and exponents, in the next proposition we give a formula for the age in two special cases: degree-regular graphons and graphons with a separable form. These correspond to the cases where one exponential term vanishes. Further, we show that the age is finite unless the graphon is constant.

Proposition 5.4. *Suppose that $\mathcal{R}$ is a balanced stirring rule of order $k\geq 3$. Then the following hold for all $W\in\mathcal{W}_0$. Write $d:=\|W\|_1$. If $W=d$, then* $\mathrm{age}(W)=\infty$. *Otherwise, we have* $\mathrm{age}(W)<\infty$ *and in particular:*

(i) *if W is degree-regular, then*

$$\mathrm{age}(W)=\frac{1}{(k)_2-2}\cdot\min\left\{\ln\frac{d}{d-\operatorname{ess\,inf}(W)}\,,\ \ln\frac{1-d}{\operatorname{ess\,sup}(W)-d}\right\};$$

(ii) *if W is of the form $W(x,y)=f(x)+f(y)$, then*

$$\mathrm{age}(W)=\frac{1}{(k-1)_2}\cdot\min\left\{\ln\frac{d}{d-\operatorname{ess\,inf}(W)}\,,\ \ln\frac{1-d}{\operatorname{ess\,sup}(W)-d}\right\}.$$

Proof. $W=d$ is clearly a fixed point, so we have $\mathrm{age}(W)=\infty$. We now show that $\mathrm{age}(W)=\infty$ only if $W=d$. Suppose that $\mathrm{age}(W)=\infty$. Since $(k-1)_2\neq(k)_2-2$, the expression in (30) can stay in $[0,1]$ for all $x,y\in\Omega$ only if the coefficients of both exponential terms are zero. The first one implies $\deg_W\equiv d$ and then the second one implies $W\equiv d$.

Hence, we further assume that W is not constant, which implies $\operatorname{ess\,inf} W < d < \operatorname{ess\,sup} W$. In the case (i), the equation (30) simplifies to

$$d + (W(x,y) - d) \cdot e^{t((k)_2 - 2)} ,$$

and then the formula for age indeed expresses the time when the values of $\Phi^{-t}W$ first cross one of the two thresholds 0 and 1.

The case (ii) is equivalent to $W(x,y) = \deg_W(x) + \deg_W(y) - d$. Hence (30) simplifies to

$$d + (W(x,y) - d) \cdot e^{t(k-1)_2} ,$$

which implies the age in the same way as in the case (i). □

6. Complementing flip processes

The *complementing rule of order k* is the rule of order k where the replacement graph is the complement of the drawn graph. Our main result about complementing rules is Proposition 6.1, which explicitly describes their trajectories, destinations, speed of convergence and age.

Proposition 6.1. *Suppose that $\mathcal{R}$ is the complementing rule of order $k \geq 2$. Then for each graphon $W \in \mathcal{W}_0$ and $t \in I_W$ we have*

$$\Phi^t W(x,y) = \frac{1}{2} + \left(W(x,y) - \frac{1}{2} \right) \cdot e^{-2(k)_2 t} . \tag{31}$$

Hence, we have $\operatorname{dest}(W) \equiv 1/2$ *and*

$$\tau^-_{\mathcal{R}}(\delta) = \tau^+_{\mathcal{R}}(\delta) = \frac{-\ln(2\delta)}{2(k)_2} \tag{32}$$

for each $0 < \delta < \frac{1}{2}$. Furthermore, writing $i_W := \operatorname{ess\,inf}(W)$ and $s_W := \operatorname{ess\,sup}(W)$, we have

$$\operatorname{age}(W) = \begin{cases} \frac{1}{2(k)_2} \min\left\{ \ln \frac{1}{1-2i_W} \,,\, \ln \frac{1}{2s_W - 1} \right\} & \text{if } i_W < 1/2 < s_W \,, \\ \frac{1}{2(k)_2} \ln \frac{1}{1-2i_W} & \text{if } i_W < 1/2 \text{ and } s_W \leq 1/2 \,, \\ \frac{1}{2(k)_2} \ln \frac{1}{2s_W - 1} & \text{if } i_W \geq 1/2 \text{ and } s_W > 1/2 \,, \\ \infty & \text{if } W \equiv 1/2. \end{cases} \tag{33}$$

Proof. We calculate the velocity using the notation preceding the formula (15). Note that $ab \in \mathbb{H}$ if and only if $ab \notin \mathbb{F}$, so

$$\mathbb{P}\left(ab \in \mathbb{H} \,|\, U_a = x, U_b = y\right) = \mathbb{P}\left(ab \notin \mathbb{F} \,|\, U_a = x, U_b = y\right) = 1 - W(x,y) .$$

Hence, by (15) we have

$$\mathfrak{V}_{\mathcal{R}}W(x,y) = \sum_{1\le a\neq b\le k} (1 - W(x,y) - W(x,y)) = (k)_2(1-2W(x,y)) \,.$$

Write $f_{(x,y)}(t)$ for the right-hand side of (31). Observe that for each $t \ge 0$ we have

$$\frac{f_{(x,y)}(t+\varepsilon) - f_{(x,y)}(t)}{\varepsilon} \to f'_{(x,y)}(t) = (k)_2(1-2W(x,y))e^{-2(k)_2 t} = (k)_2\left(1 - 2f_{(x,y)}(t)\right)$$

as $\varepsilon \to 0$, where the convergence is uniform over all choices of $(x,y) \in \Omega^2$ by Taylor's theorem because W is bounded. Since the trajectory $\Phi^{\cdot}W$ is unique by Theorem 3.2, we have $\Phi^t W(x,y) = f_{(x,y)}(t)$ as desired.

Clearly, the unique destination is the constant-$\frac{1}{2}$ graphon. By (22) we have $d_\square(\Phi^t W, \frac{1}{2}) \le \left\|\Phi^t W - \frac{1}{2}\right\|_\infty \le \frac{1}{2}e^{-2(k)_2 t}$, with both equalities attained at the constant-0 and constant-1 graphons. Hence, we have $\tau^-_{\mathcal{R}}(\delta) = \tau^+_{\mathcal{R}}(\delta) = \frac{-\ln(2\delta)}{2(k)_2}$ for each $\delta \in (0, \frac{1}{2})$. Note that (31) tells us that for a point (x,y) where $W(x,y) < \frac{1}{2}$, the value of $\Phi^{-t}W$ reaches 0 at $t = -\frac{1}{2(k)_2}\ln(1-2W(x,y))$, and for a point (x,y) where $W(x,y) > \frac{1}{2}$, the value of $\Phi^{-t}W$ reaches 1 at $t = -\frac{1}{2(k)_2}\ln(2W(x,y)-1)$. This implies (33). □

We conjecture that the complementing flip process exhibits the fastest convergence among flip processes whose trajectories converge to a single destination. Observe that the convergence time of the complementing flip process (see (32)) is half that of an ignorant process with output density 1/2 (cf. (23)).

Conjecture 6.2. *Suppose that $\mathcal{R}$ is a convergent flip process of order k with a single destination. Then for each $\delta \in (0, \frac{1}{2})$ we have $\tau^-_{\mathcal{R}}(\delta) \ge \frac{-\ln(2\delta)}{2(k)_2}$.*

7. Extremist flip processes

In this section we define the extremist rules and prove results about their behaviour.

Definition 7.1 (Extremist rule)**.** *Let $k \in \mathbb{N}$. The* extremist *rule $\mathcal{R}$ of order k is defined by*

$$\mathcal{R}_{F,H} = \begin{cases} 1 & \text{if } e(F) < \frac{1}{2}\binom{k}{2} \text{ and } H = \overline{K_k}, \\ 1 & \text{if } e(F) > \frac{1}{2}\binom{k}{2} \text{ and } H = K_k, \\ 1 & \text{if } e(F) = \frac{1}{2}\binom{k}{2} \text{ and } H = F, \\ 0 & \text{otherwise .} \end{cases} \tag{34}$$

A possible variation of the extremist rule is to permit the replacement of graphs with $\binom{k}{2}/2$ edges by a clique or empty graph with equal probabilities. In fact, most proofs from this section (except for Theorem 7.6) work for this variation, but for the sake of simplicity we shall not consider it.

7.1. **Behaviour on constants.** We begin by considering the trajectories for extremist rules on constant graphons. Recall that a trajectory starting with a constant graphon takes values only among constant graphons and the velocity is a constant function (see [12, Corollary 6.6]). Therefore, we shall identify the constant-p kernel with the number p. We prove the following proposition, which implies that the trajectory of an extremist rule which starts from the constant-p graphon converges to 0 if $p < 1/2$, to 1 if $p > 1/2$ and is fixed for $p = 1/2$.

Proposition 7.2. *If $\mathcal{R}$ is the extremist rule of order $k \geq 3$, then we have $\mathfrak{V}_{\mathcal{R}}(p) = 0$ for $p \in \{0, 1/2, 1\}$, $\mathfrak{V}_{\mathcal{R}}(p) < 0$ for $p \in (0, 1/2)$ and $\mathfrak{V}_{\mathcal{R}}(p) > 0$ for $p \in (1/2, 1)$.*

Proof. Since $\mathfrak{V}_{\mathcal{R}}(p)$ is a constant kernel, we have $\mathfrak{V}_{\mathcal{R}}(p) \equiv \int_{\Omega^2} \mathfrak{V}_{\mathcal{R}}(p)$. Therefore, by (17) and denoting by $\mathbb{G}$ the random graph $\mathbb{G}(k, p)$, we have

$$\mathfrak{V}_{\mathcal{R}}(p) = 2\left[\tbinom{k}{2}\mathbb{P}\left(e(\mathbb{G}) > \tbinom{k}{2}/2\right) + \tbinom{k}{2}/2 \cdot \mathbb{P}\left(e(\mathbb{G}) = \tbinom{k}{2}/2\right)\right] - (k)_2 p\,.$$

Writing $n = \binom{k}{2}$ and $B_{n,p}$ for a binomial random variable with parameters n and p, we have

$$\begin{aligned}\mathfrak{V}_{\mathcal{R}}(p) &= 2n\mathbb{P}(B_{n,p} > n/2) + n\mathbb{P}(B_{n,p} = n/2) - 2np \\ &= n\big(\mathbb{P}(B_{n,p} > n/2) + \mathbb{P}(B_{n,p} \geq n/2) - 2p\big)\,.\end{aligned}$$

It is clear that $\mathfrak{V}_{\mathcal{R}}(p) = 0$ for $p \in \{0, 1/2, 1\}$. To prove the proposition, it is enough to show that for the function

$$f_n(p) := \mathbb{P}(B_{n,p} > n/2) + \mathbb{P}(B_{n,p} \geq n/2)\,,$$

its derivative f_n' is strictly increasing on $(0, 1/2)$ and strictly decreasing on $(1/2, 1)$. Indeed, this implies that $p \mapsto f_n(p)$ is strictly convex on $(0, 1/2)$ and strictly concave on $(1/2, 1)$. Then, these properties are inherited by $p \mapsto \mathfrak{V}_{\mathcal{R}}(p)$, thereby implying that $\mathfrak{V}_{\mathcal{R}}(p)$ is negative on $(0, 1/2)$ and positive on $(1/2, 1)$. Observe that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p}\left(\mathbb{P}(B_{n,p} = k)\right) &= \frac{\mathrm{d}}{\mathrm{d}p}\left(\binom{n}{k}p^k(1-p)^{n-k}\right) \\ &= \binom{n}{k}\left(kp^{k-1}(1-p)^{n-k} - (n-k)p^k(1-p)^{n-k-1}\right) \\ &= n\left[\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k} - \binom{n-1}{k}p^k(1-p)^{n-k-1}\right] \\ &= n\left[\mathbb{P}(B_{n-1,p} = k-1) - \mathbb{P}(B_{n-1,p} = k)\right]\,.\end{aligned}$$

Therefore, we have

$$\begin{aligned} f_n'(p) &= n \sum_{k>n/2} \left[\mathbb{P}(B_{n-1,p} = k-1) - \mathbb{P}(B_{n-1,p} = k)\right] \\ &\quad + n \sum_{k\geq n/2} \left[\mathbb{P}(B_{n-1,p} = k-1) - \mathbb{P}(B_{n-1,p} = k)\right] \\ &= n \cdot \left(\mathbb{P}\left(B_{n-1,p} = \left\lfloor \frac{n-1}{2} \right\rfloor\right) + \mathbb{P}\left(B_{n-1,p} = \left\lceil \frac{n-1}{2} \right\rceil\right)\right) \\ &= n \cdot \begin{cases} 2\mathbb{P}(B_{n-1,p} = \lfloor \frac{n-1}{2} \rfloor), & n \text{ odd} \\ \mathbb{P}(B_{n-1,p} = \lfloor \frac{n-1}{2} \rfloor)\left(1 + \frac{p}{1-p}\right) & n \text{ even} \end{cases} \\ &= (1 + \mathbb{I}_{n \text{ odd}})\, n \cdot \binom{n-1}{\lfloor \frac{n-1}{2} \rfloor} (p(1-p))^{\lfloor \frac{n-1}{2} \rfloor}, \end{aligned}$$

Now $k \geq 3$ implies that $\lfloor \frac{n-1}{2} \rfloor \geq 1$, so the polynomial $f_n'(p)$ inherits from the parabola $p(1-p)$ the property of being strictly increasing on $(0, 1/2)$ and strictly decreasing on $(1/2, 1)$. This completes our proof. □

7.2. **Density thresholds.** In the previous subsection we saw that for constant graphons we have $\mathrm{dest}_{\mathcal{R}}(W) = 0$ as soon as the density of W is below $1/2$. In this subsection we consider when we start with an arbitrary initial graphon W and investigate what density guarantees $\mathrm{dest}_{\mathcal{R}}(W) = 0$. We define the *density threshold* for the extremist rule $\mathcal{R}$ of order $k \geq 3$ to be

$$\theta(k) := \sup\{p \in [0,1] : \forall W \in \mathcal{W}_0 \ \|W\|_1 \leq p \text{ implies } \mathrm{dest}_{\mathcal{R}}(W) = 0\}.$$

In the following proposition, we collect a few useful properties related to the density threshold.

Proposition 7.3. *The following holds for the extremist rule $\mathcal{R}$ of order $k \geq 3$.*

(i) *For every $W \in \mathcal{W}_0$ and every $t \in I_W$ we have $\Phi^t(1-W) = 1 - \Phi^t W$. In particular, if $\mathrm{dest}_{\mathcal{R}}(W) = 0$, then $\mathrm{dest}_{\mathcal{R}}(1-W) = 1$.*

(ii) *For each $W \in \mathcal{W}_0$ satisfying $\|W\|_1 > 1 - \theta(k)$ we have $\mathrm{dest}_{\mathcal{R}}(W) = 1$.*

(iii) *$\theta(k) \leq 1/2$.*

Proof. First we show (i). From formula (15) it is easy to see that for any graphon W we have $\mathfrak{V}(1-W) = -\mathfrak{V} W$, so $\frac{\mathrm{d}}{\mathrm{d}t}(1 - \Phi^t W) = -\mathfrak{V}\,\Phi^t W = \mathfrak{V}(1 - \Phi^t(W))$, that is, $t \mapsto 1 - \Phi^t W$ is a trajectory. Since it starts at the same graphon as the trajectory $t \mapsto \Phi^t(1-W)$, by uniqueness (see Theorem 3.2) it follows that $\Phi^t(1-W) = 1 - \Phi^t W$ for every $t \in I_W$. The second conclusion is obvious.

To prove (ii), let $W \in \mathcal{W}_0$ satisfy $\|W\|_1 > 1 - \theta(k)$. Since $\|1 - W\|_1 = 1 - \|W\|_1 < \theta(k)$, we have $\mathrm{dest}_{\mathcal{R}}(1-W) = 0$ and therefore (i) implies $\mathrm{dest}_{\mathcal{R}}(W) = 1$.

Part (iii) follows from the fact that the constant-1/2 graphon W satisfies $\mathrm{dest}_{\mathcal{R}}(W) = 1/2$. □

We prove the following lower bound on the density threshold. In particular, we show that the density threshold is strictly positive for $k \geq 5$.

Theorem 7.4. *Let $k \geq 5$ be an integer and define $\sigma(k) := \max\{2^{-4-24/(k-5)}, 1/2 - \sqrt{48/k}\}$ for $k \geq 6$ and $\sigma(5) = 1/30$. For the extremist rule $\mathcal{R}$ of order k we have that*

$$\theta(k) \geq \sigma(k) > 0\,.$$

Proof. The following is our key claim.

Claim 7.5. *Given a graphon $U \in \mathcal{W}_0$ satisfying $\|U\|_1 \leq \sigma(k)$, we have*

$$\int \mathfrak{V}_{\mathcal{R}} U \,\mathrm{d}\pi^2 \leq -\binom{k}{2} \|U\|_1\,. \tag{35}$$

Before we prove this claim, let us first demonstrate how it implies our theorem. Fix $W \in \mathcal{W}_0$ satisfying $\|W\|_1 \leq \sigma(k)$. Consider the function $f(t) := \left\|\Phi^t W\right\|_1$ defined for $t \in [0, \infty)$. Aiming to apply Lemma 2.2 with $D = \sigma(k)$, we note that $f(0) \leq \sigma(k)$ and $f(t) \geq 0$ for all $t \geq 0$. We argue that $f'(t) = \int \mathfrak{V}_{\mathcal{R}} \Phi^t W \,\mathrm{d}\pi^2$. Indeed, we have that

$$f'(t) = \lim_{\varepsilon \to 0} \frac{\left\|\Phi^{t+\varepsilon} W\right\|_1 - \left\|\Phi^t W\right\|_1}{\varepsilon} = \lim_{\varepsilon \to 0} \int \frac{\Phi^{t+\varepsilon} W - \Phi^t W}{\varepsilon} \,\mathrm{d}\pi^2\,.$$

Now by Lemma 3.7 the integrand is $O(\varepsilon)$-close to $\mathfrak{V}_{\mathcal{R}} \Phi^t W$ in the L^∞-norm (and thus in the L^1-norm), so in the limit we get $f'(t) = \int \mathfrak{V}_{\mathcal{R}} \Phi^t W \,\mathrm{d}\pi^2$. Together with (35) we get that $f'(t) \leq -\binom{k}{2} f(t)$ whenever $f(t) \leq \sigma(k)$, so by Lemma 2.2 we have $f(t) \leq f(0) e^{-\binom{k}{2} t}$ for all $t \in [0, \infty)$. Therefore, it follows that $\mathrm{dest}_{\mathcal{R}}(W) = 0$ and we conclude that $\theta(k) \geq \sigma(k) > 0$.

It remains to prove Claim 7.5.

Proof of Claim 7.5. Set $p := \|U\|_1$. Sample $\mathbb{F} \sim \mathbb{G}(k, U)$ and sample $\mathbb{H}$ according to the distribution $\mathbb{P}(\mathbb{H} = H \,|\, \mathbb{F} = F) = \mathcal{R}_{F,H}$. We first obtain an upper bound for $\mathbb{E}e(\mathbb{H})$. Since $\mathbb{E}(e(\mathbb{H}) \,|\, \mathbb{F} = F) = 0$ for any $F \in \mathcal{H}_k$ with fewer than $\frac{1}{2}\binom{k}{2}$ edges, we have

$$\mathbb{E}e(\mathbb{H}) \leq \binom{k}{2} \mathbb{P}\left(e(\mathbb{F}) \geq \frac{1}{2}\binom{k}{2}\right)\,. \tag{36}$$

Now suppose that we have

$$\mathbb{P}\left(e(\mathbb{F}) \geq \frac{1}{2}\binom{k}{2}\right) \leq \frac{p}{2}\,. \tag{37}$$

Then by (17) and (36) we conclude that

$$\int \mathfrak{V}_{\mathcal{R}} U \,\mathrm{d}\pi^2 = 2\mathbb{E}e(\mathbb{H}) - (k)_2 \cdot p \leq p\left(\binom{k}{2} - (k)_2\right) = -p\binom{k}{2}\,,$$

which yields (35). Hence, it remains to prove (37).

We begin with $k = 5$ and $p \le 1/30$. Note that if $\mathbb{F}$ has at least $\frac{1}{2}\binom{k}{2} = 5$ edges then it has a matching with 2 edges. There are $\binom{5}{2}\binom{3}{2}/2 = 15$ such matchings in K_5 and each of them appears in $\mathbb{F}$ with probability p^2 (as edges occur independently with probability p), so we have

$$\mathbb{P}\left(e(\mathbb{F}) \ge \frac{1}{2}\binom{k}{2}\right) \le \mathbb{P}\left(\mathbb{F} \text{ contains a 2-matching}\right) \le 15p^2 \le \frac{p}{2}\,.$$

Next, we consider $k \ge 6$ and $p \le 2^{-4-24/(k-5)}$. Let r be the chromatic index of K_k. It is well-known that $r = \binom{k}{2}/\left\lfloor \frac{k}{2} \right\rfloor \in \{k-1, k\}$. Hence, we can partition the edges of K_k into disjoint maximum matchings $M_1, \dots, M_r$. If $e(\mathbb{F}) \ge \binom{k}{2}/2$ then there exists $i \in [r]$ such that

$$e(\mathbb{F} \cap M_i) \ge \left\lceil \frac{\binom{k}{2}}{2r} \right\rceil = \left\lceil \frac{\lfloor k/2 \rfloor}{2} \right\rceil = \left\lfloor \frac{k+2}{4} \right\rfloor =: k'\,.$$

Therefore, using $r \le k \le 2^{k/2}$ and $\binom{\lfloor k/2 \rfloor}{k'} \le 2^{k/2}$, we have

$$\mathbb{P}\left(e(\mathbb{F}) \ge \frac{1}{2}\binom{k}{2}\right) \le \sum_{i \in [r]} \mathbb{P}\left(e(\mathbb{F} \cap M_i) \ge k'\right) \le r\binom{\lfloor k/2 \rfloor}{k'} p^{k'} \le 2^k p^{(k-1)/4} \le \frac{p}{2}\,.$$

Finally, we consider $2^{-4-24/(k-5)} \le p \le 1/2 - \sqrt{48/k}$; here we may assume that $k \ge 192$. We consider the function $f(G) := e(G)/(v(G)-1)$. Since f varies by at most 1 under the editing of edges incident to any fixed vertex, by [18, Theorem 10.3] we obtain

$$\mathbb{P}\left(e(\mathbb{F}) \ge \frac{1}{2}\binom{k}{2}\right) = \mathbb{P}\left(f(\mathbb{F}) \ge \mathbb{E}f(\mathbb{F}) + (\tfrac{1}{2} - p)k/2\right) \le \exp\left(-\frac{(1/2-p)^2 k}{8}\right)\,.$$

Now we have $e > 2$ and $\frac{(1/2-p)^2 k}{8} \ge 6 \ge 5 + \frac{24}{k-5}$, so we obtain

$$\mathbb{P}\left(e(\mathbb{F}) \ge \binom{k}{2}/2\right) \le 2^{-5-24/(k-5)} \le \frac{p}{2}$$

as required. ∎

This completes the proof of Theorem 7.4. □

Now we show that the results of Theorem 7.4 cannot be extended to $k \in \{3, 4\}$.

Theorem 7.6. $\theta(3) = \theta(4) = 0$.

Proof. Let $\mathcal{R}$ be the extremist rule of order $k \in \{3, 4\}$. Define

$$B_1 := \left\{(F, a, b) : F \in \mathcal{H}_k, e(F) < \frac{1}{2}\binom{k}{2}, ab \in E(F)\right\} \quad \text{and}$$

$$B_2 := \left\{(F, a, b) : F \in \mathcal{H}_k, e(F) > \frac{1}{2}\binom{k}{2}, ab \notin E(F)\right\}\,.$$

By (14) and (34), for each $W \in \mathcal{W}_0$ and each $(x,y) \in \Omega^2$ we have

$$\mathfrak{V}_{\mathcal{R}} W(x,y) = \sum_{(F,a,b)\in B_2} t_{\mathrm{ind}}^{(x,y)}(F^{a,b}, W) - \sum_{(F,a,b)\in B_1} t_{\mathrm{ind}}^{(x,y)}(F^{a,b}, W)\,. \tag{38}$$

We shall construct a sequence $(W_n)_{n\in\mathbb{N}}$ of graphons so that $\|W_n\|_1 \to 0$ as $n \to \infty$ and for each $n \in \mathbb{N}$ we have $\left\|\Phi^t W_n\right\|_\square \nrightarrow 0$ as $t \to \infty$. Fix an arbitrary sequence $(\Omega_n)_{n\in\mathbb{N}}$ of subsets of Ω such that for each $n \in \mathbb{N}$ we have $\pi(\Omega_n) = 1/n$ and $\Omega_{n+1} \subseteq \Omega_n$. For each $n \in \mathbb{N}$ set $S_n := ((\Omega \setminus \Omega_n) \times \Omega_n) \cup (\Omega_n \times (\Omega \setminus \Omega_n))$ and $\mathcal{A}_n := \{W \in \mathcal{W}_0 : W\restriction_{S_n} = 1\}$. For each $n \in \mathbb{N}$ define the graphon $W_n := \mathbb{I}_{S_n}$. Note that $\|W_n\|_1 = \frac{2}{n}\left(1 - \frac{1}{n}\right) \to 0$ as $n \to \infty$.

Fix $n \in \mathbb{N}$. We have $W_n \in \mathcal{A}_n$. By Theorem 3.2 there is a unique trajectory $\Phi^t W_n$ on $[0,\infty)$. Let $U \in \mathcal{A}_n$. For any $(F,a,b) \in B_1 \cup B_2$ and $(x,y) \in S_n$ we have $t_{\mathrm{ind}}^{(x,y)}(F^{a,b}, U) = 0$, so by (38) we have $(\mathfrak{V}_{\mathcal{R}} U)\restriction_{S_n} = 0$. Now by applying Proposition 3.5 with $\mathcal{A}_n$ we have $\Phi^t W_n \in \mathcal{A}_n$ for all $t \in [0,\infty)$. We have $\left\|\Phi^t W_n\right\|_\square \geq \frac{2}{n}\left(1 - \frac{1}{n}\right)$ for all $t \in [0,\infty)$; in particular, $\left\|\Phi^t W_n\right\|_\square \nrightarrow 0$ as $t \to \infty$. Hence, we have $\theta(3) = \theta(4) = 0$. □

We also prove an upper bound on the density threshold.

Theorem 7.7. *There exists $K \in \mathbb{N}$ such that for every $k \geq K$, setting $C_k := (k)_2 2^{\binom{k}{2}-1}$, we have $\theta(k) \leq \frac{1}{2} - \frac{1}{2\cdot 10^5 C_k}$.*

Proof. Let C be the constant returned by Theorem 2.14 and set $K := 400C^2$. Let $\mathcal{R}$ be the extremist rule of order $k \geq K$ and set $f(k) := \frac{1}{2\cdot 10^5 C_k}$. Suppose for a contradiction that $\theta(k) > 1/2 - f(k)$.

Set $\delta := \frac{1}{2}\sqrt{1 - 2\theta(k) + 2f(k)}$ and $\gamma := 1/2 - \delta$. Fix a subset $\Omega' \subseteq \Omega$ such that $\pi(\Omega') = \gamma$. Set $S := ((\Omega \setminus \Omega') \times \Omega') \cup (\Omega' \times (\Omega \setminus \Omega'))$. Let $W := \mathbb{I}_S$. Note that $\|W\|_1 = \theta(k) - f(k) < \theta(k)$. Sample $\mathbb{F} \sim \mathbb{G}(k, W)$ and sample $\mathbb{H}$ according to the distribution $\mathbb{P}(\mathbb{H} = H \,|\, \mathbb{F} = F) = \mathcal{R}_{F,H}$. Note that

$$\mathbb{E}e(\mathbb{H}) \geq \binom{k}{2}\mathbb{P}\left(e(\mathbb{F}) > \frac{1}{2}\binom{k}{2}\right). \tag{39}$$

Set $Y := e(\mathbb{F})$. By the structure of W we have $B \sim \mathrm{Bin}(k,\gamma)$ such that $Y = B(k-B)$. Write $B = \sum_{i=1}^k B_i$ as the sum of k independent random variables $B_i \sim \mathrm{Ber}(\gamma)$. Set $X_i := B_i - \gamma$ for $i \in [k]$ and $X := \sum_{i=1}^k X_i$. We have

$$\begin{aligned}
\mathbb{P}\left(Y > \frac{1}{2}\binom{k}{2}\right) = \mathbb{P}\left(\left|B - \frac{k}{2}\right| < \frac{\sqrt{k}}{2}\right) &\geq \mathbb{P}\left(|X| < \frac{\sqrt{k}}{2} - k\delta\right) \\
&\geq \mathbb{P}\left(\frac{|X|}{\sqrt{\mathrm{Var}(X)}} < \frac{1 - \sqrt{4kf(k)}}{\sqrt{1 - 4f(k)}}\right) \\
&\geq \mathbb{P}\left(\frac{|X|}{\sqrt{\mathrm{Var}(X)}} < 0.99\right).
\end{aligned}$$

Now by Theorem 2.14 we have

$$\mathbb{P}\left(\frac{|X|}{\sqrt{\mathrm{Var}(X)}} < 0.99\right) \geq \mathbb{P}\left(Z < 0.99\right) - \frac{1}{9} \geq 0.55,$$

where $Z \sim N(0,1)$ is a standard Gaussian random variable. Then by (17) and (39) we have

$$\int \mathfrak{V}\, W \,\mathrm{d}\pi^2 \geq (k)_2 \left(\mathbb{P}\left(Y > \frac{1}{2}\binom{k}{2}\right) - \|W\|_1\right) \geq \frac{(k)_2}{100}. \tag{40}$$

Set $T := \frac{1}{200(k)_2 C_k}$. By (21) in Lemma 3.7 we have

$$\int \left(\Phi^T W - W - T\mathfrak{V}_{\mathcal{R}} W\right) \,\mathrm{d}\pi^2 \leq \left\|\Phi^T W - W - T\mathfrak{V}_{\mathcal{R}} W\right\|_\infty \leq C_k (k)_2^2 T^2,$$

so by (40) we have

$$\begin{aligned}\left\|\Phi^T W\right\|_1 = \int \Phi^T W \,\mathrm{d}\pi^2 &\geq \int W \,\mathrm{d}\pi^2 + T\int \mathfrak{V}\, W \,\mathrm{d}\pi^2 - C_k (k)_2^2 T^2 \\ &\geq \|W\|_1 + \frac{1}{2\cdot 10^4 C_k} - \frac{1}{4\cdot 10^4 C_k} \geq \frac{1}{2} + \frac{1}{8\cdot 10^4 C_k} > 1 - \theta(k).\end{aligned}$$

Hence, by Proposition 7.3(ii) we have $\mathrm{dest}(W) = \mathrm{dest}(\Phi^T W) = 1$. On the other hand, $\|W\|_1 < \theta(k)$ implies that $\mathrm{dest}(W) = 0$. This is a contradiction, so we must have $\theta(k) \leq 1/2 - f(k)$. □

7.3. **Trajectory rescaling for large k.** Observe that the order of a rule can significantly restrict the maximum rate of change in its trajectories. Indeed, suppose that $\mathcal{R}$ is a rule of order $k \in \mathbb{N}$. Then, by (20) we have $-(k)_2 \leq \mathfrak{V}_{\mathcal{R}} W \leq (k)_2$ for all graphons W. To adjust for the fact that larger values of k permit greater rates of change, we shall introduce the following limit notion of rescaling. Let $(\mathcal{R}_k)_{k\geq 2}$ be a sequence of rules such that $\mathcal{R}_k$ is a rule of order k for each $k \geq 2$. We say that $(\mathcal{R}_k)_{k\geq 2}$ is *rescalable* at a graphon W if the Banach-space limit

$$\lim_{k\to\infty} \Phi_{\mathcal{R}_k}^{t/(k)_2} W$$

exists for all $t \in [0,\infty)$. Then, we write $\Psi^t(W) := \lim_{k\to\infty} \Phi_{\mathcal{R}_k}^{t/(k)_2} W$ and call this a *rescaled trajectory*. Let $\mathcal{B} \subseteq \mathcal{W}_0$ be the set of graphons at which $(\mathcal{R}_k)_{k\geq 2}$ is rescalable. We define the *rescaling core* of $(\mathcal{R}_k)_{k\geq 2}$, which we denote by $\mathcal{B}^\bullet$, to be the set of elements in the interior of $\mathcal{B}$ (with respect to the topology induced by the cut norm) at which the maps $W \mapsto \Psi^t(W)$, for $t \geq 0$, are continuous (with respect to the topology induced by the cut norm). In particular, the rescaling core of $(\mathcal{R}_k)_{k\geq 2}$ is a set of graphons W such that $\Psi^t(W)$ is a good approximation for $\Phi_{\mathcal{R}_k}^{t/(k)_2}(\tilde{W})$ when k is sufficiently large and $\tilde{W}$ is sufficiently close to W. (Let us note that in principle, we might investigate rescaled trajectories outside of the rescaling core as well, and this indeed seems to be doable using

the main result from [13]. However, the favourable continuity property is lost in this case, and in particular such rescaled trajectories would have no connection to flip processes on finite graphs, as provided by Theorem 1.1.)

Theorem 7.8 characterizes the rescaling core of the sequence of extremist rules and explicitly describes the rescaled trajectories.

Theorem 7.8. *The rescaling core of the sequence $(\mathcal{R}_k)_{k\geq 2}$ of extremist rules is the set of graphons whose density is not $\frac{1}{2}$. The following holds for each graphon $W \in \mathcal{W}_0$.*

(i) *If $\|W\|_1 > \frac{1}{2}$ then $\Psi^t W = 1 - e^{-t} \cdot (1 - W)$ for each $t \geq 0$.*
(ii) *If $\|W\|_1 < \frac{1}{2}$ then $\Psi^t W = e^{-t} \cdot W$ for each $t \geq 0$.*

We only prove Theorem 7.8(i). Indeed, then Theorem 7.8(ii) follows by symmetry. Further, the set of graphons of density other than $\frac{1}{2}$ is open and the maps $W \mapsto \Psi^t(W)$ are continuous. So, it only needs to be argued that the rescaling core does not contain any graphon of density $\frac{1}{2}$. This is because such a graphon can be approximated by graphons of density more than $\frac{1}{2}$ and less than $\frac{1}{2}$ at the same time, which are mapped by $\Psi^3(\cdot)$ to a graphon of density more than 0.9 and less than 0.1, respectively, demonstrating a discontinuity.

Proof of Theorem 7.8(i). Fix $\varepsilon > 0$ and consider an extremist rule $\mathcal{R}_k$ of sufficiently large order k. We write $\mathfrak{V}_k$ and Φ_k for the velocity operator and the trajectory of $\mathcal{R}_k$, respectively. Let $\mathcal{U}$ denote the set of all graphons with density at least $\frac{1}{2} + \varepsilon$. We begin with the following claim, which approximates $\mathfrak{V}_k$ for graphons in $\mathcal{U}$.

Claim 7.9. *If $U \in \mathcal{U}$ then we have*

$$\|\mathfrak{V}_k U - (k)_2 \cdot (1 - U)\|_\infty = o_k(k^2) \, .$$

Proof. Set $d' := \|U\|_1$. Observe that we have

$$\sum_{F \in \mathcal{H}_k : ab \notin E(F)} T_F^{a,b} W(x,y) = \sum_{F \in \mathcal{H}_k : ab \notin E(F)} t_{\mathrm{ind}}^{(x,y)}(F^{a,b}, W) = 1 - W(x,y) \, . \tag{41}$$

Indeed, recall the integral definition of rooted densities $t_{\mathrm{ind}}^{(x,y)}(G^{a,b}, W)$ given in Definition 2.5. Consider a graph $F \in \mathcal{H}_k$ with a nonedge $cd \neq ab$ and its counterpart $F \cup cd$. The integrands for F and $F \cup cd$ differ only in the multiplicative factor corresponding to cd, which are $1 - W(x_c, x_d)$ and $W(x_c, x_d)$ for F and $F \cup cd$ respectively. These two factors sum to 1, so adding the two integrands together has the effect of 'cancelling' the factor corresponding to cd. Repeating this 'pairing' argument over all pairs $cd \neq ab$, we find that the sum of integrands over all $F \in \mathcal{H}_k$ with $ab \notin E(F)$ is precisely $1 - W(x,y)$.

For $1 \le a \ne b \le k$ and a labelled graph J on $[k] \setminus \{a,b\}$, let $\mathcal{F}_J^{a,b} \subset \mathcal{H}_k$ be the set of all graphs on $[k]$ from which the removal of a and b yields J. Observe that we have

$$t_{\mathrm{ind}}(J,U) = \sum_{F \in \mathcal{F}_J^{a,b}} t_{\mathrm{ind}}^{(x,y)}(F^{a,b},U)\,. \tag{42}$$

For $\alpha,\beta \in [0,1]$ and $\ell \in \mathbb{N}$, let $\mathcal{H}_\ell^{\alpha,\beta} \subset \mathcal{H}_\ell$ be the set of all graphs whose density is $\alpha \pm \beta$. Since $t_{\mathrm{ind}}(J,U)$ represents the probability that $\mathbb{G}(k-2,U) = J$, Lemma 2.8 tells us that

$$\sum_{J \in \mathcal{H}_{k-2} \setminus \mathcal{H}_{k-2}^{d',\varepsilon/3}} t_{\mathrm{ind}}(J,U) \le \mathbb{P}\left(\delta_\square(\mathbb{G}(k-2,U),U) \ge \frac{\varepsilon}{2}\right) = o_k(1/k^2) \tag{43}$$

Since any graph $F \in \mathcal{H}_k^{d',\varepsilon/2}$ has density greater than $\frac{1}{2}$, we have $\mathcal{R}_{F,K_k} = 1$. On the other hand, since k is sufficiently large and at most $2k = o_k(k^2)$ edges are incident to a or b, if a graph F does not have density $d' \pm \varepsilon/2$, then by removing vertices a and b from F we obtain a graph J which does not have density $d' \pm \varepsilon/3$ and $F \in \mathcal{F}_J^{a,b}$. Hence, we may rewrite (14) as

$$\begin{aligned}
\mathfrak{V}_k U(x,y) &= \sum_{1 \le a \ne b \le k} \sum_{F \in \mathcal{H}_k : ab \notin E(F)} (T_F^{a,b}U)(x,y) \pm 2 \sum_{1 \le a \ne b \le k} \sum_{F \in \mathcal{H}_k \setminus \mathcal{H}_k^{d',\varepsilon/2}} (T_F^{a,b}U)(x,y) \\
&\boxed{\text{by (41)}} \quad = (k)_2 \cdot (1 - U(x,y)) \pm 2 \sum_{1 \le a \ne b \le k} \sum_{J \in \mathcal{H}_{k-2} \setminus \mathcal{H}_{k-2}^{d',\varepsilon/3}} \sum_{F \in \mathcal{F}_J^{a,b}} t_{\mathrm{ind}}^{(x,y)}(F^{a,b},U) \\
&\boxed{\text{by (42) and (43)}} \quad = (k)_2 \cdot (1 - U(x,y)) + o_k(1)\,,
\end{aligned}$$

finishing the proof. ∎

Now we apply Claim 7.9 to show that trajectories which start in $\mathcal{U}$ are confined to $\mathcal{U}$.

Claim 7.10. *If $U \in \mathcal{U}$, then we have $\Phi_k^t U \in \mathcal{U}$ for all $t \ge 0$.*

Proof. Let $U \in \mathcal{U}$ and suppose that $\Phi_k^t U \notin \mathcal{U}$ for some $t \ge 0$. Set $T := \inf\{t \ge 0 : \Phi_k^t U \notin \mathcal{U}\}$. By the continuity of $\Phi_k^t U$, the graphon $\Phi_k^T U$ has density $\frac{1}{2} + \varepsilon$. Set $S = \left((k)_2^3 2^{\binom{k}{2}}\right)^{-1}$ and let $0 < s \le S$. By (21) for s and Claim 7.9, we have

$$\left\| \frac{\Phi_k^{T+s}U - \Phi_k^T U}{s} - (k)_2 \cdot (1 - \Phi_k^T U) \right\|_\infty = o_k(k^2).$$

Then, we obtain

$$\left\|\Phi_k^{T+s}U\right\|_1 = \left\|\Phi_k^T U\right\|_1 + s((k)_2 \left\|1 - \Phi_k^T U\right\|_1 - o_k(k^2)) \ge \left\|\Phi_k^T U\right\|_1 = \frac{1}{2} + \varepsilon,$$

so we have $\Phi_k^{T+s}U \in \mathcal{U}$. This contradicts the definition of T, so $\Phi_k^t U \in \mathcal{U}$ for all $t \ge 0$. ∎

Let $W \in \mathcal{U}$. We first consider the integral equation $\Gamma_k^\tau := W + \int_{s=0}^{\tau} \mathfrak{Z}_k(\Gamma^s)\,\mathrm{d}s$, where $\mathfrak{Z}_k(U) = (k)_2 \cdot (1-U)$. The solution of this integral equation is

$$\Gamma_k^\tau = 1 - e^{-(k)_2\tau} \cdot (1-W)\ .$$

Recall that in fact $\Phi_k^\cdot W$ is the solution of the integral equation (2), that is,

$$\Phi_k^\tau W = W + \int_0^\tau \mathfrak{V}_k\left(\Phi_k^s W\right)\,\mathrm{d}s\ .$$

Now $\mathfrak{Z}_k(\cdot)$ is $(k)_2$-Lipschitz, both $\Gamma_k^\cdot$ and $\Phi_k^\cdot W$ are confined to $\mathcal{U}$ and Claim 7.9 tells us that $\|\mathfrak{V}_k(\cdot) - \mathfrak{Z}_k(\cdot)\|_\infty = o_k(k^2)$ on $\mathcal{U}$, so Lemma 2.3 gives us

$$\|\Phi_k^\tau W - \Gamma_k^\tau\|_\infty \le o_k(k^2)\tau \cdot e^{(k)_2\tau}\ .$$

Hence, we conclude that for any fixed $t \ge 0$, we have that $\Phi_k^{t/(k)_2} W$ converges uniformly to $1 - e^{-t}\cdot(1-W)$ as $k \to \infty$. □

7.4. **Stabilization in an extremist flip process.** Let us consider a flip process of order k with rule $\mathcal{R}$. Let G be an n-vertex graph ($n \ge k$), and $G_0 := G, G_1, G_2, \ldots$ be an evolution of the graph G with respect to the flip process with rule $\mathcal{R}$. The *stabilization time* $\kappa_{\mathcal{R}}(G)$ is a random variable defined as $\kappa_{\mathcal{R}}(G) = \min\{\ell : G_\ell = G_{\ell+1} = G_{\ell+2} = \ldots\}$ if it exists and $\kappa_{\mathcal{R}}(G) = \infty$ otherwise. For example, if $\mathcal{R}_{F,F} = 0$ for all $F \in \mathcal{H}_k$ then each step of the flip process changes the graph and $\kappa_{\mathcal{R}}(G) = \infty$ almost surely.

We now consider the concept of stabilization time in the context of an extremist flip process of a given order $k \ge 3$. Suppose that the initial graph has n vertices. It can be shown that the stabilization time is finite almost surely, and further, if $n \ge n_0(k)$ is sufficiently large, stabilization occurs (almost surely) at the moment the flip process reaches the edgeless or the complete graph (which are graphs at which stabilization obviously occurs).

We leave it as an interesting question to find universal bounds on stabilization times.

Question 7.11. *For each $k \ge 3$ with the corresponding extremist rule $\mathcal{R}_k$, we seek a function $b_k : \mathbb{N} \to \mathbb{N}$ growing as slowly as possible so that*

$$\min_{G \in \mathcal{H}_n} \mathbb{P}\left(\kappa_{\mathcal{R}_k}(G) \le b_k(n)\right) \to 1, \qquad \text{as } n \to \infty\ .$$

It seems very likely that quasirandom graphs of edge density around $\frac{1}{2}$ have the largest stabilization time, and that we have to take $b_k(n) \approx \exp(n^C)$ for some $C \in (0,2]$. In particular, the number of steps goes far beyond $\Theta(n^2)$, and so this question is outside the framework of our graphon trajectories.[d]

[d]Recall Theorem 1.1.

8. Monotone flip processes

Broadly speaking, monotone flip processes are those where the number of edges does not decrease in any step. Formally, we define a few different notions of monotonicity.

Definition 8.1. *Let $\mathcal{R}$ be a rule of order $k \in \mathbb{N}$. We say that $\mathcal{R}$ is*

- non-decreasing *if for each $F \in \mathcal{H}_k$ we have $\mathcal{R}_{F,H} = 0$ whenever H is not a (labelled) supergraph of F.*
- increasing *if it is non-decreasing and in addition for each $F \in \mathcal{H}_k \setminus \{K_k\}$ there exists a proper supergraph H of F with $\mathcal{R}_{F,H} > 0$.*
- density-non-decreasing *if for each $F \in \mathcal{H}_k$ we have $\sum_{H \in \mathcal{H}_k} \mathcal{R}_{F,H} e(H) \geq e(F)$.*
- density-increasing *if it is density-non-decreasing and further for each $F \in \mathcal{H}_k \setminus \{K_k\}$ the above inequality is strict.*

In particular, a non-decreasing rule is density-non-decreasing and an increasing rule is density-increasing.

The following proposition asserts that non-decreasing and density-increasing flip processes are convergent, but this need not be true for density-non-decreasing flip processes.

Proposition 8.2.

(i) *Each trajectory of a non-decreasing rule $\mathcal{R}$ has a destination.*

(ii) *Let $\mathcal{R}$ be a density-increasing rule. The destination of each trajectory is the constant-1 graphon. The age of each graphon except the constant-1 graphon is finite.*

(iii) *There is a density-non-decreasing rule with a non-convergent trajectory.*

Proof of Proposition 8.2(i). Let $W \in \mathcal{W}_0$ be a graphon. Observe that since $\mathcal{R}$ is non-decreasing, the negative terms of (14) have coefficient zero. Hence, the velocity is pointwise non-negative. In other words, for each point (x, y), the function $t \mapsto \Phi^t W(x, y)$ is non-decreasing. In this case it is obvious that the destination is the pointwise limit of $\Phi^t W$. □

Proof of Proposition 8.2(ii). Suppose $\mathcal{R}$ is of order k. Let

$$\delta = \min_{F \in \mathcal{H}_k \setminus \{K_k\}} \sum_{H \in \mathcal{H}_k} \mathcal{R}_{F,H}(e(H) - e(F)).$$

Since $\mathcal{R}$ is density-increasing, we have $\delta > 0$. We shall show that for every graphon W we have

$$\int \mathfrak{V}\, W \, \mathrm{d}\pi^2 \geq 2\delta(1 - t(K_k, W)) \,. \tag{44}$$

Let $(\mathbb{F}, \mathbb{H})$ be a pair of random graphs in $\mathcal{H}_k$ such that $\mathbb{F} \sim \mathbb{G}(k, W)$ and $\mathbb{P}(\mathbb{H} = H \,|\, \mathbb{F} = F) = \mathcal{R}_{F,H}$ for all $F, H \in \mathcal{H}_k$. For $F \neq K_k$ we have

$$\mathbb{E}(e(\mathbb{H}) \,|\, \mathbb{F} = F) = \sum_{H \in \mathcal{H}_k} \mathcal{R}_{F,H} e(H) \geq e(F) + \delta. \tag{45}$$

Now by (16) we have

$$\begin{aligned}\int \mathfrak{V}\, W \,\mathrm{d}\pi^2 &= 2\mathbb{E}\left[e(\mathbb{H}) - e(\mathbb{F})\right] = 2\sum_{F\in\mathcal{H}_k} t(F,W)\mathbb{E}\left(e(\mathbb{H}) - e(\mathbb{F}) \,|\, \mathbb{F} = F\right)\\ &= 2\sum_{F\in\mathcal{H}_k} t(F,W)\left[\mathbb{E}\left(e(\mathbb{H}) \,|\, \mathbb{F} = F\right) - e(F)\right]\\ &\overset{(45)}{\geq} 2\sum_{F\neq K_k} t(F,W)\delta = 2\left(1 - t(K_k,W)\right)\delta\,,\end{aligned}$$

which shows (44).

The growth given by (44) allows us to prove the statement. Indeed, let U be an arbitrary graphon. Let $f : [-\mathrm{age}(U), +\infty) \to [0,1]$ be defined by $f(t) := \int \Phi^t U \,\mathrm{d}\pi^2$. Then (44) gives

$$\frac{\mathrm{d}}{\mathrm{d}t} f(t) \geq 2\delta(1 - t(K_k, \Phi^t U)) \overset{\text{Prop 2.13}}{\geq} 2\delta(1 - f(t)^{k/2})\,.$$

Firstly, this implies that $f(t) \to 1$ as $t \to \infty$, i.e., that the destination of U is the constant-1 graphon. Secondly, it gives that $f(t)$ is increasing, and hence for $t < 0$, we have that $\frac{\mathrm{d}}{\mathrm{d}t} f(t) \geq 2\delta(1 - f(0)^{2/k})$. That means that unless $f(0) = 1$, the derivative of f in the negative domain is bounded from below by a positive constant. Since $f \in [0,1]$, this means that the age is finite. □

Sketch of proof of Proposition 8.2(iii). We build on Section 6.10 of [12] where a flip process with a periodic trajectory was constructed. We briefly recall this construction. Fix a partition $\Omega = \Omega_1 \cup \Omega_2$ with $\pi(\Omega_1) = \pi(\Omega_2) = 1/2$. We take k large and $\alpha > 0$ small (probably $k = 2000$ and $\alpha = 0.01$ would work). Fix $r = 0.1$, $a := 0.2$, $b := 0.8$. Let $\mathsf{circle}(a, b; r)$ be the circle centered at the point (a, b) and of radius r. For input $F \in \mathcal{H}_k$ the rule is as follows. If F consists of two components C_1 and C_2, each of order $(\frac{1}{2} \pm \alpha)k$, say C_1 of density d_1 and C_2 of density d_2 with the condition that the point (d_1, d_2) is within distance of α from $\mathsf{circle}(a, b; r)$ then we add or delete edges into C_1 and C_2 so that the expected change of edges in C_1 and C_2 is given by the first and the second coordinate of $(g + h)(d_1, d_2)$, respectively. Here, $g, h : \mathbb{R}^2 \setminus \{(a, b)\} \to \mathbb{R}^2$ are defined as follows,

$$\begin{aligned} g(x,y) &= \frac{1}{\sqrt{(-y+b)^2 + (x-a)^2}} \cdot (-y + b, x - a)\,,\\ h(x,y) &= \left(\frac{r}{\sqrt{(x-a)^2 + (y-b)^2}} - 1\right) \cdot (x - a, y - b)\,.\end{aligned}$$

No edges between C_1 and C_2 are added. For other Fs, the rule is idle. Observe that values of g are tangent vectors of circles centered at (a, b). Furthermore, h vanishes on $\mathsf{circle}(a, b; r)$ and pushes towards $\mathsf{circle}(a, b; r)$ elsewhere.

Fix $(x, y) \in \mathsf{circle}(a, b; r)$. Consider a graphon U taking value x on $\Omega_1 \times \Omega_1$, value y on $\Omega_2 \times \Omega_2$ and 0 elsewhere. If it were true that for each such graphon we would have with probability 1 that the drawn F has the two components as above and $d_1 = x$ and $d_2 = y$, then $\mathsf{circle}(a, b; r)$ would indeed be a periodic trajectory of the flip process. However, because the law of large numbers is not perfect, the above statement holds only with probability $1 - o_k(1)$ and only with $d_1 = x + o_k(1)$ and $d_2 = y + o_k(1)$. In particular, we cannot conclude that $\mathsf{circle}(a, b; r)$ is a periodic trajectory. This is where the correcting function h comes into the picture. In particular, since h pushes the trajectory towards $\mathsf{circle}(a, b; r)$ in case of any substantial deviations, in [12] it is argued using the Poincaré-Bendixson Theorem that there exists a periodic trajectory close to $\mathsf{circle}(a, b; r)$.

The above flip process is clearly not density-non-decreasing. However, we can modify it as follows. Partition $\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_{\text{reservoir}}$ with $\pi(\Omega_1) = \pi(\Omega_2) = 1/4$ and $\pi(\Omega_{\text{reservoir}}) = 1/2$. Now, the rule is as follows. If the drawn graph F is such that it consists of three components, say C_1, C_2 and $C_{\text{reservoir}}$, such that C_1 and C_2 have orders $(\frac{1}{4} \pm \alpha)k$ and $C_{\text{reservoir}}$ has order $(\frac{1}{2} \pm \alpha)k$ and the densities of C_1 and C_2 are as above[e] and the density of $C_{\text{reservoir}}$ is between 0.2 and 0.8 then we add or delete edges from C_1 and C_2 according to $(g + h)(d_1, d_2)$ and then add or delete edges from $C_{\text{reservoir}}$ so that the total number of edges in this step does not change. If F is not of this form then we have an idle step.

Since the number of edges does not change, the flip process is non-density-decreasing. Starting with a graphon W whose values on $\Omega_i \times \Omega_j$ $(i, j \in [2])$ are defined as for U, and further which is constant-$\frac{1}{2}$ on $\Omega^2_{\text{reservoir}}$ and zero on $(\Omega_1 \cup \Omega_2) \times \Omega^2_{\text{reservoir}}$ we see that the densities on Ω_1^2 and on Ω_2 evolve close to $\mathsf{circle}(a, b; r)$ while $\Omega_{\text{reservoir}}$ just serves to keep the overall density constant. In particular, this trajectory does not have a destination. □

Our next result says that for a non-decreasing rule the limit of the destinations of a convergent sequence of graphons is greater than or equal to the destination of the limit.

Proposition 8.3. *Let $\mathcal{R}$ be a non-decreasing rule of order $k \in \mathbb{N}$ and $(U_n)_{n\in\mathbb{N}}$ be a sequence of graphons which converges to a graphon U in the cut norm distance. Suppose that the sequence $(\text{dest}(U_n))_{n\in\mathbb{N}}$ converges in the cut norm distance to a graphon S. Then $S \geq \text{dest}(U)$.*

Remark 8.4. *Observe that the converse inequality does not hold in general. Indeed, for each $k \geq 3$ consider a flip process of order k in which a drawn graph containing a single edge and $k - 2$ isolated vertices is replaced by K_k while the step is idle for other drawn graphs. Let $(U_n)_{n\in\mathbb{N}}$ be the sequence of graphons where U_n is the constant-$\frac{1}{n}$ graphon. On the one hand, this sequence of graphons converges to the constant-0 graphon U; here we*

[e]Note that in this setting, the components C_1, C_2 and $C_{\text{reservoir}}$ are uniquely determined.

have $\mathrm{dest}(U) = 0$. *On the other hand,* $(\mathrm{dest}(U_n))_{n\in\mathbb{N}}$ *is a sequence of constant-*1 *graphons, which clearly converges to the constant-*1 *graphon.*

Nevertheless, we show in Theorem 10.7 that we do have equality for removal flip processes (which are also monotone, though non-increasing).

Proof of Proposition 8.3. We shall prove that for arbitrary sets $A, B \subset \Omega$ and an arbitrary $\varepsilon > 0$ we have $\int_{A\times B} \mathrm{dest}(U) \leq \int_{A\times B} S + \varepsilon$ (how this implies the proposition, is an exercise in measure theory which we leave to the reader). To this end, let T be such that $d_\square(\mathrm{dest}(U), \Phi^T U) < \varepsilon/2$. Define $\delta := \varepsilon/(4\exp(C_{\square,k}T))$, where $C_{\square,k}$ is from Theorem 3.8. Then by Theorem 3.8, for every U_i for which $d_\square(U_i, U) < \delta$, we have $d_\square(\Phi^T U_i, \Phi^T U) < \varepsilon/2$. In particular, this means that $\int_{A\times B} \mathrm{dest}(U) \leq \int_{A\times B} \Phi^T U_i + \varepsilon$. Since we have a non-decreasing flip process, $\Phi^T U_i \leq \mathrm{dest}(U_i)$ and hence $\int_{A\times B} \mathrm{dest}(U) \leq \int_{A\times B} \mathrm{dest}(U_i) + \varepsilon$. As $\mathrm{dest}(U_1), \mathrm{dest}(U_2), \ldots \xrightarrow{\square} S$, we get $\int_{A\times B} \mathrm{dest}(U) \leq \int_{A\times B} S + \varepsilon$ as needed. □

9. Component completion flip processes

In this section we define the component completion rule and prove Theorem 9.6, which characterizes the destinations of the component completion rule. The component completion rule completes the components of the drawn graph F by adding edges between all non-edge pairs of vertices in the same connected component of F. We give a formal definition below.

Definition 9.1 (Component completion graph)**.** *Let* $k \in \mathbb{N}$ *and* $F \in \mathcal{H}_k$. *For distinct* $a, b \in [k]$ *write* $a \sim_F b$ *to mean that there is a path between* a *and* b *in* F. *The* component completion graph $\mathcal{C}(F)$ *of* F *is the labelled graph on vertex set* $[k]$ *with edge set* $\{ab : a, b \in [k], a \neq b, a \sim_F b\}$.

Definition 9.2 (Component completion rule)**.** *Let* $k \in \mathbb{N}$. *A rule* $\mathcal{R}$ *of order* k *is* component completion *if for all* $F, H \in \mathcal{H}_k$ *we have*

$$\mathcal{R}_{F,H} = \begin{cases} 1 & \text{if } H = \mathcal{C}(F), \\ 0 & \text{otherwise .} \end{cases} \tag{46}$$

To prepare for the statement and proof of Theorem 9.6, we give definitions of connectedness, component decomposition and component completeness for graphons.

Definition 9.3 (Connectedness)**.** *A graphon* W *is* disconnected *if* $W = 0$ *or there is a subset* $A \subseteq \Omega$ *with* $\pi(A) \in (0, 1)$ *such that* $W = 0$ *a.e. on* $A \times (\Omega \setminus A)$. *Otherwise,* W *is* connected.

Definition 9.4 (Component decomposition)**.** *A* component decomposition *of a graphon* W *is a partition* $\Omega = \bigsqcup_{i=0}^{K} \Omega_i$ *into measurable sets* Ω_i *with* $0 \leq K \leq \infty$ *such that the following hold.*

(GC1) $\alpha_i := \pi(\Omega_i) > 0$ *for all* $i \geq 1$ *but* Ω_0 *may be empty.*
(GC2) For all $i \geq 1$ *the graphon* $W{\restriction_{\Omega_i^2}}$ *on* $\big(\Omega_i, (\alpha_i^{-1}\pi){\restriction_{\Omega_i}}\big)$ *is connected.*
(GC3) $W = 0$ *a.e. on* $\Omega^2 \setminus \Big(\bigsqcup_{i=1}^K \Omega_i^2\Big)$.

Definition 9.5 (Component completeness)**.** *A graphon* W *is* component complete *if it has a component decomposition* $\Omega = \bigsqcup_{i=0}^K \Omega_i$ *such that* $W{\restriction_{\Omega_i^2}} = 1$ *for all* $i \geq 1$.

Theorem 9.6 characterizes the destinations of the component completion rule.

Theorem 9.6. *If* $\mathcal{R}$ *is a component completion rule of order* $k \geq 3$*, then the following hold.*

(i) *Every trajectory of* $\mathcal{R}$ *has a destination.*
(ii) *Every destination of* $\mathcal{R}$ *is component complete. Moreover, the destination of a graphon with component decomposition* $\Omega = \bigsqcup_{i=0}^K \Omega_i$ *is component complete with the same component decomposition.*
(iii) *Every component complete graphon is a destination of* $\mathcal{R}$.

To prepare for the proof of Theorem 9.6, we state two lemmas. We begin with Lemma 5.5 from [14] about component decomposition for graphons.

Lemma 9.7. *Every graphon has a component decomposition.*

Next, we give a technical lemma which, roughly speaking, says that every graphon with zero density of 3-vertex induced paths is $\{0,1\}$-valued and in fact complete if connected. Its proof is given at the end of this section.

Lemma 9.8. *Let* W *be a graphon which satisfies*

$$\int_{\Omega^3} (1 - W(x,y))W(x,z)W(y,z)\,\mathrm{d}\pi^3 = 0. \tag{47}$$

Then $W(x,y) \in \{0,1\}$ *a.e. on* Ω^2*. If in addition* W *is connected, then* $W = 1$.

Now we shall prove Theorem 9.6. Let us briefly describe the key ideas. The action of the non-decreasing rule $\mathcal{R}$ is driven by the addition of edges between non-edge pairs of vertices which are connected by a path. Hence, the component decomposition of the initial graphon remains intact. Since our result is concerned with destinations of trajectories, which are a form of limit behaviour, 3-vertex induced paths are particularly relevant because every induced path between non-adjacent vertices contains a 3-vertex induced path. Lemma 9.8 implies that the components of a destination must be complete.

Proof of Theorem 9.6. Since $\mathcal{R}$ is a non-decreasing rule, every trajectory of $\mathcal{R}$ has a destination by Proposition 8.2(i). Let

$$B := \{(F,a,b) : F \in \mathcal{H}_k, ab \notin E(F), a \sim_F b\}.$$

By (14) and (46), for each $W \in \mathcal{W}_0$ and each $(x,y) \in \Omega^2$ we have

$$\mathfrak{V}_{\mathcal{R}} W(x,y) = \sum_{(F,a,b)\in B} t_{\mathrm{ind}}^{(x,y)}(F^{a,b}, W). \tag{48}$$

Let $U \in \mathcal{W}_0$, with a fixed component decomposition $\Omega = \bigsqcup_{i=0}^{K} \Omega_i$ of U given by Lemma 9.7. We set $\Lambda_i := (\Omega_i \times (\Omega \setminus \Omega_i)) \cup ((\Omega \setminus \Omega_i) \times \Omega_i)$ for $i \in [K]$ and $\Lambda_0 := (\Omega_0 \times \Omega) \cup (\Omega \times \Omega_0)$. Set $\Lambda := \Omega^2 \setminus \left(\bigsqcup_{i=1}^{K} \Omega_i^2\right) = \bigsqcup_{i=0}^{K} \Lambda_i$. For each $i \in [K]$ we show that the component completion rule does not change the value of W at points $(x,y) \in \Lambda_i$, that is, pairs of vertices between Ω_i and its complement. This is intuitive because the rule would behave in the same way on graphs.

For $i \geq 0$ set $\mathcal{A}_i := \{W \in \mathcal{W}_0 : W{\restriction}_{\Lambda_i} = 0\}$. We show that a trajectory starting at a graphon $W \in \mathcal{A}_i$ stays in $\mathcal{A}_i$ for every $i \geq 0$. Indeed for any $(F,a,b) \in B$ and $(x,y) \in \Lambda_i$ we have that $t_{\mathrm{ind}}^{(x,y)}(F^{a,b}, W) = 0$ exactly because any copy of $F^{a,b}$ sampled from W with $x = a$ and $y = b$ sends the vertices a and b into different components. Hence, by (48) we have $(\mathfrak{V}_{\mathcal{R}} W){\restriction}_{\Lambda_i} = 0$. Then, by applying Proposition 3.5 with $\mathcal{A}_i$ for each $i \in [K] \cup \{0\}$ we have $\Phi^t U = 0$ on Λ for all $t \in [0,\infty)$. Since $\Phi^t U \xrightarrow{\square} \mathrm{dest}(U) =: S$, it follows that $S = 0$ a.e. on Λ. By (48), for all $t \in [0,\infty)$ we have $\mathfrak{V}_{\mathcal{R}} \Phi^t U \geq 0$ pointwise. Since $S = U + \int_{[0,\infty)} \mathfrak{V}_{\mathcal{R}} \Phi^s U \, \mathrm{d}s$, we have $S \geq U$. Hence, for all $i \geq 1$ the graphon $S{\restriction}_{\Omega_i^2}$ on $\left(\Omega_i, (\alpha_i^{-1}\pi){\restriction}_{\Omega_i}\right)$ remains connected.

Define

$$Q := \left\{(x,y) \in \Omega^2 : (1 - S(x,y)) \int_{z\in\Omega} S(x,z)S(y,z) > 0\right\}.$$

We aim to prove that Q has measure zero. To this end, consider any $(x,y) \in Q$. Let H be the labelled graph on $\{1,2,3\}$ with edge set $\{13,23\}$. Observe that we have $t_{\mathrm{ind}}^{(x,y)}(H^{1,2}, S) > 0$ and

$$t_{\mathrm{ind}}^{(x,y)}(H^{1,2}, S) = \sum_{F} t_{\mathrm{ind}}^{(x,y)}(F^{1,2}, S),$$

where the sum is over all labelled graphs $F \in \mathcal{H}_k$ which induce a labelled copy of H on $\{1,2,3\}$. Hence, there is a labelled graph $F \in \mathcal{H}_k$ which induces a labelled copy of H on $\{1,2,3\}$ such that $t_{\mathrm{ind}}^{(x,y)}(F^{1,2}, S) > 0$. This implies that $\mathfrak{V}_{\mathcal{R}} S(x,y) > 0$. Now since S satisfies $\mathfrak{V}_{\mathcal{R}} S = 0$ by Proposition 3.9 and $\mathfrak{V}_{\mathcal{R}} S(x,y) > 0$ for all $(x,y) \in Q$, the set $Q \subseteq \Omega^2$ has measure zero. Hence, we have (47) for S and $S{\restriction}_{\Omega_i^2}$ for all $i \geq 1$ and so by Lemma 9.8 we have $S = 1$ a.e. on Ω_i^2 for all $1 \leq i \leq K$. Therefore, S is a component complete graphon with component decomposition $\Omega = \bigsqcup_{i=0}^{K} \Omega_i$.

Finally, by (ii) a component complete graphon is clearly its own destination. □

To close this section, we provide the proof of Lemma 9.8.

Proof of Lemma 9.8. Define the following sets.

$$\begin{aligned}
M &:= \{(x,y) \in \Omega^2 : W(x,y) \in (0,1)\},\\
N_1 &:= \left\{(x,y) \in \Omega^2 : \int_{z\in\Omega} W(x,z)W(y,z)\,\mathrm{d}\pi > 0\right\},\\
N_2 &:= \left\{(x,y) \in \Omega^2 : \int_{z\in\Omega} W(x,z)(1-W(y,z))\,\mathrm{d}\pi > 0\right\},\\
N &:= N_1 \cup N_2,\\
P &:= \{x \in \Omega : (x,y) \in M \setminus N \text{ for some } y \in \Omega\}.
\end{aligned}$$

For each $(x,y) \notin N$ we have $\int_{z\in\Omega} W(x,z)\,\mathrm{d}\pi = 0$. By Fubini's theorem we have

$$0 \le \int_{M\setminus N} W(x,y)\,\mathrm{d}\pi^2 \le \int_P \int_\Omega W(x,y)\,\mathrm{d}\pi(y)\,\mathrm{d}\pi(x) = 0.$$

Hence, we have $\pi^2(M \setminus N) = 0$. By (47) and Fubini's theorem we have $\pi^2(M \cap N_1) = \pi^2(M \cap N_2) = 0$, so we have $\pi^2(M) = 0$, i.e. $W(x,y) \in \{0,1\}$ a.e. on Ω^2.

Now suppose that W is connected. For each $x \in \Omega$ define the following.

$$\begin{aligned}
S_x &:= \{y \in \Omega : W(x,y) > 0\}, \qquad S_x^+ := S_x \cup \{x\},\\
P_x &:= \{(y,z) \in S_x^+ \times (\Omega \setminus S_x^+) : W(y,z) > 0\},\\
I(x) &:= \int_\Omega W(x,y)\,\mathrm{d}\pi(y).
\end{aligned}$$

In addition, we define the following.

$$\begin{aligned}
S &:= \{(x,y) \in \Omega^2 : W(x,y) > 0\}, \qquad J := \{x \in \Omega : I(x) > 0\},\\
P &:= \{(x,y,z) \in \Omega^3 : x \in J, (y,z) \in P_x\},\\
Q &:= \{(x,y,z) \in \Omega^3 : x \in J, (y,z) \in S_x^2 \setminus S\}.
\end{aligned}$$

Set $f(x,y,z) := (1-W(x,y))W(x,z)W(y,z)$ for $(x,y,z) \in \Omega^3$.

Since W is connected, we have $\pi^2(S) = \int_S \mathrm{d}\pi^2 > 0$. Now by Fubini's theorem we have $\int_\Omega I(x)\,\mathrm{d}\pi(x) = \int_{\Omega^2} W(x,y)\,\mathrm{d}\pi^2 > 0$, so we have $\pi(J) > 0$. Let $x \in J$. Since $I(x) > 0$, we have $\pi(S_x) > 0$. Suppose that $\pi(S_x^+) < 1$. Since W is connected, we have $\pi^2(P_x) > 0$. By Fubini's theorem we have

$$\int_P \mathrm{d}\pi^3 = \int_J \int_{P_x} \mathrm{d}\pi^2(y,z)\,\mathrm{d}\pi(x) > 0,$$

so we have $\pi^3(P) > 0$. Since $f(x,z,y) > 0$ for all $(x,y,z) \in P$, we have $\int_{\Omega^3} f(x,z,y)\,\mathrm{d}\pi^3 > 0$. This contradicts (47), so we have $\pi(S_x) = \pi(S_x^+) = 1$.

Now suppose that $\pi^2(\Omega^2 \setminus S) > 0$. Since $\pi(S_x^2) = 1$, we have $\pi^2(S_x^2 \setminus S) > 0$. By Fubini's theorem we have

$$\int_Q \mathrm{d}\pi^3 = \int_J \int_{S_x^2 \setminus S} \mathrm{d}\pi^2(y,z)\,\mathrm{d}\pi(x) > 0,$$

so we have $\pi^3(Q) > 0$.

Since $f(y,z,x) > 0$ for all $(x,y,z) \in Q$, we have $\int_Q f(y,z,x)\,\mathrm{d}\pi^3 > 0$. This contradicts (47), so $\pi^2(S) = 1$. Now since $W(x,y) \in \{0,1\}$ a.e. on Ω^2, in fact we have $W = 1$. □

10. Removal flip processes

In a removal flip process, we have a graph $F \in \mathcal{H}_k$ which we remove if it is contained in the drawn graph while drawing an F-free graph leads to an idle step. A formal definition follows.

Definition 10.1. *Let $k \in \mathbb{N}$ and $F \in \mathcal{H}_k$. A rule $\mathcal{R}$ of order k is* F-removal *if for each $H \in \mathcal{H}_k$ which contains F as a labelled subgraph we have $\mathcal{R}_{H,H-F} = 1$ while for each H which does not contain F as a labelled subgraph we have $\mathcal{R}_{H,H} = 1$. Of course, all the remaining entries of the matrix $\mathcal{R}$ are 0.*

Remark 10.2. *Observe that Definition 10.1 is concerned with the removal of* non-induced *copies of F. While we could define even flip processes removing induced copies, such a concept would not have the nice properties we study here. For example, the dual*[f] *to the flip process from Remark 8.4 shows that for such flip processes we do not have Theorem 10.7.*

Remark 10.3. *Consider Definition 10.1 for the labelled two-edge path F with $E(F) = \{12, 23\}$. Observe that only in one out of three samples of P_2 will that path actually be removed; if a triangle is drawn, then two of its three edges are removed. While this may seem to be at odds with the spirit of P_2-removal, this is not actually the case; indeed, eventually all copies of P_2 are removed.*

Remark 10.4. *It is not entirely apparent how to define removal rules which remove a family of graphs. For example, how should we define a removal rule of order 5 which removes the 4-leaf star and the pentagon at the same time? One possibility would be to remove all edges that belong to some 4-leaf star or to some pentagon. We leave it to the reader whether they find this definition natural and appealing.*

Given a graph $F \in \mathcal{H}_k$ we say that a graphon W is *F-free* if $t(F,W) = 0$. In the rest of this section we write $(i,j) \in E(F)$ for any *ordered* pair of vertices that forms an edge of a

[f] The *dual* of a flip process $\mathcal{R}$ is a flip process $\mathcal{R}^*$ of the same order in which $\mathcal{R}^*_{F,H}$ is defined using the graph complements $\overline{F}$ and $\overline{H}$, $\mathcal{R}^*_{F,H} := \mathcal{R}_{\overline{F},\overline{H}}$.

graph F. We substitute into (14) and see that for the F-removal rule $\mathcal{R}$, we have

$$\mathfrak{V}_{\mathcal{R}}W(x,y) = \sum_{(i,j)\in E(F)} \sum_{G\in\mathcal{H}_k:F\subseteq G} t^{x,y}_{\mathrm{ind}}(G^{i,j},W)\cdot(-1) = -\sum_{(i,j)\in E(F)} t^{x,y}(F^{(i,j)},W). \tag{49}$$

Most interesting features about removal flip processes concern their destinations and the way they are approached. Observe that a removal flip process is non-increasing and thus by (the dual version of) Proposition 8.2(i) each graphon has a destination. We have the following fact about these destinations.

Fact 10.5. *If $\mathcal{R}$ is the F-removal rule and $e(F)\geq 1$, then*

$$\{\mathrm{dest}_{\mathcal{R}}(W) : W\in\mathcal{W}_0\} = \{W\in\mathcal{W}_0 : W \text{ is } F\text{-free}\}\,.$$

Proof. Indeed, if W is not F-free, then $\mathfrak{V}\,W\neq 0$ and Proposition 3.9 tells us that W is not a destination. The converse implication is clear. □

We give a useful identity. For the sake of brevity, we write $\Phi^{\infty}W = \mathrm{dest}(W)$.

Lemma 10.6. *Let $k\in\mathbb{N}$, $F\in\mathcal{H}_k$ and $\mathcal{R}$ be the F-removal rule. For every graphon W, every $T_1\in(-\mathrm{age}(W),\infty)$ and $T_2\in(T_1,\infty]$ and every pair of sets $S_1,S_2\subset\Omega$ we have*

$$\int_{S_1\times S_2}\left(\Phi^{T_1}W-\Phi^{T_2}W\right) = \sum_{(i,j)\in E(F)}\int_{t=T_1}^{T_2}\int_{(x,y)\in S_1\times S_2} t^{x,y}(F^{(i,j)},\Phi^t W)\,.$$

Proof. By Remark 3.4, the left-hand side of the equation is

$$\Phi^{T_1}W-\Phi^{T_2}W = \int_{t=T_1}^{T_2}(-\mathfrak{V}\,\Phi^t W)\,.$$

Therefore, except when $T_2=\infty$, the Lemma follows by applying Fubini's Theorem together with (49). For $T_2=\infty$ we have

$$\int_{S_1\times S_2}\lim_{T\to\infty}\int_{T_1}^{T}(-\mathfrak{V}\,\Phi^t W) = \lim_{T\to\infty}\int_{S_1\times S_2}\int_{T_1}^{T}(-\mathfrak{V}\,\Phi^t W),$$

by the Monotone Convergence Theorem (since the velocity is always non-positive for a removal rule). Therefore, we obtain the desired outcome for $T_2=\infty$ by applying (49) and Fubini's Theorem inside the limit on the right-hand side of the equation above. □

10.1. **Destination of a limit.** Our next theorem says that the destination is continuous for removal rules. Recall that in Remark 8.4 we saw that this feature does not hold in the more general class of monotone rules.

Theorem 10.7. *Let $k\in\mathbb{N}$, $F\in\mathcal{H}_k$ and $\mathcal{R}$ be an F-removal rule. Suppose that $e(F)>0$. Then the function $U\mapsto\mathrm{dest}_{\mathcal{R}}(U)$ is continuous as a function from $(\mathcal{W}_0,d_{\square})$ to itself.*

Theorem 10.7 is an important stepping stone towards our proof of Theorem 10.9, which shows that convergence time is finite and uniformly bounded. The proof of Theorem 10.7 is quite involved. In particular, it uses Szemerédi's regularity lemma, but the advantage is that then the proof of Theorem 10.9 is short and conceptually simple. On the other hand, Theorem 10.9 is very similar in spirit to the classical removal lemma. In particular, we ask whether Theorem 10.9 could be proven using the removal lemma instead of the Szemerédi's regularity lemma.

Proof. The statement is obvious when F has at most 2 vertices. So, let us assume that $v(F) \geq 3$.

Given a graphon W, we write $W_t = \Phi^t W$ and $W_\infty = \mathrm{dest}(W)$. We will show that for every $\gamma > 0$ (without loss of generality, $\gamma < 1$) there exists δ such that if $d_\square(W, U) < \delta$ then $d_\square(W_\infty, U_\infty) < \gamma$. With foresight of the cooking process, we set

$$\beta := \frac{\gamma}{8k^2}, \quad \alpha := \frac{\beta^{e(F)}}{2k^k}.$$

Let M_0 be given by Theorem 2.11 with error parameter α. Let $\varepsilon := \beta^{e(F)}/(2k^2(M_0)^k)$. Fix an arbitrary graphon U and pick T such that

$$d_\square(U_T, U_\infty) < \varepsilon/2 \,. \tag{50}$$

Let $\delta := \varepsilon/(2\exp(C_{\square,k}T))$, where $C_{\square,k}$ comes from Theorem 3.8.

From now on we fix a graphon W satisfying $d_\square(W, U) < \delta$. Theorem 3.8 implies that

$$d_\square(W_T, U_T) \leq \exp(C_k T) \cdot d_\square(W, U) < \varepsilon/2 \,,$$

which we combine with (50) to get that

$$d_\square(W_T, U_\infty) < \varepsilon \,. \tag{51}$$

We will show that

$$f := d_\square(W_T, W_\infty) < \gamma/2 \tag{52}$$

and combine with (51) to finish the proof.

First notice that since U_∞ is a destination, Fact 10.5 implies that $t(F, U_\infty) = 0$ and therefore Lemma 2.6 yields

$$t(F, W_T) < k^2 d_\square(W_T, U_\infty) \overset{(51)}{\leq} k^2\varepsilon \,. \tag{53}$$

The strategy is to obtain a contradiction by showing that if W_T is not too close to its destination, then it has substantial density $t(F, W_T)$. Since the F-removal rule is non-increasing, we have $W_T \geq W_\infty$ and in particular

$$f = d_\square(W_T, W_\infty) = \int_{\Omega^2} (W_T - W_\infty) \, .$$

We apply Theorem 2.11 to find an α-regular partition $\mathcal{P}$ of the graphon $W_T - W_\infty$ with $M \leq M_0$ parts.

Given $\mathbf{S} = (S_1, \ldots, S_k) \in \mathcal{P}^k$, recall Definition 2.5 and set

$$w_{\mathbf{S}} := \int_{t=T}^{\infty} t^{\mathbf{S}}(F, W_t) \, .$$

From Lemma 10.6 we get that

$$f = \int_{\Omega^2} (W_T - W_\infty) = \sum_{(i,j) \in E(F)} \int_{t=T}^{\infty} \int_{(x,y) \in \Omega^2} t^{x,y}(F^{(i,j)}, W_t) \tag{54}$$

$$= \sum_{(i,j) \in E(F)} \int_{t=T}^{\infty} \sum_{\mathbf{S} \in \mathcal{P}^k} t^{\mathbf{S}}(F, W_t) = \sum_{(i,j) \in E(F)} \sum_{\mathbf{S} \in \mathcal{P}^k} w_{\mathbf{S}} \, . \tag{55}$$

By (54), there exists $(i, j) \in E(F)$ such that

$$\sum_{\mathbf{S} \in \mathcal{P}^k} w_{\mathbf{S}} \geq \frac{f}{k^2} \, . \tag{56}$$

Let $\mathcal{B}_2 \subset \mathcal{P}^2$ be the set of (ordered) pairs that are either not α-regular in the graphon $W_T - W_\infty$ or have density below $\beta = \frac{\gamma}{8k^2}$. Let $\mathcal{B}_k \subset \mathcal{P}^k$ be the set of k-tuples $\mathbf{S}$ such that there exists $(i, j) \in E(F)$ such that $(S_i, S_j) \in \mathcal{B}_2$. We have

$$\begin{aligned}
\sum_{\mathbf{S} \in \mathcal{B}_k} w_{\mathbf{S}} &\leq \sum_{(S_1', S_2') \in \mathcal{B}_2} \sum_{(i,j) \in E(F)} \sum_{\mathbf{S} : S_i = S_1', S_j = S_2'} w_{\mathbf{S}} \\
&= \sum_{(S_1', S_2') \in \mathcal{B}_2} \sum_{(i,j) \in E(F)} \int_{t=T}^{\infty} \int_{(x,y) \in S_1' \times S_2'} t^{x,y}(F^{(i,j)}, W_t) \\
&\overset{\boxed{\text{Lemma 10.6}}}{=} \sum_{(S_1', S_2') \in \mathcal{B}_2} \int_{S_1' \times S_2'} (W_T - W_\infty) \, .
\end{aligned}$$

Since, by α-regularity of $\mathcal{P}$, the total measure of rectangles $S_1' \times S_2'$ that are not α-regular is at most $\alpha M^2 \cdot M^{-2} = \alpha$, and the integral over the remaining rectangles is at most β, we obtain

$$\sum_{\mathbf{S} \in \mathcal{B}_k} w_{\mathbf{S}} \leq \alpha + \beta \leq \frac{\gamma}{4k^2} \, . \tag{57}$$

We are now in a position to complete the proof by showing (52). Suppose for contradiction that $f \geq \gamma/2$. Then from (56) and (57) we have

$$\sum_{\mathbf{S} \in \mathcal{P}^k \setminus \mathcal{B}_k} w_{\mathbf{S}} \geq \frac{\gamma}{4k^2} > 0 \,.$$

In particular, that means that $\mathcal{P}^k \setminus \mathcal{B}_k \neq \emptyset$. Let us take an arbitrary $\mathbf{S} \in \mathcal{P}^k \setminus \mathcal{B}_k$. By the definition of $\mathcal{B}_k$, for each $(a,b) \in E(F)$ we have, in the graphon $W_T - W_\infty$, that (S_a, S_b) is an α-regular pair of density at least β. Lemma 2.12 gives that

$$t(F, W_T) \geq t^{\mathbf{S}}(F, W_T - W_\infty) \geq \left(\beta^{e(F)} - \alpha k^k\right) \left(\frac{1}{M_0}\right)^k \geq \frac{\beta^{e(F)}}{2M_0^k} \geq k^2 \varepsilon \,,$$

a contradiction to (53). Hence, we conclude that $f < \gamma/2$, as was needed. □

Remark 10.8. *One of the main steps in the proof above is similar to the standard deduction of the removal lemma from Szemerédi's regularity lemma. That is, substantial density of F in W_T is deduced from a single copy of F in the cluster graph ('$\mathcal{P}^k \setminus \mathcal{B}_k \neq \emptyset$'). We think that finding a way to instead directly apply the removal lemma is an interesting problem in its own right. Note that such a direct application of the removal lemma would likely yield an improvement in the implicit epsilon–delta quantification of continuity in Theorem 10.7 since it would allow the use of the bounds famously obtained by Fox* [11].

10.2. **Speed of removal flip processes.** Generally, removal flip processes may converge to the destination very slowly. For example, consider the triangle removal flip process. A well known construction of Behrend (see [29, Corollary 2.5.2]) implies that for every $n \in \mathbb{N}$ there exists a graphon L_m with $t(K_3, L_m) = O(1/m^2)$, $\|L_m\|_1 = \exp(-\Theta(\sqrt{\log m}))$, and whose destination is $\operatorname{dest}(L_m) = 0$. Obviously, for the velocity of the triangle removal flip process we have, for each $T \geq 0$, that $\|\mathfrak{V}\, \Phi^T L_m\|_1 \leq \|\mathfrak{V}\, L_m\|_1 = 6t(K_3, L_m) = O(1/m^2)$. This example shows that the convergence time[g] $\tau_{\mathcal{R}}^-(\delta)$ cannot be bounded from above by any polynomial in $1/\delta$. So far the reasoning for the slow speed of convergence of the triangle removal flip process[h] was based on lower bound constructions for the triangle removal lemma. But a priori we cannot rule out even the case that $\tau_{\mathcal{R}}^-(\delta) = \infty$. This is because the removal lemma says 'if the graph is far from all triangle-free graphs then it has many triangles' and not 'if the graph is far from its destination then it has many triangles'.

Our result shows that the convergence times of each removal rule are finite and uniformly bounded. Theorem 10.7 plays a key role in the proof.

[g]Due to monotonicity, $\tau_{\mathcal{R}}^- = \tau_{\mathcal{R}}^+$ for removal processes.

[h]which we make formal and more general in Proposition 10.10

Theorem 10.9. *Let $k \in \mathbb{N}$, $F \in \mathcal{H}_k$ and $\mathcal{R}$ be the F-removal rule. Suppose that $e(F) > 0$. Then for every $\varepsilon > 0$ there exists $T \geq 0$ so that for every graphon W we have $\|\Phi^T W - \mathrm{dest}(W)\|_1 < \varepsilon$.*

Proof. Suppose that this does not hold. Then there exists $\varepsilon > 0$ and a sequence of graphons $(W_n)_{n\in\mathbb{N}}$ so that for every $n \in \mathbb{N}$, we have

$$\|\Phi^n W_n - \mathrm{dest}(W_n)\|_1 \geq \varepsilon \,. \tag{58}$$

By the Lovász–Szegedy compactness theorem (see [18, Theorem 9.23]) there exist indices $n_1 < n_2 < \cdots$, measure preserving bijections $\psi_1, \psi_2, \ldots$, and a graphon U so that writing $U_i := W_{n_i}^{\psi_i}$ we have

$$\|U_i - U\|_\square \to 0 \,. \tag{59}$$

By Theorem 10.7, we have

$$\|\mathrm{dest}(U_i) - \mathrm{dest}(U)\|_\square \to 0 \,. \tag{60}$$

Take T such that $\big\|\Phi^T U - \mathrm{dest}(U)\big\|_\square < \varepsilon/4$. Let $C_{\square,k}$ be as in Theorem 3.8. By (59) and (60) we can choose $\ell \geq T$ such that

$$\|U_\ell - U\|_\square < \varepsilon/(4\exp(C_{\square,k}T)) \quad \text{and} \quad \|\mathrm{dest}(U_\ell) - \mathrm{dest}(U)\|_\square < \varepsilon/4. \tag{61}$$

Since $n_\ell \geq \ell \geq T$, we have

$$\begin{aligned}
\|\Phi^{n_\ell} U_\ell - \mathrm{dest}(U_\ell)\|_1 &= \int_{\Omega^2} (\Phi^{n_\ell} U_\ell - \mathrm{dest}(U_\ell)) \\
&\leq \int_{\Omega^2} \big(\Phi^T U_\ell - \mathrm{dest}(U_\ell)\big) \\
&\leq \big\|\Phi^T U_\ell - \Phi^T U\big\|_\square + \big\|\Phi^T U - \mathrm{dest}(U)\big\|_\square + \|\mathrm{dest}(U_\ell) - \mathrm{dest}(U)\|_\square \\
&\boxed{\text{(61) and Theorem 3.8}} \quad < \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} < \varepsilon \,.
\end{aligned}$$

Noting that $\Phi^{n_\ell} U_\ell = (\Phi^{n_\ell} W_{n_\ell})^{\psi_\ell}$ and $\mathrm{dest}(U_\ell) = (\mathrm{dest}(W_{n_\ell}))^{\psi_\ell}$ (see Section 6.3 in [12]), we obtain a contradiction to (58). □

We do not obtain any explicit bounds on the convergence time T in Theorem 10.9. In lieu of the Lovász–Szegedy compactness theorem, one could probably use the machinery of a regularity lemma (the Frieze-Kannan version seems to be sufficient) in the proof. Further, to obtain an explicit bound in Theorem 10.9, one would need to make the ε–δ quantification of continuity in Theorem 10.7 explicit as well. Recall that Theorem 10.7 uses Szemerédi's regularity lemma (Theorem 2.11) with its infamous tower-type dependency. Putting them together, we would likely get that $T \leq \mathsf{tower}((1/\varepsilon)^{O(1)})$; even if we successfully replace Szemerédi's regularity lemma with the removal lemma as mentioned in Remark 10.8, we

would still get only $T \leq \mathsf{tower}(O(\log(1/\varepsilon)))$. In our next proposition, however, we prove that a polynomial upper bound exists if and only if F is bipartite. To this end, some preparatory notation is needed.

Suppose that F is a bipartite graph. Let m_F be the smallest number such that for each graphon W we have

$$t(F,W) \geq \|W\|_1^{m_F} \ . \tag{62}$$

We call m_F the *Sidorenko power of* F. It is easy to show that m_F exists and $m_F \in [e(F), (\frac{v(F)}{2})^2]$. The Sidorenko conjecture (see [18]) asserts that $m_F = e(F)$ for every bipartite F.

Proposition 10.10. *Let $k \in \mathbb{N}$, $F \in \mathcal{H}_k$ and $\mathcal{R}$ be the F-removal rule. Suppose that $e(F) > 0$. Then the following hold.*

(i) *If F is bipartite, then the destination of every graphon is the constant-0 graphon.*
(ii) *Suppose that F is bipartite and its Sidorenko power is m_F. Then Theorem 10.9 holds with $T = \varepsilon^{-m_F+1}$ when $e(F) \geq 2$ and with $T = \ln(1/\varepsilon)$ when $e(F) = 1$.*
(iii) *If F is not bipartite then no polynomial bound on T in terms of ε exists, that is, for every M there is $\varepsilon_0 > 0$ so that for all $\varepsilon \in (0, \varepsilon_0]$ we have $\tau_{\mathcal{R}}^-(\varepsilon) \geq \varepsilon^{-M}$.*

Proof. First suppose that F is bipartite. Let W be a graphon. By (49) and (62) we have

$$\|\mathfrak{V}\, W\|_1 = (k)_2 \cdot t(F,W) \geq \|W\|_1^{m_F} \ .$$

Then Proposition 3.9 tells us any destination is constant-0, proving (i).

Turning to (ii), we first consider the case $e(F) \geq 2$, which entails $m_F \geq 2$; we shall return to the case $e(F) = 1$ later. The required conclusion is trivial for the constant-0 graphon, so consider $W \not\equiv 0$. We claim that the (continuous) function $f(t) := \left\|\Phi^t W\right\|_1$ satisfies

$$-f'(t) \geq f(t)^{m_F}. \tag{63}$$

Indeed, Lemma 10.6 and (62) imply

$$\begin{aligned} f(t) - f(t+\varepsilon) = \int_{\Omega^2} \left(\Phi^t W - \Phi^{t+\varepsilon} W\right) &= \sum_{(i,j)\in E(F)} \int_{\tau=t}^{t+\varepsilon} \int_{(x,y)\in\Omega^2} t^{x,y}(F^{(i,j)}, \Phi^\tau W) \\ &\overset{(62)}{\geq} (k)_2 \int_t^{t+\varepsilon} \|\Phi^\tau W\|_1^{m_F} \geq (k)_2 \varepsilon f(t+\varepsilon)^{m_F} \geq \varepsilon f(t+\varepsilon)^{m_F} \ . \end{aligned}$$

Dividing by ε and taking the limit $\varepsilon \to 0$ we obtain (63), proving our claim. Now (63) implies that the function $g(t) := f(t)^{-(m_F-1)}$ satisfies

$$g'(t) = -(m_F - 1) f(t)^{-m_F} f'(t) \geq m_F - 1,$$

where we used the fact that f takes positive values. Consequently, $g(t) \geq g(0)+(m_F-1)t$, and by substituting g we obtain

$$f(t) \leq \big((m_F-1)t + f(0)^{1-m_F}\big)^{-\frac{1}{m_F-1}} \leq t^{-\frac{1}{m_F-1}} ,$$

and the desired conclusion in Theorem 10.9 follows. Similarly, for the case $e(F) = 1$ we have the differential inequality $-f'(t) \geq f(t)$ with auxiliary function $g(t) = -\ln f(t)$. This gives $f(t) \leq f(0)\exp(-t)$, which then implies the desired conclusion.

To deduce (iii), we recall one of the main results of [2]. Namely, Lemma 3.4 in [2] asserts that if F is not bipartite, then there exists $c > 0$ such that for all sufficiently small $\alpha > 0$ and all sufficiently large integers n there exists an n-vertex graph G that is α-far from being F-free and contains at most $(\frac{\alpha}{c})^{c\ln(\frac{c}{\alpha})}n^{v(F)}$ copies of F. Here, *α-far* means that more than αn^2 edges need to be removed from G so that it becomes F-free. Since the number of non-injective homomorphisms of F into G is $O(n^{v(F)-1})$, by taking n large enough (depending on α), we can assume that the homomorphism density $t(F,G)$ is at most $2(\frac{\alpha}{c})^{c\ln(\frac{c}{\alpha})}$. Let W be a graphon representation of G. Recalling that $t(F,W) = t(F,G)$, we have $t(F,W) \leq 2\left(\frac{\alpha}{c}\right)^{c\ln(\frac{c}{\alpha})}$. Also note that since $t(F,\Phi^{\cdot}W)$ is nonincreasing for the F-removal flip process, Lemma 10.6 implies

(64)
$$\big\|W-\Phi^tW\big\|_1 = (v(F))_2\int_{x=0}^{t} t(F,\Phi^xW) \leq t(v(F))_2\cdot t(F,W) \leq t(v(F))_2\cdot 2\left(\frac{\alpha}{c}\right)^{c\ln(\frac{c}{\alpha})} .$$

The notion of being α-far translates in a straightforward way into a graphon representation. In this translation, the L^1-norm is the graphon counterpart to measuring the number of edges removed. Let us use this intuition to show that

$$\|W-\mathrm{dest}(W)\|_1 \geq \alpha . \tag{65}$$

Indeed, it follows from Proposition 6.5 of [12] that $\mathrm{dest}(W)$ is a step graphon on steps $(\Omega_v : v \in V(G))$ corresponding to the representation of G. Let $G' \subset G$ be the subgraph of G containing edges in the support of $\mathrm{dest}(W)$, that is, $ij \in E(G')$ if and only if $\mathrm{dest}(W)\restriction_{\Omega_i\times\Omega_j} > 0$. Obviously, $\frac{e(G)-e(G')}{n^2} \leq \|W-\mathrm{dest}(W)\|_1$. At the same time, $\mathrm{dest}(W)$ is F-free, and therefore so is G'. Since G is α-far from being F-free, we have $\frac{e(G)-e(G')}{n^2} > \alpha$. This allows us to deduce (65).

Thus, for every $t \geq 0$ we have

$$\big\|\Phi^tW-\mathrm{dest}(W)\big\|_1 \geq \|W-\mathrm{dest}(W)\|_1 - \big\|W-\Phi^tW\big\|_1 \overset{(65),\,(64)}{\geq} \alpha - t\cdot 2(v(F))_2\left(\frac{\alpha}{c}\right)^{c\ln(\frac{c}{\alpha})} .$$

Given $M > 0$, taking $\alpha > 0$ small enough, we have $2(v(F))_2(\frac{\alpha}{c})^{c\ln(\frac{c}{\alpha})} \leq (\alpha/2)^{M+1}$. Setting $\varepsilon = \alpha/2$ and $t = \varepsilon^{-M}$ we have $\big\|\Phi^tW-\mathrm{dest}(W)\big\|_1 \geq \varepsilon$, whence $\tau^-_{\mathcal{R}}(\varepsilon) \geq \varepsilon^{-M}$. This completes the proof of (iii). □

10.3. **Graphons with the same destination.** Fact 10.5 (see also Proposition 10.10) tells us that for non-bipartite graphs F, the F-removal rule has a nontrivial set of destinations. An intriguing problem for any such F is to determine equivalence classes of graphons with the same destination. A particularly important equivalence class is the class of graphons whose destination is constant-0. We do not have even a guess what the answer might be.

Question 10.11. *Suppose that F is a fixed graph. Determine all graphons whose destination in the F-removal flip process is the constant-0 graphon.*

In this section we consider the simplest nontrivial example, $F = K_3$, and we study the destinations of step graphons $\{W_{\alpha,\beta} : \alpha, \beta \in [0,1]\}$ with two steps Ω_1 and Ω_2 of equal size such that $W_{\alpha,\beta} = \alpha$ on $\Omega_1^2 \cup \Omega_2^2$ and $W_{\alpha,\beta} = \beta$ elsewhere. It is easy to see (using [12, Corollary 5.6 and Remark 3.8]) that the trajectory starting at $W_{\alpha,\beta}$ stays a step graphon with respect to the partition $\Omega_1 \cup \Omega_2$ and becomes a constant graphon if and only if $\alpha = \beta$. An easy symmetry argument (using [12, Remark 3.8], say), also implies that the value on Ω_1^2 stays equal to the value on Ω_2^2. To sum up, for all $\alpha, \beta \in [0,1]$ there are some functions $a, b : [0,\infty) \to [0,1]$ such that we have

$$\alpha = \beta \text{ implies } a(t) = b(t) \text{ for every } t \geq 0 \,, \tag{66}$$

$$\alpha \neq \beta \text{ implies } a(t) \neq b(t) \text{ for every } t \geq 0 \,, \tag{67}$$

and for all $t \geq 0$ we have

$$\Phi^t W_{\alpha,\beta} = W_{a(t),b(t)} \,. \tag{68}$$

It is very intuitive that $\mathrm{dest}(W_{\alpha,\beta}) = W_{0,\gamma}$, with $\gamma > 0$ whenever α/β is small enough and $\gamma = 0$ when α/β is large enough. The following result says that the critical point is $\alpha = \beta$.

Proposition 10.12. *Consider the K_3-removal flip process. For a step graphon $W_{\alpha,\beta}$ as defined above we have*

$$\mathrm{dest}(W_{\alpha,\beta}) = \begin{cases} W_{0,\beta\exp[-2\alpha/(\beta-\alpha)]} & \text{if } \alpha < \beta \,, \\ W_{0,0}, & \text{if } \alpha \geq \beta \,. \end{cases}$$

Proof. Since $\mathfrak{V}\, W(x,y) = -6t^{x,y}(K_3^{\bullet\bullet}, W)$, it is easy to check that the functions a and b from (68) satisfy the system of differential equations

$$a' = -a^3/2 - ab^2/2,$$
$$b' = -ab^2$$

with the initial condition $a(0) = \alpha, b(0) = \beta$. Observe that since the functions a and b take non-negative values, it follows easily from the system of differential equations that in the case $\alpha = 0$ we have $a \equiv 0$ and $b \equiv \beta$. A similar argument tells us that in the case $\beta = 0$ we

get $b \equiv 0$ and $a(t) = (t + 1/\alpha^2)^{-1/2}$. Now if $\alpha = \beta > 0$, then by (66) we actually have a single equation $a' = -a^3$, which has an explicit solution $a(t) = b(t) = (2(t + 1/(2\alpha^2))^{-1/2}$. Hence, we may assume that $\alpha \neq \beta$, $\alpha > 0, \beta > 0$. This implies that a, b take positive values for every $t \geq 0$ (see [12, Proposition 6.7]) and hence are strictly decreasing continuous functions. In particular $\lim_{x\to\infty} a(x)$ is nonnegative; it is actually zero, in view of the differential equation for a'. Hence the inverse function a^{-1} is defined on $(0, \alpha]$. Define a function $f : (0, \alpha] \to (0, 1]$ by $f(x) := b(a^{-1}(x))$. Note that

$$\operatorname{dest}(W_{\alpha,\beta}) = W_{0,\lim_{x\to 0} f(x)}. \tag{69}$$

In view of (69), it remains to determine $\lim_{x\to 0} f(x)$. To do so, we shall seek to understand the function f. Since $\alpha \neq \beta$, note that (67) and continuity implies that

$$\text{either} \quad f(x) < x \ \forall x \in (0, \alpha] \quad \text{or} \quad f(x) > x \ \forall x \in (0, \alpha]\,. \tag{70}$$

We have

$$\begin{aligned} f'(x) &= \frac{b'\left(a^{-1}(x)\right)}{a'\left(a^{-1}(x)\right)} = -\frac{a\left(a^{-1}(x)\right) b(a^{-1}(x))^2}{a\left(a^{-1}(x)\right)^3/2 + a\left(a^{-1}(x)\right) b\left(a^{-1}(x)\right)^2/2} \\ &= \frac{2xf(x)^2}{x^3 + xf(x)^2} = \frac{2}{(x/f(x))^2 + 1}\,. \end{aligned}$$

Hence f satisfies a differential equation $f' = 2/(x^2/f^2 + 1)$ for $x \in (0, \alpha]$ with initial condition $f(\alpha) = \beta$. Denoting $v(x) = f(x)/x$, we have $f' = v'x + v$, whence

$$\begin{aligned} xv' = f' - v &= \frac{2}{v^{-2} + 1} - v = \frac{-v(v-1)^2}{1 + v^2}\,, \\ \frac{v'}{v} &= -\frac{v^2 - 2v + 1}{x(1 + v^2)} = -\frac{1}{x} + \frac{2v}{x(1 + v^2)}\,. \end{aligned} \tag{71}$$

Since $(\ln v)' = v'/v$ and $(2/(v-1))' = -2v'/(v-1)^2$, we get

$$\begin{aligned} \ln v(x) &= \ln v(\alpha) + \int_\alpha^x \left(\frac{2v}{x(1+v^2)} - \frac{1}{x}\right) \mathrm{d}x \overset{(71)}{=} \ln\frac{f(\alpha)}{\alpha} + \ln\frac{\alpha}{x} + \int_\alpha^x \frac{-2v'}{(v-1)^2} \\ \boxed{f(\alpha)=\beta} \quad &= \ln\frac{\beta}{x} + \frac{2}{v(x) - 1} - \frac{2}{v(\alpha) - 1}. \end{aligned}$$

Since $v(x) = f(x)/x$ and $v(\alpha) = \beta/\alpha$, it follows that any solution f satisfies equation

$$\ln\frac{f(x)}{x} - \frac{2}{f(x)/x - 1} = \ln\frac{\beta}{x} + \frac{2}{1 - \beta/\alpha}. \tag{72}$$

So, writing $y = f(x)/x - 1$, $F(y) := \ln(1 + y) - 2/y$ and defining a constant

$$c_0 := \ln\beta + \tfrac{2}{1-\beta/\alpha},$$

equation (72) reads

$$F(y) = c_0 - \ln x.$$

Elementary calculus shows that F is strictly increasing and continuous on $(-1, 0)$ and on $(0, \infty)$ and undefined elsewhere (which is not a problem, since by (70) we have $y \in (-1, 0) \cup (0, \infty)$ for $x \in (0, \alpha]$). Moreover, $F((-1, 0)) = F((0, \infty)) = \mathbb{R}$. Hence the inverse of F consists of two increasing bijections: $F_-^{-1} : \mathbb{R} \to (-1, 0)$ and $F_+^{-1} : \mathbb{R} \to (0, \infty)$. By (70), and noting that $f(\alpha) = \beta$ we get that if $\alpha < \beta$, then $y(\alpha) = \beta/\alpha - 1 > 0$ and thus $y > 0$ for all x; similarly, if $\alpha > \beta$, then $y < 0$ for all x. Consequently

$$f(x) = \begin{cases} x\left(1 + F_-^{-1}(c_0 - \ln x)\right), & \text{if } \alpha > \beta, \\ x\left(1 + F_+^{-1}(c_0 - \ln x)\right), & \text{if } \alpha < \beta. \end{cases} \tag{73}$$

We shall now determine $\lim_{x \downarrow 0} f(x)$; in view of (69), this will complete the proof. It clearly equals zero when $\alpha > \beta$, since F_-^{-1} is bounded. In the case $\alpha < \beta$, we have that $F_+^{-1}(x) \to \infty$, as $x \to \infty$, so we bound $F_+^{-1}(x)$ for large x by choosing an arbitrary constant $\varepsilon > 0$ and using inequalities

$$\ln(1 + y) - \varepsilon =: L_\varepsilon(y) \leq F(y) \leq L_0(y) = \ln(1 + y)$$

that hold for large enough y. We note that for large enough x

$$e^x - 1 = L_0^{-1}(x) \leq F_+^{-1}(x) \leq L_\varepsilon^{-1}(x) = e^{x+\varepsilon} - 1,$$

which implies that $xF_+^{-1}(c_0 - \ln x) \in [e^{c_0}, e^{c_0+\varepsilon}]$. Since $\varepsilon > 0$ was arbitrary, we obtain that

$$\lim_{x \downarrow 0} x(1 + F_+^{-1}(c_0 - \ln x)) = e^{c_0} = \beta e^{-\frac{2\alpha}{\beta - \alpha}},$$

completing the proof. □

10.4. **Stabilization in a removal flip process.** In Section 7.4 we argued that extremist flip processes may take very long to stabilize. On the other hand, removal flip processes stabilize substantially faster. Indeed, an F-removal flip process started on an n-vertex graph stabilizes no later than the moment that each of its $(n)_k$-many distinct k-tuples of vertices has been sampled at least once. Then, the solution to the classical coupon collector's problem tells us that $\Theta(n^k \log n)$ is asymptotically almost surely an upper bound on the stabilization time. On the other hand, suppose that the host graph consists of a single copy of F. Then that copy of F needs to be sampled; this event does not happen asymptotically almost surely before $\omega(n^k)$ steps. While the upper-bound seems closer to the truth, we were not able to obtain an optimal upper-bound on the stabilization time.

Question 10.13. *For each graph F with the corresponding F-removal flip process $\mathcal{R}_F$, we seek a function $b_F : \mathbb{N} \to \mathbb{N}$ growing as slowly as possible so that*

$$\min_{G \in \mathcal{H}_n} \mathbb{P}\left(\kappa_{\mathcal{R}_F}(G) \leq b_F(n)\right) \to 1, \qquad \text{as } n \to \infty .$$

The only nontrivial case studied in the literature seems to be $F = K_3$. In particular, it is possible that by using results from [3] one can obtain the order of magnitude of b_{K_3}.

Since the stabilization time in Question 10.13 is clearly superquadratic for general F, the limitations of Theorem 1.1 kick in. That means that *a priori* we cannot connect the concept of destinations to the structure of the *final graph*, that is, the graph after stabilization occurs. Our next theorem shows that, in the case of removal flip processes, destinations serve as good approximations for the final graphs.

Theorem 10.14. *Given a graph F, for every $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that the following holds for all $n \geq n_0$. Suppose that G^* is the final graph in the F-removal flip process started at an n-vertex graph G. Let W_G and W_{G^*} be the graphon representations of G and G^*, respectively. Then with probability at least $1-\varepsilon$ we have $d_\square(\mathrm{dest}(W_G), W_{G^*}) < \varepsilon$.*

The proof uses several tools, including Theorem 10.9 and the graph removal lemma, the latter of which we now recall.

Proposition 10.15 (Graph removal lemma)**.** *Given a graph F, for every $\alpha > 0$ there exists $\beta > 0$ with the following property. Suppose that G is an n-vertex graph whose density of F is less than β. Then we can remove at most αn^2 edges from G and make it F-free.*

Proof of Theorem 10.14. Let $\varepsilon > 0$ and F be a graph. Observe that the statement is trivial when $e(F) = 0$, so we shall assume that $e(F) > 0$ and $v(F) \geq 2$. Proposition 10.15 returns $\beta > 0$ for input $\alpha := \varepsilon/(4e(F))$. Set $\gamma := \min\{\alpha, \beta/k^2\}$. Theorem 10.9 returns $T \geq 0$ for error parameter γ. Lastly, let n_0 be sufficiently large so that we have $1 - \exp(-\Omega(n_0^2)) \geq 1 - \varepsilon$ in Theorem 1.1 with error parameter γ and input $T \geq 0$.

Suppose now G is an n-vertex graph for $n \geq n_0$. We run the F-removal flip process on G and obtain a sequence of graphs $G_0 := G, G_1, G_2, \ldots, G_{\kappa(G)} =: G^*$. Let $\mathcal{B}$ be the event that for all $\ell \in \mathbb{N} \cap [0, Tn^2]$ we have $d_\square(W_{G_\ell}, \Phi^{\ell/n^2} W_{G_0}) < \gamma$. Theorem 1.1 tells us that the probability of this event is at least $1 - \varepsilon$. Assume that $\mathcal{B}$ holds. We shall prove that $d_\square(\mathrm{dest}(W_G), W_{G^*}) < \varepsilon$.

We have $d_\square\left(\mathrm{dest}(W_G), \Phi^T W_G\right) \leq \gamma$ and $d_\square\left(\Phi^T W_G, W_{G_{Tn^2}}\right) \leq \gamma$ by Theorem 10.9 and Theorem 1.1 respectively. Putting them together, we obtain

$$d_\square\left(\mathrm{dest}(W_G), W_{G_{Tn^2}}\right) \leq d_\square\left(\mathrm{dest}(W_G), \Phi^T W_G\right) + d_\square\left(\Phi^T W_G, W_{G_{Tn^2}}\right) \leq 2\gamma. \tag{74}$$

Since $\mathrm{dest}(W_G)$ is F-free by Fact 10.5, it follows from (74) and Lemma 2.6 that $t(F, W_{G_{Tn^2}}) < \beta$. Then, Proposition 10.15 tells us that we can find a set $R \subset E(G_{Tn^2})$ of at most

$\varepsilon n^2/(4e(F))$ edges whose removal would destroy all the copies of F in G_{Tn^2}. In particular, every copy of F in G_{Tn^2} has an edge in R. Let $X \subseteq [Tn^2, \dots, \kappa(G)]$ be the indices of the non-idle steps. In particular, $X \neq \emptyset$ if and only if $\kappa(G) > Tn^2$. Observe that in each non-idle step indexed in X a copy of F in G_{Tn^2} is removed, so at least one edge from R is removed. Hence, we have

$$e(G_{Tn^2}) - e(G^*) = e(F) \cdot |X| \leq e(F) \cdot |R| \leq \frac{\varepsilon n^2}{4},$$

which implies $d_\square\left(W_{G_{Tn^2}}, W_{G^*}\right) \leq \varepsilon/4$. Combining this with (74), we obtain

$$d_\square\left(\mathrm{dest}(W_G), W_{G^*}\right) \leq d_\square\left(\mathrm{dest}(W_G), W_{G_{Tn^2}}\right) + d_\square\left(W_{G_{Tn^2}}, W_{G^*}\right) < \varepsilon,$$

as was needed. □

References

[1] D. Achlioptas, R. M. D'Souza, and J. Spencer. Explosive percolation in random networks. *Science*, 323(5920):1453–1455, 2009.

[2] N. Alon. Testing subgraphs in large graphs. *Random Structures Algorithms*, 21(3-4):359–370, 2002. Random structures and algorithms (Poznan, 2001).

[3] T. Bohman, A. Frieze, and E. Lubetzky. Random triangle removal. *Adv. Math.*, 280:379–438, 2015.

[4] T. Bohman and P. Keevash. The early evolution of the H-free process. *Invent. Math.*, 181(2):291–336, 2010.

[5] T. Bohman and P. Keevash. Dynamic concentration of the triangle-free process. *Random Structures Algorithms*, 58(2), 2021.

[6] B. Bollobás. To prove and conjecture: Paul Erdős and his mathematics. *Amer. Math. Monthly*, 105(3):209–237, 1998.

[7] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[8] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[9] Paul Erdos. On the number of complete subgraphs contained in certain graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 7(3):459–464, 1962.

[10] G. Fiz Pontiveros, S. Griffiths, and R. Morris. The triangle-free process and the Ramsey number $R(3, k)$. *Mem. Amer. Math. Soc.*, 263(1274):v+125, 2020.

[11] J. Fox. A new proof of the graph removal lemma. *Ann. of Math. (2)*, 174(1):561–579, 2011.

[12] F. Garbe, J. Hladký, M. Šileikis, and F. Skerman. From flip processes to dynamical systems on graphons. Accepted to *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, available on the journal website.

[13] J. Hladký, Ch. Pelekis, and M. Šileikis. A limit theorem for small cliques in inhomogeneous random graphs. *J. Graph Theory*, 97(4):578–599, 2021.

[14] S. Janson. Connectedness in graph limits, 2008. arXiv:0802.3795.

[15] G. Katona. A theorem of finite sets. In *Theory of graphs (Proc. Colloq., Tihany, 1966)*, pages 187–207, 1968.

[16] J. B. Kruskal. The number of simplices in a complex. In *Mathematical optimization techniques*, pages 251–278. Univ. of California Press, Berkeley, Calif., 1963.

[17] L. Lovász. *Combinatorial problems and exercises*. North-Holland Publishing Co., Amsterdam, second edition, 1993.

[18] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[19] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[20] L. Lovász and B. Szegedy. Szemerédi's lemma for the analyst. *Geom. Funct. Anal.*, 17(1):252–270, 2007.

[21] D. Osthus and A. Taraz. Random maximal H-free graphs. *Random Structures Algorithms*, 18(1):61–82, 2001.

[22] N. Pippenger and J. Spencer. Asymptotic behavior of the chromatic index for hypergraphs. *J. Combin. Theory Ser. A*, 51(1):24–42, 1989.

[23] O. Riordan and L. Warnke. Achlioptas process phase transitions are continuous. *Ann. Appl. Probab.*, 22(4):1450–1464, 2012.

[24] O. Riordan and L. Warnke. The evolution of subcritical Achlioptas processes. *Random Structures Algorithms*, 47(1):174–203, 2015.

[25] V. Rödl. On a packing and covering problem. *European J. Combin.*, 6(1):69–78, 1985.

[26] J. Spencer. Asymptotic packing via a branching process. *Random Structures Algorithms*, 7(2):167–172, 1995.

[27] T. Tao. *Nonlinear dispersive equations*, volume 106 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2006. Local and global analysis.

[28] L. Warnke. The C_ℓ-free process. *Random Structures Algorithms*, 44(4):490–526, 2014.

[29] Y. Zhao. Graph theory and additive combinatorics. preprint, February 14, 2022.

Email address: `araujo|hladky|hng|matas@cs.cas.cz`

Institute of Computer Science of the Czech Academy of Sciences, Pod Vodárenskou věží 2, 182 00, Praha 8, Czechia.

Part IV

Limits of other discrete structures

LIMITS OF LATIN SQUARES

FREDERIK GARBE, ROBERT HANCOCK, JAN HLADKÝ, MARYAM SHARIFZADEH

Abstract. We develop a limit theory of Latin squares, paralleling the recent limit theories of dense graphs and permutations. We introduce a notion of density, an appropriate version of the cut distance, and a space of limit objects — so-called Latinons. Key results of our theory are the compactness of the limit space and the equivalence of the topologies induced by the cut distance and the left-convergence. Last, using Keevash's recent results on combinatorial designs, we prove that each Latinon can be approximated by a finite Latin square.

An extended abstract of this work appeared in the proceedings of EuroComb2019, [18].

1. Introduction

The purpose of this paper is to lay down the basics of a combinatorial theory of limits of Latin squares. This is a further addition to a potent approach to discrete structures using analytic tools. In combinatorics, this research trend blossomed in the early 2000's with the theories of dense graph limits initiated by Borgs, Chayes, Lovász, Sós, Szegedy, and Vesztergombi, [9, 10, 39], and sparse graph limits initiated by Benjamini and Schramm [6]. After that, many other theories followed, and Razborov provided an alternative, more syntactic limit framework, [43].

Recall that a *Latin square* is an $n \times n$ matrix filled with values of $[n] \coloneqq \{1, \dots, n\}$ in such a way that each row contains each value exactly once, and each column contains each value exactly once, as well. Let us emphasise that the rows, columns, and symbols in the matrix are ordered (from top to bottom, from left to right, and from small to large). In the same spirit, a *permutation* is a bijection from the naturally ordered set $[n]$ into itself.

In many aspects, our theory parallels those of limits of dense graphs and of permutations. In order to make a comparison, let us quickly recall the key features of these theories. For foundations of these two respective theories, see [39] and [24]. So, by a 'structure', we mean either a finite graph or a finite permutation.[1]

FG, RH, JH: Research supported by GAČR project 18-01472Y and RVO: 67985840; MSh: Research supported by the European Union's Horizon 2020 research and innovation programme under the Marie Curie Individual Fellowship agreement No 752426. RH: This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 648509) and from the MUNI Award in Science and Humanities of the Grant Agency of Masaryk University. This publication reflects only its authors' view; the European Research Council Executive Agency is not responsible for any use that may be made of the information it contains.

[1]But the same could be said about limit theories of many other discrete structures such as uniform hypergraphs.

(F1) Both theories introduce a certain space of analytically defined limit objects (*graphons* and *permutons*).

(F2) Both theories introduce a notion of *densities* $t(\circ, \star)$ of a structure $\circ$ both into a structure $\star$, or into a limit object $\star$.

(F3) Both theories also provide more global, 'cut-like', parameters on structures and on limit objects.

There are three key features in both theories.

(F4) Compactness theorem: Each sequence of structures contains a subsequence that converges (more precisely, 'left-converges') to a suitable limit object with respect to all densities.

(F5) Equivalence of left-convergence and convergence with respect to 'cut-distance': A sequence of structures converges with respect to all densities if and only if it converges with respect to the above 'cut-like' parameters.

(F6) Sampling lemma: The densities of structures in a limit object provide a notion of random sampling of a structure S_n of order n from that limit object. The sampling lemma says that S_n is typically very close to the limit object. More specifically, almost surely, the sequence $(S_n)_n$ converges to the original limit object. As a consequence of the sampling lemma, each limit object can be approximated by a structure with an arbitrary precision.

In our work, we introduce a class of analytic objects, so-called *Latinons*. We also introduce certain notions of densities for Latin squares and Latinons (however, with a slight difference compared to (F2) which we shall return to later) and 'cut-like' parameters, and establish (F5). Of course, the most important result is the compactness theorem (F4), which we establish as well. The situation with (F6) is more complicated. Given a Latinon L, we do not have any reasonable way of generating a random Latin square of order n from L (and probably no such way exists, see Remark 3.4). However, we can still establish the approximability consequence of (F6) (which is the most important part of (F6)).

Theories of limits of sequences of particular discrete structures have many applications, but two particular directions stand out. First and foremost, they (and the more syntactic theory of flag algebras, [43]) have led to solutions to many asymptotic problems in extremal combinatorics. As some representative examples for the class of graphs, let us mention the solution to the triangle versus edge density problem [44], the solution to the Erdős problem on the number of pentagons in triangle-free graphs [22, 21, 35], or progress on the Sidorenko conjecture [37]. For permutations, we are aware of only one application so far, but it is a remarkable one. Namely, Král' and Pikhurko [32] used the theory of permutation limits to characterise sequences of quasirandom permutations. We expect the theory we develop in this article to have applications mostly in this area, that is, to answer

asymptotic extremal questions about Latin squares. The set of tools we provide in this article should be complete for such purposes. Indeed, very shortly after we published this preprint at arXiv, Cooper, Král', Lamaison, Mohr [15] used our theory to characterise sequences of quasirandom Latin squares; see Section 12.4. A second line of research where limits of discrete structures opened new possibilities is more probabilistic. Namely, Chatterjee and Varadhan [11] developed a method of expressing certain density-based large deviation events of Erdős–Rényi random graphs. This theory in particular allows us to approximately count graphs with a prescribed list of density constraints. A similar programme was carried out for permutations as well, [30]. We do not offer such tools for Latin squares in the present article, but in Section 12.2 we discuss a reasonable approach.

1.1. **Work on limits of related discrete structures.** The crucial feature of the structures whose limit theory we introduce is that they are ordered; that is, the rows go from top to bottom, the columns go from left to right, and the entries go from small to large. Arguably, the first class of ordered structures for which a similar limit theory was developed are posets, [25, 23]. However, the most related work is a recent paper of Ben-Eliezer, Fischer, Levi, and Yoshida [4] (see also conference version [5]). In that paper, a theory of limits of ordered graphs, that is, graphs whose vertices are put into a linear order, is developed. The limit objects they obtain (which they call 'orderons') is less complex than our Latinons. More precisely, the fact that they are dealing with 0/1-matrices (adjacency matrices of ordered graphs) whereas we are dealing with more general matrices (the values are the set $[n]$) is reflected in that their limit objects are $[0,1]$-valued, while ours are distribution-valued. The necessity of such a more complex limit concept is best illustrated by examples, which we provide in Section 3.1. Let us point out that the key difference between [4] and our paper is how we view and treat the limit structure. We believe that our approach is conceptually simpler. We give more details in Section 8.1.

We already mentioned that our Latinons are distribution-valued, a feature that causes substantial technical complications. While indeed most limit theories lead to real-valued limit objects, the more general setting was also considered. In this direction, the most relevant paper is [33] which introduces graphons taking values from the dual of a separable Banach space. Indeed, our Latinons take values which are probability measures on $[0,1]$, and these can be viewed as elements in the dual of the space of continuous functions on $[0,1]$ (equipped with the uniform norm). However, the notion of densities considered in [33] is substantially different and we could not find a way to directly relate our setting to that of [33].

1.2. **Representation of a Latin square in a 3-dimensional matrix.** Suppose that L is a Latin square of order n. Then we can associate L with the following 3-uniform hypergraph. It has $3n$ vertices, which we label as $(k,\texttt{row})$, $(k,\texttt{column})$, $(k,\texttt{value})$, $k \in$

$[n]$. The hypergraph is partite, that is, only edges of the form `row-column-value` are present. More precisely, $(i, \texttt{row}), (j, \texttt{column}), (k, \texttt{value})$ forms an edge if and only if $L_{i,j} = k$. Observe that the hypergraph has n^2 edges, and that

(L1) for every $i, j \in [n]$ there exists a unique $k \in [n]$ such that $(i, \texttt{row}), (j, \texttt{column}), (k, \texttt{value})$ forms an edge;

(L2) for every $i, k \in [n]$ there exists a unique $j \in [n]$ such that $(i, \texttt{row}), (j, \texttt{column}), (k, \texttt{value})$ forms an edge;

(L3) for every $j, k \in [n]$ there exists a unique $i \in [n]$ such that $(i, \texttt{row}), (j, \texttt{column}), (k, \texttt{value})$ forms an edge.

Actually, this defines a one-to-one correspondence between Latin squares and 3-uniform hypergraphs with the above properties. Thus, our limit theory can be viewed as a limit theory of certain 3-uniform hypergraphs of density of relative exponent 2/3. Our understanding of general limit theories in such intermediate densities is limited,[2] so in this sense the fact that we have a complete theory may be suprising. We shall return to this in Section 3.1.4, and shall also use the hypergraph representation in our proof of the Latinon approximability result, Theorem 3.14.

Let us also note that a counterpart to the above construction of 3-hypergraphs can also be made in uniformity 2, where it corresponds to permutations, and in uniformity $\ell > 2$, where it leads to hypergraphs with $\Theta(n^{\ell-1})$ edges. These hypergraphs were first considered from this perspective by Linial and Luria [36], who call them *higher dimensional permutations*. Thus it would be natural to ask how much of our theory generalises to this setting. We address this in Section 12.3.

1.3. **Organisation of the paper.** In Section 2 we introduce and summarise necessary basics from measure theory and the theories of limits of graphs and permutations. In Section 3 we introduce the basic concepts of our theory, i.e. the definition of the limit objects, the notion of density and state our main compactness result and approximation results, Theorem 3.13 and Theorem 3.14. We further give some interesting illustrated examples. In Section 4 we introduce an analogue of the cut distance for Latinons and state related counting lemmas which imply the equivalence of the two topologies generated by left-convergence and the cut distance, respectively. We also state a sampling lemma for Latinons. In Section 5 we recall some results of the theory of dense graph limits generalised to vectors of bigraphons which we need to prove our results. In Section 6 we introduce a main concept used by our proofs, an approximation of a Latinon by a vector of bigraphons. In Section 7 we use this approximation to prove the counting lemma stated in Section 4. In Section 8 we prove the compactness result stated in Section 3. In Section 9 we prove the sampling lemma for Latinons, which we in turn use to prove the inverse counting lemma

[2] See [8, 7] for some progress in the case of graphs.

in Section 10. In Section 11 we show how to construct an approximating sequence of Latin squares for a given Latinon. Finally, we discuss some open questions in Section 12. The appendix contains proofs of two auxiliary results which are pretty standard, but which we did not find in the literature. The proof of the equivalent definition of a Latinon can be found in Appendix A. In Appendix B we prove a result concerning a decomposition of $K_{n,n,n}$ into triangles which satisfies certain requirements on densities.

2. Basic definitions and notation

2.1. **General notation.** For a set S with a strict partial order $<$ and a natural number k we write $S^k_<$ for the set of all k-tuples of S which are in increasing order according to $<$, i.e.

$$S^k_< \coloneqq \{(x_1, \dots, x_k) \in S^k \mid x_i < x_{i+1} \text{ for all } i \in [k-1]\} .$$

For a set S we denote by $\mathbb{1}_S$ the indicator function, i.e.

$$\mathbb{1}_S(x) \coloneqq \begin{cases} 1, & \text{if } x \in S; \\ 0, & \text{otherwise.} \end{cases}$$

For every $k \in \mathbb{N}$ we define a partition $\mathfrak{J}_k$ of $[0,1]$ into k parts given by $\mathsf{J}_{k,i} \coloneqq [\frac{i-1}{k}, \frac{i}{k})$ for $i \in [k-1]$ and $\mathsf{J}_{k,k} \coloneqq [\frac{k-1}{k}, 1]$. For every $d \in \mathbb{N}$ we define a partition $\mathfrak{D}_d$ of $[0,1)$ into 2^d parts given by $\mathsf{D}_{d,s} \coloneqq [\frac{s-1}{2^d}, \frac{s}{2^d})$ for $i \in [2^d]$. We also write $\mathfrak{D}_* \coloneqq \bigcup_{d\in\mathbb{N}} \mathfrak{D}_d$.

2.2. **Probability theory.** We will need the following version of the Chernoff bound, see e.g. [26, Corollary 2.3].

Lemma 2.1. *If $X \sim Bi(N,p)$, $\mu = Np$, and $\varepsilon > 0$, then*

$$\mathbb{P}[|X - \mu| \geq \varepsilon\mu] \leq 2\exp(-\varepsilon^2\mu/3) .$$

Further, we recall McDiarmid's concentration inequality, [41].

Lemma 2.2. *Suppose that $r \in \mathbb{N}$, and that Z is a random variable on a discrete product probability space $\Lambda \coloneqq \prod_{i=1}^r \Lambda_i$. Suppose that $\mathbf{c} \in \mathbb{R}^r$ is a vector with the following property. For each two vectors $\mathbf{x}, \mathbf{x}' \in \Lambda$ that differ on one coordinate, say the i-th, we have $|Z(\mathbf{x}) - Z(\mathbf{x}')| \leq \mathbf{c}_i$. Then for each $t > 0$ we have that*

$$\mathbb{P}\left[|Z - \mathbb{E}[Z]| > t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^r \mathbf{c}_i^2}\right) .$$

2.3. **Measure theory.** When working with measure spaces, we will suppress referring to the underlying sigma-algebra, unless it needs to be mentioned explicitly. Recall that measurable sets of zero measure are called *null* and their complements are called *conull*. We denote the Lebesgue measure on $\mathbb{R}$ by λ and by $\lambda^{\otimes n}$ the Lebesgue measure on $\mathbb{R}^n$. Throughout the paper Ω always denotes an arbitrary separable atomless probability space (Ω, μ). The actual choice of Ω will not be important; recall that every separable atomless probability space is almost isomorphic to $([0,1], \lambda)$. This means that there exists a conull set $\Omega' \subset \Omega$, a conull set $X \subset [0,1]$ and a measure preserving bijection $b : \Omega' \to X$ such that its inverse b^{-1} is also measurable. So while we could have worked only with $([0,1], \lambda)$ in the entire paper, we prefer to work with (Ω, μ) at the places where we want to emphasise that there is no natural linear order on that probability space. We write $\mu^{\otimes n}$ for the n-th power of μ, which is a probability measure on Ω^n.

Let $\mathcal{B}(X)$ denote the space of all Borel probability measures on a metric space X and let $\mathcal{B}^0(X)$ denote the space of all Borel signed measures of total measure between $[-1,1]$ on a metric space X. Recall that there is a natural notion of a sigma-algebra on $\mathcal{B}(X)$ and $\mathcal{B}^0(X)$; see [27, Section 17.E]. We will not need details, except that this justifies that given a function $f : \Omega \to \mathcal{B}(X)$ we can ask whether f is measurable or not. We write $\mathcal{B} \coloneqq \mathcal{B}([0,1])$ and $\mathcal{B}^0 \coloneqq \mathcal{B}^0([0,1])$. For $p \in [0,1]$, we write $\boldsymbol{\delta}_p \in \mathcal{B}$ for the Dirac measure on p.

2.3.1. *Construction of measures from semirings.* Quite a bit of the technical work deals with constructing measures, in particular in the proof of compactness, where we construct the so-called measure representation of the limit Latinon (see Section 3.2). Construction of measures on a sigma-algebra is often conveniently done by extending a premeasure. We recall the necessary definitions.

Definition 2.3 (Semiring)**.** *For a given set Ω, a set S of subsets of Ω is called a* semiring *if*

(i) $\emptyset \in S$,
(ii) $A, B \in S \Rightarrow A \cap B \in S$,
(iii) if $A, B \in S$, then there exist disjoint sets $K_1, \ldots, K_n \in S$ such that $A \setminus B = \bigcup_{i=1}^n K_i$.

Definition 2.4 (Premeasure)**.** *Suppose that S is a semiring on Ω. We say that $\mu : S \to [0,\infty)$ is a* premeasure *if for all sets $A \in S$ for which there exists a countable decomposition $A = \bigcup_{i=1}^\infty A_i$ in disjoint sets $A_i \in S$, $i \in \mathbb{N}$, we have $\mu(A) = \sum_{i=1}^\infty \mu(A_i)$.*

Theorem 2.5 (Carathéodory's extension theorem)**.** *Let S be a semiring on Ω and let $\mu : S \to [0,1]$ be a premeasure on S. Then there exists a measure $\mu' : \sigma(S) \to [0,1]$, where $\sigma(S)$ is the sigma-algebra generated by S, such that μ' is the extension of μ.*

2.3.2. *Weak convergence.* Let us also recall the notion of weak convergence of measures. Suppose that (X, ρ) is a compact metric space. We say that a sequence of Borel measures $\nu_1, \nu_2, \dots$ on X *weak converges* to a Borel measure ν if for each continuous function $f : X \to \mathbb{R}$ we have that the sequence $\int f \mathsf{d}\nu_1, \int f \mathsf{d}\nu_2, \dots$ converges to $\int f \mathsf{d}\nu$. It is well known that with the compactness assumption as above, if we have a sequence of Borel measures $\nu_1, \nu_2, \dots$ that are uniformly bounded, say $\nu_n(X) \leq C$ for all n, then there exists a subsequence $\nu_{n_1}, \nu_{n_2}, \dots$ and a Borel measure ν so that $\nu_{n_1}, \nu_{n_2}, \dots$ weak converges to ν. This follows immediately from Prokhorov's theorem [42].

2.3.3. *Disintegration theorem.* The disintegration theorem is a standard result in measure theory. Let Y and X be two Radon spaces. Let $\alpha \in \mathcal{B}(Y)$, let $\pi : Y \to X$ be a Borel-measurable function, and let $\beta \in \mathcal{B}(X)$ be the pushforward measure $\beta = \alpha \circ \pi^{-1}$. Then there exists a β-almost everywhere uniquely determined family of measures $\{\alpha_x \in \mathcal{B}(Y)\}_{x \in X}$ such that

- the function $x \mapsto \alpha_x$ is measurable,
- for β-almost all $x \in X$, $\alpha_x(Y \setminus \pi^{-1}(x)) = 0$,
- for every measurable function $s : Y \to \mathbb{R}$, we have

$$\int_{y \in Y} s(y) \mathsf{d}\alpha(y) = \int_{x \in X} \int_{y \in \pi^{-1}(x)} s(y) \mathsf{d}\alpha_x(y) \mathsf{d}\beta(x) \, .$$

The collection $\{\alpha_x \in \mathcal{B}(Y)\}_{x \in X}$ is called the *disintegration of α according to π*. Quite often, Y will be of the form $Y = X \times A$ and π will be the natural projection. We then say that $\{\alpha_x \in \mathcal{B}(Y)\}_{x \in X}$ is the *disintegration of α according to the first coordinate*.

2.4. **Graphons.** In this paper, we will borrow many tools from the theory of dense graph limits. A *graphon* W is a measurable function $W : \Omega^2 \to [0, 1]$ that is symmetric, i.e., $W(x, y) = W(y, x)$. Graphons arise as limits of sequences of graphs, and could be informally understood as limits of their adjacency matrices rescaled into a square of unit area. The symmetricity property is then inherited from the fact that an adjacency matrix of a graph is symmetric. We will need a non-symmetric version of this, called bigraphons, and first introduced in [40]. A *bigraphon* is a measurable function $W : \Omega^2 \to [0, 1]$. We can think of a bigraphon as a limit of matrices of bipartite graphs $(G_n)_n$ on $n + n$ vertices, where the two parts are distinguished, say *the first part* and *the second part*. Then W can be thought of as a limit of the 'bipartite adjacency matrices' B_n, in which the entry on position (i, j) is 1 if and only if the i-th vertex in the first part of G_n is adjacent to the j-th vertex in the second part. The reason why bigraphons will be more relevant in this paper is that the 'first part' and the 'second part' will encode the rows and the columns of the Latin square respectively. Let $\mathcal{W}_0$ be the space of all bigraphons. A *distribution-valued*

bigraphon is a measurable function $W : \Omega^2 \to \mathcal{B}$. A *distribution-valued signed bigraphon* is a measurable function $W : \Omega^2 \to \mathcal{B}^0$.

2.4.1. *Cut norm and cut distance.* Given a function W with domain Ω_1^2 and functions $\alpha, \beta : \Omega_2 \to \Omega_1$, we define $W^{\alpha,\beta}$ to be the function with domain Ω_2^2 given by $W^{\alpha,\beta}(x,y) \coloneqq W(\alpha(x), \beta(y))$ for all $x, y \in \Omega_2$. We also use the shorthand $W^\alpha \coloneqq W^{\alpha,\alpha}$. We use S_Ω to denote the set of all invertible measure preserving maps $\Omega \to \Omega$.

For $X \in L^\infty(\Omega^2)$, we define the *cut norm* of X,

$$\|X\|_\square \coloneqq \sup_{S,T \subseteq \Omega} \left| \int_{S \times T} X(x,y) \mathsf{d}\mu(x) \mathsf{d}\mu(y) \right| .$$

The *cut distance* between bigraphons $U, W \in \mathcal{W}_0$ is defined as

$$\delta_\square(U,W) \coloneqq \inf_{\varphi \in S_\Omega} \|U - W^\varphi\|_\square .$$

2.4.2. *Degrees in a bigraphon.* Suppose that $U : \Omega^2 \to [0,1]$ is a bigraphon. Given $x \in \Omega$, we define the *degree of* x in U by $\deg_U(x) \coloneqq \int_\Omega U(x,y) \mathsf{d}y$. (So, we choose the convention of integrating over the second parameter.) We prove that the degree distribution in a cut distance convergent sequence of graphons is continuous. While this fact is not difficult, we were not able to find an explicit reference in the literature.

Lemma 2.6. *Suppose that U and W are two bigraphons with $\|U - W\|_\square < \varepsilon$. Then the measure of $x \in \Omega$ for which $|\deg_U(x) - \deg_W(x)| > \sqrt{\varepsilon}$ is at most $2\sqrt{\varepsilon}$.*

Proof. Let $A \coloneqq \{x \in \Omega \mid \deg_U(x) - \deg_W(x) > \sqrt{\varepsilon}\}$ and $B \coloneqq \{x \in \Omega \mid \deg_W(x) - \deg_U(x) > \sqrt{\varepsilon}\}$. We have

$$\begin{aligned} \int_{A \times \Omega} U(x,y) \mathsf{d}(x,y) &= \int_A \deg_U(x) \mathsf{d}x \\ &\geq \int_A \deg_W(x) \mathsf{d}x + \mu(A)\sqrt{\varepsilon} = \int_{A \times \Omega} W(x,y) \mathsf{d}(x,y) + \mu(A)\sqrt{\varepsilon} . \end{aligned}$$

As $\|U - W\|_\square < \varepsilon$, we in particular get $\int_{A\times\Omega} U(x,y)\mathsf{d}(x,y) < \int_{A\times\Omega} W(x,y)\mathsf{d}(x,y) + \varepsilon$. We conclude that $\mu(A) < \sqrt{\varepsilon}$. Similarly, we can get that $\mu(B) < \sqrt{\varepsilon}$. □

From this result, we immediately obtain the following.

Lemma 2.7. *Suppose that $W_1, W_2, \ldots$ is a sequence of bigraphons converging to a bigraphon W in the cut distance. Then for each open interval I, we have*

$$\lim_n \mu(\{x \in \Omega \mid \deg_{W_n}(x) \in I\}) = \mu(\{x \in \Omega \mid \deg_W(x) \in I\}) .$$

2.5. **Permutons.** We need to recall the notion of permutons, which is actually used as a building block in our definition of Latinons. For us, a *permutation of size* n is a bijection $\pi : [n] \to [n]$. Suppose that ρ is a bijection of size k for some $k \leq n$. Then, the *density* of ρ in π, denoted by $t(\rho, \pi)$ is the probability that when picking a random k-set K of $[n]$, the *restriction of* π *on* K *is* ρ. By that we mean that for any $a, b \in [k]$, and for the corresponding a-th smallest element k_a of K and the b-th smallest element k_b of K, we have $\rho(a) < \rho(b)$ if and only if $\pi(k_a) < \pi(k_b)$. A *permuton* is a probability measure on $[0,1]^2$ with uniform marginals on both axes. The reader can find more information about permutons in [32] and [24]. Yet, we find it instructive to recall how each finite permutation can be represented as a permuton, and also that each sequence of permutations contains a subsequence 'converging' to a permuton.[3] So, suppose that $\pi : [n] \to [n]$ is a bijection and consider the partition $\mathfrak{J}_n$. Then we can represent π naturally as a permuton by defining it as an n-multiple of the Lebesgue measure on each square $\mathsf{J}_{n,j} \times \mathsf{J}_{n,\pi(j)}$ and zero elsewhere. Next, suppose that $(\pi_n : [n] \to [n])_n$ is a sequence of permutations and that $(\nu_n)_n$ are the associated permutons. Then by the compactness of the weak topology, there exists a subsequence $\nu_{n_1}, \nu_{n_2}, \ldots$ to which we have a weak limit, say ν. It is easy to check that such a measure ν must be a probability measure with uniform marginals, and hence a permuton. In this case, we say that ν is a limit of $\pi_{n_1}, \pi_{n_2}, \ldots$. The reason why the weak convergence is the right concept is that it preserves densities. More precisely, $\nu_{n_1}, \nu_{n_2}, \ldots$ weak converges to ν, if and only if for each fixed permutation ρ, we have $\lim_i t(\rho, \pi_{n_i}) = t(\rho, \nu)$. Here, the density of a permutation ρ of size k into a permuton ν is defined in the following way. Sample a k-tuple of independent points $(x_1, y_1), \ldots, (x_k, y_k) \in [0,1]^2$ according to the law of ν. Let $t(\rho, \nu)$ be the probability that for each $a, b \in [k]$ and for the corresponding a-th smallest element x_{a^*} of $\{x_1, \ldots, x_k\}$ and the b-th smallest element x_{b^*} of $\{x_1, \ldots, x_k\}$, we have $\rho(a) < \rho(b)$ if and only if $y_{a^*} < y_{b^*}$.

3. Latinons and the statements of the main results

In this section, we introduce the new key definitions. Recall that for us, a *Latin square of order* n is a matrix $L \in [n]^{n \times n}$ such that $L_{i,j} \neq L_{i,k}$ and $L_{j,i} \neq L_{k,i}$ for all $1 \leq i \leq n$ and $1 \leq j < k \leq n$. Our first goal is to generalise the density notion of permutations to two dimensions.

Let $A \in \mathbb{R}^{k \times \ell}$ be a matrix and let $O_<(A) = (a_1, \ldots, a_r)$ be the increasing ordering of $E(A) = \{a_{i,j} \mid 1 \leq i \leq k, 1 \leq j \leq \ell\}$ according to the natural order $<$. We denote by $p_A : E(A) \to [r]$ the function which assigns to each $a_{i,j}$ its index in $O_<(A)$. For example

[3] We explain later which notion of convergence is suitable for permutons.

for $A = \begin{pmatrix} 5 & 6 \\ 4 & 8 \end{pmatrix}$ we have $p_A(4) = 1, p_A(5) = 2, p_A(6) = 3, p_A(8) = 4$ and for $A' = \begin{pmatrix} 5 & 5 \\ 4 & 8 \end{pmatrix}$ we have $p_{A'}(4) = 1, p_{A'}(5) = 2, p_{A'}(8) = 3$.

Definition 3.1 (Structural equivalence)**.**

(i) We say that two matrices $A, B \in \mathbb{R}^{k\times\ell}$ are structurally equivalent*, written $A \equiv B$, if $p_A(a_{i,j}) = p_B(b_{i,j})$ for all $1 \le i \le k$ and $1 \le j \le \ell$. Matrices of different orders are never considered structurally equivalent.*

(ii) For a matrix $A \in \mathbb{R}^{k\times\ell}$ and a set $S \subseteq \mathbb{R}$ we define

$$\mathcal{R}^A(S) \coloneqq \{M \in S^{k\times\ell} \mid M \equiv A\} .$$

Remark 3.2. *Note that given k and ℓ, there are only finitely many structural equivalence classes of matrices in $\mathbb{R}^{k\times\ell}$.*

We will be interested in matrices A in $[k\ell]^{k\times\ell}$ such that $|E(A)| = k\ell$. In particular, this means that each value from $[k\ell]$ appears exactly once in A. We denote this set of matrices by $\mathcal{R}(k,\ell)$ and refer to these as *patterns*. So, the right way of thinking of a pattern is that it is a matrix with a unique smallest entry, a unique second smallest entry etc. By structural equivalence we do not really worry about the actual values, but the relative position of those entries. Given $A \in \mathcal{R}(k,\ell)$ and $B \in [0,1]^{k\times\ell}$ such that $B \equiv A$, we call B a *realisation* of A.

Let L be a Latin square of order n and $I, J \subseteq [n]$. We denote by $L|_{(I,J)}$ the submatrix which results from L after deleting all lines with indices $[n]\setminus I$ and all columns with indices $[n] \setminus J$. This notation allows us to introduce the key notion of densities in a finite Latin square, very much in parallel to the notion of densities in a finite permutation. Denote by $\mathbf{x} \in [n]^k_<$ the set of all $\mathbf{x} \in [n]^k$ such that if $\mathbf{x} = (x_1, \dots, x_k)$ then $x_1 < x_2 < \cdots < x_k$.

Definition 3.3 (Densities in Latin squares)**.** *Let L be a Latin square of order n and $A \in \mathcal{R}(k,\ell)$ for some $k, \ell \in [n]$. We define the* density *of the* pattern *A in L by*

$$t(A, L) \coloneqq \frac{1}{\binom{n}{k}\binom{n}{\ell}} \sum_{\mathbf{x}\in[n]^k_<, \mathbf{y}\in[n]^\ell_<} \mathbb{1}_{\mathcal{R}^A([n])}(L_n|_{(\mathbf{x},\mathbf{y})}) .$$

Remark 3.4. *Given any sequence $(L_n)_n$ of Latin squares of growing orders, observe that if A is a matrix in $[k\ell]^{k\times\ell}$ such that there is a repeated entry (i.e. $A \notin \mathcal{R}(k,\ell)$), then the sum in Definition 3.3 is at most $O(n^{k+\ell-1})$ and so $t(A, L_n) \to 0$ as $n \to \infty$.*

So, (when $k = \ell$) restricting a large Latin square to a random $k \times k$ submatrix, we typically get a structure in which each entry appears only once, and such a structure in particular does not correspond to any Latin square. This is in contrast with limits of dense graphs and limits of permutations, where in both cases by restricting a large structure to a

random sample of size k we get a substructure which we can interpret in that class (i.e., a graph or a permutation). This is also related to the fact that we do not think that there is a sensible notion of an 'L-random Latin square' (for a Latinon L) while there is a notion of a 'W-random graph' (for a graphon W) and of a 'ν-random permutation' (for a permuton ν, see Section 2.5).

In particular, we obtain for each fixed $k, \ell \in \mathbb{N}$ that

$$\lim_{n\to\infty} \sum_{A\in\mathcal{R}(k,\ell)} t(A, L_n) = 1 \,. \tag{1}$$

We are now ready to define our limit objects, Latinons. As these definitions are rather technical, in Section 3.1, we give some basic examples of sequences of finite Latin squares and the corresponding limit object. These examples in particular explain why our Latinons are distribution-valued, and also the role of the function f.

Definition 3.5 (Semilatinon and Latinon)**.** *A* semilatinon *is a pair (W, f) such that $W : \Omega^2 \to \mathcal{B}$ is a distribution-valued bigraphon and $f : \Omega \to [0,1]$ is measure preserving function. A* Latinon *is a semilatinon which satisfies the following properties.*

- *(i) The measure $\mu^1_{W,x}$ on $[0,1]^2$ defined by $\mu^1_{W,x}(S \times T) \coloneqq \int_{y\in f^{-1}(S)} W(x,y)(T)\mathsf{d}y$ for all measurable $S, T \subseteq [0,1]$ is a permuton for almost all $x \in \Omega$.*[4]
- *(ii) The measure $\mu^2_{W,y}$ on $[0,1]^2$ defined by $\mu^2_{W,y}(S \times T) \coloneqq \int_{x\in f^{-1}(S)} W(x,y)(T)\mathsf{d}x$ for all measurable $S, T \subseteq [0,1]$ is a permuton for almost all $y \in \Omega$.*

We denote the space of all Latinons by $\mathcal{L}$.

Condition (i) (respectively (ii)) are limit counterparts to the fact that given a Latin square, we can take its slice at any given row (resp. column) and get a permutation.

Remark 3.6. *Note that the conditions of Definition 3.5 ensure for a Latinon (W, f) that there is no set $X \subset \Omega$ of positive measure such that $W|_{X\times\Omega}$ or $W|_{\Omega\times X}$ is identically equal to $\boldsymbol{\delta}_p$ for some $p \in [0,1]$.*

We identify Latinons if they agree everywhere except a nullset. So we are actually working with the quotient space under this relation, but we will for convenience still speak of Latinons instead of equivalence classes.

Given a function $f : \Omega \to [0,1]$ we can introduce a strict partial order $<_f$ on Ω by defining $x <_f y$ if and only if $f(x) < f(y)$. We write $\mathbf{x} \in \Omega^k_{<_f}$ to denote $\mathbf{x} = (x_1, \ldots, x_k) \in \Omega^k$ such that $\mathbf{x}$ is increasingly ordered according to $<_f$, i.e. $f(x_1) < f(x_2) < \cdots < f(x_k)$. Furthermore for measures $\mu_{(1,1)}, \ldots, \mu_{(k,\ell)}$ on $[0,1]$ we write $\bigotimes_{(i,j)\in[k]\times[\ell]} \mu_{(i,j)}$ for

[4] $\mu^1_{W,x}$ is defined here only on sets of the form $S \times T$. This is sufficient as such sets generate the Borel sigma-algebra on $[0,1]^2$.

the corresponding product measure on $[0,1]^{k\times\ell}$. Now we can extend our notion of density to Latinons.

Definition 3.7 (Densities in Latinons)**.** *Let (W,f) be a Latinon over the ground space Ω and $A\in\mathcal{R}(k,\ell)$. We denote by $t(A,(W,f))$ the* density *of the* pattern *A in (W,f) and define it to be*

$$t(A,(W,f)) \coloneqq k!\ell! \int_{\mathbf{x}\in\Omega^k_{<_f}} \int_{\mathbf{y}\in\Omega^\ell_{<_f}} \left(\bigotimes_{(i,j)\in[k]\times[\ell]} W(x_i,y_j) \right) (\mathcal{R}^A([0,1]))\mathtt{dydx}\,.$$

Definition 3.8 (Sampling from a Latinon)**.** *Let (W,f) be a Latinon over the ground space Ω and $A\in\mathcal{R}(k,\ell)$. We say that $B\in[0,1]^{k\times\ell}$ is* sampled *from (W,f) if B is a matrix sampled according to the following random selection. We repeatedly choose $\mathbf{x}=(x_1,\dots,x_k)\in\Omega^k$ and $\mathbf{y}=(y_1,\dots,y_\ell)\in\Omega^\ell$ independently and uniformly at random until we find $\mathbf{x}$ and $\mathbf{y}$ which are increasingly ordered according to $<_f$. For each pair (i,j) with $1\le i\le k$ and $1\le j\le\ell$ we sample a real value $b_{i,j}$ from the distribution $W(x_i,y_j)$ and set $B=(b_{i,j})$.*

Informally, we see that $t(A,(W,f))$ is the probability that $B\equiv A$, where B is sampled from (W,f). Also note that according to our earlier definitions B is a realisation of the pattern A. The sampling view gives us the following counterpart to (1).

Fact 3.9. *Suppose that (W,f) is a Latinon. Then $\sum_{A\in\mathcal{R}(k,\ell)} t(A,L)=1$.*

The following definition will help us to associate a finite Latin square with a Latinon. The similarities to associating a permuton to a finite permutation as was done in Section 2.5 are obvious.

Definition 3.10. *Let L_n be a Latin square of order n. Given a measure preserving function $f:\Omega\to[0,1]$ we derive a Latinon (W_{L_n},f) over the ground space Ω from L_n by setting*

$$(W_{L_n}(x,y))(S) \coloneqq n\cdot\lambda\left(S\cap\left[\frac{L_{nxy}-1}{n},\frac{L_{nxy}}{n}\right]\right),\ \text{for } x,y\in\Omega \text{ and } S\subseteq[0,1]\,,$$

where $L_{nxy}\coloneqq L_n(\lceil n\cdot f(x)\rceil,\lceil n\cdot f(y)\rceil)$.

Note that this indeed defines a Latinon and in fact Latinon-representations of Latin squares have almost the same densities.

Proposition 3.11. *Let L_n be a Latin square of order n and $f:\Omega\to[0,1]$ a measure preserving function. Then for every $k,\ell\in\mathbb{N}$ and every $k\times\ell$ pattern A we have*

$$|t(A,(W_{L_n},f))-t(A,L_n)|\le\tfrac{2(k+\ell)^2}{n}.$$

Proof. Consider the probability that $B \equiv A$, where $B \in [0,1]^{k\times\ell}$ is a matrix sampled according to the following random selection. Sample as one would for $t(A,(W_{L_n},f))$ in Definition 3.7, except rather than just sampling so that $\mathbf{x}$ and $\mathbf{y}$ are increasingly ordered according to $<_f$, but also so that for each i,j we have $\lceil n\cdot f(x_i)\rceil \neq \lceil n\cdot f(x_j)\rceil$ (and similarly for the y_i). The crucial observation is that this probability is equal to $t(A,L_n)$ by definition of (W_{L_n},f). So define S_1 and S_2 to be subsets of $[0,1]^{k+\ell}$ such that

$$S_1 \coloneqq \left\{\mathbf{x}\in[0,1]^k_{<_f},\mathbf{y}\in[0,1]^\ell_{<_f} \mid \lceil n\cdot f(x_i)\rceil \neq \lceil n\cdot f(x_j)\rceil, \lceil n\cdot f(y_i)\rceil \neq \lceil n\cdot f(y_j)\rceil \text{ for } i\neq j\right\},$$

$$S_2 \coloneqq \left\{\mathbf{x}\in[0,1]^k_{<_f},\mathbf{y}\in[0,1]^\ell_{<_f}\right\}\setminus S_1 .$$

We thus obtain

$$t(A,(W_{L_n},f)) = k!\ell!\left(\int_{S_1} t(A,L_n)\mathsf{d}(\mathbf{x},\mathbf{y}) + \int_{S_2}\left(\bigotimes_{(i,j)\in[k]\times[\ell]} W(x_i,y_j)\right)(\mathcal{R}^A([0,1]))\mathsf{d}(\mathbf{x},\mathbf{y})\right),$$

and also by noting that $\lambda^{\otimes k+\ell}(S_1)+\lambda^{\otimes k+\ell}(S_2) = 1/(k!\ell!)$, we have

$$\begin{aligned}|t(A,(W_{L_n},f)) - t(A,L_n)| &\leq |k!\ell!\lambda^{\otimes k+\ell}(S_1)-1|\cdot t(A,L_n) + k!\ell!\lambda^{\otimes k+\ell}(S_2) \leq 2k!\ell!\lambda^{\otimes k+\ell}(S_2)\\ &= 2\left(1-\frac{(n-1)\cdots(n-k+1)}{n^{k-1}}\cdot\frac{(n-1)\cdots(n-\ell+1)}{n^{\ell-1}}\right)\\ &\leq 2\left(\frac{n^{k+\ell-2}-(n-k-\ell+2)^{k+\ell-2}}{n^{k+\ell-2}}\right) \leq \frac{2(k+\ell)^2}{n}.\end{aligned}$$

□

We can now define a notion of left-convergence for Latin squares and Latinons.

Definition 3.12 (Left-convergence)**.** *(i) Let $(L_n)_{n\in\mathbb{N}}$ be a sequence of Latin squares and let (W,f) be a Latinon. We say that (W,f) is the limit of $(L_n)_{n\in\mathbb{N}}$, written $L_n \overset{\text{left}}{\to} (W,f)$, if*
$\lim_{n\to\infty} t(A,L_n) = t(A,(W,f))$ for every $k,\ell\in\mathbb{N}$ and $A\in\mathcal{R}(k,\ell)$.

(ii) Let $(L_n)_{n\in\mathbb{N}}$ be a sequence of Latinons and let L be a Latinon. We say that L is the limit of $(L_n)_{n\in\mathbb{N}}$, written $L_n \overset{\text{left}}{\to} L$, if $\lim_{n\to\infty} t(A,L_n) = t(A,L)$ for every $k,\ell\in\mathbb{N}$ and $A\in\mathcal{R}(k,\ell)$.

Note that by Definition 3.10 and Proposition 3.11 we have that a sequence of Latin squares (of growing orders) converges in the sense of Definition 3.12(i) if and only if their Latinon-representations converge in the sense of Definition 3.12(ii).

Note that we require the convergence only for countably many sequences of densities, but by Remark 3.2 this implies convergence for the densities of all $A\in\mathbb{R}^{k\times\ell}$. Note also, that by a standard abstract compactness argument, if $(L_n)_{n\in\mathbb{N}}$ is a sequence of Latin squares of growing orders, then there exists a subsequence $(L_{n_i})_i$ so that for every $k,\ell\in\mathbb{N}$

and $A \in \mathcal{R}(k,\ell)$ we have that $\lim_{i\to\infty} t(A, L_{n_i})$ exists. A similar type of left-convergence is at the heart of flag algebras, [43]. However, here we aim to obtain Latinons as a more explicit representation of the limit. One of the main results of this paper is the following compactness result.

Theorem 3.13 (Compactness for Latinons)**.** *Let $(L_n)_{n\in\mathbb{N}}$ be a sequence of Latinons over arbitrary ground spaces. There exists a subsequence $(L_{n_i})_{i\in\mathbb{N}}$ and a Latinon (W,f) over the ground space Ω such that*

$$L_{n_i} \overset{\text{left}}{\to} (W,f)\ .$$

It is worth recalling that in Section 2.5 we sketched an analogous compactness result for permutons, which follows immediately from the compactness of the weak topology of probability measures on $[0,1]^2$. Although we can represent a Latinon by a probability measure (see Definition 3.16), the proof of Theorem 3.13 is much more complicated. The reason for this is that it is not enough to view coordinates of different values with global lenses, as our example in Section 3.1.2 shows. (That is, the probability measure needs to be on $\Omega^2 \times [0,1]$ rather than $[0,1]^3$.)

Recall that by Proposition 3.11 this applies to sequences of Latin squares as well. Theorem 3.13 can be accompanied by the following approximation result.

Theorem 3.14. *For every Latinon $(W,f) \in \mathcal{L}$ there exists a sequence $(L_n)_{n\in\mathbb{N}}$ of Latin squares such that $L_n \overset{\text{left}}{\to} (W,f)$.*

3.1. **Examples.** For this section only, for convenience, we will assume that all finite $n\times n$ Latin squares have row numbers, column numbers and values in $[0,n-1]$ (rather than $[n]$).

3.1.1. *Standard cyclic Latin squares.* One of the most natural examples of a sequence of Latin squares comes from the Cayley tables of cyclic groups. For each $n\in\mathbb{N}$ define the Latin square L_n by $L_n(x,y) \coloneqq x+y \mod n$ for all $x,y\in[0,n-1]$. We define the *standard cyclic Latinon* (L,f) as a Latinon on $\Omega=[0,1]$ by letting $L:[0,1]^2\to\mathcal{B}$ be defined by $L(x,y)\coloneqq \boldsymbol{\delta}_{x+y \bmod 1}$ for all $x,y\in[0,1]$, and by setting $f:[0,1]\to[0,1]$ to be the identity.

It can easily be shown that (L,f) is a Latinon, and in Section 4.2 we will show that $(L_n)_n$ converges to (L,f) as n tends to infinity. As mentioned above L_n is a Cayley table of a cyclic group of order n. By thinking of x,y and $x+y$ as points on the unit circle, one can also consider (L,f) to be the Cayley table of the group $\mathbb{R}/\mathbb{Z}$.

3.1.2. *Regular versus random alternation.* Consider the partition $\mathfrak{J}_n$ which we defined earlier as the equipartition of $[0,1]$ into n intervals. Let $C_n : [0,1]^2 \to [0,1]$ be the 'chessboard graphon', that is C_n is either constant 0 or 1 on each $\mathsf{J}_{n,i}\times\mathsf{J}_{n,j}$ depending on the parity of $i+j$. Let R_n be a graphon in which a constant 0 or 1 on each $\mathsf{J}_{n,i}\times\mathsf{J}_{n,j}$

0	1	2	3	4	5
1	2	3	4	5	0
2	3	4	5	0	1
3	4	5	0	1	2
4	5	0	1	2	3
5	0	1	2	3	4

FIGURE 1. The Latin square L_6 and the standard cyclic Latinon (L, f) in which one should think of x from 0 to 1 as running from left to right, y from 0 to 1 as running from top to bottom, and within the image, the entries are scaled from white being 0 to black being 1.

$(i \leq j)$ is chosen uniformly at random. Recall that R_n can be thought of as the adjacency matrix of an Erdős–Rényi random graph $\mathbb{G}(n, \frac{1}{2})$ (except the values on the diagonal). Thus, basic theory of graphons tells us that $(C_n)_n$ converges to the complete balanced bipartite graphon while $(R_n)_n$ converges to the constant-one-half graphon. Our examples below can be thought of as a Latin square counterpart to this difference between 'regular fifty-fifty' versus 'random fifty-fifty'.

For each n, define an $n \times n$ matrix H_n,

$$H_n(i,j) \coloneqq \begin{cases} i+j \mod n & \text{if } i+j \equiv 0 \mod 2, \\ -i-j \mod n & \text{if } i+j \equiv 1 \mod 2. \end{cases}$$

It is easy to check that for n even, H_n is actually a Latin square.

Let us now consider a matrix P_n which should be informally thought of in the following way:

$$P_n(i,j) \coloneqq \begin{cases} i+j \mod n & \text{with probability } 1/2, \\ -i-j \mod n & \text{with probability } 1/2. \end{cases}$$

One could say that P_n is a Latin square 'in expectation'. By that we mean that for each $k \in \{0, \ldots, n-1\}$, on each row i, there is exactly one column j so that $k = i + j \mod n$ and exactly one column j so that $k = -i - j \mod n$, indeed giving that the number of times the symbol k appears is 1 in expectation. As it turns out, it is indeed possible to construct deterministic Latin squares which behave like P_n. We omit the details here since this construction is described in Section 11. Now, $(P_n)_n$ exhibits the problem that $\lim_{n\to\infty} P_n(xn, yn)/n$ does not exist for $x, y \in [0,1]$. We expect the two accumulation points to be $x + y \mod 1$ and $-x - y \mod 1$. A solution to this problem is to let the limit object be a distribution-valued bigraphon. A candidate for the limit of $(P_n)_n$ could

therefore be a distribution-valued bigraphon P defined on $[0,1]^2$, where one copy of $[0,1]$ would represent rows and the other copy would represent columns,

$$P(x,y) \coloneqq \frac{1}{2}\boldsymbol{\delta}_{(x+y \bmod 1)} + \frac{1}{2}\boldsymbol{\delta}_{(-x-y \bmod 1)} \,.$$

However, the sequence $(H_n)_n$ shows that such a definition is still not general enough. In the limit and for values $(x,y) \in [0,1]^2$ we cannot distinguish the odd from the even case. Since in a square $[(x-\varepsilon)n,(x+\varepsilon)n] \times [(y-\varepsilon)n,(y+\varepsilon)n]$ half of the time the value $H_n(i,j)$ will be $i+j \mod n$ and half of the time $-i-j \mod n$, one could guess that H_n has the same limit as P_n. This is however not correct, as the densities of substructures in $(H_n)_n$ and $(P_n)_n$ converge to different limits; one can check that for the 2×3 pattern

$$A \coloneqq \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

we have that $\lim_{n\to\infty} t(A,H_n) > \lim_{n\to\infty} t(A,P_n)$. We therefore need to be able to encode 'local information' in the limit object, which suggests to consider a limit object like $H : \Omega^2 \to \mathcal{B}$, where in this case $\Omega = [0,1] \times \{\text{odd}, \text{even}\}$ and

$$H\left((x,\eta_x),(y,\eta_y)\right) = \begin{cases} x+y \mod 1 & \text{if } \eta_x = \eta_y, \\ -x-y \mod 1 & \text{if } \eta_x \neq \eta_y. \end{cases}$$

The Latinons for the above two sequences can be described as follows. For the sequence (H_n), we set (H,h) to be the Latinon on $\Omega = [0,1] \times \{\text{odd, even}\}$, where $H : ([0,1] \times \{\text{odd, even}\})^2 \to \mathcal{B}$ and $h : [0,1] \times \{\text{odd, even}\} \to [0,1]$ are defined by

$$h(x,a) \coloneqq x \,, \quad H((x,a),(y,b)) \coloneqq \begin{cases} x+y \mod 1 & \text{if } a = b, \\ -x-y \mod 1 & \text{if } a \neq b. \end{cases}$$

For the sequence (P_n), we set (P,p) to be the Latinon on $\Omega = [0,1]$, where p is the identity and $P : [0,1]^2 \to \mathcal{B}$ is defined by

$$P(x,y) \coloneqq \begin{cases} x+y \mod 1 & \text{with probability } 1/2, \\ -x-y \mod 1 & \text{with probability } 1/2. \end{cases}$$

The sequence $(H_n)_n$, which shows that in addition to the 'global position' of each row/column, some local information ('parity-like') must be recorded as well, nicely illustrates the main difference between the theory of Latinons and that of permutons. Indeed, recall that permutons record only the global position by being a measure on $[0,1]^2$. Also, recall the proof of the compactness result for permutons, as sketched in Section 2.5. The approach using the weak topology, which is the key tool in the proof, can only measure the behaviour of the finite permutations on a global scale, that is, it tracks quantities such as 'how many elements around $0.4n$ are mapped by π_n to a value of around $0.9n$'. In other

words, one cannot create any parity-like-based constructions of finite permutations that would drastically change subpermutation densities.

3.1.3. *Very Local Cyclic Latin squares.* Let us give a sequence of Latin squares which are constructed in a similar way to standard cyclic Latin squares, but change their values in a more discontinuous fashion. For every $n = k^2$ for some $k \in \mathbb{N}$, define the Latin square C_n by setting

$$C_n((i-1)k + x, (j-1)k + y) \coloneqq i + j + (x + y - 2)k \mod n$$

for every $i, j, x, y \in [k]$. Now by considering the partition of such a Latin square into $k \times k$ blocks $B_{i,j}$, $i, j \in [k]$, we see that the blocks differ very little from each other (since the role of i, j is rather inferior in the definition of C_n) whereas within a block the numbers change much more. One possible limit representation is a Latinon (C, f) on $\Omega = [0, 1] \times [0, 1]$ with $C((i, x), (j, y)) = (x + y) \mod 1$ and f being the identity. We omit a proof, which could be done using the same techniques which we use in Section 4.2 to prove that (L, f) is a correct limit to standard cyclic Latin squares.

3.1.4. *Columns and values swap.* Recall that swapping row numbers and column numbers, row numbers and values, and column numbers and values of a Latin square again results into a Latin square. It is not hard to see by the definition of left-convergence that if a sequence $(L_n)_n$ of Latin squares converges, so does the sequence $(M_n)_n$ with swapped rows and columns, i.e. $M_n(x, y) = L_n(y, x)$ for all $x, y \in [0, n-1]$. However, the situation presents itself differently in the case of swapping column numbers and values. For a finite $n \times n$ Latin square L, we denote by L' the finite Latin square obtained from L by swapping column numbers and values; that is for all $x, y, z \in [0, n-1]$ we have that if $L(x, y) = z$, then $L'(x, z) = y$. In this section we prove the following.

Lemma 3.15. *There exist two sequences of finite Latin squares $(J_n)_{n\in 3\mathbb{N}}$ and $(K_n)_{n\in 3\mathbb{N}}$ so that by interlacing them, we get a left-convergent sequence, yet by interlacing the respective sequences formed by column-value swaps, $(J'_n)_{n\in 3\mathbb{N}}$ and $(K'_n)_{n\in 3\mathbb{N}}$ we do not get a left-convergent sequence.*

Proof. Suppose n is divisible by 3. Let $I_1 \coloneqq [0, \frac{n}{3}-1]$, $I_2 \coloneqq [\frac{n}{3}, \frac{2n}{3}-1]$ and $I_3 \coloneqq [\frac{2n}{3}, n-1]$. Set $x^* \coloneqq x \mod (\frac{n}{3})$ and $y^* \coloneqq y \mod (\frac{n}{3})$. We define two Latin squares J_n and K_n by

$$J_n(x,y) \coloneqq \begin{cases} 3(x^*+y^*) \mod n & \text{if } x,y \in I_1, \text{ or } x \in I_2 \text{ and } y \in I_3, \text{ or } x \in I_3 \text{ and } y \in I_2; \\ 3(x^*+y^*+1) \mod n & \text{if } x \in I_2 \text{ and } y \in I_1, \text{ or } x \in I_1 \text{ and } y \in I_2, \text{ or } x,y \in I_3; \\ 3(x^*+y^*+2) \mod n & \text{otherwise;} \end{cases}$$

$$K_n(x,y) \coloneqq \begin{cases} 3(x^*+y^*) \mod n & \text{if } x,y \in I_1, \text{ or } x,y \in I_2, \text{ or } x,y \in I_3; \\ 3(x^*+y^*+1) \mod n & \text{if } x \in I_1 \text{ and } y \in I_3, \text{ or } x \in I_2 \text{ and } y \in I_1, \text{ or } x \in I_3 \text{ and } y \in I_2; \\ 3(x^*+y^*+2) \mod n & \text{otherwise.} \end{cases}$$

Let M_n be the restriction of either of J_n or K_n to $I_a \times I_b$ for any fixed $a, b \in [3]$. It is easy to see that $(M_n)_{n\in 3\mathbb{N}}$ converges to the standard cyclic Latinon L. Since this is true for any a, b for both J_n and K_n, it follows that both sequences $(J_n)_{n\in 3\mathbb{N}}$ and $(K_n)_{n\in 3\mathbb{N}}$ converge to the same Latinon. (This Latinon is formed of nine blocks of equal size in a 3×3 array, which each themselves are a copy of L scaled down by three.)

Now define J'_n and K'_n to be the respective Latin squares obtained from J_n and K_n by swapping column numbers and values as described above, and let A be the pattern $\begin{pmatrix} 2 & 1 \end{pmatrix}$. A simple calculation shows that

$$t(A, J'_n) = \frac{9\left(\sum_{i=0}^{\frac{n}{3}-1} \binom{\frac{n}{3}-i}{2} + \binom{i}{2}\right) + 2n(\frac{n}{3})^2}{n\binom{n}{2}},$$

whereas

$$t(A, K'_n) = \frac{9\left(\sum_{i=0}^{\frac{n}{3}-1} \binom{\frac{n}{3}-i}{2} + \binom{i}{2}\right) + n(\frac{n}{3})^2}{n\binom{n}{2}}.$$

From this we observe that

$$\lim_{n\to\infty} |t(A, J'_n) - t(A, K'_n)| = \lim_{n\to\infty} \frac{n(\frac{n}{3})^2}{n\binom{n}{2}} = \frac{2}{9},$$

and thus $(J'_n)_{n\in 3\mathbb{N}}$ and $(K'_n)_{n\in 3\mathbb{N}}$ do not converge to the same Latinon. □

3.2. **Two cryptomorphic definitions of Latinons.** We give two alternative definitions of a Latinon. The first one is as a measure on $\Omega^2 \times [0,1]$. In the second one, we do not index the rows (and columns) of a Latinon by Ω together with the measure preserving function $f : \Omega \to [0,1]$ but rather by $[0,1] \times \Gamma$ (and without any measure preserving function), where Γ is a standard probability space.

Definition 3.16 (Latinons as measures)**.** *A* Latinon in measure representation *is a pair* (ν, f) *where* $f : \Omega \to [0,1]$ *is a measure preserving function and* ν *is a Borel measure on* $\Omega^2 \times [0,1]$ *which has uniform marginals, that is, for all subsets* $X, Y \subseteq \Omega$ *and* $Z \subseteq [0,1]$

we have $\nu(X \times Y \times [0,1]) = \mu(X)\mu(Y)$, $\nu(\Omega \times Y \times Z) = \mu(Y)\lambda(Z)$, *and* $\nu(X \times \Omega \times Z) = \mu(X)\lambda(Z)$.

Definition 3.17 (Latinons indexed by global and local coordinates)**.** *Suppose that* Γ *is a standard measure space with a probability measure* γ *and* U *is a function* $U : ([0,1] \times \Gamma)^2 \to \mathcal{B}$. U *is said to be a* Latinon indexed by global and local coordinates *if for almost every* $(a,x),(b,y) \in [0,1] \times \Gamma$ *the measures* $\mu^1_{a,x}$ *on* $[0,1]^2$ *defined by* $\mu^1_{a,x}(S \times T) := \int_{b\in S}\int_{y\in\Gamma} U\big((a,x),(b,y)\big)(T)\mathsf{d}y\mathsf{d}b$ *and* $\mu^2_{b,y}$ *on* $[0,1]^2$ *defined by* $\mu^2_{b,y}(S \times T) := \int_{a\in S}\int_{x\in\Gamma} U\big((a,x),(b,y)\big)(T)\mathsf{d}x\mathsf{d}a$ *are permutons.*

We shall see in the sections below that one can indeed transform a Latinon from one representation to another. It is important that we could translate the notion of density to any of these alternative definitions. To this end, we only need to be able to sample random rows and columns.

- In case of a representation as a measure, we use the measure μ.
- In case of a representation indexed by global and local coordinates, we use the product measure $\lambda \times \gamma$.

3.2.1. *The role of* $[0,1] \times \Gamma$ *in the global-local representation.* We saw in Section 3.1 that sometimes we need to distinguish different types of rows/columns in a matrix even when these rows/columns are very close together. Because of this necessity f (in Definition 3.5) need not be injective. Definition 3.17 provides an alternative way of recording different types of local behaviour. That is, the component $[0,1]$ records the global position of rows/columns of some large Latin square while the coordinate Γ is used to store different types of rows/columns with a similar global position. Sometimes the local coordinate is irrelevant, e.g. in the standard cyclic Latinon from Section 3.1.1, whereas for (H,h) from Section 3.1.2 the space Γ would be partitioned into two pieces of measure half corresponding to odd and even.[5] The example from Section 3.1.3 shows a limit graphon whose values need not depend on the global coordinate but are highly nonconstant on the local coordinate.

3.2.2. *From a standard representation to a representation by a measure.* Suppose that (W,f) is a standard representation of a Latinon. Then define ν by setting $\nu(X \times Y \times T) := \int_{x\in X}\int_{y\in Y} W(x,y)(T)\mathsf{d}y\mathsf{d}x$ for each $X,Y \subset \Omega$ and $T \subset [0,1]$. Clearly, the uniform marginal condition is satisfied.

[5]Note that we do not have a universal partition of Γ into the odd and even part; the partition may change as we vary the global coordinate.

3.2.3. *From a representation by a measure to a standard representation.* Suppose that (ν, f) is a Latinon as in Definition 3.16. Now, it is enough to define $\{W(x,y)\}_{x,y\in\Omega}$ to be the disintegration of ν according to Ω^2. It is straightforward to check that Latinons are transformed correctly in this construction.

3.2.4. *From a standard representation to a representation with global and local coordinates.* The key to the whole transformation is Proposition 3.18 below. Proposition 3.18 looks like a standard result in real analysis. However, we were not able to find it anywhere. We thank Jan Grebík for suggesting its proof, and Martin Doležal for discussions about its subtleties. The proof is included in the appendix.

Proposition 3.18. *Suppose that (Ω, μ) is a separable atomless probability measure space, and that $f : \Omega \to [0,1]$ is a measure preserving function. Then there exists a measure preserving function $h : [0,1] \times [0,1] \to \Omega$ such that for almost every $(a,x) \in [0,1] \times [0,1]$ we have $f(h(a,x)) = a$.*

Given Proposition 3.18, it is easy to perform the desired transformation of the domain Ω^2 to $([0,1]^2)^2$. Indeed, given a Latinon (M, f) in standard representation, we take the function $h : [0,1]^2 \to \Omega$ from Proposition 3.18 and define $U : ([0,1]^2)^2 \to \mathcal{B}$ by $U((a,x),(b,y)) \coloneqq M(h(a,x),h(b,y))$.

3.2.5. *From a representation with global and local coordinates to a standard representation.* Suppose that $U : ([0,1] \times \Gamma)^2 \to \mathcal{B}$ is a Latinon as in Definition 3.17. Define a function $g : [0,1] \times \Gamma \to [0,1]$ by $(a,x) \mapsto a$. Clearly, this function is measure preserving. If $\Omega = [0,1] \times \Gamma$ (as a measure space), then (U, g) is obviously a Latinon as in Definition 3.5. In the general case, we first fix a measure preserving bijection $\phi : \Omega \to [0,1] \times \Gamma$, and set $f : \Omega \to [0,1]$, $f(z) \coloneqq g(\phi(z))$ and $W : \Omega^2 \to \mathcal{B}$, $W(u,v) \coloneqq U(\phi(u),\phi(v))$. It is straightforward to check that Latinons are transformed correctly in this construction.

3.3. **A symmetric limit theory of Latin squares.** As we explained in Section 1.2, each Latin square of order n can be represented using a 3-uniform tripartite hypergraph on $n+n+n$ vertices satisfying 1.2–1.2, and that, in fact, this representation is a bijection. The way we introduce convergence in Definition 3.12 is *a priori* asymmetric in that the rows and the columns play a different role to the values. Indeed, we saw in Section 3.1.4 that swapping columns and values can turn a convergent sequence of Latin square into a divergent one.

In this section, we briefly discuss a potential limit theory motivated purely by the representation of a Latin square as a 3-uniform tripartite hypergraph satisfying 1.2–1.2. While the extra symmetries may make such a theory seem more appealing in the context of hypergraphs we do not see how to introduce any meaningful notion of densities to derive

combinatorial applications. Such a limit theory is also briefly mentioned (in the context of quasirandomness) in [16, Section 3.1].

So, suppose that L is a Latin square encoded by a 3-uniform hypergraph H as described in Section 1.2. For $i, j, k \in [n]$ we write $(i, j, k) \in H$ if $(i, \texttt{row}), (j, \texttt{column}), (k, \texttt{value})$ forms an edge. We can represent L by a Borel measure κ_L on $[0, 1]^3$ by setting

$$\kappa_L(B) := n \cdot \lambda^{\otimes 3}\left(B \cap \bigcup_{(i,j,k)\in H} [\tfrac{i-1}{n}, \tfrac{i}{n}] \times [\tfrac{j-1}{n}, \tfrac{j}{n}] \times [\tfrac{k-1}{n}, \tfrac{k}{n}]\right)$$

for each Borel set $B \subset [0, 1]^3$. In particular, κ_L is a Borel probability measure and properties 1.2–1.2 tell us that the marginals on each of the three coordinates are permutons. We call all probability measures on $[0, 1]^3$ with the above marginal condition *symmetric-Latinons*. Recall that the space of Borel probability measures on $[0, 1]^3$ equipped with the weak topology is compact, and hence so is the space of all symmetric-Latinons. That is, if we take the weak topology as 'the cut distance'[6] for the general limit framework (F1)–(F6) then compactness is guaranteed. To complete this program, we would need a notion of densities (as in (F2) and (F5)) and a sampling lemma (as in (F6)). This could be an opening for future research, but quite possibly no combinatorially meaningful interpretation exists.

Note that relying solely on the weak topology is similar to that in the theory of permutation limits, [24].

4. A cut distance, counting lemmas, and equivalence of convergence

In this section we introduce a cut distance for Latinons $\delta_{\boxtimes}(\cdot, \cdot)$, which is a counterpart to the cut distance for graphons. The definition is given in Definition 4.3. Before giving preliminary definitions that are needed to this end, let us explain on a high level, the features we want this cut distance to have. Suppose that we have Latinons $L_1 = (W, f)$ and $L_2 = (U, g)$ over the ground space Ω. Firstly, suppose that for each interval $(a, b) \subset [0, 1]$ we have that the law of W on $f^{-1}((a, b))$ is the same as the law of U on $g^{-1}((a, b))$.[7] Informally, this means that for almost every $x \in [0, 1]$, W on the fibre $f^{-1}(x)$ can be obtained by permuting U on the fibre $g^{-1}(x)$ in a measure preserving fashion. In such a scenario, we clearly want to have $\delta_{\boxtimes}(L_1, L_2) = 0$. We want $\delta_{\boxtimes}(L_1, L_2)$ to be small if and only if we are in a situation which is only a slightly perturbed version of the above scenario, in that

[cd1] the laws of W and U on the individual fibres need not be equal, but it suffices to be approximately equal for most fibres,

[6] metrized for example by the Lévy–Prokhorov metric

[7] Note that these laws are on $\mathcal{B}$-valued functions.

[cd2] instead of pairing up a fibre $f^{-1}(x)$ with $g^{-1}(x)$, we may pair it up with $g^{-1}(x')$ if $x \approx x'$ (for most choices of x).

Definition 4.1 quantifies [cd1]. In Definition 4.2 we then introduce 'order fixing bigraphons'. These order fixing bigraphons are then used in the terms $\left\|O^f - O^{g\circ\varphi}\right\|_\square$ and $\left\|O^f - O^{g\circ\psi}\right\|_\square$ in the main formula (2) to quantify [cd2].

Definition 4.1 (Cut norm for distribution-valued signed bigraphons)**.** *Let $W : \Omega^2 \to \mathcal{B}^0$. We define*

$$\|W\|_{\boxtimes} \coloneqq \sup_{\substack{R,C\subseteq\Omega,\\ V\subseteq[0,1]\ interval}} \left|\int_{x\in R}\int_{y\in C} W(x,y)(V)\mathsf{d}x\mathsf{d}y\right| .$$

Definition 4.2 (Order fixing bigraphon)**.** *Define $O : [0,1]^2 \to [0,1]$ to be the bigraphon such that*

$$O(x,y) = \begin{cases} 1, & \text{if } x < y; \\ 0, & \text{otherwise.} \end{cases}$$

Definition 4.3 (Cut distance for Latinons)**.** *Let $L_1 = (W, f)$ and $L_2 = (U, g)$ be Latinons over the ground space Ω. We define*

$$\delta_{\boxtimes}(L_1, L_2) \coloneqq \inf_{\varphi,\psi\in S_\Omega} \left(\left\|O^f - O^{g\circ\varphi}\right\|_\square + \left\|O^f - O^{g\circ\psi}\right\|_\square + \left\|W - U^{\varphi,\psi}\right\|_{\boxtimes}\right) . \tag{2}$$

(Here, $W - U^{\varphi,\psi}$ is a difference between two distribution-valued bigraphons, and hence a distribution-valued signed bigraphon.)

It is easy to verify that $\delta_{\boxtimes}(\cdot,\cdot)$ is a pseudometric, i.e., it is symmetric and obeys the triangle inequality.

Remark 4.4. *Definition 4.3 can also be extended to semilatinons, in which case it also obeys the triangle inequality.*

The two key lemmas below state that the topologies given by left-convergence and the cut distance are equivalent. We will prove Lemma 4.5 in Section 7 and Lemma 4.6 in Section 10.

Lemma 4.5 (Counting lemma for Latinons)**.** *Let $k, \ell \in \mathbb{N}$. Then there exists a constant $c_{k,\ell}$ such that for every $A \in \mathcal{R}(k,\ell)$ and Latinons L_1, L_2 over the ground space Ω we have*

$$|t(A, L_1) - t(A, L_2)| < c_{k,\ell} \cdot \delta_{\boxtimes}(L_1, L_2)^{1/(2k\ell)} .$$

Lemma 4.6 (Inverse counting lemma for Latinons)**.** *Let L_1 and L_2 be Latinons over the ground space Ω such that $\delta_{\boxtimes}(L_1, L_2) \geq d > 0$. Then for every $k > 2^{(120/d)^{10}}$ there exists a pattern $A \in \mathcal{R}(k,k)$ such that $|t(A, L_1) - t(A, L_2)| \geq \frac{1}{2(k^{2!})}$.*

By combining Lemma 4.5 and Lemma 4.6, we immediately get the following.

Theorem 4.7 (Equivalence of convergence)**.** *Let $(L_n)_{n\in\mathbb{N}}$ be a sequence of Latinons and L be a Latinon. The following are equivalent.*

(i) $(L_n)_{n\in\mathbb{N}} \overset{\text{left}}{\to} L$.

(ii) $(L_n)_{n\in\mathbb{N}} \overset{\delta_{\boxtimes}}{\to} L$.

4.0.1. *Latinons on different ground spaces.* We said that throughout the paper, we will work with a fixed probability space (Ω, μ) on which all the Latinons will live. There could be a situation where we need to work with an additional probability space, say (Ω', μ'). Even in this situation, we can define the cut distance between a Latinon L over the ground space Ω and a Latinon K over the ground space Ω', as follows,

$$\delta^*_{\boxtimes}(L, K) \coloneqq \inf_{\psi:\Omega\to\Omega' \text{ measure preserving bijection}} \delta_{\boxtimes}(L, K^\psi) \,.$$

Here, we use the notation that if $K = (U, g)$ is a Latinon, then $K^\psi \coloneqq (U^\psi, g \circ \psi)$.

Firstly, note that transforming K into K^ψ does not change the densities.

Secondly, it is easy to check that $\delta^*_{\boxtimes}(\cdot,\cdot)$ is an extension of $\delta_{\boxtimes}(\cdot,\cdot)$, that is, if $\Omega' = \Omega$, then $\delta^*_{\boxtimes}(L, K) = \delta_{\boxtimes}(L, K)$.

Thirdly, it is straightforward to check that $\delta^*_{\boxtimes}(\cdot,\cdot)$ is symmetric and it satisfies the triangle inequality, i.e., if in addition H is a Latinon over a ground space Ω'' then we have $\delta^*_{\boxtimes}(L, H) \le \delta^*_{\boxtimes}(L, K) + \delta^*_{\boxtimes}(K, H)$.

In view of Theorem 4.7 and Proposition 3.11 it is also reasonable to define a cut distance between a Latinon $L = (W, f)$ over the ground space Ω and a Latin square L_n via its Latinon-representation $L' \coloneqq (W_{L_n}, f)$: we define $\delta_{\boxtimes}(L, L_n) \coloneqq \delta_{\boxtimes}(L, L')$.

4.1. **A sampling lemma.** The key result for proving the inverse counting lemma is a sampling lemma.[8] The sampling lemma, stated below as Lemma 4.9, tells us that a pattern sampled from a Latinon (as in Definition 3.8) is close in the cut distance.

Recall that we have a way of creating a Latinon from a Latin square (Definition 3.10). We also require an analogy for general matrices; since the object we start with is not a Latin square, we cannot hope to produce a Latinon, so we instead produce a semilatinon. The key result is that the semilatinon that will be associated with a sampled matrix from a Latinon is close to the Latinon in the cut distance with high probability.

Definition 4.8. *Suppose that $k \in \mathbb{N}$.*

(i) For a distribution-valued matrix $A \in \mathcal{B}^{k\times k}$ together with a measure preserving map $f : \Omega \to [0, 1]$ we define an associated semilatinon $L^A = (W^A, f)$ in the following way. Given the partition $\mathfrak{J}_k$, set $W^A(x, y) \coloneqq A_{i,j}$ for $x \in f^{-1}(\mathsf{J}_{k,i})$ and $y \in f^{-1}(\mathsf{J}_{k,j})$.

[8] This approach parallels that for graphons; see Section 10.6 of [38].

(ii) For a matrix $B \in [0,1]^{k\times k}$ we define the associated semilatinon $L^B = L^A$, where $A \in \mathcal{B}^{k\times k}$ is the distribution-valued matrix defined by $A_{i,j} = \boldsymbol{\delta}_{B_{i,j}}$ for $i,j \in [k]$.

(iii) For a pattern $C \in \mathcal{R}(k,k)$, define the matrix $B \in [0,1]^{k\times k}$ by $B_{i,j} \coloneqq C_{i,j}/k^2$. Then the associated semilatinon L^C of the pattern C is given by L^B defined in (ii).

We can now state the sampling lemma. Note that by Remark 4.4, the cut distance can also be applied to semilatinons.

Lemma 4.9 (Sampling lemma)**.** *Suppose that $L = (W, f)$ is a Latinon and $k \in \mathbb{N}$ is arbitrary. If $A \in [0,1]^{k\times k}$ is sampled from L according to Definition 3.8, then with probability at least $1 - 30\exp(-k^{1/8}/10)$ we have*

$$\delta_{\boxtimes}\left(L, (W^A, f)\right) \leq \frac{30}{\log(k)^{1/8}}\ .$$

We prove the sampling lemma in Section 9.

4.2. **Return to the standard cyclic Latinon example.** Recall the standard cyclic Latinon (L, f) and the Latin squares L_n from Section 3.1. We can now proceed with the proof that $L_n \overset{\text{left}}{\to} (L, f)$. First let $L'_n \coloneqq (W_{L_n}, f)$ denote the Latinon-representation of L_n. By Proposition 3.11 and Lemma 4.5, it suffices to show that $L'_n \overset{\delta\boxtimes}{\to} (L, f)$.

By taking ϕ and ψ to be the identity, we obtain

$$\delta_{\boxtimes}(L'_n, (L, f)) \leq \sup_{\substack{R,C,V\subseteq[0,1]\\ V\text{interval}}} \int_{x\in R}\int_{y\in C} W_{L_n}(x,y)(V) - L(x,y)(V)\mathsf{d}y\mathsf{d}x. \tag{3}$$

Next by the uniform marginals property of Latinons, we have for any interval $V = [z_1, z_2]$,

$$\int_{x\in[0,1]}\int_{y\in[0,1]} W_{L_n}(x,y)(V) - L(x,y)(V)\mathsf{d}y\mathsf{d}x \leq z_2 - z_1. \tag{4}$$

We have one of the following three conditions:

- There exists an integer $a \in [0, n-1]$ such that $\frac{a}{n} \leq z_1 < z_2 \leq \frac{a+1}{n}$;
- There exists an integer $a \in [1, n-1]$ such that $\frac{a-1}{n} \leq z_1 \leq \frac{a}{n} \leq z_1 \leq \frac{a+1}{n}$;
- We can write V as $[z_1, \frac{a}{n}] \cup [\frac{a}{n}, \frac{b}{n}] \cup [\frac{b}{n}, z_2]$ for integers $a, b \in [0, n-1]$, $a < b$.

Thus via (3) and (4) we obtain

(5)

$$\delta_{\boxtimes}(L'_n, (L, f)) \leq \frac{2}{n} + \sup_{\substack{R,C\subseteq[0,1]\\ a,b\in[0,n-1]}} \int_{x\in R}\int_{y\in C} W_{L_n}(x,y)\left(\left[\frac{a}{n}, \frac{b}{n}\right]\right) - L(x,y)\left(\left[\frac{a}{n}, \frac{b}{n}\right]\right)\mathsf{d}y\mathsf{d}x,$$

where $a, b \in [0, n-1]$ and $a < b$. For each $r, c \in [0, n-1]$ define

$$A_{r,c} \coloneqq \left\{(x,y) \mid x \in \mathsf{J}_{n,r}, y \in \mathsf{J}_{n,c}, x+y \leq \frac{(r+c \mod n)+1}{n}\right\};$$
$$B_{r,c} \coloneqq \left\{(x,y) \mid x \in \mathsf{J}_{n,r}, y \in \mathsf{J}_{n,c}, x+y > \frac{(r+c \mod n)+1}{n}\right\}.$$

We have

$$\int_{(x,y)\in A_{r,c}} W_{L_n}(x,y)\,(\mathsf{J}_{n,v})\,\mathsf{d}y\mathsf{d}x = \begin{cases} \frac{1}{2n^2} & \text{if } v = r+c \mod n; \\ 0 & \text{otherwise;} \end{cases}$$

$$\int_{(x,y)\in B_{r,c}} W_{L_n}(x,y)\,(\mathsf{J}_{n,v})\,\mathsf{d}y\mathsf{d}x = \begin{cases} \frac{1}{2n^2} & \text{if } v = r+c \mod n; \\ 0 & \text{otherwise;} \end{cases}$$

$$\int_{(x,y)\in A_{r,c}} L(x,y)\,(\mathsf{J}_{n,v})\,\mathsf{d}y\mathsf{d}x = \begin{cases} \frac{1}{2n^2} & \text{if } v = r+c \mod n; \\ 0 & \text{otherwise;} \end{cases}$$

$$\int_{(x,y)\in B_{r,c}} L(x,y)\,(\mathsf{J}_{n,v})\,\mathsf{d}y\mathsf{d}x = \begin{cases} \frac{1}{2n^2} & \text{if } v = r+c-1 \mod n; \\ 0 & \text{otherwise.} \end{cases}$$

Combining the above we have

$$\int_{(x,y)\in D_{r,c}} W_{L_n}(x,y)\,(\mathsf{J}_{n,v}) - L(x,y)\,(\mathsf{J}_{n,v})\,\mathsf{d}y\mathsf{d}x$$
$$= \begin{cases} \frac{1}{2n^2} & \text{if } D = B \text{ and } v = r+c \mod n; \\ -\frac{1}{2n^2} & \text{if } D = B \text{ and } v = r+c-1 \mod n; \\ 0 & \text{otherwise.} \end{cases}$$

Now by a telescoping sum we obtain

$$\int_{(x,y)\in D_{r,c}} W_{L_n}(x,y)\left(\left[\frac{a}{n},\frac{b}{n}\right]\right) - L(x,y)\left(\left[\frac{a}{n},\frac{b}{n}\right]\right)\mathsf{d}y\mathsf{d}x$$
$$= \begin{cases} \frac{1}{2n^2} & \text{if } D = B \text{ and } b = r+c-1 \mod n; \\ -\frac{1}{2n^2} & \text{if } D = B \text{ and } a = r+c-1 \mod n; \\ 0 & \text{otherwise.} \end{cases}$$

From this it is clear that the supremum in (5) is bounded by $\frac{1}{2n}$ (since at most there are n areas of weight $\frac{1}{2n^2}$ in any choice of R and C). Thus we have $\lim_{n\to\infty} \delta_{\boxtimes}(L'_n, (L,f)) \to 0$ as required.

5. Vectors of bigraphons

In our proofs we will often reduce the assertion to a claim about a vector of bigraphons. We therefore recall several results of this theory which we will use later on. We can define a metric $\delta_{\square}^{\mathbb{N}}$ on $\mathcal{W}_0^{\mathbb{N}}$ by setting

$$\delta_{\square}^{\mathbb{N}}((U_n)_{n\in\mathbb{N}},(W_n)_{n\in\mathbb{N}}) := \inf_{\varphi\in S_\Omega} \sum_{n=1}^{\infty} \frac{1}{2^n}\cdot \|U_n - W_n^{\varphi}\|_{\square} \ . \tag{6}$$

Note that this is not the same as $\sum_{n=1}^{\infty}\frac{1}{2^n}\delta_{\square}(U_n,W_n)$ and therefore compactness of $(\mathcal{W}_0^{\mathbb{N}},\delta_{\square}^{\mathbb{N}})$ does not follow immediately from Tychonoff's theorem applied to $(\mathcal{W}_0,\delta_{\square})$.

Theorem 5.1 (Generalised compactness for bigraphons)**.** *$(\mathcal{W}_0^{\mathbb{N}},\delta_{\square}^{\mathbb{N}})$ is compact.*

Theorem 5.1 is not really new. In fact, most proofs of the Lovász–Szegedy compactness theorem generalise. Nevertheless, we provide a sketch of the generalisation of the original proof of Lovász and Szegedy. To this end, we need the following 'multicolour' version of the (weak) regularity lemma. Given a measurable partition $\mathcal{P}$ of Ω and a bigraphon W, we write $W^{\bowtie\mathcal{P}}$ for the *stepping* of W according to $\mathcal{P}$. For each $A,B\in\mathcal{P}$, $W^{\bowtie\mathcal{P}}$ on $A\times B$ is defined to be a constant $\frac{1}{\mu(A)\mu(B)}\cdot\int_{A\times B} W(x,y)\mathsf{d}(x,y)$. Recall that $\mathcal{P}$ is an *equipartition* if each cell has measure $\frac{1}{|\mathcal{P}|}$.

Next, let us state a weak regularity lemma for tuples of bigraphons.

Lemma 5.2. *Suppose that $r,m\in\mathbb{N}$ are arbitrary. Let $(W_1,\dots,W_m)$ be an m-tuple of bigraphons on Ω^2. Let $\mathcal{P}^*$ be an arbitrary partition of Ω. Then,*

(i) there exists a partition $\mathcal{P}$ of Ω into at most $r\cdot|\mathcal{P}^*|$ classes that refines $\mathcal{P}^*$ and such that $\big\|W_i-(W_i)^{\bowtie\mathcal{P}}\big\|_{\square}<\sqrt{\frac{2m}{\log(r)}}$ for each $i\in[m]$, and

(ii) there exists an equipartition $\mathcal{K}$ of Ω into at most $r^2\cdot|\mathcal{P}^*|$ classes that refines $\mathcal{P}^*$ and such that $\big\|W_i-(W_i)^{\bowtie\mathcal{K}}\big\|_{\square}<\sqrt{\frac{2m}{\log(r)}}+\frac{2}{r}$ for each $i\in[m]$.

We consider Lemma 5.2 standard enough to omit giving a full proof, but still provide a summary. Let us start with Part (i). Like in the usual proofs of the weak regularity lemma (for a single graph/graphon), the central quantity is that of an index of a partition. That is, we start with a trivial partition $\mathcal{P}:=\mathcal{P}^*$, and record m indices of that partition with respect to all bigraphons $(W_1,\dots,W_m)$. Now, if $\big\|W_i-(W_i)^{\bowtie\mathcal{P}}\big\|_{\square}\geq\sqrt{\frac{2m}{\log(r)}}$ for some i, then there exists a refinement of $\mathcal{P}$ in which each cluster is split into at most 2, and such that i-th index goes up by at least $\frac{m}{\log(r)}$. Also, a refinement of a partition never decreases an index, which in particular applies to the remaining $m-1$ indices. Since the sum of the indices is between 0 and m, we conclude that we end up with a regular partition in at most $m/\frac{m}{\log(r)}=\log(r)$ steps. So the number, say N, of clusters at the end of this process will be at most $|\mathcal{P}^*|\cdot 2^{\log(r)}=r\cdot|\mathcal{P}^*|$.

For the proof of Part (ii) we use the proof for Part (i), but need to turn the partition into an equipartition. We cut each cell X of $\mathcal{P}$ arbitrarily into $\lfloor \frac{\mu(X)}{rN} \rfloor$ cells, of individual measure $\frac{1}{rN}$ each; of course, there is a possible leftover of measure less than $\frac{1}{rN}$. We aggregate all the leftover sets; this aggregate set A has measure less than $N \cdot \frac{1}{rN} = \frac{1}{r}$. We cut A arbitrarily into $\frac{\mu(A)}{rN}$ cells, of individual measure $\frac{1}{rN}$ each (note that $\mu(A)$ is an integer multiple of $\frac{1}{rN}$, so this is really possible). Clearly, we now have an equipartition $\mathcal{K}$ with exactly $rN \leq r^2 \cdot |\mathcal{P}^*|$ parts. Furthermore, it is a property of the cut norm distance that it does not increase due to refinements. So, the only increase in the cut norm distance that could have been caused by the leftover A, that is, $\left\|W_i - (W_i)^{\bowtie\mathcal{K}}\right\|_\square \leq \left\|W_i - (W_i)^{\bowtie\mathcal{P}}\right\|_\square + 2\mu(A) < \sqrt{\frac{2m}{\log(r)}} + \frac{2}{r}$, as needed.

Sketch of proof of Theorem 5.1. Let us consider an arbitrary sequence $\mathbf{W}^1, \mathbf{W}^2, \ldots$ of countable tuples of bigraphons, $\mathbf{W}^i = (W_1^i, W_2^i, \ldots)$. For a given i and $m \in \mathbb{N}$, let us write $\mathbf{W}^i_{\restriction m}$ for the m-tuple $\mathbf{W}^i_{\restriction m} \coloneqq (W_1^i, W_2^i, \ldots, W_{m-1}^i, W_m^i)$. Let us take a sequence $(\varepsilon_\ell)_{\ell\in\mathbb{N}}$ of positive numbers that tend to 0. Set $\mathbf{j}_0 \coloneqq (1, 2, \ldots)$ and $\mathcal{P}_{0,q} \coloneqq \{\Omega\}$ for each $q \in \mathbb{N}$. Set $r_0 \coloneqq 1$ and $r_{i+1} \coloneqq r_i \cdot 2^{2/\varepsilon_i^3}$.

Now, for $\ell = 1, 2, \ldots$ we do the following. Set $m \coloneqq \lfloor 1/\varepsilon_\ell \rfloor$. For each $q = 1, 2, \ldots$, we apply Lemma 5.2 with a prepartition $\mathcal{P}_{\ell-1,q}$, m-tuple of bigraphons that is q-th with respect to $\mathbf{j}_{\ell-1}$, that is, $\mathbf{W}^{\mathbf{j}_{\ell-1}(q)}_{\restriction m}$, with prepartition $\mathcal{P}_{\ell-1,q}$. Lemma 5.2 yields a partition $\mathcal{Q}_{\ell,q}$, $|\mathcal{Q}_{\ell,q}| \leq r_i \cdot |\mathcal{P}_{\ell-1,q}| \leq r_\ell$, which refines $\mathcal{P}_{\ell-1,q}$ and with the property that $\delta_\square\left(W_i^{\mathbf{j}_{\ell-1}(q)}, (W_i^{\mathbf{j}_{\ell-1}(q)})^{\bowtie\mathcal{Q}_{\ell,q}}\right) < \varepsilon_\ell$ for each $i \in [\lceil 1/\varepsilon_\ell \rceil]$. We now prepare ourselves for the step $\ell+1$. We pass to an infinite subsequence $\mathbf{j}_\ell \subset \mathbf{j}_{\ell-1}$ so that in this subsequence all the partitions $\mathcal{Q}_{\ell,q}$ have the same size, say R, and so that we have the following.

- Order the cells of $\mathcal{Q}_{\ell,q}$ arbitrarily as $Q_{\ell,q;1}, \ldots, Q_{\ell,q;R}$. Then, as $q \to \infty$, the vectors $(\mu(Q_{\ell,q;1}), \ldots, \mu(Q_{\ell,q;R}))$ converge,
- For each $i \in [\lceil 1/\varepsilon_\ell \rceil]$, the bigraphons $(W_i^{\mathbf{j}_\ell(q)})^{\bowtie\mathcal{Q}_{\ell,q}}$, when viewed as $R \times R$ matrices ordered as above, converge, as $q \to \infty$.

Note that such a subsequence exists, since the above convergence notions refer to the (compact) spaces $[0,1]^R$ and $[0,1]^{R^2}$, respectively. We can proceed with the step $\ell + 1$.

Now, the diagonalisation is done as in the Lovász–Szegedy proof. □

We will also need to consider homomorphism densities of oriented graphs in m-tuples of bigraphons. This extension of the concept of a homomorphism to tuples plays a prominent role for example in [52]. Recall that in an oriented graph, each edge has an initial and a terminal vertex; parallel or counterparalled edges and self-loops are not allowed.

Definition 5.3 (Densities in vectors of bigraphons)**.** *Let $\mathbf{W} = (W_1, \ldots, W_m)$ be an m-tuple of bigraphons on Ω^2, let F be an oriented graph and let $\alpha : E(F) \to [m]$ be any map.*

We then set

$$t_\alpha(F, \mathbf{W}) \coloneqq \int_{\mathbf{x}\in\Omega^{V(F)}} \prod_{(x_i,x_j)\in E(F)} W_{\alpha(x_i,x_j)}(x_i, x_j)\mathsf{d}\mathbf{x} \ .$$

We will make use of a counting lemma which is the following slightly generalised version of a lemma from [52, Lemma 5.3].

Lemma 5.4 (Counting lemma for bigraphons)**.** *Let* $\mathbf{U} = (U_1, \dots, U_m)$ *and* $\mathbf{W} = (W_1, \dots, W_m)$ *be two* m*-tuples of bigraphons. Suppose that* $\|W_i - U_i\|_\square \leq \varepsilon$ *for each* $i \in [m]$*. Then for any oriented graph* F *and any map* $\alpha : E(F) \to [m]$ *we have*

$$|t_\alpha(F, \mathbf{U}) - t_\alpha(F, \mathbf{W})| \leq \varepsilon \cdot e(F) \ ,$$

where $e(F)$ *is the number of edges in* F*.*

Furthermore, we will make use of the following 'First sampling lemma' for bigraphons in a form analogous to that stated in [38, Lemma 10.6] for graphons. Note that considering bigraphons does not affect the argument of the proof, so the analogous statement holds. Given a bigraphon U on $[0,1]^2$, $k \in \mathbb{N}$, and $S, T \in [0,1]^k$, we define $U[S,T]$ to be the bigraphon on $[0,1]^2$ defined in the following way. For every $i, j \in [k]$ and $x \in \mathsf{J}_{k,i}$, $y \in \mathsf{J}_{k,j}$ we set $U[S,T](x,y) \coloneqq U(S_i, T_j)$.

Lemma 5.5 (First sampling lemma for bigraphons)**.** *Let* $k \geq 1$*, and let* U *be a bigraphon on* $[0,1]^2$*. If* S, T *are chosen independently uniformly at random from* $[0,1]^k$*, then with probability at least* $1 - 4\exp(-\sqrt{k}/10)$*,*

$$|\|U[S,T]\|_\square - \|U\|_\square| \leq \tfrac{8}{k^{1/4}} \ .$$

We will need an ordered version which fits better to Definition 4.8. It follows immediately from Lemma 5.5 by noting that $\left\|U^{\varphi,\psi}\right\|_\square = \|U\|_\square$ for every $\varphi, \psi \in S_{[0,1]}$.

Lemma 5.6 (First sampling lemma for bigraphons (ordered version))**.** *Let* $k \geq 1$*, and let* U *be a bigraphon on* $[0,1]^2$*. If* S, T *are chosen independently uniformly at random from* $[0,1]^k_<$*, then with probability at least* $1 - 4\exp(-\sqrt{k}/10)$*,*

$$|\|U[S,T]\|_\square - \|U\|_\square| \leq \tfrac{8}{k^{1/4}} \ .$$

6. Compressions of Latinons

Several of our proofs are based on the idea that we can approximate a Latinon by a vector of bigraphons, where the length of the vector depends on the precision of the approximation. We will also consider vectors of infinite length which contain all of the finite approximations.

Definition 6.1 (Compressions of Latinons)**.** *Let $d \in \mathbb{N}$. For a distribution-valued bigraphon W on Ω^2 we define the* compression of depth d *of W to be the 2^d-tuple of bigraphons on Ω^2*

$$\mathbf{W}^d = (W_{d,1}, \dots, W_{d,2^d}) ,$$

where for every $s \in [2^d]$ we have

$$W_{d,s}(x,y) = W(x,y)(\mathsf{D}_{d,s}) \text{ for all } x,y \in \Omega .$$

We define the infinite compression *of W to be the vector $\mathbf{W}^{\mathbb{N}} \in \mathcal{W}_0^{\mathbb{N}}$ with*

$$\mathbf{W}^{\mathbb{N}} = (\mathbf{W}^1, \mathbf{W}^2, \dots) .$$

For a Latinon $L = (W, f)$ we define the compression of depth d *of L to be the $2^d + 1$-tuple of bigraphons on Ω^2*

$$\mathbf{L}^d = (O^f, \mathbf{W}^d) .$$

We define the infinite compression *of L to be the vector $\mathbf{L}^{\mathbb{N}} \in \mathcal{W}_0^{\mathbb{N}}$ with*

$$\mathbf{L}^{\mathbb{N}} = (O^f, \mathbf{W}^{\mathbb{N}}) .$$

The next observation states that most times when working with a Latinon, it does not make a difference whether we work with open, closed or half open intervals in the definition of a compression.

Observation 6.2. *Suppose that $L = (W, f)$ is a Latinon. Then there exists a nullset $S \subseteq \Omega^2$ such that for every $(x,y) \in \Omega^2 \setminus S$ we have that $W(x,y)((a,b)) = W(x,y)([a,b])$ for all $a, b \in \mathbb{Q} \cap [0,1]$ with $a < b$.*

Proof. It is enough to prove that for each $t \in [0,1]$, the set of $(x,y) \in \Omega^2$ for which $W(x,y)(\{t\}) > 0$ is null. Indeed, once this is settled, it is enough to take S to be the union of these exceptional nullsets, over all $t \in \mathbb{Q} \cap [0,1]$. But the former easily follows from the uniform marginals property (Definition 3.5),

$$\int_{x \in \Omega} \int_{y \in \Omega} W(x,y)(\{t\}) \mathsf{d}x \mathsf{d}y = \lambda(\{t\}) = 0 .$$

□

Thus, with the help of Observation 6.2, we have the following nestedness property.

Observation 6.3. *Suppose that $L = (W, f)$ is a Latinon. Then there exists a nullset $S \subseteq \Omega^2$ such that for each $d \in \mathbb{N}$ and each $(x,y) \in \Omega^2 \setminus S$ we have for its compressions of depth d and $d+1$ and each $i \in [2^d]$ that $W_{d,i}(x,y) = W_{d+1,2i-1}(x,y) + W_{d+1,2i}(x,y)$.*

In particular, for $(x,y) \in \Omega^2 \setminus S$ we have $\sum_{i=1}^{2^d} W_{d,i}(x,y) = 1$.

Next, we show how to proceed in the other direction, from compressions to Latinons.

Definition 6.4. *Suppose that $d \in \mathbb{N}$. Given a 2^d-tuple of bigraphons $\mathbf{W} = (W_1, \dots, W_{2^d})$ on Ω^2 with*

$$\sum_{i=1}^{2^d} W_i(x,y) = 1 \tag{7}$$

for each $(x,y) \in \Omega^2$, we can define a distribution-valued bigraphon, denoted by $\nabla(\mathbf{W})$, for any $(x,y) \in \Omega^2$ and measurable set $X \subset [0,1]$, by

$$\nabla(\mathbf{W})(x,y)(X) \coloneqq \sum_{i=1}^{2^d} W_i(x,y) \cdot 2^d \cdot \lambda(X \cap \mathsf{D}_{d,i}) \,.$$

Note that (7) ensures that the values of $\nabla(\mathbf{W})$ are indeed probability distributions.

The following observations are immediate.

Fact 6.5. *Suppose that (W,f) is a Latinon and $d \in \mathbb{N}$. Let $(O^f, W_1, \dots, W_{2^d})$ be the compression of depth d of W. Then $(\nabla(W_1, \dots, W_{2^d}), f)$ is a Latinon. Further, its compression of depth d is $(O^f, W_1, \dots, W_{2^d})$.*

Fact 6.5 allows to extend the symbol $\nabla(\cdot)$ even to compressions of Latinons.

Lemma 6.6 (Latinon-distance of compressions)**.** *Let $d \in \mathbb{N}$.*

(i) If W is a distribution-valued bigraphon with uniform marginals and $\mathbf{W}^d$ its compression of depth d, then

$$\left\| W - \nabla(\mathbf{W}^d) \right\|_{\boxtimes} \leq \frac{1}{2^{d-1}} \,.$$

(ii) If L is a Latinon and $\mathbf{L}^d$ its compression of depth d, then

$$\delta_{\boxtimes}\left(L, \nabla(\mathbf{L}^d)\right) \leq \frac{1}{2^{d-1}} \,.$$

Proof. For (i) let $S, T \subseteq \Omega$ and $V \subseteq [0,1]$ be an interval with its smallest value in $\mathsf{D}_{d,s}$ and its largest in $\mathsf{D}_{d,t}$. We then have

$$\begin{aligned}
&\int_S \int_T W(x,y)(V) \mathsf{d}y \mathsf{d}x \\
&= \int_S \int_T \sum_{i=s+1}^{t-1} W_{d,i}(x,y) \mathsf{d}y \mathsf{d}x + \int_S \int_T W(x,y)\left(V \cap (\mathsf{D}_{d,s} \cup \mathsf{D}_{d,t})\right) \mathsf{d}y \mathsf{d}x \\
&= \int_S \int_T \mathbf{W}^d(x,y)(V) \mathsf{d}y \mathsf{d}x \pm \tfrac{2}{2^d} \,,
\end{aligned}$$

where we use that W has uniform marginals for the last equality.

Next, observe that (ii) follows immediately from (i). □

The next fact helps to compare compressions of Latinons by comparing the individual bigraphons via cut distance.

Proposition 6.7. *Let $d \in \mathbb{N}$, let W_1, W_2 be distribution-valued bigraphons with uniform marginals, and let $\mathbf{W_1}^d = (W_1^{d,1}, \dots, W_1^{d,2^d})$ and $\mathbf{W_2}^d = (W_2^{d,1}, \dots, W_2^{d,2^d})$ be their compressions of depth d. Then*

$$\left\| \mathbf{W_1}^d - \mathbf{W_2}^d \right\|_{\boxtimes} \leq \sum_{i=1}^{2^d} \left\| W_1^{d,i} - W_2^{d,i} \right\|_{\square} .$$

Proof. Set $\mathbf{U}^d := \mathbf{W_1}^d - \mathbf{W_2}^d$ and $U^{d,i} := W_1^{d,i} - W_2^{d,i}$. For $S, T \subseteq \Omega$ and an interval $V \subseteq [0,1]$ we then have

$$\begin{aligned}
\left| \int_S \int_T \mathbf{U}^d(x,y)(V) \mathsf{d}x\mathsf{d}y \right| &= \left| \int_S \int_T \sum_{i=1}^{2^d} \mathbf{U}^d(x,y) \left(V \cap \mathsf{D}_{d,i}\right) \mathsf{d}x\mathsf{d}y \right| \\
&= \left| \sum_{i=1}^{2^d} 2^d \cdot \lambda\left(V \cap \mathsf{D}_{d,i}\right) \cdot \int_S \int_T U^{d,i}(x,y) \mathsf{d}x\mathsf{d}y \right| \\
&\leq \sum_{i=1}^{2^d} \left\| W_1^{d,i} - W_2^{d,i} \right\|_{\square} .
\end{aligned}$$

□

We conclude the section by showing that the infinity-distance of infinite compressions bounds the Latinon-distance of the finite compressions.

Proposition 6.8. *Let $d \in \mathbb{N}$. If L_1 and L_2 are Latinons, then*

$$\delta_{\boxtimes}(\mathbf{L_1}^d, \mathbf{L_2}^d) < 2^{2^{d+1}} \cdot \delta_{\square}^{\mathbb{N}}(\mathbf{L_1}^{\mathbb{N}}, \mathbf{L_2}^{\mathbb{N}}).$$

Proof. We write $\mathbf{L_1}^{\mathbb{N}} = (O^f, W_{1,1}, \dots)$, $\mathbf{L_2}^{\mathbb{N}} = (O^g, U_{1,1}, \dots)$ and $\delta = \delta_{\square}^{\mathbb{N}}(\mathbf{L_1}^{\mathbb{N}}, \mathbf{L_2}^{\mathbb{N}})$. Now there exists $\varphi \in S_\Omega$ such that

$$\tfrac{1}{2} \left\| O^{f \circ \varphi} - O^g \right\|_{\square} + \sum_{d=1}^{\infty} \sum_{k=1}^{2^d} \tfrac{1}{2^{2^d + k - 1}} \left\| W_{d,k}^{\varphi} - U_{d,k} \right\|_{\square} < 2\delta ,$$

and therefore by Proposition 6.7,

$$\delta_{\boxtimes}(\mathbf{L_1}^d, \mathbf{L_2}^d) \leq 2 \left\| O^{f \circ \varphi} - O^g \right\|_{\square} + \sum_{k=1}^{2^d} \left\| W_{d,k}^{\varphi} - U_{d,k} \right\|_{\square} < 2^{2^{d+1}} \delta .$$

□

7. Proof of the counting lemma

In this section, we prove the counting lemma, Lemma 4.5. We make use of the compressions of Latinons defined in Section 6.

Proof of Lemma 4.5. Let $L_1 = (W, f)$ and $L_2 = (U, g)$ be Latinons, $A \in \mathcal{R}(k, \ell)$ and $d \in \mathbb{N}$. We write $\mathbf{L_1}^d = (O^f, W_{d,1}, \dots, W_{d,2^d})$ and $\mathbf{L_2}^d = (O^g, U_{d,1}, \dots, U_{d,2^d})$ for the corresponding compressions of depth d. The calculations below are valid for each $d \in \mathbb{N}$, which we will choose only in the last step of the proof.

Note that by Definition 4.3 there exist $\varphi, \psi \in S_\Omega$ such that

$$\delta_{\boxtimes}(L_1, L_2) \geq \left\| O^f - O^{g\circ\varphi} \right\|_\square ,$$
$$\delta_{\boxtimes}(L_1, L_2) \geq \left\| O^f - O^{g\circ\psi} \right\|_\square ,$$

and

$$\begin{aligned}\delta_{\boxtimes}(L_1, L_2) &\geq \left\| W - U^{\varphi,\psi} \right\|_{\boxtimes} \\ &\geq \sup_{R,C\subseteq\Omega} \left| \int_{x\in R}\int_{y\in C} W(x,y)(\mathsf{D}_{d,i}) - U^{\varphi,\psi}(x,y)(\mathsf{D}_{d,i})\mathsf{d}x\mathsf{d}y \right| \\ &= \sup_{R,C\subseteq\Omega} \left| \int_{x\in R}\int_{y\in C} W_{d,i}(x,y) - U_{d,i}(\varphi(x),\psi(y))\mathsf{d}x\mathsf{d}y \right| = \left\| W_{d,i} - U^{\varphi,\psi}_{d,i} \right\|_\square ,\end{aligned}$$

for every $i \in [2^d]$. We now define two tuples, $\mathbf{W} = (O^f, O^f, W_{d,1}, \dots, W_{d,2^d})$ and $\mathbf{U} = (O^{g\circ\varphi}, O^{g\circ\psi}, U^{\varphi,\psi}_{d,1}, \dots, U^{\varphi,\psi}_{d,2^d})$. We also define an oriented graph F with vertex set

$$\{a_i \mid i \in [k]\} \cup \{b_i \mid i \in [\ell]\}$$

and edge set

$$\{(a_i, b_j) \mid i \in [k], j \in [\ell]\} \cup \{(a_i, a_j) \mid 1 \leq i < j \leq k\} \cup \{(b_i, b_j) \mid 1 \leq i < j \leq \ell\} .$$

Furthermore, given a matrix $P = (p_{i,j})_{i\in[k],j\in[\ell]} \in \mathcal{R}^A([2^d])$, we define a map $\alpha_P : E(F) \to [2 + 2^d]$ by $(a_i, a_j) \mapsto 1$, $(b_i, b_j) \mapsto 2$, and $(a_i, b_j) \mapsto p_{i,j} + 2$ for $i \in [k], j \in [\ell]$. We then get by Lemma 5.4 and the above inequalities that

$$|t_{\alpha_P}(F, \mathbf{W}) - t_{\alpha_P}(F, \mathbf{U})| \leq |F| \cdot \delta_{\boxtimes}(L_1, L_2) = (k\ell + \tbinom{k}{2} + \tbinom{\ell}{2}) \cdot \delta_{\boxtimes}(L_1, L_2) . \tag{8}$$

We will make use of the following notation. Let S be a set and $(m_{i,j}) = M \in S^{k\times\ell}$. Furthermore given a partition $\mathcal{Q} = (Q_1, \dots, Q_d)$ of a collection of subsets of S, we write $M \perp \mathcal{Q}$ if there are no $(i_1, j_1), (i_2, j_2) \in [k] \times [\ell]$ and $s \in [d]$ such that $(i_1, j_1) \neq (i_2, j_2)$ and $m_{i_1,j_1}, m_{i_2,j_2} \in Q_s$.

Claim 7.1. *We have that* $\sum_{P\in\mathcal{R}^A([2^d])} t_{\alpha_P}(F,\mathbf{W})$ *is equal to*

$$\int_{\mathbf{x}\in\Omega^k_{<_f}}\int_{\mathbf{y}\in\Omega^\ell_{<_f}}\left(\bigotimes_{(i,j)\in[k]\times[\ell]} W(x_i,y_j)\right)(\{M\in\mathcal{R}^A([0,1])\mid M\perp\mathfrak{D}_d\})\mathsf{dydx}\,.$$

Proof of Claim 7.1. Indeed,

$$\begin{aligned}
&\sum_{P\in\mathcal{R}^A([2^d])} t_{\alpha_P}(F,\mathbf{W})\\
&=\sum_{P\in\mathcal{R}^A([2^d])}\int_{\mathbf{x}\in\Omega^k}\int_{\mathbf{y}\in\Omega^\ell}\prod_{i=1}^{k}\prod_{j=1}^{\ell} W_{d,p_{i,j}}(x_i,y_j)\mathbb{1}_{\Omega^k_{<_f}}(\mathbf{x})\mathbb{1}_{\Omega^\ell_{<_f}}(\mathbf{y})\mathsf{dydx}\\
&=\sum_{P\in\mathcal{R}^A([2^d])}\int_{\mathbf{x}\in\Omega^k_{<_f}}\int_{\mathbf{y}\in\Omega^\ell_{<_f}}\prod_{i=1}^{k}\prod_{j=1}^{\ell} W_{d,p_{i,j}}(x_i,y_j)\mathsf{dydx}\\
&=\sum_{P\in\mathcal{R}^A([2^d])}\int_{\mathbf{x}\in\Omega^k_{<_f}}\int_{\mathbf{y}\in\Omega^\ell_{<_f}}\prod_{i=1}^{k}\prod_{j=1}^{\ell} W(x_i,y_j)(\mathsf{D}_{d,p_{i,j}})\mathsf{dydx}\\
&=\int_{\mathbf{x}\in\Omega^k_{<_f}}\int_{\mathbf{y}\in\Omega^\ell_{<_f}}\left(\bigotimes_{(i,j)\in[k]\times[\ell]} W(x_i,y_j)\right)\Big(\bigcup_{P\in\mathcal{R}^A([2^d])}(\mathsf{D}_{d,p_{i,j}})_{(i,j)\in[k]\times[\ell]}\Big)\mathsf{dydx}\\
&=\int_{\mathbf{x}\in\Omega^k_{<_f}}\int_{\mathbf{y}\in\Omega^\ell_{<_f}}\left(\bigotimes_{(i,j)\in[k]\times[\ell]} W(x_i,y_j)\right)(\{M\in\mathcal{R}^A([0,1])\mid M\perp\mathfrak{D}_d\})\mathsf{dydx}\,.
\end{aligned}$$

□

Claim 7.2. *We have that* $\sum_{P\in\mathcal{R}^A([2^d])} t_{\alpha_P}(F,\mathbf{U})$ *is equal to*

$$\int_{\mathbf{x}\in\Omega^k_{<_g}}\int_{\mathbf{y}\in\Omega^\ell_{<_g}}\left(\bigotimes_{(i,j)\in[k]\times[\ell]} U(x_i,y_j)\right)(\{M\in\mathcal{R}^A([0,1])\mid M\perp\mathfrak{D}_d\})\mathsf{dydx}\,.$$

Proof of Claim 7.2. The calculations are similar as in Claim 7.1, but in addition use that φ and ψ are measure preserving bijections.

$$\begin{aligned}
&\sum_{P\in\mathcal{R}^A([2^d])} t_{\alpha_P}(F,\mathbf{U})\\
&= \sum_{P\in\mathcal{R}^A([2^d])} \int_{\mathbf{x}\in\Omega^k}\int_{\mathbf{y}\in\Omega^\ell} \prod_{i=1}^{k}\prod_{j=1}^{\ell} U_{d,p_{i,j}}(\varphi(x_i),\psi(y_j))\mathbb{1}_{\Omega^k_{<_g}}(\varphi(\mathbf{x}))\mathbb{1}_{\Omega^\ell_{<_g}}(\psi(\mathbf{y}))\mathsf{d}\mathbf{y}\mathsf{d}\mathbf{x}\\
&= \sum_{P\in\mathcal{R}^A([2^d])} \int_{\mathbf{x}\in\Omega^k}\int_{\mathbf{y}\in\Omega^\ell} \prod_{i=1}^{k}\prod_{j=1}^{\ell} U_{d,p_{i,j}}(x_i,y_j)\mathbb{1}_{\Omega^k_{<_g}}(\mathbf{x})\mathbb{1}_{\Omega^\ell_{<_g}}(\mathbf{y})\mathsf{d}\mathbf{y}\mathsf{d}\mathbf{x}\\
&= \int_{\mathbf{x}\in\Omega^k_{<_g}}\int_{\mathbf{y}\in\Omega^\ell_{<_g}} \left(\bigotimes_{(i,j)\in[k]\times[\ell]} U(x_i,y_j)\right)(\{M\in\mathcal{R}^A([0,1]) \mid M\perp\mathfrak{D}_d\})\mathsf{d}\mathbf{y}\mathsf{d}\mathbf{x}\,.
\end{aligned}$$

□

Claim 7.3. *We have*

$$\int_{x_1\in\Omega}\int_{y_1\in\Omega}\int_{y_2\in\Omega} W(x_1,y_1)(\mathsf{D}_{d,r})W(x_1,y_2)(\mathsf{D}_{d,r})\mathsf{d}y_2\mathsf{d}y_1\mathsf{d}x_1 = \left(\tfrac{1}{2^d}\right)^2$$

and

$$\int_{x_1\in\Omega}\int_{x_2\in\Omega}\int_{y_1\in\Omega} W(x_1,y_1)(\mathsf{D}_{d,r})W(x_2,y_1)(\mathsf{D}_{d,r})\mathsf{d}y_1\mathsf{d}x_2\mathsf{d}x_1 = \left(\tfrac{1}{2^d}\right)^2\,.$$

Proof of Claim 7.3. It is enough to prove the first equality, the second equality follows by symmetry of the following argument.

$$\begin{aligned}
&\int_{x_1\in\Omega}\int_{y_1\in\Omega}\int_{y_2\in\Omega} W(x_1,y_1)(\mathsf{D}_{d,r})W(x_1,y_2)(\mathsf{D}_{d,r})\mathsf{d}y_2\mathsf{d}y_1\mathsf{d}x_1\\
&= \int_{x_1\in\Omega}\int_{y_1\in\Omega} W(x_1,y_1)(\mathsf{D}_{d,r})\int_{y_2\in\Omega} W(x_1,y_2)(\mathsf{D}_{d,r})\mathsf{d}y_2\mathsf{d}y_1\mathsf{d}x_1\\
&= \int_{x_1\in\Omega}\int_{y_1\in\Omega} W(x_1,y_1)(\mathsf{D}_{d,r})\mu^1_{W,x_1}(\Omega\times\mathsf{D}_{d,r})\mathsf{d}y_1\mathsf{d}x_1\\
&= \int_{x_1\in\Omega} \mu^1_{W,x_1}(\Omega\times\mathsf{D}_{d,r})\int_{y_1\in\Omega} W(x_1,y_1)(\mathsf{D}_{d,r})\mathsf{d}y_1\mathsf{d}x_1\\
&= \int_{x_1\in\Omega} \mu^1_{W,x_1}(\Omega\times\mathsf{D}_{d,r})\mu^1_{W,x_1}(\Omega\times\mathsf{D}_{d,r})\mathsf{d}x_1\\
&= \int_{x_1\in\Omega} \tfrac{1}{2^d}\tfrac{1}{2^d}\mathsf{d}x_1\\
&= \left(\tfrac{1}{2^d}\right)^2\,,
\end{aligned}$$

where we use the fact that the 2-dimensional measure μ^1_{W,x_1} has uniform marginals for almost every $x_1 \in \Omega$, as (W, f) is a Latinon. □

Claim 7.4. *We have*

$$\int_{x_1\in\Omega}\int_{x_2\in\Omega}\int_{y_1\in\Omega}\int_{y_2\in\Omega} W(x_1,y_1)(\mathsf{D}_{d,r})W(x_2,y_2)(\mathsf{D}_{d,r})\mathsf{d}y_2\mathsf{d}y_1\mathsf{d}x_2\mathsf{d}x_1 = \left(\tfrac{1}{2^d}\right)^2 .$$

Proof of Claim 7.4.

$$\begin{aligned}
&\int_{x_1\in\Omega}\int_{x_2\in\Omega}\int_{y_1\in\Omega}\int_{y_2\in\Omega} W(x_1,y_1)(\mathsf{D}_{d,r})W(x_2,y_2)(\mathsf{D}_{d,r})\mathsf{d}y_2\mathsf{d}y_1\mathsf{d}x_2\mathsf{d}x_1\\
=&\int_{x_1\in\Omega}\int_{y_1\in\Omega} W(x_1,y_1)(\mathsf{D}_{d,r})\int_{x_2\in\Omega}\int_{y_2\in\Omega} W(x_2,y_2)(\mathsf{D}_{d,r})\mathsf{d}y_2\mathsf{d}x_2\mathsf{d}y_1\mathsf{d}x_1\\
=&\int_{x_1\in\Omega}\int_{y_1\in\Omega} W(x_1,y_1)(\mathsf{D}_{d,r})\int_{x_2\in\Omega}\mu^1_{W,x_2}(\mathsf{D}_{d,r})\mathsf{d}x_2\mathsf{d}y_1\mathsf{d}x_1\\
=&\tfrac{1}{2^d}\int_{x_1\in\Omega}\int_{y_1\in\Omega} W(x_1,y_1)(\mathsf{D}_{d,r})\mathsf{d}y_1\mathsf{d}x_1\\
=&\tfrac{1}{2^d}\int_{x_1\in\Omega}\mu^1_{W,x_1}(\mathsf{D}_{d,r})\mathsf{d}x_1\\
=&\left(\tfrac{1}{2^d}\right)^2 ,
\end{aligned}$$

where we use the fact that the 2-dimensional measure $\mu^1_{W,x}$ has uniform marginals for almost every $x \in \Omega$, as (W, f) is a Latinon. □

We therefore get

(9)
$$\begin{aligned}
&\left|t(A,L_1)-k!\ell!\sum_{P\in\mathcal{R}^A([2^d])}t_{\alpha_P}(F,\mathbf{W})\right|\\
\overset{C7.1}{=}&k!\ell!\int_{\mathbf{x}\in\Omega^k_{<_f}}\int_{\mathbf{y}\in\Omega^\ell_{<_f}}\left(\bigotimes_{(i,j)\in[k]\times[\ell]}W(x_i,y_j)\right)(\{M\in\mathcal{R}^A([0,1])\mid M\not\perp\mathfrak{D}_d\})\mathbf{dydx}\\
\leq&k!\ell!\sum_{r\in[2^d]}\sum_{\substack{(i_1,j_1)\\ \neq(i_2,j_2)\\ \in[k]\times[\ell]}}\int_{\substack{\mathbf{x}\in\Omega^k\\ \mathbf{y}\in\Omega^\ell}}W(x_{i_1},y_{j_1})(\mathsf{D}_{d,r})W(x_{i_2},y_{j_2})(\mathsf{D}_{d,r})\prod_{\substack{(i,j)\in[k]\times[\ell]\\ \neq(i_1,j_1),(i_2,j_2)}}W(x_i,y_j)([0,1])\mathbf{dydx}\\
=&k!\ell!\sum_{r\in[2^d]}\sum_{\substack{i_1,i_2\in[k],j_1,j_2\in[\ell]\\ i_1\neq i_2,j_1\neq j_2}}\int_{\substack{(x_{i_1},x_{i_2})\in\Omega^2\\ (y_{j_1},y_{j_2})\in\Omega^2}}W(x_{i_1},y_{j_1})(\mathsf{D}_{d,r})W(x_{i_2},y_{j_2})(\mathsf{D}_{d,r})\mathsf{d}(y_{j_1},y_{j_2})\mathsf{d}(x_{i_1},x_{i_2})\\
+&k!\ell!\sum_{r\in[2^d]}\sum_{\substack{i_1,i_2\in[k],j_1,j_2\in[\ell]\\ i_1=i_2,j_1\neq j_2}}\int_{\substack{x_{i_1}\in\Omega\\ (y_{j_1},y_{j_2})\in\Omega^2}}W(x_{i_1},y_{j_1})(\mathsf{D}_{d,r})W(x_{i_1},y_{j_2})(\mathsf{D}_{d,r})\mathsf{d}(y_{j_1},y_{j_2})\mathsf{d}(x_{i_1})\\
+&k!\ell!\sum_{r\in[2^d]}\sum_{\substack{i_1,i_2\in[k],j_1,j_2\in[\ell]\\ i_1\neq i_2,j_1=j_2}}\int_{\substack{(x_{i_1},x_{i_2})\in\Omega^2\\ y_{j_1}\in\Omega}}W(x_{i_1},y_{j_1})(\mathsf{D}_{d,r})W(x_{i_2},y_{j_1})(\mathsf{D}_{d,r})\mathsf{d}y_{j_1}\mathsf{d}(x_{i_1},x_{i_2})\\
\overset{C7.3,7.4}{=}&k!\ell!\cdot 2^d\cdot k\ell(k\ell-1)\cdot(\tfrac{1}{2^d})^2=k!\ell!\cdot k\ell(k\ell-1)\cdot\tfrac{1}{2^d}\ .
\end{aligned}$$

A similar calculation to (9) which uses Claim 7.2 instead of Claim 7.1 yields

$$\left|t(A,L_2)-k!\ell!\sum_{P\in\mathcal{R}^A([2^d])}t_{\alpha_P}(F,\mathbf{U})\right|\leq k!\ell!\cdot k\ell(k\ell-1)\cdot\tfrac{1}{2^d}\ . \tag{10}$$

We are now in a position to bound $|t(A,L_1)-t(A,L_2)|$. Hence, take d such that

$$\tfrac{1}{\delta_{\boxtimes}(L_1,L_2)^{1/(2k\ell)}}\leq 2^d\leq\tfrac{2}{\delta_{\boxtimes}(L_1,L_2)^{1/(2k\ell)}}$$

and set $c_{k,\ell} \coloneqq 2k!\ell!k\ell(k\ell-1)+2^{k\ell}k!\ell!(k\ell+\binom{k}{2}+\binom{\ell}{2})$. Starting with the triangle inequality, we get

$$\begin{aligned}
&|t(A,L_1)-t(A,L_2)|\\
\leq&\left|t(A,L_1)-k!\ell!\sum_{P\in\mathcal{R}^A([2^d])}t_{\alpha_P}(F,\mathbf{W})\right|\\
&+k!\ell!\sum_{P\in\mathcal{R}^A([2^d])}|t_{\alpha_P}(F,\mathbf{W})-t_{\alpha_P}(F,\mathbf{U})|\\
&+\left|k!\ell!\sum_{P\in\mathcal{R}^A([2^d])}t_{\alpha_P}(F,\mathbf{U})-t(A,L_2)\right|\\
\overset{C7.1,\ (8),C7.2}{\leq}&2k!\ell!k\ell(k\ell-1)/2^d+|\mathcal{R}^A([2^d])|k!\ell!(k\ell+\tbinom{k}{2}+\tbinom{\ell}{2})\delta_{\boxtimes}(L_1,L_2)\\
\leq&2k!\ell!k\ell(k\ell-1)/2^d+2^{dk\ell}k!\ell!(k\ell+\tbinom{k}{2}+\tbinom{\ell}{2})\delta_{\boxtimes}(L_1,L_2)\\
\leq&2k!\ell!k\ell(k\ell-1)\delta_{\boxtimes}(L_1,L_2)^{1/(2k\ell)}+2^{k\ell}k!\ell!(k\ell+\tbinom{k}{2}+\tbinom{\ell}{2})\sqrt{\delta_{\boxtimes}(L_1,L_2)}\\
\leq&c_{k,\ell}\delta_{\boxtimes}(L_1,L_2)^{1/(2k\ell)}\ .
\end{aligned}$$

□

8. Proof of compactness

In this section, we prove Theorem 3.13. In fact, Theorem 3.13 follows from Lemma 4.5, remarks in Section 4.0.1 and the proposition below, which we prove instead.

Proposition 8.1. *The space $(\mathcal{L},\delta_{\boxtimes})$ is compact.*

Proposition 8.1 follows immediately from Lemma 8.3 and Lemma 8.6 below. To state these lemmata, we introduce a function ι which maps a Latinon to its infinite compression (see Definition 6.1), i.e.

$$\iota:(\mathcal{L},\delta_{\boxtimes})\to(\mathcal{W}_0^{\mathbb{N}},\delta_{\square}^{\mathbb{N}}),\ L\mapsto\mathbf{L}^{\mathbb{N}}\ .$$

As we show in the lemma below, the function ι is injective.

Lemma 8.2. *The function ι is injective.*

Proof. Suppose that $L=(W,f),L'=(W',f')\in\mathcal{L}$ are such that $\delta_{\square}^{\mathbb{N}}(\iota(L),\iota(L'))=0$. We need to show that $\delta_{\boxtimes}(L,L')<\alpha$ for an arbitrary given $\alpha>0$. Let us write $h=\lceil 4/\alpha\rceil$. Let us write $\iota(L)=(O^f,W_{1,1},W_{1,2},\dots)$ and $\iota(L')=(O^{f'},W'_{1,1},W'_{1,2},\dots)$. Let $\varphi\in S_\Omega$ be

as in (6) such that

$$\frac{1}{2}\left\|O^f - O^{f'\circ\varphi}\right\|_\square + \sum_{\ell=1}^{\infty}\sum_{q=1}^{2^\ell} \frac{1}{2^{2^\ell+q-1}} \cdot \left\|W_{\ell,q} - (W'_{\ell,q})^\varphi\right\|_\square < \frac{\alpha}{2^{3+2^{h+1}}} \ . \tag{11}$$

Taking $\varphi = \psi$ in (2), we see (by (11)) that the sum of the first and the second term is at most $\frac{4\alpha}{2^{3+2^{h+1}}} < \alpha/4$. To conclude that $\delta_\boxtimes(L, L') < \alpha$, it thus remains to bound $\|W - (W')^\varphi\|_\boxtimes$. So, let $R, C \subseteq \Omega$ be arbitrary, and let $V = [v_1, v_2] \subseteq [0,1]$ be an interval. Let $V' = [w_1, w_2]$ be an interval whose endpoints w_1, w_2 are dyadic rationals with dyadic denominators at most 2^h and with $|w_i - v_i| < \alpha/4$ (clearly, such a dyadic approximation exists). By the uniform marginals property, we have

$$\begin{aligned}&\int_{x\in R}\int_{y\in C}(W - (W')^\varphi)(x,y)(V)\mathsf{d}y\mathsf{d}x\\ &= \int_{x\in R}\int_{y\in C}(W - (W')^\varphi)(x,y)(V')\mathsf{d}y\mathsf{d}x \ \pm \ (|w_1 - v_1| + |w_2 - v_2|) \ ,\end{aligned}$$

and so it suffices to prove that $|\int_{x\in R}\int_{y\in C}(W - (W')^\varphi)(x,y)(V')\mathsf{d}y\mathsf{d}x| < \alpha/4$, which we do now. Let us write $Q \subset [2^h]$ for the indices of dyadic intervals at depth h encoding V', i.e., $Q \coloneqq \{i \in [2^h] \mid \mathsf{D}_{h,i} \subset V'\}$. Then we have

$$\begin{aligned}\left|\int_{x\in R}\int_{y\in C}(W - (W')^\varphi)(x,y)(V')\mathsf{d}y\mathsf{d}x\right| &= \left|\int_{x\in R}\int_{y\in C}\sum_{i\in Q}(W_{h,i} - (W'_{h,i})^\varphi)(x,y)\mathsf{d}y\mathsf{d}x\right|\\ &\le \sum_{i\in Q}\left|\int_{x\in R}\int_{y\in C}(W_{h,i} - (W'_{h,i})^\varphi)(x,y)\mathsf{d}y\mathsf{d}x\right|\\ &\le \sum_{i\in Q}\left\|W_{h,i} - (W'_{h,i})^\varphi\right\|_\square \overset{(11)}{<} \alpha/4 \ ,\end{aligned}$$

proving the statement. □

Lemma 8.3. *The space $(\iota(\mathcal{L}), \delta_\square^{\mathbb{N}})$ is compact.*

Proof. Let $(L_n)_{n\in\mathbb{N}}$ be a sequence of Latinons. We have to show that $(\iota(L_n))_{n\in\mathbb{N}}$ contains a convergent subsequence with respect to $\delta_\square^{\mathbb{N}}$. For each $n \in \mathbb{N}$ we write $\iota(L_n) = (O^n, W^n_{1,1}, W^n_{1,2}, \dots)$. By Theorem 5.1 there exist bigraphons $\widetilde{O}, \widetilde{W}_{d,k}$, $d \in \mathbb{N}, k \in [2^d]$, such that, after passing to a subsequence, we have that

$$(O^n, W^n_{1,1}, W^n_{1,2}, W^n_{2,1}, W^n_{2,2}, W^n_{2,3}, \dots) \overset{\delta_\square^{\mathbb{N}}}{\to} \widetilde{\mathbf{W}} = (\widetilde{O}, \widetilde{W}_{1,1}, \widetilde{W}_{1,2}, \widetilde{W}_{2,1}, \widetilde{W}_{2,2}, \widetilde{W}_{2,3}, \dots) \ . \tag{12}$$

Again by passing to a subsequence we can additionally assume that for every $n \in \mathbb{N}$,

$$\delta_\square^{\mathbb{N}}(\iota(L_n), \widetilde{\mathbf{W}}) < \tfrac{1}{2^{n+2}} \ , \tag{13}$$

and in particular we can fix $\varphi_n \in S_\Omega$ for every $n \in \mathbb{N}$ such that

$$\tfrac{1}{2}\left\|O^{n,\varphi_n} - \widetilde{O}\right\|_\square + \sum_{d=1}^{\infty}\sum_{k=1}^{2^d} \tfrac{1}{2^{2^d+k-1}}\left\|W_{d,k}^{n,\varphi_n} - \widetilde{W}_{d,k}\right\|_\square < \tfrac{1}{2^{n+1}}\,. \tag{14}$$

We have to show that there exists a Latinon $L = (\nu, f)$ such that $\iota(L) = \widetilde{\mathbf{W}} = (\widetilde{O}, \widetilde{W}_{1,1}, \widetilde{W}_{1,2}, \widetilde{W}_{2,1}, \dots)$.

First, we derive a measure preserving map $f : \Omega \to [0,1]$ from $\widetilde{O}$ by setting $f(x) \coloneqq \deg_{\widetilde{O}}(x)$. Note that then $O^f = \widetilde{O}$ up to a null set. Since the degree distribution function of each O^{n,φ_n} is measure preserving, Lemma 2.7 implies that this property is inherited to the limit, i.e. f is indeed measure preserving.

For the definition of the Latinon-measure ν note that the dyadic intervals $\mathfrak{D}_*$ together with $\emptyset$ form a semiring which we call R, and that R generates the Borel sigma-algebra on $[0,1)$.[9] Consequently,

$$R' \coloneqq \{S \times T \times J \mid S, T \subseteq \Omega \text{ Borel measurable}, J \in \mathfrak{D}_*\} \cup \{\emptyset\}$$

is a semiring which generates the Borel sigma-algebra on $\Omega^2 \times [0,1)$. We define

$$\nu^*(S \times T \times \mathsf{D}_{d,s}) \coloneqq \int_S \int_T \widetilde{W}_{d,s}(x,y)\mathsf{d}x\mathsf{d}y$$

for all $S, T \subset \Omega$, $d \in \mathbb{N}$ and $s \in [2^d]$. We want to show that ν^* defines a premeasure on $\Omega^2 \times [0,1)$. This is shown by the following claim.

Claim 8.4. *For every partition $S \times T \times \mathsf{D}_{d',k'} = \bigcup_{i=1}^{\infty} S_i \times T_i \times \mathsf{D}_{d_i,k_i}$, where $S, T, S_i, T_i \subseteq \Omega$ and $\mathsf{D}_{d',k'}, \mathsf{D}_{d_i,k_i} \in \mathfrak{D}_*$, we have*

$$\nu^*(S \times T \times \mathsf{D}_{d',k'}) = \sum_{i=1}^{\infty} \nu^*(S_i \times T_i \times \mathsf{D}_{d_i,k_i})\,.$$

Proof of Claim 8.4. Let $S \times T \times \mathsf{D}_{d',k'} = \bigcup_{i=1}^{\infty} S_i \times T_i \times \mathsf{D}_{d_i,k_i}$ be as in the assumption. We need to show that for an arbitrary $\varepsilon > 0$ (which we consider fixed from now on) there exists $N_0 \in \mathbb{N}$ such that for every $N > N_0$ we have

$$\left|\nu^*(S \times T \times \mathsf{D}_{d',k'}) - \sum_{i=1}^{N} \nu^*(S_i \times T_i \times \mathsf{D}_{d_i,k_i})\right| < \varepsilon\,. \tag{15}$$

We define a map $g : \mathsf{D}_{d',k'} \to \mathbb{N}$ by

$$g(p) \coloneqq \min\left\{ r \in \mathbb{N} \mid \mu^{\otimes 2}\left(S \times T \setminus \bigcup_{i \in [r] \text{ s.t. } p \in \mathsf{D}_{d_i,k_i}} S_i \times T_i\right) < \varepsilon/4 \right\}. \tag{16}$$

[9] See Section 2.3.1 for constructions of measures from semirings.

Note that g is measurable and finite everywhere and that $\sum_{i=1}^{\infty} \lambda(g^{-1}(i)) = \lambda(\bigcup_{i\in\mathbb{N}} g^{-1}(i)) = \lambda(\mathsf{D}_{d',k'}) \leq 1$ is convergent. Therefore there exists $N_0 \in \mathbb{N}$ such that

$$\sum_{i=N_0+1}^{\infty} \lambda(g^{-1}(i)) < \varepsilon/4 \,. \tag{17}$$

We set $E \coloneqq \bigcup_{i=N_0+1}^{\infty} g^{-1}(i)$. We now prove the following claim.

Claim 8.5. *Suppose that (U,h) is an arbitrary Latinon. Then*

$$\sum_{i=N_0+1}^{\infty} \int_{S_i}\int_{T_i} U(x,y)(\mathsf{D}_{d_i,k_i})\mathsf{d}x\mathsf{d}y < \varepsilon/2 \,.$$

Proof. U induces a measure on $\Omega^2 \times [0,1]$ with uniform marginals. Let $\{\xi_p\}_{p\in[0,1]}$ be the family on Ω^2 given by applying the disintegration theorem to this measure with respect to the third coordinate. It follows from the uniform marginals property that almost every measure $\{\xi_p\}_{p\in[0,1]}$ is a probability measure. Then

$$\begin{aligned}
&\sum_{i=N_0+1}^{\infty} \int_{S_i}\int_{T_i} U(x,y)(\mathsf{D}_{d_i,k_i})\mathsf{d}x\mathsf{d}y \\
&= \sum_{i=N_0+1}^{\infty} \int_{[0,1]} \mathbb{1}_{\mathsf{D}_{d_i,k_i}}(p)\cdot\xi_p(S_i\times T_i)\mathsf{d}p = \sum_{i=N_0+1}^{\infty} \int_{\mathsf{D}_{d',k'}} \mathbb{1}_{\mathsf{D}_{d_i,k_i}}(p)\cdot\xi_p(S_i\times T_i)\mathsf{d}p \\
&= \int_{\mathsf{D}_{d',k'}} \xi_p\left(\bigcup_{i>N_0 \text{ s.t. } p\in\mathsf{D}_{d_i,k_i}} S_i\times T_i\right)\mathsf{d}p \\
&= \int_{E} \xi_p\left(\bigcup_{i>N_0 \text{ s.t. } p\in\mathsf{D}_{d_i,k_i}} S_i\times T_i\right)\mathsf{d}p + \int_{\mathsf{D}_{d',k'}\setminus E} \xi_p\left(\bigcup_{i>N_0 \text{ s.t. } p\in\mathsf{D}_{d_i,k_i}} S_i\times T_i\right)\mathsf{d}p \\
&\leq \int_{E} \xi_p\left(\Omega^2\right)\mathsf{d}p + \int_{\mathsf{D}_{d',k'}\setminus E} \xi_p\left(\bigcup_{i>N_0 \text{ s.t. } p\in\mathsf{D}_{d_i,k_i}} S_i\times T_i\right)\mathsf{d}p \overset{(16),(17)}{<} \varepsilon/4+\varepsilon/4 = \varepsilon/2 \,.
\end{aligned}$$

□

Let us now return to proving (15). Suppose that $N > N_0$ is arbitrary. Choose $M \in \mathbb{N}$ such that $\frac{N+1}{M} < \varepsilon/2$ and $d_{max} = \max(\{d_i \mid i \in [N]\} \cup \{d'\})$. By (14) there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $d \leq d_{max}$ and $k \in [2^d]$,

$$\left|\int_S\int_T W_{d,k}^{n,\varphi_n}(x,y)\mathsf{d}x\mathsf{d}y - \int_S\int_T \widetilde{W}_{d,k}(x,y)\mathsf{d}x\mathsf{d}y\right| < \tfrac{1}{M} \,. \tag{18}$$

Recalling the definition of ν^*, we then get for $n \geq n_0$ that

$$\begin{aligned}
&\left|\nu^*(S \times T \times \mathsf{D}_{d',k'}) - \sum_{i=1}^{N} \nu^*(S_i \times T_i \times \mathsf{D}_{d_i,k_i})\right| \\
\leq\ &\left|\nu^*(S \times T \times \mathsf{D}_{d',k'} - \int_S \int_T W^{n,\varphi_n}_{d',k'}(x,y)\mathsf{d}x\mathsf{d}y\right| \\
+\ &\left|\int_S \int_T W^{n,\varphi_n}_{d',k'}(x,y)\mathsf{d}x\mathsf{d}y - \sum_{i=1}^{N} \int_{S_i} \int_{T_i} W^{n,\varphi_n}_{d_i,k_i}(x,y)\mathsf{d}x\mathsf{d}y\right| \\
+\ &\sum_{i=1}^{N} \left|\int_{S_i} \int_{T_i} W^{n,\varphi_n}_{d_i,k_i}(x,y)\mathsf{d}x\mathsf{d}y - \nu^*(S_i \times T_i \times \mathsf{D}_{d_i,k_i})\right| \\
\overset{2\times(18)}{\leq}\ &\tfrac{1}{M} + \left|\sum_{i=N+1}^{\infty} \int_{S_i} \int_{T_i} W^{n,\varphi_n}_{d_i,k_i}(x,y)\mathsf{d}x\mathsf{d}y\right| + \tfrac{N}{M} \overset{C8.5}{<} \varepsilon\ ,
\end{aligned}$$

indeed proving (15). □

By Claim 8.4 we have that ν^* defines a premeasure on R'. Therefore by Carathéodory's extension theorem (Theorem 2.5) there exists a unique measure ν on $\Omega^2 \times [0,1]$ such that $\nu(S \times T \times \mathsf{D}_{d,s}) = \int_S \int_T \widetilde{W}_{d,s}(x,y)(\mathsf{D}_{d,s})\mathsf{d}x\mathsf{d}y$ for all $d \in \mathbb{N}$ and $s \in [2^d]$. Lastly, we need to check that ν has uniform marginals; this will also verify that the total mass of ν is 1. Note that it is sufficient to restrict ourselves to sets of the semiring R'. Let $S, T \subseteq \Omega$ and $J \in \mathfrak{D}_*$. By making use of the uniform marginals of W^n we get

$$\begin{aligned}
\nu(S \times T \times [0,1]) &= \lim_{n\to\infty} \int_S \int_T W^{n,\varphi_n}(x,y)([0,1])\mathsf{d}x\mathsf{d}y = \lim_{n\to\infty} \mu(S)\mu(T) = \mu(S)\mu(T)\ , \\
\nu(\Omega \times T \times J) &= \lim_{n\to\infty} \int_\Omega \int_T W^{n,\varphi_n}(x,y)(J)\mathsf{d}x\mathsf{d}y = \lim_{n\to\infty} \mu(T)\lambda(J) = \mu(T)\lambda(J)\ , \\
\nu(S \times \Omega \times J) &= \lim_{n\to\infty} \int_S \int_\Omega W^{n,\varphi_n}(x,y)(J)\mathsf{d}x\mathsf{d}y = \lim_{n\to\infty} \mu(S)\lambda(J) = \mu(S)\lambda(J)\ .
\end{aligned}$$

Hence ν is a Latinon according to Definition 3.16. □

To state Lemma 8.6, we need to recall that ι^{-1} is well-defined by Lemma 8.2.

Lemma 8.6. *The map $\iota^{-1} : (\iota(\mathcal{L}), \delta_\square^{\mathbb{N}}) \to (\mathcal{L}, \delta_{\boxtimes})$ is continuous.*

Proof. Let $\varepsilon > 0$. We fix $d \in \mathbb{N}$ large enough such that $\frac{1}{2^{d-2}} < \varepsilon/2$ and then choose $\delta > 0$ small enough such that $2^{2^{d+1}} \cdot \delta < \varepsilon/2$. Let L_1 and L_2 be a pair of Latinons such that $\delta_\square^{\mathbb{N}}(\iota(L_1), \iota(L_2)) < \delta$. By Lemma 6.6 and Proposition 6.8 we get that

$$\delta_{\boxtimes}(L_1, L_2) \leq \delta_{\boxtimes}(L_1, \mathbf{L_1}^d) + \delta_{\boxtimes}(\mathbf{L_1}^d, \mathbf{L_2}^d) + \delta_{\boxtimes}(\mathbf{L_2}^d, L_2) \leq \tfrac{1}{2^{d-2}} + 2^{2^{d+1}}\delta < \varepsilon\ .$$

□

Proof of Proposition 8.1. By Lemma 8.3, the space $(\iota(\mathcal{L}), \delta_{\square}^{\mathbb{N}})$ is compact. By Lemma 8.6, ι^{-1} is continuous. Since a continuous image of a compact space is compact, we conclude that $(\mathcal{L}, \delta_{\boxtimes})$ is compact. □

8.1. **An additional application of the method: compactness of orderons.** Here, we show how a simplified version of the above approach can be used to treat orderons. An *orderon*, as defined [4], is a symmetric function $W : ([0,1] \times [0,1])^2 \to [0,1]$. Here, the first component $[0,1]$ encodes the global position and the second component $[0,1]$ is used to encode local information. Definition 2.7 of [4] gives a definition of 'CS-distance' between two orderons U and W. The philosophy behind the 'CS-distance' is to modify the usual cut distance — in which one would consider all measure preserving bijections ψ from $[0,1] \times [0,1]$ to itself and take the infimum of $\|U - W^{\psi}\|_{\square}$ — by penalising those ψ that shuffle the global coordinates substantially. So, instead of repeating the definition from [4], let us give an alternative but equivalent one:[10] define

$$\delta_{\triangle}(U,W) := \inf_{\psi} \left\|O^{\pi_1} - O^{\pi_1 \circ \psi}\right\|_{\square} + \left\|U - W^{\psi}\right\|_{\square} ,$$

where ψ is as above, $\pi_1 : [0,1] \times [0,1] \to [0,1]$ is the projection on the first coordinate and O is the order fixing bigraphon.

Let us show that this alternative view together with the strategy we used to prove Proposition 8.1 allows us to easily establish compactness of orderons (with respect to $\delta_{\triangle}$); the original proof occupies Section 4 of [4]. Indeed, suppose that we are given a sequence of orderons $W_1, W_2, W_3, \ldots$. We represent these orderons as pairs

$$(O^{\pi_1}, W_1), (O^{\pi_1}, W_2), (O^{\pi_1}, W_3), \ldots .$$

Applying Theorem 5.1 (on pairs rather than on infinite vectors), we get that there exists a sequence of indices $i_1 < i_2 < \ldots$, a bigraphon $\widetilde{O}$, a graphon $\widetilde{W}$, and a sequence of measure preserving bijections $\psi_{i_1}, \psi_{i_2}, \ldots$ such that

$$\left\|O^{\pi_1 \circ \psi_{i_n}} - \widetilde{O}\right\|_{\square} \to 0 \text{ , and}$$
$$\left\|W_{i_n}^{\psi_{i_n}} - \widetilde{W}\right\|_{\square} \to 0 .$$

We might stop here and say that the pair $(\widetilde{O}, \widetilde{W})$ is the sought after accumulation point. But let us do one more step and transform $(\widetilde{O}, \widetilde{W})$ back into the global-local representation which is used in [4]. We proceed as described in Section 3.2.4. That is, we appeal to Proposition 3.18 with a function $f : [0,1] \times [0,1] \to [0,1]$, $f((a,x)) := \deg_{\widetilde{O}}((a,x))$ and get

[10] 'Equivalent' meaning that it generates the same topology as Definition 2.7 of [4]. This is routine to check.

a function $h : [0,1] \times [0,1] \to [0,1] \times [0,1]$. Then $W^*((a,x),(b,y)) \coloneqq \widetilde{W}(h(a,x),h(b,y))$ is the accumulation point in the global-local representation.

9. Proof of the sampling lemma for Latinons

In this section we prove the sampling lemma, Lemma 4.9. The proof splits into two parts. In the first part, we show that we can approximately mimic sampling from a Latinon L on Ω^2 by sampling from a suitably chosen Latinon (U, id) on $[0,1]^2$. In the second part, we show that a pattern sampled from a Latinon (U, id) on $[0,1]^2$ is close to (U, id) in the cut distance for Latinons. We start with the first part contained in the following lemma.

Lemma 9.1. *Suppose that $L = (W, f)$ is a Latinon on Ω^2, $k \in \mathbb{N}$, and $\varepsilon > 0$. Then there exists a Latinon $L' = (U, id)$ on $[0,1]^2$ such that*

(i) $\delta^*_{\boxtimes}(L, L') < \varepsilon$ and
(ii) there exists a coupling $\mathcal{C}$ of sampling a $k \times k$ pattern A from L and sampling a $k \times k$ pattern B from L' such that $\mathbb{P}_{\mathcal{C}}[A \neq B] < \varepsilon k^2$.

Proof. Choose $m \in \mathbb{N}$ such that $2/m < \varepsilon$. Given the partition $\mathfrak{J}_m$, for each $i \in [m]$ we choose an arbitrary measure preserving bijection $\varphi_i : f^{-1}(\mathsf{J}_{m,i}) \to \mathsf{J}_{m,i}$ (where the domain is equipped with the restriction of the measure μ and the codomain is equipped with the Lebesgue measure). Define $\varphi : \Omega \to [0,1]$ by $\varphi(x) = \varphi_i(x)$, if $x \in \mathsf{J}_{m,i}$. We define U to be the distribution-valued bigraphon on $[0,1]^2$ with $U(x,y) = W(\varphi^{-1}(x), \varphi^{-1}(y))$. Note that then for $L' = (U, id)$ we have

$$\delta^*_{\boxtimes}(L, L') \leq 2 \left\| O^{id} - O^{f \circ \varphi^{-1}} \right\|_{\square} + \left\| U - W^{\varphi^{-1}} \right\|_{\boxtimes} \leq 2/m + 0 < \varepsilon \,,$$

as $f \circ \varphi^{-1}$ only reorders elements inside each $\mathsf{J}_{m,i}$. We define the coupling for the sampling process in the following way. We sample k-sets $R, C \subseteq \Omega$ and order them according to $<_f$, written $(r_1, \ldots, r_k)$ and $(c_1, \ldots, c_k)$. We then consider the sets $\varphi(R)$ and $\varphi(C)$ and reorder the k-tuples according to $<$, written $(\varphi(r_{\sigma(1)}), \ldots, \varphi(r_{\sigma(k)}))$ and $(\varphi(c_{\tau(1)}), \ldots, \varphi(c_{\tau(k)}))$. We then sample values $a_{i,j} = b_{\sigma(i),\tau(j)}$ from $W(r_i, c_j) = U(\varphi(r_i), \varphi(c_j))$. Thus the event $A \neq B$ can only happen if $\tau \neq id$ or $\sigma \neq id$, which in turn can only happen if two elements of R (or C) lie in the same set $f^{-1}(\mathsf{J}_{m,i})$ for some $i \in [k]$. Hence

$$\mathbb{P}_{\mathcal{C}}[A \neq B] \leq 2 \cdot \binom{k}{2} \cdot m \cdot (1/m)^2 \leq k^2/m < \varepsilon k^2 \,.$$

□

We can now move to the second part of the argument in which show that a pattern sampled from a Latinon (U, id) on $[0,1]^2$ is close to (U, id) in the cut distance. To this end we first need the following auxiliary lemma.

Lemma 9.2. *Suppose that $k \in \mathbb{N}$. Suppose that $L = (W, id)$ is a semilatinon on $[0,1]^2$ which is constant on every set $\mathsf{J}_{k,i} \times \mathsf{J}_{k,j}$ for each $i, j \in [k]$. For each $i, j \in [k]$ sample from $W(\mathsf{J}_{k,i}, \mathsf{J}_{k,j})$ a value $a_{i,j}$ and call the resulting $k \times k$ matrix A. Then with probability at least $1 - \exp(-5k)$ we have*

$$\left\|W - W^A\right\|_{\boxtimes} \leq 6/k^{1/4} .$$

Proof. We choose $d = \log(k^{1/4})$ and $\varepsilon = 2/\sqrt{k}$. For every $s \in [2^d]$ and i, j let $X^s_{i,j} \in \{0,1\}$ be the random variable $X^s_{i,j} = \mathbb{1}_{a_{i,j} \in \mathsf{D}_{d,s}}$. Let $S, T \subseteq [k]$. For every $S, T \subseteq [k]$ and $s \in [2^d]$ we have by Lemma 2.2 (with $Z = \sum_{i \in S} \sum_{j \in T} X^s_{i,j}$ and $t = \varepsilon k^2$) that with probability at least $1 - \exp(-2\varepsilon^2 k^2)$

$$\left| \sum_{i \in S} \sum_{j \in T} (X^s_{i,j} - \mathbb{E}[X^s_{i,j}]) \right| \leq \varepsilon k^2 . \tag{19}$$

Hence we have with probability at least $1 - 2^{2k} 2^d \exp(-2\varepsilon^2 k^2) > 1 - \exp(-5k)$ that for every $s \in [2^d]$ and $S, T \subseteq [k]$

$$\left| \int_{\bigcup_{i \in S} \mathsf{J}_{k,i}} \int_{\bigcup_{j \in T} \mathsf{J}_{k,j}} W^A_{d,s}(x,y) - W_{d,s}(x,y) \mathsf{d}x \mathsf{d}y \right| = \left| \tfrac{1}{k^2} \sum_{i \in S} \sum_{j \in T} (X^s_{i,j} - \mathbb{E}[X^s_{i,j}]) \right| < \varepsilon .$$

Note that, since $W^A_{d,s} - W_{d,s}$ is a step-function, the maximum $\left\|W^A_{d,s} - W_{d,s}\right\|_{\square}$ is attained on $\bigcup_{i \in S} \mathsf{J}_{k,i} \times \bigcup_{j \in T} \mathsf{J}_{k,j}$ for some $S, T \subseteq [k]$ and therefore

$$\left\|W^A_{d,s} - W_{d,s}\right\|_{\square} < \varepsilon .$$

By Lemma 6.6 and Lemma 6.7 we therefore get that with probability at least $1 - \exp(-5k)$

$$\begin{aligned} \left\|W^A - W\right\|_{\boxtimes} &\leq \left\|W^A - (\mathbf{W^A})^{\mathbf{d}}\right\|_{\boxtimes} + \left\|(\mathbf{W^A})^{\mathbf{d}} - \mathbf{W^d}\right\|_{\boxtimes} + \left\|\mathbf{W^d} - \mathbf{W}\right\|_{\boxtimes} \\ &\leq \tfrac{4}{2^d} + 2^d \varepsilon \leq 6/k^{1/4} . \end{aligned}$$

□

We also need the following definition of sampling a distribution-valued step-bigraphon, which can be seen as an intermediate step of sampling a pattern.

Definition 9.3. *Let U be a distribution-valued bigraphon on $[0,1]^2$ and $R, C \in [0,1]^n_<$ be k-tuples. We write $R = (r_1, \ldots, r_k)$, $C = (c_1, \ldots, c_k)$. We define $U[R, C]$ to be the distribution-valued bigraphon defined by $U[R, C](x, y) \coloneqq U(r_i, c_j)$, if $x \in \mathsf{J}_{k,i}$ and $y \in \mathsf{J}_{k,j}$.*

Now we can prove the second part of the overall argument. Recall that by $W^{\bowtie \mathcal{P}}$ we denote the step bigraphon derived from a bigraphon W with respect to the partition $\mathcal{P}$.

Lemma 9.4. *Suppose that (U, id) is a Latinon on $[0,1]^2$ and $k \in \mathbb{N}$. If A is a $k \times k$ pattern sampled from (U, id), then with probability at least $1 - 29\exp(-k^{1/8}/10)$ we have*

$$\delta_{\boxtimes}((U, id), (W^A, id)) < 29/\sqrt{\log(k)}\,.$$

Proof. Set $m, r = k^{1/8}$ and $d = \log(\log(k)^{1/4})$ (we omit rounding symbols and assume that m, k and d are integers). Let $L^d = (O, U_{d,1}, \ldots, U_{d,2^d})$ be the compression of L of depth d. Consider the partition $\mathfrak{J}_r$. By Lemma 5.2(ii) we find an equipartition $\mathcal{P}$ of $[0,1]$ refining $\mathfrak{J}_r$, i.e. $\mathsf{J}_{r,j} = P_{j,1} \cup \cdots \cup P_{j,m}$, and such that, for every $1 \leq i \leq 2^d$,

$$\left\|U_{d,i} - (U_{d,i})^{\bowtie\mathcal{P}}\right\|_{\square} \leq \sqrt{\tfrac{2^{d+1}}{\log(m)}} + \tfrac{2}{m}\,. \tag{20}$$

This defines a step Latinon $L^d_{\mathcal{P}} = (U_1, id)$ where $U_1 = \nabla(((U_{d,i})^{\bowtie\mathcal{P}})_{i\in[2^d]})$. Now, let $R, C \in [0,1]^n_<$ be k-sets chosen uniformly at random. We write $L[R,C]$ for the semilatinon $(U[R,C], id)$, $L^d[R,C]$ for the semilatinon $(\mathbf{U}^d[R,C], id)$ and $L^d_{\mathcal{P}}[R,C]$ for the semilatinon (U_2, id), where $U_2 = U_1[R,C]$. Furthermore, let A be the pattern generated by sampling $A_{i,j}$ from $U(r_i, c_j)$ for each $i, j \in [k]$. We then have with probability at least

$$1 - (2^d \cdot 4\exp(-\sqrt{k}/10) + 20\exp(-k^{1/8}/10) + \exp(-5k)) > 1 - 29\exp(-k^{1/8}/10)$$

that Claims 9.5, 9.6, 9.7 (stated and proven below) and Lemma 9.2 hold. Hence with probability at least $1 - 29\exp(-k^{1/8}/10)$ we have

$$\begin{aligned}
\delta_{\boxtimes}((U, id), (W^A, id)) &\leq \delta_{\boxtimes}(L, L^d) && \text{(Lemma 6.6)}\\
&+ \delta_{\boxtimes}(L^d, L^d_{\mathcal{P}}) && \text{(Claim 9.5)}\\
&+ \delta_{\boxtimes}(L^d_{\mathcal{P}}, L^d_{\mathcal{P}}[R,C]) && \text{(Claim 9.7)}\\
&+ \delta_{\boxtimes}(L^d_{\mathcal{P}}[R,C], L^d[R,C]) && \text{(Claim 9.6)}\\
&+ \delta_{\boxtimes}(L^d[R,C], L[R,C]) && \text{(Lemma 6.6)}\\
&+ \delta_{\boxtimes}(L[R,C], (W^A, id)) && \text{(Lemma 9.2)}\\
&\leq \tfrac{1}{2^{d-2}} + 2^{d+1}\left(\sqrt{\tfrac{2^{d+1}}{\log(m)}} + \tfrac{2}{m}\right) + \tfrac{4}{k^{1/16}} + \tfrac{8\cdot 2^d}{k^{1/4}} + \tfrac{6}{k^{1/4}}\\
&\leq \tfrac{29}{\log(k)^{1/8}}\,.
\end{aligned}$$

Claim 9.5. *We have*

$$\delta_{\boxtimes}(L^d, L^d_{\mathcal{P}}) \leq 2^d \cdot \left(\sqrt{\tfrac{2^{d+1}}{\log(m)}} + \tfrac{2}{m}\right)\,.$$

Proof. The Latinons L^d and $L^d_{\mathcal{P}}$ share the same order-fixing bigraphon O^f. Thus, their distance can be bounded using Proposition 6.7 with (20) for each $i \in [2^d]$. □

Claim 9.6. *We have with probability at least* $1 - 2^d \cdot 4\exp(-\sqrt{k}/10)$ *that*

$$\delta_{\boxtimes}(L^d_{\mathcal{P}}[R,C], L^d[R,C]) \leq \tfrac{8\cdot 2^d}{k^{1/4}} + 2^d\left(\sqrt{\tfrac{2^{d+1}}{\log(m)}} + \tfrac{2}{m}\right) .$$

Proof. Let us fix an $i \in [2^d]$. By Lemma 5.6 we have with probability at least $1 - 4\exp(-\sqrt{k}/10)$ that

$$\left|\left\|U_{d,i}[R,C] - (U_{d,i})^{\bowtie\mathcal{P}}[R,C]\right\|_{\square} - \left\|U_{d,i} - (U_{d,i})^{\bowtie\mathcal{P}}\right\|_{\square}\right| \leq \tfrac{8}{k^{1/4}} ,$$

and hence together with (20) that

$$\begin{aligned}
&\left\|U_{d,i}[R,C] - (U_{d,i})^{\bowtie\mathcal{P}}[R,C]\right\|_{\square}\\
\leq&\left|\left\|U_{d,i}[R,C] - (U_{d,i})^{\bowtie\mathcal{P}}[R,C]\right\|_{\square} - \left\|U_{d,i} - (U_{d,i})^{\bowtie\mathcal{P}}\right\|_{\square}\right| + \left\|U_{d,i} - (U_{d,i})^{\bowtie\mathcal{P}}\right\|_{\square}\\
\leq&\tfrac{8}{k^{1/4}} + \sqrt{\tfrac{2^{d+1}}{\log(m)}} + \tfrac{2}{m} .
\end{aligned}$$

Since this holds for any $i \in [2^d]$ we get by the union bound and Proposition 6.7 that with probability at least $1 - 2^d \cdot 4\exp(-\sqrt{k}/10)$ we have

$$\delta_{\boxtimes}(L^d_{\mathcal{P}}[R,C], L^d[R,C]) \leq \sum_{i=1}^{2^d}\left\|U_{d,i}[R,C] - (U_{d,i})^{\bowtie\mathcal{P}}[R,C]\right\|_{\square} \leq \tfrac{8\cdot 2^d}{k^{1/4}} + 2^d\left(\sqrt{\tfrac{2^{d+1}}{\log(m)}} + \tfrac{2}{m}\right) .$$

□

Claim 9.7. *We have with probability at least* $1 - 20\exp(-k^{1/8}/10)$ *that*

$$\delta_{\boxtimes}(L^d_{\mathcal{P}}, L^d_{\mathcal{P}}[R,C]) \leq \tfrac{4}{k^{1/16}} .$$

Proof. Recall that $L^d_{\mathcal{P}} = (U_1, id)$ and $L^d_{\mathcal{P}}[R,C] = (U_2, id)$. Furthermore we write $R = (r_1, \ldots, r_k)$ and $C = (c_1, \ldots, c_k)$. For $r_i \in R$ and $c_i \in C$ we assign an interval of measure $1/k$ by setting $\mathsf{J}_{k,r_i} \coloneqq \mathsf{J}_{k,c_i} \coloneqq \mathsf{J}_{k,i}$. Also we want to count how many elements of R fall into the intervals of $\mathfrak{J}_r$ and their refinement $\mathcal{P}$ by setting $R_i \coloneqq R \cap \mathsf{J}_{r,i}$ and $R_{i,j} \coloneqq R \cap P_{i,j}$. Similarly, we write $C_i \coloneqq C \cap \mathsf{J}_{r,i}$ and $C_{i,j} \coloneqq C \cap P_{i,j}$. Finally we write $R'_{i,j} \coloneqq \bigcup_{x\in R_{i,j}} \mathsf{J}_{k,x}$ and $C'_{i,j} \coloneqq \bigcup_{x\in C_{i,j}} \mathsf{J}_{k,x}$. The key observation is that by the definition of the sampling process the two distribution-valued bigraphons U_1 and U_2 are very similar. Both are step Latinons and U_1 takes the same value on $P_{i,j} \times P_{s,t}$ as U_2 on $R'_{i,j} \times C'_{s,t}$. The only difference lies in the size and the position of the steps. It will turn out that with high probability most of $R'_{i,j}$ and $C'_{i,j}$ actually lies in $\mathsf{J}_{r,i}$ and is of similar measure as $P_{i,j}$. Hence a suitable reordering inside each $\mathsf{J}_{r,i}$ matches U_1 to U_2 up to a small error set of $[0,1]^2$.

We first observe that with high probability the chosen elements of R distribute among the partite sets $P_{i,j}$ as expected. By Lemma 2.1 with $\varepsilon = 1/k^{5/16}$ and $\mu = \lambda(P_{i,j}) \cdot k = k/(mr) = k^{3/4}$ we have with probability at least $(1 - rm \cdot \exp(-k^{1/8}/3)) > 1 - 10\exp(-k^{1/8}/10)$ that

$$|R_{i,j}| = \lambda(P_{i,j}) \cdot k \pm k^{11/16}\lambda(P_{i,j}) = k^{3/4} \pm k^{7/16} \text{ for every } i \in [r], j \in [m] . \tag{21}$$

Now, we can observe that most of $R'_{i,j}$ is contained in $\mathsf{J}_{r,i}$. In the worst case each of the at most rm intervals before $R'_{i,j}$ are too short (or too long) by the maximum size, by (21) i.e. $rm \cdot k^{7/16} \cdot 1/k$, and hence

$$\lambda(R'_{i,j} \cap \mathsf{J}_{r,i}) = \lambda\left(\left(\bigcup_{x \in R_{i,j}} \mathsf{J}_{k,x}\right) \cap \mathsf{J}_{r,i}\right) \geq \lambda(P_{i,j}) - rm/k^{9/16} = 1/k^{1/4} - 1/k^{5/16} .$$

Thus for each i, j we can choose sets $A_{i,j} \subseteq R'_{i,j} \cap \mathsf{J}_{r,i}$ and $B_{i,j} \subseteq P_{i,j} \subseteq \mathsf{J}_{r,i}$ of measure $1/k^{1/4} - 1/k^{5/16}$ and a bijection $\varphi_{i,j} : A_{i,j} \to B_{i,j}$. Let $\varphi : [0,1] \to [0,1]$ be defined by

$$x \mapsto \begin{cases} \varphi_{i,j}(x), \text{ if } x \in A_{i,j}, \\ x, \text{ otherwise.} \end{cases}$$

Note that then, since φ only permutes elements inside each $\mathsf{J}_{r,i}$, that

$$\|O - O^{\varphi}\|_{\square} \leq r \cdot 1/r^2 = 1/r = 1/k^{1/8} .$$

With probability at least $1 - 10\exp(-k^{1/8}/10)$ we can repeat the same procedure for C to get sets $A'_{i,j} \subseteq C'_{i,j} \cap \mathsf{J}_{r,i}$ and $B'_{i,j} \subseteq P_{i,j} \subseteq \mathsf{J}_{r,i}$ of measure $1/k^{1/4} - 1/k^{5/16}$ and a bijection ψ defined in an analogous way such that

$$\left\|O - O^{\psi}\right\|_{\square} \leq 1/k^{1/8} .$$

Note that then $U_1^{\varphi,\psi}(x,y) = U_2(x,y)$ for every $x \in \bigcup_{i \in [r], j \in [m]} A_{i,j}$ and $y \in \bigcup_{i \in [r], j \in [m]} A'_{i,j}$. Thus $U_1^{\varphi,\psi}$ and U_2 agree everywhere on $[0,1]^2$ except on a set of measure at most $2 \cdot rm \cdot 1/k^{5/16} = 2/k^{1/16}$ and we have

$$\left\|U_1^{\varphi,\psi} - U_2\right\|_{\square} \leq 2/k^{1/16} .$$

Hence in total with probability at least $1 - 20\exp(-k^{1/8}/10)$

$$\delta_{\boxtimes}((U_1, id), (U_2, id)) \leq \tfrac{2}{k^{1/16}} + \tfrac{2}{k^{1/8}} < \tfrac{4}{k^{1/16}} .$$

□

□

We can now prove Lemma 4.9.

Proof of Lemma 4.9. Let $k \in \mathbb{N}$. By Lemma 9.1 with $\varepsilon = \exp(-k)/k^2$ there exists a Latinon $L' = (U, id)$ on $[0,1]^2$ such that

$$\delta^*_{\boxtimes}(L, L') < \exp(-k)/k^2 ,$$

and a coupling $\mathcal{C}$ of sampling a $k \times k$ pattern A from L and a $k \times k$ pattern B from L' such that

$$\mathbb{P}_{\mathcal{C}}[A \neq B] < \varepsilon k^2 = \exp(-k) .$$

By Lemma 9.4 we then have with probability at least $1 - 29\exp(-k^{1/8}/10) - \exp(-k) > 1 - 30\exp(-k^{1/8}/10)$ that

$$\delta^*_{\boxtimes}(L, (W^A, id)) \leq \delta^*_{\boxtimes}(L, L') + \delta^*_{\boxtimes}(L', (W^B, id)) \leq \exp(-k) + \tfrac{29}{\log(k)^{1/8}} < \tfrac{30}{\log(k)^{1/8}} \ .$$

□

10. Proof of the inverse counting lemma

In this section we prove the inverse counting lemma, Lemma 4.6. To this end, we need to draw a connection between sampled realisations and patterns. The following definition will be useful.

Definition 10.1. *Given $\varepsilon > 0$, we say that a finite multiset $S \subset [0,1]$ is ε-spread if for each interval $I \subset [0,1]$ we have that $|I \cap S| = (\lambda(I) \pm \varepsilon)|S|$.*

Our first observation is that sampled realisations are almost always spread.

Lemma 10.2. *Let L be a Latinon on ground space Ω, and $k \in \mathbb{N}$. Let S be the (random) multiset of k^2 values obtained from L by $k \times k$-sampling. Then we have with probability at least $1 - \exp(-k^{0.1})$ that S is $(4k^{-0.4})$-spread.*

Proof. Let us fix an arbitrary set $Y \subset [0,1]$. Let $(i,j) \in [k]^2$ be arbitrary. Now, consider the sampling from L for the value v corresponding to the i-th sampled row and j-th sampled column.[11] By the uniform marginals property we have

$$\mathbb{P}[v \in Y] = \lambda(Y) \ .$$

Thus, by the linearity of expectation we have for the multiset of all k^2 sampled values that

$$\mathbb{E}[|S \cap Y|] = \lambda(Y)k^2 \ .$$

In the rest of the proof, we argue that this value is concentrated. To this end we want to use McDiarmid's inequality, Lemma 2.2. We need to set up a suitable product probability space. There is an option which reflects our sampling. The first k coordinates will be copies of Ω representing the selection of rows, the next k coordinates will be copies of Ω representing the selection of columns, and the last k^2 coordinates should model the choice of the individual entries of the sampled matrix M. We have to be careful because these last probability spaces need to be fixed from the beginning in the setting of McDiarmid's inequality. Let $\Lambda = \Omega^k \times \Omega^k \times [0,1]^{k \times k}$. For $(i,j) \in [k]^2$, and $\mathbf{w} \in \Lambda$, let q_{ij} be defined as follows. Take $\mathcal{D}_{ij} \coloneqq L(\mathbf{w}_i, \mathbf{w}_{k+j})$. That is, $\mathcal{D}_{ij}$ is the distribution on $[0,1]$ at the i-th sampled row and the j-th sampled column. Let $F_{\mathcal{D}_{ij}}$ be the cumulative distribution function of $\mathcal{D}_{ij}$. Now, let q_{ij} be the infimum of t's for which $F_{\mathcal{D}_{ij}}(t) > \mathbf{w}_{k+i+kj}$. This construction

[11] Notice that we are really talking about the row that was sampled as the i-th, not about the row that has the i-th smallest global position. Same for the column.

ensures that the matrix $(q_{ij})_{i,j}$ has the same distribution as the matrix sampled from L (before reordering of rows and columns).

We can now define a map $Z : \Lambda \to \mathbb{R}$, $Z(\mathbf{w}) \coloneqq \sum_{i,j} \mathbb{1}_{q_{ij} \in Y}$. The above justifies that Z satisfies the requirements of Lemma 2.2 with $\mathbf{c}$ being k on the first $2k$ coordinates and being 1 on the remaining k^2 coordinates. Thus, Lemma 2.2 tells us that for each $t > 0$ we have

$$\mathbb{P}\left[\left|Z - \lambda(Y)k^2\right| > t\right] \leq \exp\left(-\frac{2t^2}{2k \cdot k^2 + k^2 \cdot 1^2}\right) \leq \exp\left(-\frac{2t^2}{3k^3}\right) . \tag{22}$$

Now, let us take $t \coloneqq \lceil k^{1.6} \rceil$, and let us apply the above for intervals $Y_{\ell_1,\ell_2} \coloneqq \left[\frac{\ell_1}{t}, \frac{\ell_2}{t}\right]$, $0 \leq \ell_1 < \ell_2 \leq t$, $\ell_1, \ell_2 \in \mathbb{N}$. The union bound over (22) tells us that

(23)

$$\begin{aligned}\mathbb{P}\left[\text{for each } \ell_1 \text{ and } \ell_2 \text{ we have } \left||S \cap Y_{\ell_1,\ell_2}| - \lambda(Y_{\ell_1,\ell_2})k^2\right| \leq \lceil k^{1.6} \rceil\right] &\geq 1 - t^2 \exp\left(-\frac{2t^2}{3k^3}\right) \\ &\geq 1 - \exp(-k^{0.1}) .\end{aligned}$$

Now, we claim that if the good event from (23) is satisfied, then S is $(40k^{-0.5})$-spread. Indeed, let $I \subset [0,1]$ be an arbitrary interval. We now choose a minimal interval $Y_{\ell_1^+,\ell_2^+}$ which contains I. Note that $\lambda(Y_{\ell_1^+,\ell_2^+}) \leq \lambda(I) + \frac{2}{\lceil k^{1.6} \rceil}$. Then we have

$$\begin{aligned}|S \cap I| \leq |S \cap Y_{\ell_1^+,\ell_2^+}| &\leq \lambda(Y_{\ell_1^+,\ell_2^+})k^2 + \lceil k^{1.6} \rceil \leq \left(\lambda(I) + \frac{2}{\lceil k^{1.6} \rceil}\right) k^2 + \lceil k^{1.6} \rceil \\ &\leq \left(\lambda(I) + 4k^{-0.4}\right) k^2 .\end{aligned}$$

Similarly, taking a maximal interval $Y_{\ell_1^-,\ell_2^-}$ which is contained in I, we can obtain that $|S \cap I| \geq \left(\lambda(I) - 4k^{-0.4}\right) k^2$. This concludes the proof. □

Our second observation is that realisations which are spread and belong to the same pattern are close in the cut norm.

Lemma 10.3. *Let $A \in \mathcal{R}(k,k)$ be a $k \times k$ pattern and $A^* \in \mathcal{R}^A([0,1])$ be a realisation. Let $L^A = (W^A, id)$ and $L^{A^*} = (W^{A^*}, id)$ be the associated semilatinons with ground space $[0,1]$. If A^* is ε-spread, then*

$$\left\|W^A - W^{A^*}\right\|_{\boxtimes} \leq 4\varepsilon + \tfrac{4}{k^2} . \tag{24}$$

Proof. Observe that $A_{i,j} = |\{A^*\} \cap [0, A^*_{i,j}]| = k^2(A^*_{i,j} \pm \varepsilon)$ for each fixed $i, j \in [k]$, where the first equality holds because of $A^* \in \mathcal{R}^A([0,1])$ and the second, as A^* is ε-spread.

In order to verify (24) using Definition 4.1, let $R, C \subseteq [0,1]$ be arbitrary and $V \coloneqq [a,b] \subseteq [0,1]$ an arbitrary interval. Set $Y \coloneqq \{(i,j) \in [k]^2 \mid k^{-2}A_{i,j} \in V, A^*_{i,j} \notin V\}$ and $Y^* \coloneqq \{(i,j) \in [k]^2 \mid A^*_{i,j} \in V,\ k^{-2}A_{i,j} \notin V\}$. Since $A^*_{i,j} = k^{-2}A_{i,j} \pm \varepsilon$, the set

$Y' \coloneqq \{(i,j) \in [k]^2 \mid \ k^{-2}A_{i,j} \geq a, A^*_{i,j} < a\}$ has size at most $\varepsilon k^2 + 1$. Similarly the set $Y'' \coloneqq \{(i,j) \in [k]^2 \mid \ k^{-2}A_{i,j} \leq b, A^*_{i,j} > b\}$ has size at most $\varepsilon k^2 + 1$. Then $|Y| \leq |Y'| + |Y''| \leq 2\varepsilon k^2 + 2$. We similarly obtain $|Y^*| \leq 2\varepsilon k^2 + 2$.

Therefore

$$\left| \int_{x \in R} \int_{y \in C} W^A(x,y)[V] - W^{A^*}(x,y)[V] \mathsf{d}x \mathsf{d}y \right| \leq \tfrac{1}{k^2}(|Y| + |Y^*|) \leq \tfrac{1}{k^2}(4\varepsilon k^2 + 4) \leq 4\varepsilon + \tfrac{4}{k^2} \ .$$

□

We now have all the tools to prove Lemma 4.6.

Proof of Lemma 4.6. We set $\varepsilon \coloneqq 4k^{-0.4}$. For patterns $A \in \mathcal{R}(k,k)$ and realisations $A^* \in [0,1]^{k \times k}$ we will work with the associated semilatinons $L^A = (W^A, id)$ and $L^{A^*} = (W^{A^*}, id)$ over the ground space $[0,1]$. Furthermore we set

$$\begin{aligned}
\mathcal{A}_1 &\coloneqq \{A \in \mathcal{R}(k,k) \mid \delta^*_{\boxtimes}(L^A, L_1) < d/2\} \ , \\
\mathcal{A}_2 &\coloneqq \{A \in \mathcal{R}(k,k) \mid \delta^*_{\boxtimes}(L^A, L_2) < d/2\} \ , \\
\mathcal{A}_1^{*,\varepsilon} &\coloneqq \{A^* \in [0,1]^{k \times k} \mid A^* \text{ is } \varepsilon\text{-spread and } \delta^*_{\boxtimes}(L^{A^*}, L_1) < d/4\} \ , \\
\mathcal{A}_2^{*,\varepsilon} &\coloneqq \{A^* \in [0,1]^{k \times k} \mid A^* \text{ is } \varepsilon\text{-spread and } \delta^*_{\boxtimes}(L^{A^*}, L_2) < d/4\} \ .
\end{aligned}$$

Note that the triangle inequality for $\delta^*_{\boxtimes}$ implies that $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$. Furthermore, as the choice of ε and the condition on the size of k guarantee that $4\varepsilon + 4/k^2 < d/4$, we have by Lemma 10.3 that the following holds. If $A \in \mathcal{R}(k,k)$ and $A^* \in \mathcal{R}^A([0,1])$ is ε-spread and such that $\delta^*_{\boxtimes}(L^{A^*}, L_1) < d/4$, then $\delta^*_{\boxtimes}(L^A, L_1) < 4\varepsilon + 4/k^2 + d/4 \leq d/2$. Therefore we have that for every $A^* \in \mathcal{A}_1^{*,\varepsilon}$ there exists $A \in \mathcal{A}_1$ such that $A^* \equiv A$. Hence we have for uniformly chosen $<_f$-ordered k-sets $S, T \subseteq \Omega$ that

$$\sum_{A \in \mathcal{A}_1} t(A, L_1) \geq \mathbb{P}[L[S,T] \in \mathcal{A}_1^{*,\varepsilon}] \ ,$$

and, as the size of k guarantees $\frac{30}{\log(k)^{1/10}} < d/4$, it follows from Lemma 4.9 and Lemma 10.2 that

$$\mathbb{P}[L[S,T] \in \mathcal{A}_1^{*,\varepsilon}] \geq 1 - \exp(-k^{0.1}) - \exp(-4k/\log(k)^{0.25}) \geq \tfrac{3}{4} \ .$$

Similarly we have

$$\sum_{A \in \mathcal{A}_2} t(A, L_2) \geq \mathbb{P}[L[S,T] \in \mathcal{A}_2^{*,\varepsilon}] \geq \tfrac{3}{4} \ ,$$

and therefore

$$\sum_{A \in \mathcal{A}_1} t(A, L_2) \leq 1 - \sum_{A \in \mathcal{A}_2} t(A, L_2) \leq \tfrac{1}{4} \ .$$

Thus

$$\sum_{A \in \mathcal{A}_1} (t(A, L_1) - t(A, L_2)) \geq \tfrac{1}{2}$$

and by averaging, and since $|\mathcal{A}_1| \leq k^2!$, there exists a pattern $A \in \mathcal{A}_1$ such that

$$|t(A, L_1) - t(A, L_2)| \geq \left(\sum_{A \in \mathcal{A}_1} t(A, L_1) - \sum_{A \in \mathcal{A}_1} t(A, L_2) \right) / (k^2!) \geq \tfrac{1}{2(k^2!)} \ .$$

□

11. Approximating Latinons

In this section, we prove Theorem 3.14. In fact, Theorem 3.14 is a direct corollary of the lemma below, which we prove instead.

Lemma 11.1. *Let L be a Latinon on Ω. For every $\varepsilon > 0$ and $n_0 \in \mathbb{N}$ there exists $n \geq n_0$ and a Latin square L_n of order n with the property that*

$$\delta^*_{\boxtimes}(L, L_n) < \varepsilon \ .$$

To find such a Latin square we first approximate the given Latinon L by a step Latinon.

Definition 11.2. *A step Latinon (U, f) on Ω is a Latinon such that there exists a partition $\mathcal{P} = (P_1, \ldots, P_m)$ of Ω and a partition $\mathcal{Q} = (Q_1, \ldots, Q_d)$ of $[0, 1]$ consisting of consecutive intervals such that there exist $\alpha_{i,j,k} \in [0, 1]$, $i, j \in [m], k \in [d]$, such that for every $I \subseteq Q_k$,*

$$U(x, y)[I] = \alpha_{i,j,k} \frac{\lambda(I)}{\lambda(Q_k)} \quad \text{for every } (x, y) \in P_i \times P_j \ .$$

In other words, the distribution of U is constant on each $P_i \times P_j$, and consists of a linear combination (with coefficients between 0 and 1) of Lebesgue measures on $\{Q_k\}_{k \in [d]}$.

Our first technical lemma says that each Latinon can be approximated by a step Latinon. For technical reasons, we wish to switch in this approximation from the ground space Ω to ground space $[0, 1]$. We leave the proof of the lemma for later.

Lemma 11.3. *Let L be a Latinon on Ω. For every $\varepsilon > 0$ there exists a step Latinon (U, id) on $[0, 1]$ such that its corresponding partitions $\mathcal{P}$ and $\mathcal{Q}$ consist of intervals and*

$$\delta^*_{\boxtimes}(L, (U, id)) < \varepsilon \ .$$

Then we find a Latin square which is close to the previously constructed step Latinon.

Lemma 11.4. *Let $\varepsilon > 0$ and $n_0 \in \mathbb{N}$. If (U, id) is a step Latinon on $[0, 1]$ whose corresponding partitions consist of intervals, then there exists $n \geq n_0$ and a Latin square L_n of order n with the property that*

$$\delta_{\boxtimes}((U, id), L_n) < \varepsilon \ .$$

Using both lemmas the proof of Lemma 11.1 then follows immediately.

Proof of Lemma 11.1. Given $\varepsilon > 0$, we use Lemma 11.3 with $\varepsilon/2$ in place of ε to get a step Latinon (U, id) on $[0,1]$ with $\delta_{\boxtimes}(L, (U, id)) < \varepsilon/2$. We then use Lemma 11.4 again with $\varepsilon/2$ in place of ε to get a number $n \geq n_0$ and a Latin square L_n of order n such that $\delta_{\boxtimes}((U, id), L_n) < \varepsilon/2$. Together we therefore get

$$\delta^*_{\boxtimes}(L, L_n) \leq \delta^*_{\boxtimes}(L, (U, id)) + \delta_{\boxtimes}((U, id), L_n) < \varepsilon/2 + \varepsilon/2 = \varepsilon \, .$$

□

We first show the proof of Lemma 11.3 and defer the proof of Lemma 11.4 to Section 11.1. For this we introduce the phrase *for α sufficiently smaller than β*, written $\alpha \ll \beta$, meaning that for any $\beta > 0$ there exists $\alpha_0 > 0$ such that for any $\alpha \leq \alpha_0$ the subsequent statement holds.

Proof of Lemma 11.3. First we choose constants $d, r, m \in \mathbb{N}$ such that

$$\tfrac{1}{r} \ll \tfrac{1}{d}, \tfrac{1}{m} \ll \varepsilon \, .$$

Let $L = (W, f)$ be the given Latinon. Let $L^d = (O^f, W_{d,1}, \ldots, W_{d,2^d})$ be the compression of L of depth d. Consider the partition $\mathfrak{J}_m$. By Lemma 5.2(i) we find a partition $\mathcal{P} = \{P_{j,t} \mid j \in [m], t \in [r]\}$ of Ω refining $f^{-1}(\mathfrak{J}_m)$, i.e. $f^{-1}(\mathsf{J}_{m,j,)} = P_{j,1} \cup \cdots \cup P_{j,r}$, $j \in [m]$, and such that, for every $1 \leq i \leq 2^d$,

$$\left\|W_{d,i} - (W_{d,i})^{\bowtie\mathcal{P}}\right\|_{\square} \leq \sqrt{\tfrac{2^{d+1}}{\log(r)}} \, . \tag{25}$$

Note that this defines a step Latinon $L^d_{\mathcal{P}} = (W', f)$ with $W' = \nabla((W_{d,1})^{\bowtie\mathcal{P}}, \ldots, (W_{d,2^d})^{\bowtie\mathcal{P}})$ on Ω and we get by Lemma 6.6, Proposition 6.7, and (25) that

$$\delta_{\boxtimes}(L, L^d_{\mathcal{P}}) \leq \delta_{\boxtimes}(L, L^d) + \delta_{\boxtimes}(L^d, L^d_{\mathcal{P}}) \leq \tfrac{1}{2^{d-1}} + 2^d \cdot \sqrt{\tfrac{2^{d+1}}{\log(r)}} < \varepsilon/2 \, . \tag{26}$$

We now fix arbitrarily for every $j \in [m]$ another partition $\mathcal{Q}_j = \{Q_{j,1}, \ldots, Q_{j,r}\}$ of $\mathsf{J}_{m,j}$ such that $Q_{j,t}$ is an interval and $\lambda(Q_{j,t}) = \mu(P_{j,t})$ for every $t \in [r]$. Note that this is possible, as $\sum_{t\in[r]} \mu(P_{j,t}) = \lambda(\mathsf{J}_{m,j})$ for $j \in [m]$. Then $\mathcal{Q} = \{Q_{j,t} \mid j \in [m], t \in [r]\}$ is a partition of $[0,1]$ consisting of intervals. We now use this partition to define a step Latinon (U, id) on $[0,1]$ by setting for every $x \in Q_{i,s}, y \in Q_{j,t}$ and $I \subseteq [0,1]$ that $U(x,y)(I) \coloneqq W'(x', y')(I)$ for some $x' \in P_{i,s}$ and $y' \in P_{j,t}$. Note that since $L^d_{\mathcal{P}}$ is a step Latinon this definition does not depend on the choice of x', y' and also consequently (U, id) is a step Latinon.

Next, we observe that (U, id) is close to $L^d_{\mathcal{P}}$ in the generalised cut distance $\delta^*_{\boxtimes}$ (see Section 4.0.1) which can compare Latinons defined on different ground spaces.

Claim 11.5. *We have*

$$\delta^*_{\boxtimes}(L^d_{\mathcal{P}}, (U, id)) < \varepsilon/2.$$

Proof of Claim 11.5. For every $j \in [m], t \in [r]$ let $\varphi_{j,t} : Q_{j,t} \to P_{j,t}$ be an arbitrary measure preserving bijection. These bijections altogether naturally define a measure preserving bijection $\varphi : [0,1] \to \Omega$. Note that then $(W')^{\varphi,\varphi}(x,y) = U(x,y)$ for every $x, y \in [0,1]$ by definition of U. Furthermore, we also have $\varphi(\mathsf{J}_{m,k}) = f^{-1}(\mathsf{J}_{m,k})$ and therefore $(f \circ \varphi)(\mathsf{J}_{m,k}) = \mathsf{J}_{m,k}$ for every $k \in [m]$. Thus

$$\delta^*_{\boxtimes}((U,id),L^d_{\mathcal{P}}) \leq 2\left\|O - O^{f\circ\varphi}\right\|_{\square} + \left\|U - (W')^{\varphi,\varphi}\right\|_{\boxtimes} \leq 2\sum_{k\in[m]} \lambda(\mathsf{J}_{m,k})^2 = \tfrac{2}{m} < \varepsilon/2\ .$$

□

We therefore have by (26), Claim 11.5, and the generalised triangle inequality from Section 4.0.1 that

$$\delta^*_{\boxtimes}(L,(U,id)) \leq \delta_{\boxtimes}(L,L^d_{\mathcal{P}}) + \delta^*_{\boxtimes}(L^d_{\mathcal{P}},(U,id)) < \varepsilon/2 + \varepsilon/2 = \varepsilon\ .$$

□

11.1. **Constructing Latin squares.** It is more convenient to represent the sought Latin square L_n as in Lemma 11.4 as a K_3-decomposition of the complete tripartite graph $K_{n,n,n}$. Indeed, given any Latin square L_n of order n, we view the three parts of $K_{n,n,n}$ as numbering rows, columns, and values, respectively. Then each entry of L_n corresponds to a triangle in $K_{n,n,n}$, and the collection of all the entries corresponds to a K_3-*decomposition* of $K_{n,n,n}$, that is a partition of $E(K_{n,n,n})$ into triangles. In fact, it is easy to see that once the role of the three parts of $K_{n,n,n}$ and their numberings are fixed, there is a one-to-one correspondence between Latin squares of order n and K_3-decompositions of $K_{n,n,n}$.

We will need to recall some notation and facts from the theory of designs, mainly following Keevash's breakthrough results [28, 29]. As we explained above, it suffices to recall these results in the special case of K_3-decompositions of tripartite graphs. Given a graph G, a vertex $v \in V(G)$, and $S, V_1, V_2 \subseteq V(G)$ where $V_1 \cap V_2 = \emptyset$, we write $N_G(v,S) = N_G(v) \cap S$ for the neighbourhood of v in S. We also write $\overline{N}_G(S) = \bigcap_{w\in S} N_G(w)$ for the common neighbourhood of S and $d_G(V_1,V_2) = \frac{|E(G[V_1,V_2])|}{|V_1||V_2|}$ for the density of the pair (V_1,V_2).

Definition 11.6. *Let G be a tripartite graph with partite sets V_1, V_2, V_3 and densities $d_{\{i,j\}} = \frac{|E(G[V_i,V_j])|}{|V_i||V_j|}$, $\{i,j\} \in \binom{[3]}{2}$.*

(i) We say that G is (partite) K_3-decomposable *if there exists an edge-disjoint collection of copies of K_3 in G which covers $E(G)$.*

(ii) We say that G is K_3-balanced *if for every $\{i,j,k\} = [3]$ and every $v \in V_k$, $\deg(v,V_i) = \deg(v,V_j)$.*

(iii) *Let $c > 0$ and $h \in \mathbb{N}$. We say that G is* (c,h)-typical*, if for every pair of distinct $i,j \subseteq [3]$, and every set $S \subseteq V_i \cup V_j$ with $|S| \leq h$ we have that $|\overline{N}_G(S) \cap V_k| = (1 \pm c) d_{\{i,k\}}^{|S \cap V_i|} d_{\{j,k\}}^{|S \cap V_j|} |V_k|$, where $k \in [3] \setminus \{i,j\}$.*

It is easy to see that a K_3-decomposable tripartite graph needs to be K_3-balanced. Now, the important result from the theory of designs we want to use is that for typical graphs the converse is also true. The first theorem we need is that a typical graph can be approximately decomposed such that the leftover graph has bounded degree. The existence of an approximate decomposition was shown first by Rödl in [45]. For our purposes we need a slightly stronger version, i.e. the additional degree condition for the leftover graph. This follows from a more general theorom of Alon and Yuster in [3]. For completeness we explain how to derive our version in Appendix B.

Theorem 11.7 ([45])**.** *Suppose that $\frac{1}{n} \ll c \ll \alpha \ll d \leq 1$. Let G be a tripartite graph with partite sets V_1, V_2, V_3 such that $|V_1| = |V_2| = |V_3| = n$ and $d_G(V_i, V_j) = (1 \pm c)d$ for all $\{i,j\} \subseteq [3]$. If G is $(c, 2^{4500})$-typical, then there exists a subgraph $S \subseteq E(G)$ such that*

(1) $G - S$ has a partite K_3-decomposition, and
(2) $\Delta(S) \leq \alpha n$.

The second theorem we need is that typical K_3-balanced tripartite graphs are K_3-decomposable and is a direct application of Theorem 1.7 from [29].

Theorem 11.8 ([29])**.** *Suppose that $\frac{1}{n} \ll c \ll d \leq 1$. Let G be a K_3-balanced tripartite graph with partite sets V_1, V_2, V_3 such that $|V_1| = |V_2| = |V_3| = n$ and $d_G(V_i, V_j) \geq d$ for all $\{i,j\} \in \binom{[3]}{2}$. If G is $(c, 2^{4500})$-typical, then G has a partite K_3-decomposition.*

We will also make use of the following simple lemma stating that selecting edges independently and uniformly at random from a typical graph results in a typical subgraph with high probability.

Lemma 11.9. *Suppose that $1/n \ll c_1 \ll c_2 \ll 1/h, \rho, d \leq 1$. Let G be a (c_1, h)-typical tripartite graph with vertex classes V_1, V_2, V_3 with $|V_1| = |V_2| = |V_3|$ and $d_G(V_i, V_j) = (1 \pm c_1)d$ for every $\{i,j\} \in \binom{[3]}{2}$. If we select each edge of G with probability ρ independently and uniformly at random for a subgraph H, then with probability $1 - \exp(-\Theta(n))$ we have*

(i) $d_H(V_i, V_j) = (1 \pm c_2)\rho d$ for every $\{i,j\} \in \binom{[3]}{2}$,
(ii) H is (c_2, h)-typical, and
(iii) $G - H$ is (c_2, h)-typical.

Proof. For every $\{i,j\} \in \binom{[3]}{2}$ we have that $\mathbb{E}[|H[V_i, V_j]|] = \rho(1 \pm c_1)dn^2$ and by Lemma 2.1 we then get that

$$\mathbb{P}\left[\,||H[V_i, V_j]| - \rho(1 \pm c_1)dn^2| \geq (c_2/2)\rho dn^2\right] \leq 2\exp(-c_2^2 \rho dn^2/12)\,.$$

Therefore by the union bound (i) holds with probability $1-\exp(-\Theta(n))$. Furthermore, for every $\{i,j,k\}=[3]$, and every set $S\subseteq V_i\cup V_j$ with $|S|\leq h$ we have that

$$\mathbb{E}\left[|\overline{N}_H(S)\cap V_k|\right]=(1\pm c_1)(\rho(1\pm c_1)d)^{|S|}n\ ,$$

and again by Lemma 2.1 we then get that

$$\mathbb{P}\left[||\overline{N}_H(S)|-(1\pm c_1)(\rho(1\pm c_1)d)^{|S|}n|\geq(c_2/2)(\rho d)^{|S|}n\right]\leq 2\exp(-c_2^2(\rho d)^{|S|}n/12)\ .$$

Again by using the union bound the probability that H is (c_2,h)-typical is at least $1-h\cdot\binom{3n}{h}\cdot\exp(-c_2^2(\rho d)^h n/12)=1-\exp(-\Theta(n))$. Similarly, one can show that (iii) holds with probability $1-\exp(-\Theta(n))$. Therefore with probability $1-\exp(-\Theta(n))$, (i)-(iii) hold altogether. □

With these tools at hand we can prove Lemma 11.4. As we explained at the beginning of Section 11.1, we will find a triangle decomposition of $K_{n,n,n}$ which represents the Latin square we want to construct. We partition the three copies of $[n]$, which represent the rows, columns and values, into *clusters* corresponding to the parts of $[0,1]$ defined by the given step Latinon (U,id). The construction itself has three steps. First, we remove a sparse *reservoir* $R\subset E(K_{n,n,n})$ at random. Second, we find an approximate K_3-decomposition[12] B of $K_{n,n,n}-R$ with the key property that between any triple of clusters (traversing the three colour classes of $K_{n,n,n}$) the number of triangles of B spanned by these three clusters is proportional to the value of U on the steps corresponding to these three clusters. Theorem 11.7 is used to this end. Third, we use Theorem 11.8 to find a K_3-decomposition in the graph $K_{n,n,n}-B$. Theorem 11.8 indeed applies in this setting as $K_{n,n,n}-B$ is almost identical to R (recall Footnote 12) which has excellent typicality properties because of Lemma 11.9.

Proof of Lemma 11.4. First note that by passing to a refinement we may assume that (U,id) has the same corresponding partition $\mathcal{C}=\{C_1,\dots,C_\ell\}$ for each coordinate. Since (U,id) is a step Latinon we therefore have real values $K_{i,j,k}$, $i,j,k\in[\ell]$, such that $U(x,y)(I)=K_{i,j,k}\lambda(I)/\lambda(C_k)$ for every $x\in C_i$, $y\in C_j$, and $I\subseteq C_k$. Let $\varepsilon>0$, set $h=2^{4500}$, and $K\coloneqq\min\{K_{i,j,k}\mid i,j,k\in[\ell],K_{i,j,k}>0\}$. Choose additional constants such that

$$\tfrac{1}{n}\ll c_1\ll c_2\ll\alpha\ll c\ll\rho\ll\tfrac{1}{M},\tfrac{1}{q}\ll\tfrac{1}{n_0},\tfrac{1}{\ell},K,\tfrac{1}{h},\varepsilon,\min_i\lambda(C_i)\ ,$$

and such that $m\coloneqq n/(qM)=2^d$ for some $d\in\mathbb{N}$.

In the following we want to approximate (U,id) with another step Latinon (U',id) whose corresponding partition fulfills some convenient divisibility conditions. Let $(U_{d,1},\dots,U_{d,m})$ be the compression of U of depth d, i.e. $U_{d,i}(x,y)=U(x,y)(\mathsf{D}_{d,i})$ for every $x,y\in[0,1]$,

[12] By an 'approximate K_3-decomposition' we mean that a small number of edges of $K_{n,n,n}-R$ will not be covered; here the number of edges will be chosen to be much smaller than that of R.

$i \in [m]$. We now set $U' \coloneqq \nabla((U_{d,1})^{\bowtie\mathfrak{D}_d}, \dots, (U_{d,m})^{\bowtie\mathfrak{D}_d})$ and note that (U', id) is a step Latinon. Set $X \coloneqq \{i \in [m] \mid \mathsf{D}_{d,i} \not\subseteq C_j \text{ for all } j \in [\ell]\}$ and $Y \coloneqq \bigcup_{i\in X} \mathsf{D}_{d,i}$. Since $\mathfrak{D}_d$ consists of smaller intervals than $\mathcal{C}$ we therefore have $|X| \leq \ell$ and $\lambda(Y) \leq \ell/m$. Then the 3-dimensional measures $\nabla(U_{d,1}, \dots, U_{d,m})$ and U' agree everywhere except on $Q = (Y \times [0,1] \times [0,1]) \cup ([0,1] \times Y \times [0,1]) \cup ([0,1] \times [0,1] \times Y)$, that is, for each $Z \subset [0,1]^3 \setminus Q$ we have

$$\nabla(U_{d,1}, \dots, U_{d,m})(Z) = U'(Z) \,.$$

Hence by the uniform marginals we get

$$\left\|\nabla(U_{d,1}, \dots, U_{d,m}) - U'\right\|_{\boxtimes} \leq 3\ell/m = \tfrac{3\ell q M}{n} < \varepsilon/6 \,.$$

Also by Proposition 6.6(i) we have that

$$\|U - \nabla(U_{d,1}, \dots, U_{d,m})\|_{\boxtimes} \leq \tfrac{2}{m} < \varepsilon/6$$

and therefore

$$\left\|U - U'\right\|_{\boxtimes} \leq \|U - \nabla(U_{d,1}, \dots, U_{d,m})\|_{\boxtimes} + \left\|\nabla(U_{d,1}, \dots, U_{d,m}) - U'\right\|_{\boxtimes} < \varepsilon/3 \,. \tag{27}$$

Since (U', id) is a step Latinon we therefore have real values $M_{i,j,k}$, $i,j,k \in [m]$, such that $U'(x,y)(I) = M_{i,j,k}\lambda(I)/\lambda(\mathsf{D}_{d,k})$ for every $x \in \mathsf{D}_{d,i}$, $y \in \mathsf{D}_{d,j}$, and $I \subseteq \mathsf{D}_{d,k}$. By the uniform marginals of (U', id) we also have for every $i,j \in [m]$ that

$$\sum_{k\in[m]} M_{i,j,k} = \sum_{k\in[m]} M_{k,i,j} = \sum_{k\in[m]} M_{i,k,j} = 1 \,. \tag{28}$$

From now on we will only work with (U', id) and construct a corresponding 3-partite graph. For this we equipartition each interval $\mathsf{D}_{d,i}$, $i \in [m]$, further into consecutive intervals $P_{i,1}, \dots, P_{i,M}$. Next, we define a corresponding partition of $[n]$ into sets $A_{i,r}$, $i \in [m]$, $r \in [M]$ by setting $A_{i,r} \coloneqq \{1 + ((i-1)M + r - 1)q, \dots, ((i-1)M + r)q\}$. Note that $|A_{i,r}| = q = \lambda(P_{i,r}) \cdot n$ for every $i \in [m], r \in [M]$. Consider $K_{n,n,n}$; we want to find a K_3-decomposition of $K_{n,n,n}$ whose corresponding Latin square is ε-close to (U', id) in the Latinon cut distance. For this we first include each edge of $K_{n,n,n}$ uniformly at random with probability ρ in a reservoir graph R. The following claim follows directly from Lemma 11.9. We denote the partite sets of $K_{n,n,n}$ by V_1, V_2 and V_3.

Claim 11.10. *We have with probability* $1 - o(1)$ *that*

(i) $|R[V_i, V_j]| = (1 \pm c_1)\rho n^2$ *for every* $\{i, j \subseteq [3]\}$,
(ii) R *is* (c_1, h)*-typical, and*
(iii) $G' \coloneqq G - R$ *is* (c_1, h)*-typical.*

We can therefore choose an outcome of the random selection such that Claim 11.10 (i)-(iii) hold. In the following we will denote by $A^{(a)}_{i,r}$, $i \in [m], r \in [M], a \in [3]$ the copy of $A_{i,r}$ in V_a. We now make a further random selection; every edge will be independently

at random selected for one of the graphs $G^{r,s,t}_{i,j,k}$, $i,j,k \in [m]$, $r,s,t \in [M]$. For every $\{i,j\} \subseteq [m]$ and $\{r,s\} \subseteq [M]$ we add an edge $\{u,v\} \in G'[A^{(1)}_{i,r}, A^{(2)}_{j,s}]$ to one of the graphs $G^{r,s,t}_{i,j,k}$ with the probability to be selected for $G^{r,s,t}_{i,j,k}$ being $M_{i,j,k}/M$, $k \in [m], t \in [M]$. Similarly, for every $\{i,k\} \subseteq [m]$ and $\{r,t\} \subseteq [M]$ we add an edge $\{u,v\} \in G'[A^{(1)}_{i,r}, A^{(3)}_{k,t}]$ with probability $M_{i,j,k}/M$ to the subgraph $G^{r,s,t}_{i,j,k}$, $j \in [m], s \in [M]$. Lastly, for every $\{j,k\} \subseteq [m]$ and $\{s,t\} \subseteq [M]$ we add an edge $\{u,v\} \in G'[A^{(2)}_{j,s}, A^{(3)}_{k,t}]$ with probability $M_{i,j,k}/M$ to the subgraph $G^{r,s,t}_{i,j,k}$, $i \in [m], r \in [M]$. Note that by (28) the probabilities for each edge indeed sum up to 1 and that we may assume $V(G^{r,s,t}_{i,j,k}) = A^{(1)}_{i,r} \cup A^{(2)}_{j,s} \cup A^{(3)}_{k,t}$.

Claim 11.11. *We have with probability* $1-o(1)$ *that for every* $i,j,k \in [m]$ *and* $r,s,t \in [M]$

(i) $|E(G^{r,s,t}_{i,j,k}[V_a, V_b])| = (1 \pm c_2)(1-\rho)q^2 \cdot M_{i,j,k}/M$ *for all* $\{a,b\} \subseteq [3]$ *and*

(ii) $G^{r,s,t}_{i,j,k}$ *is* (c_2, h)*-typical.*

Proof of Claim 11.11. For fixed $i,j,k \in [m]$ and $r,s,t \in [M]$ we can model the random selection of $G^{r,s,t}_{i,j,k}$ by a random selection from the (c_1,h)-typical graph $G - R$ according to the distribution $\mathrm{Bi}(n, M_{i,j,k}/M)$. Therefore by Lemma 11.9 we have that (i) and (ii) hold with probability $1 - \exp(-\Theta(n))$ for $G^{r,s,t}_{i,j,k}$. By using the union bound we get that the probability that (i) or (ii) does not hold for all $i,j,k \in [m]$ and $r,s,t \in [M]$ is at most $\sum_{i,j,k \in [m]} \sum_{r,s,t \in [M]} \exp(-\Theta(n)) \leq m^3 M^3 \exp(-\Theta(n)) = \exp(-\Theta(n))$ which proves the assertion. □

For every $i,j,k \in [m]$ and $r,s,t \in [M]$, we can now use Theorem 11.7 on $G^{r,s,t}_{i,j,k}$ to get an almost K_3-decomposition $T^{r,s,t}_{i,j,k}$ of the edge set such that the leftover graph has maximum degree at most αq. After having done this for every $G^{r,s,t}_{i,j,k}$ we denote the set of leftover edges of G' by F and note that $\Delta(F) < (mM)^2 \alpha q = \alpha m M n$. Let us write $B \coloneqq \bigcup_{i,j,k \in [m], r,s,t \in [M]} E(T^{r,s,t}_{i,j,k})$. Now, $G'' \coloneqq G - B$ is K_3-balanced, as it is the leftover of the K_3-balanced graph $K_{n,n,n}$ after removing a set of edge-disjoint triangles. Furthermore, note that $G'' = R \cup F$ is (c,h)-typical with density at least ρ, as for every $S \subseteq V_i \cup V_j$, $\{i,j\} \subseteq [3]$, with $|S| \leq h$ we have that $|\overline{N}_{R \cup F}(S)| = (1 \pm c_1)\rho^{|S|} n \pm \alpha m M n = (1 \pm c)\rho^{|S|} n$. We can therefore use Theorem 11.8 with ρ in place of d to find a K_3-decomposition D' of G'' and hence $D \coloneqq D' \cup B$ is a K_3-decomposition of $K_{n,n,n}$ corresponding to a Latin square which we denote by L_n. Note that we have for the linear 3-uniform hypergraph H which corresponds to L_n that

$$\left| E(H[A^{(1)}_{i,r}, A^{(2)}_{j,s}, A^{(3)}_{k,t}]) \right| = (1 \pm 2\rho) M_{i,j,k} q^2 / M, \text{ for all } i,j,k \in [m], r,s,t \in [M] \,. \tag{29}$$

It now only remains to show that L_n is ε-close to (U, id). We write $L \coloneqq (W_{L_n}, id)$ for the Latinon-representation of L_n over $[0,1]$, and recalling (27) and $\delta_{\boxtimes}((U,id), L_n) = \delta_{\boxtimes}((U,id), L)$, it suffices to show $\|U' - W_{L_n}\|_{\boxtimes} \leq 2\varepsilon/3$. So let $S, T \subseteq [0,1]$ and $V \subseteq [0,1]$

be an interval. Since $1/M \ll \varepsilon$ we can choose $S', T', V' \subseteq [0,1]$ such that V' is an interval and suitable index sets $Q_i^{S'}, Q_j^{T'}, Q_k^{V'} \subseteq [M]$, for $i,j,k \in [m]$, such that

- $S' = \bigcup_{i\in[m]} \bigcup_{r\in Q_i^{S'}} P_{i,r}$,
- $T' = \bigcup_{j\in[m]} \bigcup_{s\in Q_j^{T'}} P_{j,s}$,
- $V' = \bigcup_{k\in[m]} \bigcup_{t\in Q_k^{V'}} P_{k,t}$, and
- $\lambda(S \triangle S'), \lambda(T \triangle T'), \lambda(V \triangle V') < \varepsilon/3$.

We then have by the uniform marginals property of (U', id) and L that

$$\left|\int_S \int_T (U'(x,y) - W_{L_n}(x,y))(V)\mathsf{d}x\mathsf{d}y\right| < \left|\int_{S'} \int_{T'} (U'(x,y) - W_{L_n}(x,y))(V')\mathsf{d}x\mathsf{d}y\right| + \varepsilon/3\,. \tag{30}$$

Furthermore, note that $P_{i,r} \coloneqq [(\min\{A_{i,r}\} - 1)/n, \max\{A_{i,r}\}/n]$ and therefore the partitions of L and U' conveniently coincide. We then have that

$$\begin{aligned}
&\left|\int_{S'} \int_{T'} U'(x,y)(V') - W_{L_n}(x,y)(V')\mathsf{d}x\mathsf{d}y\right| \\
&= \left|\sum_{i,j,k\in[m]} \sum_{r\in Q_i^{S'}, s\in Q_j^{T'}, t\in Q_k^{V'}} \int_{P_{i,r}} \int_{P_{j,s}} U'(x,y)(P_{k,t}) - W_{L_n}(x,y)(P_{k,t})\mathsf{d}x\mathsf{d}y\right| \\
&= \left|\sum_{i,j,k\in[m]} \sum_{r\in Q_i^{S'}, s\in Q_j^{T'}, t\in Q_k^{V'}} \int_{P_{i,r}} \int_{P_{j,s}} U'(x,y)(P_{k,t}) - |E(H[A_{i,r}^{(1)}, A_{j,s}^{(2)}, A_{k,t}^{(3)}])|/q^2\mathsf{d}x\mathsf{d}y\right| \\
&\overset{(29)}{\leq} \sum_{i,j,k\in[m]} \sum_{r\in Q_i^{S'}, s\in Q_j^{T'}, t\in Q_k^{V'}} |\lambda(P_{i,s})\lambda(P_{j,t})(M_{i,j,k}\lambda(P_{k,t})/\lambda(Q_{d,k}) - (1 \pm 2\rho)M_{i,j,k}/M)| \\
&\leq \sum_{i,j,k\in[m]} M^3 \cdot (q^2/n^2) \cdot 2\rho M_{i,j,k}/M \\
&= (2\rho/m^2) \cdot \sum_{i,j,k\in[m]} M_{i,j,k} = 2\rho < \varepsilon/3\,,
\end{aligned}$$

and thus together with (30) we obtain $\|U' - W_{L_n}\|_{\boxtimes} \leq 2\varepsilon/3$ as required. □

12. Possible further work

Here, we include some questions which would be a natural continuation of the theory we introduced in this paper.

12.1. **Removal lemma for Latin squares.** The removal lemma for graphs [47, 2, 17] states that if a large n-vertex graph G contains only $o(n^{v(H)})$ copies of H then by removing a suitable set of $o(n^2)$ edges from G we obtain a graph with no copies of H at all. Similar statements were investigated for permutations. Cooper [14] formulated a first removal-type lemma for permutations and Klimošová and Král' [31] a strong version using a metric called the *Spearman's footrule distance*. In [19] the first and the third author together with Kun and Pekárková showed how to derive a non-quantitative version of the removal lemma from [31] using the theory of permutation limits. It could be that a similar statement for Latin squares, which we state as a conjecture, can be obtained using the theory we introduce here.

Conjecture 12.1. *Let $k, \ell \in \mathbb{N}$ and $A \in \mathcal{R}(k,\ell)$. For every $\varepsilon > 0$ there exists $\delta > 0$ such that the following holds. Suppose that L is a Latin square of order n with $t(A,L) < \delta$. Then there exists a Latin square $\tilde{L}$ of order n which is A-avoiding and for which we have*

$$\sum_{i,j=1}^{n} |L(i,j) - \tilde{L}(i,j)| < \varepsilon n^3 \,. \tag{31}$$

The removal lemma in [19] is deduced from another result in [19] which says that permutons avoiding a fixed patterns must have a 'one-dimensional' structure. (Note that a finitary version of this one-dimensionality result is Theorem 3 in [14]). It would be interesting to investigate counterparts of this result for Latinons.

12.2. **Entropy and counting of Latin squares.** It is a basic calculation in the theory of random graphs that for any $\varepsilon > 0$ and any graph H, asymptotically almost every (as $n \to \infty$) graph G of order n satisfies that the density of H in G is $(1\pm\varepsilon)(\frac{1}{2})^{e(H)}$. Rephrased in the limit language: if $(G_n)_n$ is a sequence of uniformly random graphs of order n, then almost surely, $(G_n)_n$ converges to a graphon $W \equiv \frac{1}{2}$.

Here, we prove a Latinon counterpart of this result.

Theorem 12.2. *Given any $\varepsilon > 0$, $k, \ell \in \mathbb{N}$ and $A \in \mathcal{R}(k,\ell)$, asymptotically almost every (as $n \to \infty$) Latin square L of order n satisfies that $t(A,L) = \frac{1}{(k\ell)!} \pm \varepsilon$. Rephrased in the limit language: if $(L_n)_n$ is a sequence of uniformly random Latin squares of order n, then almost surely, $(L_n)_n$ converges to a Latinon $W \equiv \lambda_{\restriction [0,1]}$.*

To prove Theorem 12.2 we use the following nontrivial result of Kwan and Sudakov.

Theorem 12.3 (Theorem 3 in [34])**.** *Asymptotically almost every (as $n \to \infty$) Latin square L of order n satisfies that for each set $R \subset [n]$ of row indices, for each set $C \subset [n]$ of column indices, and each set $V \subset [n]$ of values, we have*

$$\left| \{(i,j) \in R \times C \mid L_{ij} \in V\} - \frac{|R|\,|C|\,|V|}{n} \right| = O\left(n^{3/2} \log n\right) \,.$$

With this result, Theorem 12.2 follows simply by computing the cut distance between Latinons.

Proof of Theorem 12.2. That is, we show asymptotically almost surely $\delta_{\boxtimes}(W, L_n)$ tends to zero as n tends to infinity, where $W \equiv \text{Lebesgue}[0,1]$ and (L_n, id) is the Latinon-representation of a uniformly random Latin square of order n. Since we have

$$\delta_{\boxtimes}(W, L_n) \leq \|W - L_n\|_{\square} = \sup_{\substack{R,C,V \subseteq [0,1],\\ V \text{ interval}}} \left| \int_{x \in R} \int_{y \in C} (W - L_n)(x,y)(V) \mathsf{d}x \mathsf{d}y \right|$$

we may assume that the R, C, V used in the supremum are at least $\Omega(1)$ in measure; then Theorem 12.3 implies that $\int_{x \in R} \int_{y \in C} L_n(x,y)(V) \mathsf{d}x \mathsf{d}y = \lambda(R)\lambda(C)\lambda(V) + O(\frac{\log n}{\sqrt{n}})$ for all such R, C, V. Then $\delta_{\boxtimes}(W, L_n) = O(\frac{\log n}{\sqrt{n}})$ which tends to zero as n tends to infinity, as required. □

We believe that the concept of Latinons records counts of finite Latin squares. To this end, we introduce the entropy of a Latinon, thus paralleling previous work on graphons [11] and for permutons [30]. Suppose that $L = (W, f)$ is a Latinon. Suppose that for almost every $(x,y) \in \Omega^2$, the measure $W(x,y)$ has a Radon–Nikodym derivative $g(x,y,\cdot)$ with respect to the Lebesgue measure. Then define the *entropy* of L by $\text{Ent}(L) \coloneqq \int_{x \in \Omega} \int_{y \in \Omega} \int_{z \in [0,1]} g(x,y,z) \log g(x,y,z) \mathsf{d}x \mathsf{d}y \mathsf{d}z$.

The number of Latin squares of order n is $((1 \pm o(1))\frac{n}{e^2})^{n^2}$, [51]. We believe that the entropy tells us how many Latin squares there are close to a given Latinon. More precisely, we believe that there are $((1 \pm o(1)) \cdot e^{\text{Ent}(L)} \cdot \frac{n}{e^2})^{n^2}$ Latin squares of order n close to a given Latinon L. The easiest way to state this formally is the following.

Conjecture 12.4. *Suppose that $(A_i \in \bigcup_{k,\ell} \mathcal{R}(k,\ell))_{i \in I}$ is a finite collection of patterns and $(a_i)_{i \in I}$ are reals in $[0,1]$. Let $\alpha \coloneqq \sup \text{Ent}(L)$, where L ranges through all Latinons with $t(A_i, L) = a_i$ for each $i \in I$. Then for each $\varepsilon > 0$ there exists $\delta > 0$ and n_0 such that for each $n > n_0$ we have that for the number N of Latin squares K of order n satisfying $t(A_i, K) = a_i \pm \delta$ for each $i \in I$ we have*

$$\frac{1}{n^2} \cdot \log \left(\frac{N}{(\frac{n}{e^2})^{n^2}} \right) = \alpha \pm \varepsilon \, .$$

12.3. **Higher dimensional permutations.** It has been a long programme of Linial to identify higher-dimensional counterparts to common combinatorial objects. Let us introduce his concept of higher dimensional permutations here. For $n \in \mathbb{N}$, any $\{0,1\}$-sequence with n entries which contains $n-1$ zeros and 1 one is called a 0-*dimensional permutation*. Now, inductively we say that a $(d+1)$-dimensional $\{0,1\}$-array whose each coordinate is indexed by $[n]$ is a *d-dimensional permutation* of order n if restricting this array by

fixing arbitrarily one coordinate we get a $(d-1)$-dimensional permutation. So, there is a one-to-one correspondence between 1-dimensional permutations and ordinary permutations on $[n]$ (represented as the corresponding permutation matrix), and 2-dimensional permutations correspond to Latin squares. It is very likely that with the same approach that we developed in this paper, one could create a theory of limits of permutations of any fixed dimension. However, there is a substantial technical challenge. Indeed, our Latinons can be encoded as compressions of graphons. Going up, say, by one dimension, we would have to work with compressions of 3-uniform hypergraphons. It is likely that many steps in our proof would actually need a substantial revision in that setting. For example, there are substantial difficulties with any 'cut distance' for 3-uniform hypergraphs (see Section 3 of [52]).

12.4. **Quasirandomness for Latinons, now solved.** The concept of pseudorandomness is omnipresent in contemporary mathematics. In graph theory, several related concepts appeared in the late 1980's, [49, 13, 46]; all these concepts stemmed from properties of the then recent regularity lemma of Szemerédi [48]. It turns out that the (now well established) concept of quasirandomness is best formalised in the language of graphons. In that language, the most comprehensive of the above results, the famous Chung–Graham–Wilson Theorem [13] states that, among others, the following properties are equivalent for a graphon W,

(a) W is a constant-p function, almost everywhere;
(b) $t(H,W) = p^{e(H)}$ for every graph H;
(c) $t(H,W) = p^{e(H)}$ for $H = K_2$ and for $H = C_4$.

Counterparts of the Chung–Graham–Wilson Theorem were obtained for other combinatorial structures, including uniform hypergraphs [50, 1], classes of oriented graphs [12, 20], and permutations [32]. Of course, the direction (a)⇒(b)⇒(c) is trivial in the graph case as well as in counterparts for other structures. After we published this preprint on the arXiv, Cooper, Král', Lamaison and Mohr [15] characterized constant Latinons, thus proving one of the conjectures posted in the preprint.

Theorem 12.5 ([15])**.** *Suppose that $L = (W, f)$ is a Latinon. Then the following are equivalent.*

(a) $W(x,y)$ is the Lebesgue measure on $[0,1]$ for almost every (x,y),
(b) $t(A,L) = 1/(k\ell)!$ for each $k,\ell \in \mathbb{N}$ and each $A \in \mathcal{R}(k,\ell)$,
(c) $t(A,L) = 1/6!$ for each $A \in \mathcal{R}(3,2)$.

Moreover, Cooper et al. show that (c) cannot be strengthened by using patterns $\mathcal{R}(2,2)$ or $\mathcal{R}(k,1)$ for any $k \in \mathbb{N}$.

Of course, Theorem 12.5 can be equivalently formulated for sequences of Latin squares of growing orders.

Acknowledgements

We thank Stefan Glock for helpful discussions about Section 11, Leonardo Nagami Coregliano, Dan Král' and Alexander Razborov for discussions about Section 3.3, and Jan Grebík and Martin Doležal for helping us with Section 3.2.4. We thank the anonymous referees for their helpful comments. In particular for pointing out that the removal lemmas, suggested in Section 12.1 in an earlier version of the paper, were not formulated in a meaningful way.

Appendix A. Proof of Proposition 3.18

Proof of Proposition 3.18. We replace Ω by the unit interval $\mathcal{I} = [0,1]$ with the Lebesgue measure (which we still denote by μ). Again, this can be assumed without loss of generality, since every separable atomless probability space is measure isomorphic to $\mathcal{I}$. Note that the unit interval now appears in the modified setting of Proposition 3.18 in two different contexts; however the numerical values of elements in the space now replacing Ω are irrelevant. So, we shall use the symbol $\mathcal{I}$ to denote just this replacement of Ω. For example, we view f as a function from $\mathcal{I}$ to $[0,1]$. In this setting, we want to find $h : [0,1]\times[0,1] \to \mathcal{I}$ satisfying that $f(h(a,x)) = a$.

Let $\Lambda : [0,1] \to \mathcal{B}$ be a disintegration of μ through f, that is, Λ satisfies

$$\Lambda(r)(f^{-1}(r)) = 1 \quad \text{for almost every } r \in [0,1], \text{ and} \tag{32}$$

$$\int_{x\in\mathcal{I}} \Lambda(f(x))(A)d\mu = \mu(A) \quad \text{for every } A \subset \mathcal{I}. \tag{33}$$

Let $\{C_n\}_{n\in\mathbb{N}}$ be an enumeration of all non-degenerate closed intervals with rational endpoints in $\mathcal{I}$.

Let $\mathbb{i} : \Omega \to \{0,1\}^{\mathbb{N}}$ be a map defined on its n-th coordinate by $(\mathbb{i}(\omega))_n \coloneqq \mathbb{1}_{\omega\in C_n}$. Since $\{C_n\}_{n\in\mathbb{N}}$ separates points, we have that $\mathbb{i}$ is injective. Next, we use $\mathbb{i}$ to construct a linear order $\prec^*$ on $\mathcal{I}$ as follows. Suppose that $\omega_1, \omega_2 \in \mathcal{I}$ are two distinct elements. If $f(\omega_i) < f(\omega_{3-i})$ for some $i \in \{1,2\}$ then put $\omega_i \prec^* \omega_{3-i}$. It remains to order ω_1 and ω_2 in case that $f(\omega_1) = f(\omega_2)$. In that case, we order ω_1 and ω_2 in $\prec^*$ in a way that is consistent with the lexicographic ordering of $\mathbb{i}(\omega_1)$ and $\mathbb{i}(\omega_2)$.

Let $B \subset \mathcal{I}$ be the set of elements $\omega \in \mathcal{I}$ that are atoms for the measure $\Lambda(f(\omega))$. We claim that B is a Borel set. Indeed, this follows upon writing

$$B = \{\omega \in \mathcal{I} \mid \Lambda(f(\omega))(\omega) > 0\} \ ,$$

and recalling that Λ is a Borel map.

Let $\mathcal{I}' \coloneqq \mathcal{I} \setminus B$, $f' \coloneqq f \restriction \mathcal{I}'$, and let μ' be the restriction of μ on $\mathcal{I}'$. Also, for $r \in [0,1]$, let $\Lambda'(r)$ be the restriction of $\Lambda(r)$ on $\mathcal{I}'$. We then have

$$\Lambda'(r)(f^{-1}(r)) = 1 - \Lambda(r)(B) \quad \text{for almost every } r \in [0,1], \text{ and} \tag{34}$$

$$\int_{x\in\mathcal{I}'} \Lambda'(f(x))(A)d\mu' = \mu'(A) \quad \text{for every } A \subset \mathcal{I}. \tag{35}$$

For each $\omega \in \mathcal{I}$, write $[\![\omega]\!] \coloneqq \{x \in f^{-1}(f(\omega)) \mid x \preceq^* \omega\}$.

For the time being, it is convenient to work with a slight modification of the Lebesgue measure on $[0,1]^2$. Define a measure γ on $[0,1]^2$ by setting

$$\gamma(A) \coloneqq \int_x \Lambda'(x)[0,1] \ \cdot \ \lambda(\{y \in [0,1] \mid (x,y) \in A\})\mathsf{d}\lambda(x) \ .$$

We are now going to define a key map $m : \mathcal{I}' \to [0,1]^2$. Given $\omega \in \mathcal{I}'$, set

$$m(\omega) \coloneqq \left(\ f(\omega) \ , \ \frac{1}{1-\Lambda(f(\omega))(B)} \cdot \Lambda'(f(\omega))[\![\omega]\!] \ \right) \ .$$

The following claim is obvious.

Claim A.1. *The map m is measure preserving with respect to the measures μ' and γ.*

The following claim is crucial.

Claim A.2. *There exists $\mathcal{I}^* \subset \mathcal{I}'$ with $\mu(\mathcal{I}^*) = \mu(\mathcal{I}')$ so that m is injective on $\mathcal{I}^*$.*

Proof. Let us introduce a piece of notation. Suppose that $d \in \mathbb{N}$ and $\mathbf{p} \in \{0,1\}^d$. Then define

$$\bigwedge_{\mathbf{p}}(C_n)_n \coloneqq \mathcal{I}' \cap \bigcap_{i\in[d]:\mathbf{p}_i=1} C_i \setminus \left(\bigcup_{i\in[d]:\mathbf{p}_i=0} C_i \right) \ .$$

In other words, $\bigwedge_{\mathbf{p}}(C_n)_n$ are the elements of $\mathcal{I}'$ whose encoding using indicator functions of the sets $(C_n)_n$ agrees with $\mathbf{p}$ on the first d coordinates.

For each $t \in [0,1]$ and each $\mathbf{p} \in \bigcup_{d=1}^{\infty}\{0,1\}^d$, let $w(t,\mathbf{p}) \coloneqq f^{-1}(t)\left(\bigwedge_{\mathbf{p}}(C_n)_n\right)$. For a given $\mathbf{p} \in \bigcup_{d=1}^{\infty}\{0,1\}^d$, let $T_{\mathbf{p}} = \{t \in [0,1] \mid w(t,\mathbf{p}) = 0\}$. Observe that $T_{\mathbf{p}}$ is a Borel subset of $[0,1]$.

Let $L \subset \mathcal{I}'$ be defined as follows. Take $\omega \in \mathcal{I}'$. We put ω in L if and only if there exists k such that for each $\ell > k$ either $(\mathbb{i}(\omega))_\ell = 0$, or $w(f(\omega),\mathbf{p}) = 0$, where $\mathbf{p} \in \{0,1\}^\ell$ has the first $\ell - 1$ coordinates as in $\mathbb{i}(\omega)$ and the ℓ-th set to 0. L is a Borel set. Furthermore, observe that for each $t \in [0,1]$, $f^{-1}(t) \cap L$ is at most countable. Since $\Lambda(t)$ is atomless on $\mathcal{I}'$, we conclude that $\Lambda(t)(L) = 0$. Integrating, we get that $\mu(L) = 0$. Set $\mathcal{I}^* \coloneqq \mathcal{I}' \setminus L$. By the above, $\mathcal{I}^*$ has full measure in $\mathcal{I}'$.

We claim that m is injective on $\mathcal{I}^*$. Indeed, suppose that $\omega_1, \omega_2 \in \mathcal{I}^*$ are two distinct elements. If $f(\omega_1) \neq f(\omega_2)$ then clearly $m(\omega_1) \neq m(\omega_2)$. So, it remains to assume that

$f(\omega_1) = f(\omega_2) = t$ for some $t \in [0,1]$. Without loss of generality, assume that $\mathbb{i}(\omega_1)$ is smaller in the lexicographic order than $\mathbb{i}(\omega_2)$, and that the first coordinate on which $\mathbb{i}(\omega_1)$ and $\mathbb{i}(\omega_2)$ differ is, say, h. Since $\omega_2 \notin L$, we know that there exists $\ell > h$ such that the ℓ-th coordinate of $\mathbb{i}(\omega_2)$ is 1, and $w(f(\omega), \mathbf{p}) > 0$, where $\mathbf{p} \in \{0,1\}^\ell$ has the first $\ell-1$ coordinates equal to $\mathbb{i}(\omega_2)$ and the ℓ-th is 0. We have $[\![\omega_1]\!] \subset [\![\omega_2]\!]$ with $[\![\omega_1]\!] \cap f^{-1}(t) \cap \bigwedge_{\mathbf{p}}(C_n)_n = \emptyset$, yet $f^{-1}(t) \cap \bigwedge_{\mathbf{p}}(C_n)_n \subset [\![\omega_2]\!]$. Since $\Lambda(t)\left(\bigwedge_{\mathbf{p}}(C_n)_n\right) = w(t, \mathbf{p}) > 0$, we have that

$$\Lambda(t)\left([\![\omega_2]\!]\right) \geq \Lambda(t)\left([\![\omega_1]\!]\right) + \Lambda(t)\left(\bigwedge_{\mathbf{p}}(C_n)_n\right) > \Lambda(t)\left([\![\omega_1]\!]\right) \ .$$

In particular, we conclude that $m(\omega_2)$ and $m(\omega_1)$ are different on the second coordinate, as was needed. □

Now, let $J \subset [0,1]^2$ be the image of $\mathcal{I}^*$ under m. As m is measure preserving and $\mathcal{I}^*$ is conull, we conclude that J is conull in $[0,1]^2$ with respect to γ. Let $g : [0,1]^2 \to \mathcal{I}^*$ be defined as the inverse map to m on J, and arbitrarily on the rest (which is null).

We are now ready to define the map h. Suppose that $(a,x) \in [0,1]^2$ is given. If $x < 1 - \Lambda(a)(B)$ then define $h(a,x) \coloneqq g(a, \frac{x}{1-\Lambda(a)(B)})$. Otherwise, let $h(a,x)$ be the atom $\zeta \in B \cap f^{-1}(a)$ for which

$$1 - \Lambda(a)(B) + \Lambda(a)\left([\![\zeta]\!]\right) > x > 1 - \Lambda(a)(B) + \Lambda(a)\left([\![\zeta]\!] \setminus \{\zeta\}\right) \ .$$

Such an atom exists for almost every choice of $x \in [0,1]$; the exceptions are at most countably many values for which the down-interval from that atom has exactly that measure.

Obviously, h has all the properties we want. □

Appendix B. Proof of Theorem 11.7

We will derive Theorem 11.7 from the following theorem of Alon and Yuster [3]. Let $H = (V, E)$ be an r-uniform hypergraph. Recall that the *degree* of a vertex v is the number of edges in H that contain v, and that the *codegree* of two distinct vertices u and v is the number of edges in H that contain u and v. We write $\delta(H)$ for the minimum degree, $\Delta(H)$ for the maximum degree, and $\Delta_2(H)$ for the maximum codegree. Let $\mathcal{F} \subseteq 2^V$ and let $0 \leq \alpha \leq 1$. A matching M in H is $(\alpha, \mathcal{F})$-*perfect* if for each $F \in \mathcal{F}$, at least $\alpha|F|$ vertices of F are covered by M. Also we write $s(\mathcal{F}) = \min_{F \in \mathcal{F}}\{|F|\}$ and $g(H) = \Delta(H)/\Delta_2(H)$.

Theorem B.1 (Theorem 1.2 [3])**.** *For an integer $r \geq 2$, a real $C > 1$ and a real $\varepsilon > 0$ there exists a real $\mu = \mu(r, C, \varepsilon)$ and a real $K = K(r, C, \varepsilon)$ so that the following holds: If the r-uniform hypergraph $H = (V, E)$ on N vertices satisfies:*

(1) $\delta(H) \geq (1 - \mu)\Delta(H)$,
(2) $g(H) > \max\{1/\mu, K(\log(N))^6\}$,

then for every $\mathcal{F} \subseteq 2^V$ *with* $|\mathcal{F}| \leq C^{g(H)^{1/(3r-3)}}$ *and with* $s(\mathcal{F}) \geq 5g(H)^{1/(3r-3)} \log(|\mathcal{F}|g(H))$ *there is a* $(1-\varepsilon, \mathcal{F})$*-perfect matching in* H*.*

Proof of Theorem 11.7. Recall that we have for the given constants that

$$1/n \ll c \ll \alpha \ll d \, .$$

Let μ and K be the constants given by Theorem B.1 applied with $r = 3, C = 2$ and $\varepsilon = \alpha$. Therefore we may also assume that

$$1/n \ll c \ll \mu, 1/K \, .$$

We define $H = (V, E)$ to be the 3-uniform hypegraph with vertex set $V = E(G)$ and edge set $E = \{\{\{x,y\},\{y,z\},\{x,z\}\} \subseteq V \mid \{x,y\},\{y,z\},\{x,z\} \in E(G)\}$. In other words, the vertices of H are the edges of G and the hyperedges of H are the triangles of G. Furthermore, for every $x \in V(G)$ we define $F_x \coloneqq \{\{x,y\} \mid y \in N_G(x)\}$, and set $\mathcal{F} \coloneqq \{F_x \mid x \in V(G)\}$. Note that $\mathcal{F} \subseteq 2^V$, $|\mathcal{F}| = 3n$, and that by the (c,h)-typicality of G,

$$s(\mathcal{F}) = \min_{F \in \mathcal{F}}\{|F|\} = 2(1 \pm 2c)dn \, .$$

Furthermore, also by the (c,h)-typicality of G, we have that $|V| = |E(G)| = 3(1 \pm c)dn^2$, $\Delta(H) = (1 \pm c)d^2n$, $\delta(H) = (1 \pm c)d^2n$, and $\Delta_2(H) = 1$. Hence

$$g(H) = \Delta(H)/\Delta_2(H) = (1 \pm c)d^2n/1 = (1 \pm c)d^2n \, .$$

We check the conditions for applying Theorem B.1.

$$\delta(H) \geq (1-c)d^2n \geq (1-\mu)(1+c)d^2n \geq (1-\mu)\Delta(H) \, .$$

Also

$$g(H) \geq (1-c)d^2n \geq \max\{1/\mu, K(\log(3(1 \pm c)dn^2))^6\} \, .$$

Also note that

$$|\mathcal{F}| \leq 3n \leq C^{((1+c)d^2n)^{1/6}} \, ,$$

and

$$s(\mathcal{F}) \geq 2(1-2c)dn \geq n^{1/3} \geq 5((1+c)d^2n)^{1/6} \log(3n(1+c)d^2n) \geq 5g(H)^{1/6} \log(|\mathcal{F}|g(H)) \, .$$

Hence there exists a $(1-\alpha, \mathcal{F})$-perfect matching T for H. Note that elements of T correspond to edge-disjoint triangles in G and that each vertex $x \in V(G)$ lies in at least $(1-\alpha)|F_x|$ triangles of T. Therefore

$$\Delta(G - T) \leq \alpha\Delta(G) \leq \alpha n \, .$$

□

References

[1] E. Aigner-Horev, D. Conlon, H. Hàn, Y. Person, and M. Schacht. Quasirandomness in hypergraphs. *Electron. J. Combin.*, 25(3):Paper No. 3.34, 22, 2018.

[2] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster. The algorithmic aspects of the regularity lemma. *J. Algorithms*, 16(1):80–109, 1994.

[3] N. Alon and R. Yuster. On a hypergraph matching problem. *Graphs Combin.*, 21:377–384, 2005.

[4] O. Ben-Eliezer, E. Fischer, A. Levi, and Y. Yoshida. Ordered graph limits and their applications. arXiv:1811.02023.

[5] O. Ben-Eliezer, E. Fischer, A. Levi, and Y. Yoshida. Ordered graph limits and their applications. In *12th Innovations in Theoretical Computer Science Conference*, volume 185 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 42, 20. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021.

[6] I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 13, 2001.

[7] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao. An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence. *Ann. Probab.*, 46(1):337–396, 2018.

[8] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao. An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Trans. Amer. Math. Soc.*, 372(5):3019–3062, 2019.

[9] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[10] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math. (2)*, 176(1):151–219, 2012.

[11] S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011.

[12] F. R. K. Chung and R. L. Graham. Quasi-random tournaments. *J. Graph Theory*, 15(2):173–198, 1991.

[13] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[14] J. N. Cooper. A permutation regularity lemma. *Electron. J. Combin.*, 13(1):Research Paper 22, 20, 2006.

[15] W. Cooper, J, D. Král', A. Lamaison, and S. Mohr. Quasirandom Latin squares. *Random Structures Algorithms*, 61(2):298–308, 2022.

[16] L. N. Coregliano and A. A. Razborov. Natural quasirandomness properties. *Random Structures Algorithms*, 62(3):624–688, 2023.

[17] Z. Füredi. Extremal hypergraphs and combinatorial geometry. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, pages 1343–1352. Birkhäuser, Basel, 1995.

[18] F. Garbe, R. Hancock, J. Hladký, and M. Sharifzadeh. Theory of limits of sequences of Latin squares. *Acta Math. Univ. Comenian. (N.S.)*, 88(3), 2019.

[19] F. Garbe, J. Hladký, G. Kun, and K. Pekárková. On pattern-avoiding permutons. arXiv:2208.12712.

[20] S. Griffiths. Quasi-random oriented graphs. *J. Graph Theory*, 74(2):198–209, 2013.

[21] A. Grzesik. On the maximum number of five-cycles in a triangle-free graph. *J. Combin. Theory Ser. B*, 102(5):1061–1066, 2012.

[22] H. Hatami, J. Hladký, D. Král, S. Norine, and A. Razborov. On the number of pentagons in triangle-free graphs. *J. Combin. Theory Ser. A*, 120(3):722–732, 2013.

[23] J. Hladký, A. Máthé, V. Patel, and O. Pikhurko. Poset limits can be totally ordered. *Trans. Amer. Math. Soc.*, 367(6):4319–4337, 2015.

[24] C. Hoppen, Y. Kohayakawa, C. G. T. de A. Moreira, B. Ráth, and R. M. Sampaio. Limits of permutation sequences. *J. Comb. Theory, Ser. B*, 103(1):93–113, 2013.

[25] S. Janson. Poset limits and exchangeable random posets. *Combinatorica*, 31(5):529–563, 2011.

[26] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. John Wiley & Sons, 2000.

[27] A. S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[28] P. Keevash. The existence of designs. arXiv:1401.3665.

[29] P. Keevash. The existence of designs II. arXiv:1802.05900.

[30] R. Kenyon, D. Kráľ, C. Radin, and P. Winkler. Permutations with fixed pattern densities. *Random Structures Algorithms*, 56(1):220–250, 2020.

[31] T. Klimošová and D. Kráľ. Hereditary properties of permutations are strongly testable. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1164–1173. ACM, New York, 2014.

[32] D. Král' and O. Pikhurko. Quasirandom permutations are characterized by 4-point densities. *Geom. Funct. Anal.*, 23, 2013.

[33] D. Kunszenti-Kovács, L. Lovász, and B. Szegedy. Multigraph limits, unbounded kernels, and Banach space decorated graphs. *J. Funct. Anal.*, 282(2):Paper No. 109284, 44, 2022.

[34] M. Kwan and B. Sudakov. Intercalates and discrepancy in random latin squares. *Random Structures Algorithms*, 52(2):181–196, 2018.

[35] B. Lidický and F. Pfender. Pentagons in triangle-free graphs. *European J. Combin.*, 74:85–89, 2018.

[36] N. Linial and Z. Luria. An upper bound on the number of high-dimensional permutations. *Combinatorica*, 34(4):471–486, 2014.

[37] L. Lovász. Subgraph densities in signed graphons and the local Simonovits-Sidorenko conjecture. *Electron. J. Combin.*, 18(1):Paper 127, 21, 2011.

[38] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[39] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[40] L. Lovász and B. Szegedy. Regularity partitions and the topology of graphons. In *An irregular mind*, volume 21 of *Bolyai Soc. Math. Stud.*, pages 415–446. János Bolyai Math. Soc., Budapest, 2010.

[41] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

[42] Yu V Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory of Probability & Its Applications*, 1(2):157–214, 1956.

[43] A. A. Razborov. Flag algebras. *J. Symbolic Logic*, 72(4):1239–1282, 2007.

[44] A. A. Razborov. On the minimal density of triangles in graphs. *Combin. Probab. Comput.*, 17(4):603–618, 2008.

[45] V. Rödl. On a packing and covering problem. *Europ. J. Combinatorics*, 5(1):69–78, 1985.

[46] V. Rödl. On universality of graphs with uniformly distributed edges. *Discrete Math.*, 59(1-2):125–134, 1986.

[47] I. Z. Ruzsa and E. Szemerédi. Triple systems with no six points carrying three triangles. In *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II*, volume 18 of *Colloq. Math. Soc. János Bolyai*, pages 939–945. North-Holland, Amsterdam-New York, 1978.

[48] E. Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

[49] A. Thomason. Pseudorandom graphs. In *Random graphs '85 (Poznań, 1985)*, volume 144 of *North-Holland Math. Stud.*, pages 307–331. North-Holland, Amsterdam, 1987.

[50] H. Towsner. σ-algebras for quasirandom hypergraphs. *Random Structures Algorithms*, 50(1):114–139, 2017.

[51] J. H. van Lint and R. M. Wilson. *A course in combinatorics*. Cambridge University Press, Cambridge, second edition, 2001.

[52] Y. Zhao. Hypergraph limits: A regularity approach. *Random Structures Algorithms*, 47(2):205–226, 2015.

Garbe: Faculty of Informatics, Masaryk University, Brno, Czechia. *Hancock:* Institut für Informatik, University of Heidelberg, Im Neuenheimer Feld 205, 69120, Heidelberg, Germany. *Hladký:* Institute of Computer Science, Czech Academy of Sciences. Pod Vodárenskou věží 2, 182 07, Prague, Czechia. *Sharifzadeh:* Department of Mathematics and Mathematical Statistics, Umeå University, 901 87, Umeå, Sweden.

POSET LIMITS CAN BE TOTALLY ORDERED

JAN HLADKÝ, ANDRÁS MÁTHÉ, VIRESH PATEL, AND OLEG PIKHURKO

Abstract. S. Janson [*Poset limits and and exchangeable random posets*, Combinatorica **31** (2011), 529–563] defined limits of finite posets in parallel to the emerging theory of limits of dense graphs.

We prove that each poset limit can be represented as a kernel on the unit interval with the standard order, thus answering an open question of Janson. We provide two proofs: real-analytic and combinatorial. The combinatorial proof is based on a Szemerédi-type Regularity Lemma for posets which may be of independent interest.

1. Introduction

Given a class $\mathcal{C}$ of finite structures and some measure $t(F,G)$ for $F,G\in\mathcal{C}$ of how frequently F appears in G as a substructure, one can say that a sequence $\{G_n\}_{n\in\mathbb{N}}$ *converges* if $\{t(F,G_n)\}_{n\in\mathbb{N}}$ converges for every $F\in\mathcal{C}$.

For example, if $\mathcal{C}$ consists of finite graphs and t denotes the subgraph density, then we obtain the convergence of (dense) graphs whose systematic study was initiated by Lovász and Szegedy [13] and Borgs et al [4]. In particular, Lovász and Szegedy [13] showed that for every convergent sequence $\{G_n\}_{n\in\mathbb{N}}$ of graphs there is a measurable function $W:[0,1]^2\to[0,1]$ (called *graphon*) such that for every graph F the limit of $t(F,G_n)$ as $n\to\infty$ is a certain integral involving W. In fact, many other parameters of G_n can be well approximated as $n\to\infty$ if we know W. This opens a general way of bringing analytic methods into the study of large finite graphs. Connections to other areas are established by alternative representations of "graph limits": by reflection positive graph parameters (Lovász and Szegedy [13]), by positive flag algebra homomorphisms (Razborov [17]), and by partially exchangeable random arrays (Diaconis and Janson [6]).

The theory of graph limits has received a great deal of attention and has been extended to other structures as well, such as hypergraphs (Elek and Szegedy [8], see also Tao [21] and Austin [1]), permutations (Hoppen et al [9, 10]), functions on compact Abelian groups (Szegedy [19]), and other.

2010 *Mathematics Subject Classification.* 06A06, 28Axx.

Key words and phrases. Homomorphism density, ordered probability space, partially ordered set, poset kernel, Regularity Lemma.

J.H. was supported by an EPSRC fellowship.

A.M. was supported by EPSRC grant EP/G050678/1.

O.P. was supported by the European Research Council (grant agreement no. 306493) and the National Science Foundation of the USA (grant DMS-1100215).

An analogous theory for limits of *posets* (i.e. partially ordered sets) was initiated by Brightwell and Georgiou [5] and further developed by Janson [11]. Let us state some of these results.

We represent a poset as a pair $(P, \prec)$ where P is a finite *ground set* and $\prec$ is a *strict order relation* (i.e. it is transitive and no $a, b \in P$ satisfy $a \prec b$ and $b \prec a$ simultaneously).

A map $f : P \to Q$ (not necessarily injective) is a *homomorphism* from $(P, \prec)$ to $(Q, \ll)$ if we have $f(x) \ll f(y)$ for every $x, y \in P$ with $x \prec y$. The *density* $t(\,(P, \prec), (Q, \ll)\,)$ is the number of homomorphisms from $(P, \prec)$ to $(Q, \ll)$ divided by the total number of possible maps $P \to Q$. In other words, it is the probability that a random map $P \to Q$ between the ground sets preserves the order relation.

Definition 1.1. A sequence of posets $\{(P_n, \prec_n)\}_{n\in\mathbb{N}}$ *converges* if $|P_n| \to \infty$ and

$$\big\{\, t(\,(P, \prec), (P_n, \prec_n)\,)\,\big\}_{n\in\mathbb{N}} \text{ converges for every poset } (P, \prec). \tag{1}$$

Remark 1.2. It is not hard to show (cf [13, Section 2.4]) that Definition 1.1 does not change if we modify t to be the density of induced and/or injective homomorphisms.

The potential usefulness of (1) comes from the result of Janson [11, Theorem 1.7] that for each convergent sequence there is an analytic limit object as follows. (See Section 2 for an overview of the measure theory notation that we use.)

Definition 1.3. An *ordered probability space* $(S, \mathcal{F}, \mu, \lhd)$ is a probability space $(S, \mathcal{F}, \mu)$ equipped with a strict order relation $\lhd$ such that $\{(x, y) : x \lhd y\}$ is an $\mathcal{F} \otimes \mathcal{F}$-measurable subset of $S \times S$.

Definition 1.4. A *(poset) kernel* is a 5-tuple $(S, \mathcal{F}, \mu, \lhd, W)$, where $(S, \mathcal{F}, \mu, \lhd)$ is an ordered probability space and W is an $\mathcal{F} \otimes \mathcal{F}$-measurable function $S \times S \to [0, 1]$ such that, for all $x, y, z \in S$,

$$W(x, y) > 0 \quad \Rightarrow \quad x \lhd y, \tag{2}$$

$$W(x, y) > 0 \text{ and } W(y, z) > 0 \quad \Rightarrow \quad W(x, z) = 1. \tag{3}$$

In particular, it follows from Definition 1.4 that $W(x, y)W(y, x) = 0$ for every $x, y \in S$.

When no confusion arises, we may abbreviate $(P, \prec)$ to P and $(S, \mathcal{F}, \mu, \lhd, W)$ to W. Also, we will usually say "kernel" instead of "poset kernel".

Definition 1.5. The *density* of a poset $(P, \prec)$ in a kernel $(S, \mathcal{F}, \mu, \lhd, W)$ is

$$t(P, W) := \int_{S^{|P|}} \prod_{\substack{a,b\in P\\ a\prec b}} W(x_a, x_b) \prod_{a\in P} \mathrm{d}\mu(x_a). \tag{4}$$

There is some analogy between $t(P, Q)$ and $t(P, W)$. Namely, one can interpret the expression in the right-hand side of (4) as follows. Select random elements x_a of $(S, \mathcal{F}, \mu)$

indexed by P and let $x_a \ll x_b$ with probability $W(x_a, x_b)$, with all choices being mutually independent. Then $t(P, W)$ is exactly the probability that $x_a \ll x_b$ for all $a \prec b$ in P. In fact, the connection is much deeper as the following result shows.

Theorem 1.6 (Janson [11, Theorems 1.7 and 1.9(ii)])**.** *For every convergent sequence $\{P_n\}_{n\in\mathbb{N}}$ of posets there is a kernel $(S, \mathcal{F}, \mu, \lhd, U)$ such that*

$$t(P, U) = \lim_{n\to\infty} t(P, P_n), \quad \text{for every poset } P. \tag{5}$$

Moreover, we can assume in (5) that

$$(S, \mathcal{F}, \mu) = ([0, 1], \mathcal{B}, \lambda) \tag{6}$$

is the unit interval with the Lebesgue measure λ on the Borel σ-algebra $\mathcal{B}$.

Also, the converse of Theorem 1.6 was established in [11]: for every kernel U there is a sequence of posets $\{P_n\}_{n\in\mathbb{N}}$ that satisfies (5). In fact, the sampling procedure informally described after (4) yields such a sequence with probability 1.

Although we can require that (6) holds, the proof in [11] gives no control over the order relation $\lhd$. This prompted Janson [11, Problem 1.10] to ask if one can always take $([0, 1], \mathcal{B}, \lambda, <)$ with the standard order $<$ in Theorem 1.6. In a later paper [12], Janson answered his question for convergent sequences of interval orders (see also [11, Theorem 1.9(iii)] for a related result). Here we give the affirmative answer in the general case.

Theorem 1.7. *For every convergent sequence $\{P_n\}_{n\in\mathbb{N}}$ of posets there is a kernel $([0, 1], \mathcal{B}, \lambda, <, U)$ such that (5) holds.*

In fact, we provide two different proofs of Theorem 1.7.

One goes via a Regularity Lemma for posets that we prove in Section 6. Our lemma finds a partition $P = V_1 \cup \cdots \cup V_k$ which is ε-regular with respect to the underlying graph of $(P, \prec)$ and has the additional property that all $\prec$-relations between parts go "forward" only. Having such a Regularity Lemma, we follow the method of Lovász and Szegedy [13, 14] to construct a kernel U by taking the "limit" of ε-regular partitions as $\varepsilon \to \infty$. The above "forward" property allows us to ensure that $U(x, y) = 0$ whenever $x \geq y$, thus proving Theorem 1.7. We expect our Regularity Lemma to have further applications.

The other proof of Theorem 1.7 is real-analytic. Actually, we prove a somewhat stronger result (Theorem 1.9 below). In order to state it, we have to give some further definitions.

Let $(S, \mathcal{F}, \mu, \lhd, W)$ be a kernel. We call it *strict* if $W(x, y) > 0$ for every $x, y \in S$ with $x \lhd y$. (Thus a kernel is strict if the two sides of (2) are equivalent.) Kernel axioms imply that if define $\lhd' := \{(x, y) \in S^2 : W(x, y) > 0\}$ then $(S, \mathcal{F}, \mu, \lhd')$ is an ordered probability

space on which W is a strict kernel. Clearly, this change does not affect (4). Thus, we can addionally assume in Theorem 1.6 that U is strict, see [11, Remark 1.2].

Definition 1.8. An *inclusion* between ordered probability spaces $(S, \mathcal{F}, \mu, \lhd)$ and $(S', \mathcal{F}', \mu', \lhd')$ is a measure preserving function $f : S \to S'$ such that the set

$$\big\{ (x, y) \in S^2 : x \lhd y,\ f(x) \not\lhd' f(y) \big\} \tag{7}$$

has $\mu \otimes \mu$-measure zero. Additionally, if we have a kernel U on $(S', \mathcal{F}', \mu', \lhd')$ then its *pull-back along* f is the function $U^f : S^2 \to [0, 1]$, defined by $U^f(x, y) := U(f(x), f(y))$ for $x, y \in S$.

Theorem 1.9. *For every strict kernel $(S, \mathcal{F}, \mu, \lhd, W)$ such that $(S, \mathcal{F}, \mu)$ is atomless, there is a kernel $([0, 1], \mathcal{B}, \lambda, <, U)$ and an inclusion $f : (S, \mathcal{F}, \mu, \lhd) \to ([0, 1], \mathcal{B}, \lambda, <)$ such that W is equal to the pull-back U^f almost everywhere.*

Since f in Theorem 1.9 is measure preserving, we necessarily have that $t(P, U) = t(P, W)$ for every poset P, that is, U and W represent the same poset limit. Thus Theorem 1.7 follows from Theorems 1.6 and 1.9.

The notion of a pull-back plays an important role in the theory of graphons. Hopefully, our Theorem 1.9 will be generally useful when studying poset kernels. For example, if the studied kernel property behaves well with respect to taking pull-backs, then one can operate with the function U that satisfies Theorem 1.9 instead of the 5-tuple $(S, \mathcal{F}, \mu, \lhd, W)$.

Given an ordered probability space $(S, \mathcal{F}, \mu, \lhd)$, the indicator function $I_\lhd$ of the order relation $\lhd$ is clearly a strict kernel on it. Thus Theorem 1.9 has the following direct corollary.

Theorem 1.10. *Every atomless ordered probability space $(S, \mathcal{F}, \mu, \lhd)$ can be included into $([0, 1], \mathcal{B}, \lambda, <)$.* □

Theorem 1.10 can be viewed as a measure theoretic analogue of the statement that every poset can be totally ordered. While extending this to infinite partially ordered sets is an easy application of Zorn's lemma, the main content of Theorem 1.10 is that this total ordering can be done in a "measurable" way. Interestingly, the limit of totally ordered increasing posets happens to be our universal target space $([0, 1], \mathcal{B}, \lambda, <)$ with the indicator function $I_<$ as its kernel.

This paper is organised as follows. Section 2 describes the measure theory notation that we frequently use. Section 3 presents some auxiliary analytic lemmas, thus making the flow of argument in the later sections smoother. Although Theorem 1.10 is a direct consequence of Theorem 1.9, we prove it first in Section 4. Then, in Section 5, we show how

Theorem 4.1 (a version of Theorem 1.10) implies Theorem 1.9. Our Regularity Lemma for posets is stated and proved in (combinatorial) Section 6 which can be read independently of the other sections. We show how this Regularity Lemma gives an alternative proof of Theorem 1.7 in Section 7. Finally, Section 8 contains some concluding remarks, including examples that certain strengthenings of our results are not possible.

2. Measure theory notation

Let us give some notation that we are going to use frequently. We do not define many standard concepts of measure theory but refer the reader to Bogachev's book [2] whose notation we generally follow. We try to provide sufficient references and explanations so that this paper is accessible to combinatorialists who do not have a strong background in measure theory.

Let $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{R}$ be the sets of respectively natural and real numbers. When we consider a subset of $\mathbb{R}$, typically the unit interval $[0, 1]$, we will denote the σ-algebra of its Borel subsets by $\mathcal{B}$ and the Lebesgue measure by λ. For a family $\mathcal{X}$ of sets, let $\sigma(\mathcal{X})$ denote the σ-algebra generated by $\mathcal{X}$. Let I_X denote the indicator function of a set X (that is, $I_X(x)$ is 1 if $x \in X$ and 0 otherwise).

Let $(S, \mathcal{F}, \mu)$ be a probability space. As it is standard in measure theory, a real-valued function f on S is called *$\mathcal{F}$-measurable* if it is $(\mathcal{F}, \mathcal{B})$-measurable. We denote by $\mathcal{F}_\mu$ the completion of $\mathcal{F}$ with respect to the measure μ.

We say that a property holds *$(\mathcal{F}, \mu)$-almost everywhere* (and abbreviate this to *$(\mathcal{F}, \mu)$-a.e.*) if the set of points of S where it fails belongs to $\mathcal{F}_\mu$ and has μ-measure zero. When the underlying measure space is understood, we just write "a.e." In some rare cases when we consider more than one σ-algebra on the same set, the bare term "a.e." refers to the largest σ-algebra.

We call two sets or two functions *equivalent* (and use the symbol $\sim$) if they coincide a.e. The *Fréchet-Nikodym distance* between two sets $A, B \in \mathcal{F}_\mu$ is $\mu(A \triangle B)$; it is in general a pseudo-metric (it satisfies the Triangle Inequality but may evaluate to 0 for $A \neq B$). The space $(S, \mathcal{F}, \mu)$ is called *separable* if $\mathcal{F}$ has a countable subset which is dense with respect to the Fréchet-Nikodym distance.

Let $\mathcal{A} \subseteq \mathcal{F}$ be another σ-algebra and let f be an integrable real-valued function on $(S, \mathcal{F}, \mu)$. The *conditional expectation* $\mathbb{E}(f|\mathcal{A})$ is the set of all $\mathcal{A}$-measurable functions $g : S \to \mathbb{R}$ such that for every bounded $\mathcal{A}$-measurable function $h : S \to \mathbb{R}$ we have

$$\int h(x)g(x)\,\mathrm{d}\mu(x) = \int h(x)f(x)\,\mathrm{d}\mu(x). \tag{8}$$

As it is well-known, every two functions in $\mathbb{E}(f|\mathcal{A})$ are equivalent and it is enough to check (8) for $\{0,1\}$-valued h only, i.e. that

$$\int_A g(x)\,\mathrm{d}\mu(x) = \int_A f(x)\,\mathrm{d}\mu(x) \qquad \text{for all } A \in \mathcal{A}. \tag{9}$$

We refer the reader to [2, Section 10.1.1] for some basic properties of conditional expectation. We may treat $\mathbb{E}(f|\mathcal{A})$ as a single function (rather as a set of functions), when the studied property does not depend on the choice of a representative.

Let $(S', \mathcal{F}', \mu')$ be another probability space. A map $f : S \to S'$ is *measure preserving* if f is $(\mathcal{F}, \mathcal{F}')$-measurable and, for every $A \in \mathcal{F}'$, we have $\mu(f^{-1}(A)) = \mu'(A)$. The products of σ-algebras and measures are denoted by $\mathcal{F} \otimes \mathcal{F}'$ and $\mu \otimes \mu'$. We use the shorthand $\mathcal{F} \bar{\otimes} \mathcal{F}'$ for $(\mathcal{F} \otimes \mathcal{F}')_{\mu\otimes\mu'}$, the completion of $\mathcal{F} \otimes \mathcal{F}'$ with respect to $\mu \otimes \mu'$. We will be using Fubini's theorem ([2, Theorem 3.4.4]) very frequently, often without explicitly mentioning it. Let us stress that one has to be careful when dealing with products of σ-algebras and measures. For example, the product of two complete measure spaces is not complete in general. Also, see Exercises 44–45, 49–51, and 55 in [2, Section 3.10] for counterexamples to some "plausible" statements related to Fubini's theorem.

Let us give some kernel-specific definitions (when the underlying ordered probability space $(S, \mathcal{F}, \mu, \lhd)$ is understood). For $A \subseteq S$, let $A^c := S \setminus A$ denote the complement of A. For $X \in \mathcal{F} \otimes \mathcal{F}$, we define

$$\mu_\lhd(X) := \int_X I_\lhd(a,b)\,\mathrm{d}\mu(a)\,\mathrm{d}\mu(b). \tag{10}$$

For a 2-variable function $W : S^2 \to \mathbb{R}$ and $y \in S$, the *slice function* $W_y : S \to \mathbb{R}$ is defined by $W_y(x) := W(x,y)$. We call $W : S \times S \to [0,1]$ an *almost (poset) kernel* if W is $\mathcal{F} \bar{\otimes} \mathcal{F}$-measurable and the kernel axioms (2) and (3) hold for a.e. triple (x,y,z) in $(S, \mathcal{F}, \mu)^3$.

Although our Theorem 1.9 takes a kernel W as input and then produces another kernel U, we have to deal with almost kernels at intermediate stages of the proof. (For example, the pull-back U^f in Theorem 1.9 is generally an almost kernel.)

3. Auxiliary analytic lemmas

Here we present some auxiliary results that we will need later.

Janson [11] proved that one can transform an almost kernel $(S, \mathcal{B}, \lambda, \lhd, W)$ with $S \subseteq \mathbb{R}$ into a kernel $(S, \mathcal{B}, \lambda, \lhd', W')$ with $W' \sim W$ and some $\lhd'$. We show that in the special case of the unit interval with the standard order, one can also keep the order relation intact.

Lemma 3.1. *Let $([0,1], \mathcal{B}, \lambda, <, U)$ be an almost kernel. Then there is $U' \sim U$ such that $([0,1], \mathcal{B}, \lambda, <, U')$ is a kernel.*

Proof. First, we choose a $\mathcal{B} \otimes \mathcal{B}$-measurable function $U_0 \sim U$; it exists by [2, Proposition 2.1.11]. Then we proceed in a similar fashion as is done by Janson [11, Pages 547–548], so we will be rather brief. We refer the reader to [2, Section 5] for the definitions and basic properties of Lebesgue and density points.

We define $U_1 \colon [0,1]^2 \to [0,1]$ by $U_1(x,y) := U_0(x,y)$ if $(x,y) \in [0,1]^2$ is a Lebesgue point of U_0. Next, if (x,y) is a density point of the set $\{(x,y) : U_0(x,y) = 1\}$, then let $U_1(x,y) := 1$. (Recall that a density point need not belong to the set itself.) For all other pairs $(x,y) \in [0,1]^2$, we define $U_1(x,y) := 0$. Note that $U_1 \sim U_0$ and U_1 is still Borel.

We claim that U_1 is a kernel on $([0,1], \mathcal{B}, \lambda, <)$. Suppose that $U_1(x,y) > 0$. Then for every sufficiently small $\varepsilon > 0$, we have $U_0(x',y') > 0$ for most points $(x',y') \in (x \pm \varepsilon) \times (y \pm \varepsilon)$, where $(x \pm \varepsilon)$ denotes the intersection of the open interval $(x-\varepsilon, x+\varepsilon)$ with $[0,1]$. In particular, $x' < y'$ for most of these pairs and therefore $x < y$.

Now suppose that $U_1(x,y) > 0$ and $U_1(y,z) > 0$. Then for every sufficiently small $\varepsilon > 0$, we have $U_0(x',y') > 0$ for most points $(x',y') \in (x \pm \varepsilon) \times (y \pm \varepsilon)$ and $U_0(y',z') > 0$ for most points $(y',z') \in (y \pm \varepsilon) \times (z \pm \varepsilon)$. This implies that we have $U_0(x',z') = 1$ for most points $(x',z') \in (x \pm \varepsilon) \times (z \pm \varepsilon)$. Thus (x,z) is a density point of $\{(x,y) : U_0(x,y) = 1\}$ and therefore $U_1(x,z) = 1$, as required. □

Lemma 3.2 (Borgs, Chayes, and Lovász [3, Lemma 3.4])**.** *Let $(S, \mathcal{F})$ and $(S', \mathcal{F}')$ be measurable spaces, and let $W : S \times S' \to \mathbb{R}$ be a bounded $\mathcal{F} \otimes \mathcal{F}'$-measurable function. Then there exist countably generated σ-algebras $\mathcal{F}_0 \subseteq \mathcal{F}$ and $\mathcal{F}'_0 \subseteq \mathcal{F}'$ such that W is $\mathcal{F}_0 \otimes \mathcal{F}'_0$-measurable.* □

Lemma 3.3. *Let $f : S \to S'$ be an inclusion of ordered probability spaces $(S, \mathcal{F}, \mu, \lhd)$ and $(S', \mathcal{F}', \mu', \lhd')$. Let W be a kernel on S such that $W \sim \mathbb{E}(W | \mathcal{A} \otimes \mathcal{A})$, where $\mathcal{A} := f^{-1}(\mathcal{F}')$. Then there is an almost kernel U on $(S', \mathcal{F}', \mu', \lhd')$ with $W \sim U^f$.*

Proof. We construct U following the argument of Borgs, Chayes, and Lovász [3, Lemma 3.1] rather closely.

Define a function ν from $\mathcal{F}' \times \mathcal{F}' := \{A \times B \mid A, B \in \mathcal{F}'\}$ to $\mathbb{R}$ by

$$\nu(A \times B) := \int_{f^{-1}(A) \times f^{-1}(B)} W(x,y) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y), \qquad A, B \in \mathcal{F}'.$$

This is a countably additive function on the semialgebra $\mathcal{F}' \times \mathcal{F}'$ which uniquely extends to a measure ν on $\mathcal{F}' \otimes \mathcal{F}'$ by [2, Proposition 1.3.10]. It is easy to see that this measure ν is absolutely continuous with respect to $\mu' \otimes \mu'$. Hence, the Radon-Nikodym derivative

$$U := \frac{\mathrm{d}\nu}{\mathrm{d}(\mu' \otimes \mu')}$$

exists ([2, Theorem 3.2.2]). Namely, U is a $\mu' \otimes \mu'$-integrable function such that for every $X \in \mathcal{F}' \otimes \mathcal{F}'$ we have $\nu(X) = \int_X U \, \mathrm{d}(\mu' \otimes \mu')$.

The last identity implies (given that f is measure preserving and that $0 \le W \le 1$) that the set

$$\{(x,y) \in S' \times S' : U(x,y) > 1 \text{ or } U(x,y) < 0\}$$

has measure zero. By changing U on a null set, we can assume that U is $\mathcal{F}' \otimes \mathcal{F}'$-measurable (see [2, Proposition 2.1.11]) and that the values of U belong to $[0,1]$. In particular, the pull-back U^f is $\mathcal{A} \otimes \mathcal{A}$-measurable. Moreover, for any $Y \in \mathcal{A} \otimes \mathcal{A}$, say $Y = (f \times f)^{-1}(X)$, we have that

$$\int_Y U^f \,\mathrm{d}(\mu \otimes \mu) = \int_X U \,\mathrm{d}(\mu' \otimes \mu') = \int_Y W \,\mathrm{d}(\mu \otimes \mu). \tag{11}$$

By (9), we conclude that $U^f \in \mathbb{E}(W|\mathcal{A} \otimes \mathcal{A})$. Thus U^f is a.e. equal to W by the assumption of the lemma.

Let us verify that U is an almost kernel. First, consider the set

$$X := \{(x,y) \in S' \times S' : x \not\lhd' y,\ U(x,y) > 0\} \in \mathcal{F}' \otimes \mathcal{F}'$$

of points where the first kernel axiom (2) fails for U. By (11), the integral of U over X is the same as the integral of W over $Y := (f \times f)^{-1}(X)$. Since f is an inclusion, we have $\mu_\lhd(Y) = 0$, where $\mu_\lhd$ is defined by (10). Since W is a kernel, it is zero a.e. on Y. It follows that X has measure zero, that is, U satisfies (2) a.e.

Define $u(x,y,z) := U(x,y)\,U(y,z)\,(1 - U(x,z))$. Since $U^f \sim W$ and W is a kernel, we have $u^f \sim 0$. Since f is measure-preserving, we have $\int u = \int u^f = 0$. The non-negativity of u implies that $u \sim 0$, that is, U satisfies (3) a.e. □

Remark 3.4. The conditional expectation of a kernel need not be an almost kernel. For example, let $S := \{a,b,c,d\}$ with $\mathcal{F} := 2^S$ and μ being the uniform measure. Let $\mathcal{A} \subseteq \mathcal{F}$ be obtained by "gluing" b and c together. Let $a \lhd b$ and $c \lhd d$ be all order relations and let $W := I_\lhd$. Then any $U \in \mathbb{E}(W|\mathcal{A} \otimes \mathcal{A})$ satisfies $U(a,b) = 1/2$, $U(b,d) = 1/2$, and $U(a,d) = 0$ and cannot be an almost kernel. Also, pull-backs do not preserve (almost) kernels in general: for example, consider the pull-back of $I_<$ with respect to the identity inclusion of $([0,1], \mathcal{B}, \lambda, \emptyset)$ into $([0,1], \mathcal{B}, \lambda, <)$.

Lemma 3.5. *Let $(S, \mathcal{F}, \mu)$ be a probability space. Let $\mathcal{A} \subseteq \mathcal{F}$ be another σ-algebra such that $(S, \mathcal{A}, \mu)$ is separable. Let $W : S^2 \to \mathbb{R}$ be a bounded $\mathcal{F} \otimes \mathcal{F}$-measurable function. Let $g \in \mathbb{E}(W|\mathcal{A} \otimes \mathcal{F})$. Then, for a.e. $y \in S$, we have that $g_y \sim \mathbb{E}(W_y|\mathcal{A})$.*

Proof. By definition, g is $\mathcal{A} \otimes \mathcal{F}$-measurable. It follows by [2, Proposition 3.3.2] that the slice function g_y is $\mathcal{A}$-measurable for every $y \in S$.

Fix $A \in \mathcal{A}$. By the definition of conditional expectation, we have that $\int_{A \times B} g = \int_{A \times B} W$ for every $B \in \mathcal{F}$. Likewise,

$$\int_A \mathbb{E}(W_y|\mathcal{A}) = \int_A W_y, \quad \text{for every } y \in S. \tag{12}$$

By Fubini's theorem, the latter function is integrable as a function of y. Moreover,

$$\int_B \left(\int_A W_y(x)\,\mathrm{d}\mu(x) \right) \mathrm{d}\mu(y) = \int_{A\times B} W = \int_{A\times B} g = \int_B \left(\int_A g_y(x)\,\mathrm{d}\mu(x) \right) \mathrm{d}\mu(y).$$

Since $B \in \mathcal{F}$ was arbitrary, [2, Corollary 2.5.4] gives that $\int_A W_y = \int_A g_y$ for a.e. y. Let us remove all exceptional points y when A runs over a dense countable subset $\{A_1, A_2, \dots\} \subseteq \mathcal{A}$ in $(S, \mathcal{A}, \mu)$ as well as those y for which $\|g_y\|_\infty > \|W\|_\infty$ or $\|W_y\|_\infty > \|W\|_\infty$. It is easy to see that the remaining set Y has measure 1.

Fix any $y \in Y$. For *every* $A \in \mathcal{A}$ we have that

$$\left| \int_A W_y - \int_A g_y \right| \le \left| \int_A W_y - \int_{A_i} W_y \right| + \left| \int_A g_y - \int_{A_i} g_y \right| \le 4\,\|W\|_\infty\, \mu(A \triangle A_i).$$

Since the right-hand side can be made arbitrarily small by choosing a suitable A_i, we conclude that $\int_A W_y = \int_A g_y$. Since $A \in \mathcal{A}$ was arbitrary and both $\mathbb{E}(W_y|\mathcal{A})$ and g_y are $\mathcal{A}$-measurable, they coincide a.e. by (12). The lemma is proved as $\mu(S \setminus Y) = 0$. □

4. Proof of Theorem 1.10

Let $(S, \mathcal{F}, \mu, \lhd)$ be given. Lemma 3.2, when applied to the indicator function $I_\lhd$, returns two countably generated σ-algebras $\mathcal{F}_0, \mathcal{F}_0' \subseteq \mathcal{F}$. Let $\mathcal{F}' := \sigma(\mathcal{F}_0 \cup \mathcal{F}_0')$ be the σ-algebra on S generated by $\mathcal{F}_0 \cup \mathcal{F}_0'$. By enlarging a set of generators of $\mathcal{F}'$ by adding a countably many elements of $\mathcal{F}$, we can additionally make $(S, \mathcal{F}', \mu)$ atomless.

Clearly, if we prove Theorem 1.10 for this new space $(S, \mathcal{F}', \mu, \lhd)$, then the same inclusion f will work for the original one (as $\mathcal{F}' \subseteq \mathcal{F}$). Thus, without loss of generality, let us assume that $\mathcal{F}$ is countably generated. It easily follows (see e.g. Exercise 1.12.102 and its hint in [2]) that $(S, \mathcal{F}, \mu)$ is separable. Thus it is enough to prove the following theorem (whose last claim will be needed later in Section 5).

Theorem 4.1. *Let $(S, \mathcal{F}, \mu, \lhd)$ be an ordered probability space such that $(S, \mathcal{F}, \mu)$ is atomless and separable. Then there is an inclusion $f : (S, \mathcal{F}, \mu, \lhd) \to ([0,1], \mathcal{B}, \lambda, <)$ such that every set $A \in \mathcal{F}$ with $\mu_\lhd(A \times A^c) = 0$ belongs to $(f^{-1}(\mathcal{B}))_\mu$, the completion of $f^{-1}(\mathcal{B})$ with respect to the measure μ.*

So we prove Theorem 4.1 now.

Claim 1. Let $B \in \mathcal{F}$ with $\mu(B) > 0$. Then there exists $A \in \mathcal{F}$ such that $\mu_\lhd(A \times A^c) = 0$, $\mu(B \cap A) > 0$, and $\mu(B \cap A^c) > 0$.

Proof of Claim. Let $\rhd_x := \{y \in S : y \rhd x\} \in \mathcal{F}$ be the *strict upper shadow* of $x \in S$ and let

$$B' := \{x \in B : \mu(B \cap \rhd_x) > 0\}.$$

First, suppose that $\mu(B') > 0$. Clearly, $\mu_\lhd(B' \times B') \le \mu(B')^2/2$. By Fubini's theorem, there is $x \in B'$ with $\mu(B' \cap \rhd_x) \le \mu(B')/2$. Clearly, $A := \rhd_x$ has the required properties.

If $\mu(B') = 0$, then $\mu_\lhd(B \times B) = 0$ by Fubini's theorem. Since $\mathcal{F}$ is atomless, it contains $A' \subseteq B$ with $0 < \mu(A' \cap B) < \mu(B)$. The function $a(x) := \mu\big(\{y \in A' : y \lhd x\}\big)$ is $\mathcal{F}$-measurable by [2, Corollary 3.3.3]. The set $X := \{x \in S : a(x) > 0\} \in \mathcal{F}$ is clearly up-closed with respect to $\lhd$ and it intersects B in a set of measure 0 by Fubini's theorem. It is easy to see that $A := A' \cup X$ satisfies the claim. □

Let

$$\mathcal{T} := \{A \in \mathcal{F} : 0 < \mu(A) < 1,\, \mu_\lhd(A \times A^c) = 0\}.$$

By Claim 1, $\mathcal{T}$ is non-empty and, moreover, infinite. Since $(S, \mathcal{F}, \mu)$ is separable, we can choose a countable subset $\{A_1, A_2, \dots\} \subseteq \mathcal{T}$ which is dense in $\mathcal{T}$ with respect to the Fréchet-Nikodym distance.

We define f so that it satisfies the following properties:

$$\begin{array}{rcl} f(A_1^c) & \subseteq & [0,\, \mu(A_1^c)], \\ f(A_1) & \subseteq & [\mu(A_1^c),\, 1], \\ f(A_1^c \cap A_2^c) & \subseteq & [0,\, \mu(A_1^c \cap A_2^c)], \\ f(A_1^c \cap A_2) & \subseteq & [\mu(A_1^c \cap A_2^c),\, \mu(A_1^c)], \\ f(A_1 \cap A_2^c) & \subseteq & [\mu(A_1^c),\, \mu(A_1^c) + \mu(A_1 \cap A_2^c)], \\ f(A_1 \cap A_2) & \subseteq & [\mu(A_1^c) + \mu(A_1 \cap A_2^c),\, 1], \end{array} \tag{13}$$

and so on. Specifically, for a (finite or infinite) binary sequence $\mathbf{b} = (b_1, b_2, \dots)$, let

$$\begin{aligned} A_{\mathbf{b}} &:= \bigcap\{A_i^c : b_i = 0\} \cap \bigcap\{A_i : b_i = 1\}, \\ S_{\mathbf{b}} &:= A_{\mathbf{b}} \cup \bigcup\{A_{b_1,\dots,b_{i-1},0} : b_i = 1\} \;=\; \bigcup\{A_{\mathbf{b}'} : \mathbf{b}' \le_{\text{lex}} \mathbf{b}\}, \end{aligned}$$

where $\le_{\text{lex}}$ denotes the lexicographical order (which we apply only to two binary sequences of the same length). Next, for $x \in S$ define $\mathbf{b}(x) := (b_1(x), b_2(x), \dots) \in \{0,1\}^{\mathbb{N}}$ by $b_i = I_{A_i}$ for $i \in \mathbb{N}$. Thus $\mathbf{b}(x)$ is the unique infinite sequence with $x \in A_{\mathbf{b}(x)}$. Finally, we define

$$f(x) := \mu(S_{\mathbf{b}(x)}).$$

Claim 2. The function f is $\mathcal{F}$-measurable.

Proof of Claim. The function $(x, y) \mapsto I_{\le_{\text{lex}}}(\mathbf{b}(x), \mathbf{b}(y))$ is $\mathcal{F} \otimes \mathcal{F}$-measurable: the pre-image of 0 is

$$\cup_{i \in \mathbb{N}} \cup_{b_1,\dots,b_{i-1}} \big(A_{b_1,\dots,b_{i-1},1} \times A_{b_1,\dots,b_{i-1},0}\big) \in \mathcal{F} \otimes \mathcal{F}.$$

Thus $f(x) = \int I_{\text{lex}}(\mathbf{b}(y), \mathbf{b}(x))\, \mathrm{d}\mu(y)$ is $\mathcal{F}$-measurable by [2, Corollary 3.3.3]. □

Claim 3. For every $a \in [0,1]$ and every infinite $\mathbf{b}$, both sets $f^{-1}(a)$ and $A_{\mathbf{b}}$ belong to $\mathcal{F}$ and have μ-measure zero.

Proof of Claim. We have $f^{-1}(a) \in \mathcal{F}$ by Claim 2 and $A_{\mathbf{b}} \in \mathcal{F}$ because each A_i is in $\mathcal{F}$. Each of $f^{-1}(a)$ and $A_{\mathbf{b}}$ is a null set because otherwise some A_i would cut it into two parts of positive measure, which is clearly impossible. □

Claim 4. Let $|\mathbf{b}|$ denote the length of the sequence $\mathbf{b}$. Then

$$\lim_{n\to\infty} \sup_{|\mathbf{b}|=n} \mu(A_{\mathbf{b}}) = 0.$$

Proof of Claim. Assume to the contrary this lim sup is $\varepsilon > 0$. Then, by König's lemma, there exists an infinite sequence $\mathbf{b} = (b_1, b_2, \ldots)$ such that $\mu(A_{b_1,\ldots,b_n}) \geq \varepsilon$ for every n. (Note that $A_{b_1,\ldots,b_n} \supseteq A_{b_1,\ldots,b_{n+1}}$.) As $A_{\mathbf{b}} = \cap_{n=1}^{\infty} A_{b_1,\ldots,b_n}$, we conclude that $\mu(A_{\mathbf{b}}) \geq \varepsilon > 0$, contradicting Claim 3. □

Claim 5. The set $\{\mu(S_{\mathbf{b}}) : |\mathbf{b}| \text{ is finite}\}$ is dense in $[0, 1]$.

Proof of Claim. Consider the binary sequences of length n. Notice that, for any finite $\mathbf{b}$,

$$\mu(S_{\mathbf{b}}) = \sum_{\mathbf{b}' \leq_{\text{lex}} \mathbf{b}} \mu(A_{\mathbf{b}'}),$$

$\mu(S_{1,1,\ldots,1}) = 1$, and that $\mu(S_{0,0,\ldots,0}) = \mu(A_{0,0,\ldots 0})$ (which tends to 0 by Claim 3). Let $\mathbf{b}' \leq_{\text{lex}} \mathbf{b}''$ be two sequences of length n which are consecutive in $\leq_{\text{lex}}$. Then

$$\mu(S_{\mathbf{b}''}) - \mu(S_{\mathbf{b}'}) = \mu(A_{\mathbf{b}''}) \leq \sup_{|\mathbf{b}|=n} \mu(A_{\mathbf{b}}).$$

Combining this with Claim 4 gives the statement. □

Claim 6. The function f is measure preserving.

Proof of Claim. Claim 5 implies that the intervals $[0, \mu(S_{\mathbf{b}})]$, where $\mathbf{b}$ runs over finite binary sequences, generate the Borel σ-algebra. Thus is enough to show that for every finite $\mathbf{b}$ we have $\mu\left(f^{-1}([0, a])\right) = a$, where $a := \mu(S_{\mathbf{b}})$. The latter identity follows from the fact that the symmetric difference of $S_{\mathbf{b}}$ and $f^{-1}([0, a])$ is a subset of $f^{-1}(a)$ and therefore has measure zero by Claim 3. □

The set $Y := \{(x, y) \in S^2 : x \lhd y,\ f(x) > f(y)\}$ is a subset of $\cup_{i=1}^{\infty}(A_i \times A_i^c) \in \mathcal{F} \otimes \mathcal{F}$. But the latter set has $\mu\otimes\mu$-measure zero by the definition of A_i's. Thus Y also has measure zero. Next, consider the set $Y_0 := \{(x, y) \in S^2 : f(x) = f(y)\}$. Since f is $\mathcal{F}$-measurable, we have $Y_0 \in \mathcal{F} \otimes \mathcal{F}$. Every slice of Y_0 has measure zero by Claim 3. By Fubini's theorem, Y_0 has itself measure zero. We conclude that f is an inclusion.

Finally, take an arbitrary $A \in \mathcal{F}$ with $\mu_{\lhd}(A \times A^c) = 0$. For every $i \in \mathbb{N}$ there is a set $A_{n_i} \in \mathcal{T}$ such that $\mu(A_{n_i} \triangle A) < 2^{-i}$. Since $A_{n_i} \triangle f^{-1}(X) \subseteq f^{-1}(Y)$, where X is a finite

union of intervals and Y is the set of their endpoints, we have by Claim 3 that A_{n_i} is $(f^{-1}(\mathcal{B}))_\mu$-measurable. This implies that A is $(f^{-1}(\mathcal{B}))_\mu$-measurable: indeed, for

$$A' := \limsup_{i\to\infty} A_{n_i} = \bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty} A_{n_j} \in (f^{-1}(\mathcal{B}))_\mu \tag{14}$$

we have $\mu(A'\triangle A) = 0$. This finishes the proof of Theorem 4.1 (and Theorem 1.10).

5. Proof of Theorem 1.9

As in Theorem 1.10, we can assume that $\mathcal{F}$ is separable. Apply Theorem 4.1 to $(S, \mathcal{F}, \mu, \lhd)$ to obtain an inclusion $f : S \to [0,1]$. As we will see later, the same f will work in Theorem 1.9. (Thus, rather interestingly, f can be chosen independently of W in Theorem 1.9 if $\mathcal{F}$ is separable.) Let

$$\mathcal{A} := (f^{-1}(\mathcal{B}))_\mu.$$

Since f is $\mathcal{F}$-measurable, we have that $\mathcal{A} \subseteq \mathcal{F}_\mu$.

We would like to apply Lemma 3.3. In order to do so, we have to verify first that $W \sim E(W|\mathcal{A}\otimes\mathcal{A})$. (Note that $E(W|\mathcal{A}\otimes\mathcal{A}) \sim E(W|f^{-1}(\mathcal{B})\otimes f^{-1}(\mathcal{B}))$.)

Claim 1. For every y, the slice function W_y is $\mathcal{A}$-measurable.

Proof of Claim. Fix any $a \in [0,1]$. For every $y \in S$, the set

$$A := W_y^{-1}(\,(a,1]\,) = \{x \in S : W(x,y) > a\}$$

is in $\mathcal{F}$. Since W is a strict kernel, we have $\mu_\lhd(A^c \times A) = 0$ for every y. By the second part of Theorem 4.1, $A \in \mathcal{F}$ belongs in fact to $\mathcal{A}$. Since intervals $(a,1]$ generate the Borel σ-algebra, the claim follows. □

Thus W and $\mathbb{E}(W|\mathcal{A}\otimes\mathcal{F}_\mu)$ are both $\mathcal{F}\,\bar\otimes\,\mathcal{F}$-measurable. (Note that $\mathcal{F}\,\bar\otimes\,\mathcal{F} = \mathcal{F}_\mu\,\bar\otimes\,\mathcal{F}_\mu$.) Also their y-slices are equivalent for a.e. y by Lemma 3.5. By Fubini's theorem, the subset of S^2 where these two functions differ has $\mu\otimes\mu$-measure zero. In other words, W is $\mathcal{A}\,\bar\otimes\,\mathcal{F}$-measurable and, by symmetry, $\mathcal{F}\,\bar\otimes\,\mathcal{A}$-measurable.

Claim 2.

$$W \sim \mathbb{E}(W|\mathcal{A}\otimes\mathcal{A}).$$

Proof of Claim. We follow the argument of Borgs, Chayes, and Lovász [3, Section 3.3.5]. Let $\widetilde{W} \in \mathbb{E}(W|\mathcal{A}\otimes\mathcal{A})$. It is enough to prove that for every $A, B \in \mathcal{F}$,

$$\int_{A\times B} W = \int_{A\times B} \widetilde{W}.$$

Take any $g_A \in \mathbb{E}(I_A|\mathcal{A})$ and let

$$\begin{aligned} U_A(y) &:= \int_A W(x,y)\,\mathrm{d}\mu(x) = \int W(x,y)I_A(x)\,\mathrm{d}\mu(x), \\ V_A(x) &:= \int W(x,y)g_A(y)\,\mathrm{d}\mu(y). \end{aligned}$$

Clearly, g_A is $\mathcal{A}$-measurable. Since W is $\mathcal{F}\bar{\otimes}\mathcal{A}$-measurable (as it was noted after Claim 1), U_A is $\mathcal{A}$-measurable by Fubini's theorem. Similarly, V_A is also $\mathcal{A}$-measurable. Repeatedly using Fubini's theorem and (8), we get

$$\begin{aligned} \int_{A\times B} W(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) = \int U_A(y)I_B(y)\,\mathrm{d}\mu(y) = \\ \int U_A(y)g_B(y)\,\mathrm{d}\mu(y) = \int W(x,y)I_A(x)g_B(y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) = \\ \int V_B(x)I_A(x)\,\mathrm{d}\mu(x) = \int V_B(x)g_A(x)\,\mathrm{d}\mu(x) = \\ \int W(x,y)g_A(x)g_B(y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) = \int \widetilde{W}(x,y)g_A(x)g_B(y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y). \end{aligned}$$

Observe that $g_A(x)g_B(y)$ is a conditional expectation of $I_A(x)I_B(y)$ with respect to $\mathcal{A}\otimes\mathcal{A}$ while $\widetilde{W}$ is measurable in this σ-algebra. Thus we can replace $g_A(x)g_B(y)$ by $I_A(x)I_B(y)$ in the last integral, obtaining $\int_{A\times B}\widetilde{W}$ as desired. □

Thus all assumptions of Lemma 3.3 are satisfied and we obtain that $W \sim U^f$ for some almost kernel U on $([0,1],\mathcal{B},\lambda,<)$. By Lemma 3.1, we can change U on a null set so that $([0,1],\mathcal{B},\lambda,<,U)$ is a kernel. Clearly, the equivalence $W \sim U^f$ is not affected by this. This finishes the proof of Theorem 1.9.

6. A finite Szemerédi-type Regularity Lemma for posets

In this section we prove a Szemerédi-type Regularity Lemma for posets, Theorem 6.1. (See Proemel, Steger, and Taraz [16] and Patel [15] for other versions.) We then show in Section 7 that this result can be used to answer Janson's question.

Suppose that $(P,\prec)$ is a poset. For two disjoint sets $X,Y\subseteq P$ we write $X \not\prec Y$ if there are no $x\in X$ and $y\in Y$ such that $x\prec y$. An (ordered) partition $\mathcal{P}=(V_1,\dots,V_k)$ of the ground set P is a *poset partition* if

$$V_i \not\prec V_j, \quad \text{for all } 1\le j<i\le k. \tag{15}$$

In other words, every $\prec$-relation that involves vertices from two different parts goes "forward". We refer to members of $\mathcal{P}$ as *clusters*. Let us say that $\mathcal{R}$ is a *poset refinement* of $\mathcal{P}$ if $\mathcal{R}$ is a poset partition that refines $\mathcal{P}$ (that is, for each $X\in\mathcal{R}$ there exists $Y\in\mathcal{P}$

such that $X \subseteq Y$). The *restriction* of $\mathcal{P}$ to $X \subseteq P$ is $\mathcal{P}|_X = (V_1 \cap X, \ldots, V_k \cap X)$. (For notational convenience, we allow empty parts.)

Let $G = G_{P,\prec}$ be an (undirected) graph on the vertex set P with edge set

$$E(G) := \{\, \{x, y\} \,:\, x \prec y \text{ or } y \prec x \,\}.$$

Clearly, if we know G and a poset partition $\mathcal{P}$, then we can reconstruct $\prec$ except for pairs lying inside a part. The main idea behind our Regularity Lemma is to find a poset partition of P that is regular with respect to G.

The following definitions apply to $A, B \subseteq P$. The *density* of the pair (A, B) is

$$d(A,B) := \frac{e(A,B)}{|A|\,|B|} := \frac{|\{(x,y) \in A \times B : x \prec y\}|}{|A|\,|B|}, \qquad \text{if } A, B \neq \emptyset,$$

and $d(A, B) := 0$ otherwise. The pair (A, B) is called *ε-regular* if $|d(A, B) - d(X, Y)| < \varepsilon$ for each $X \subseteq A$ and $Y \subseteq B$ with $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$. When we will apply the definition of ε-regularity to (A, B), it will always be the case that $B \not\prec A$ (and we obtain the standard graph definition). Also, let

$$q(A,B) := \frac{|A|\,|B|}{|P|^2}\, d^2(A,B).$$

For disjoint sets $V_1, \ldots, V_k, U_1, \ldots, U_m \subseteq P$, we define

$$\begin{aligned} q(\,(V_1,\ldots,V_k),(U_1,\ldots,U_m)\,) &:= \sum_{i=1}^{k}\sum_{j=1}^{m} q(V_i,U_j), \\ q(\,(V_1,\ldots,V_k)\,) &:= \sum_{i<j} q(V_i,V_j). \end{aligned}$$

The function q is called the *index* and is crucial in the proof of the original Regularity Lemma. Also, let $\mathbb{I}_\varepsilon(\,(V_1, \ldots, V_k)\,)$ be the set of pairs (i, j) such that $i < j$ and (V_i, V_j) is not ε-regular.

The sizes of the clusters in our Regularity Lemma can vary vastly (at least in our proof). This is why our next definition is slightly different from the standard one. A poset partition $\mathcal{P} = (V_1, \ldots, V_k)$ of P is *ε-regular* if each $|V_i| \leq \max(\varepsilon|P|, 1)$ and

$$\sum_{(i,j)\in\mathbb{I}_\varepsilon(\mathcal{P})} |V_i|\,|V_j| \leq \varepsilon \binom{|P|}{2}.$$

Theorem 6.1 (Regularity Lemma for Posets)**.** *For each $\varepsilon > 0$ there exists a number M such that the following holds. For each poset $(P, \prec)$ with a poset partition $\mathcal{P}$ such that $|\mathcal{P}| \leq 1/\varepsilon$, there exists a poset refinement $\mathcal{R}$ of $\mathcal{P}$ which is ε-regular and has at most M parts.*

Remark 6.2. It is important for our later application in Section 7 that there is no garbage cluster in our partition.

We prove Theorem 6.1 by following Szemerédi's original proof of the Regularity Lemma for graphs [20] (a more accessible reference is for example [7, Section 7.4]). The basic idea is that if a current partition $\mathcal{P}$ is not ε-regular then we can refine it so that $q(\mathcal{P})$ increases by at least δ, where $\delta > 0$ depends on ε only. Since q is always between 0 and 1, we reach an ε-regular partition in at most $1/\delta$ refinements. The following "pumping-up" lemma estimates by how much we can increase q by subdividing one irregular pair (A, B).

Lemma 6.3. *Suppose that $(P, \prec)$ is a poset and $A, B \subseteq P$ are disjoint nonempty sets. If $B \not\prec A$ and (A, B) is not ε-regular, then there are partitions $A = Z_1 \cup Z_2$ and $B = Z_3 \cup Z_4$ such that $Z_2 \not\prec Z_1$, $Z_4 \not\prec Z_3$, and*

$$q(\,(Z_1, Z_2), (Z_3, Z_4)\,) \geq q(A, B) + \varepsilon^4 \frac{|A|\,|B|}{n^2}. \tag{16}$$

Proof. Let $d := d(A, B)$. Consider a witness of irregularity (X, Y) of the pair (A, B). Assume without loss of generality that $d(X, Y) \geq d + \varepsilon$.

Iteratively, repeat the following as long as possible: replace some $x \in X$ by some $y \in A \setminus X$ with $y \prec x$. Clearly, this operation preserves the size of X and cannot decrease $d(X, Y)$. Also, we have to stop at some point. Let Z_1 be the final X and let $Z_2 := A \setminus Z_1$.

Similarly, replace $Y \subseteq B$ by an up-closed subset $Z_4 \subseteq B$ such that $|Z_4| = |Y|$ and $d(Z_1, Z_4) \geq d(Z_1, Y) \geq d + \varepsilon$. Let $Z_3 := B \setminus Z_4$.

Of course, we have that $Z_2 \not\prec Z_1$ and $Z_4 \not\prec Z_3$. Also, (Z_1, Z_4) witnesses that the pair (A, B) is not ε-regular. Now, (16) follows from the proof of Lemma 7.4.3 in [7]. □

Proof of Theorem 6.1. Let $s := \lceil 4/\varepsilon^5 \rceil$, $k_0 := \lceil 2/\varepsilon \rceil$, and inductively for $t = 0, \ldots, s-1$, let $k_{t+1} := k_t\, 2^{k_t - 1}$. We claim that $M = k_s$ suffices.

Suppose that $n > 1/\varepsilon$ for otherwise we can let $\mathcal{R}$ be a partition into singletons.

Initially, let $\mathcal{R}_0$ be an arbitrary poset refinement of $\mathcal{P}$ such that $|\mathcal{R}_0| \leq k_0$ and each part has at most εn vertices.

Iteratively, for $t = 0, 1, \ldots$, we repeat the following procedure. Let $\mathcal{R}_t = (V_1, \ldots, V_k)$. If $\mathcal{R}_t$ is ε-regular then we stop and output $\mathcal{R}_t$; so suppose otherwise. Let $\mathcal{R}' := \mathcal{R}_t$. We modify $\mathcal{R}'$ by using another (embedded) iterative procedure. Namely, in turn for each $(i, j) \in \mathbb{I}_\varepsilon(\mathcal{R}_t)$, we take the partitions $V_i = Z_{1ij} \cup Z_{2ij}$ and $V_j = Z_{3ij} \cup Z_{4ij}$ returned by Lemma 6.3 and replace every $X \in \mathcal{R}'$ by $X \cap Z_{1ij}, \ldots, X \cap Z_{4ij}$, with these four parts coming in the specified order. Clearly, $\mathcal{R}'$ is still a poset partition. Once we have processed all elements of $\mathbb{I}_\varepsilon(\mathcal{R}_t)$, we let $\mathcal{R}_{t+1} := \mathcal{R}'$.

In order to estimate how q changes, let us re-write

$$q(\mathcal{R}_{t+1}) - q(\mathcal{R}_t) = \sum_{1\le i<j\le k} \Big(q(\mathcal{R}_{t+1}|_{V_i}, \mathcal{R}_{t+1}|_{V_j}) - q(V_i, V_j) \Big). \tag{17}$$

We can estimate each summand corresponding to $(i,j) \in \mathbb{I}_\varepsilon(\mathcal{R}_t)$ by passing from $q(V_i, V_j)$ first to $q(\,(Z_{1ij}, Z_{2ij}), (Z_{3ij}, Z_{4ij})\,)$ and then to $q(\mathcal{R}_{t+1}|_{V_i}, \mathcal{R}_{t+1}|_{V_j})$. The first step increases q as specified by Lemma 6.3. The second step has non-negative effect by [7, Lemma 7.4.2]. Each other term in the right-hand side of (17) is non-negative, again by [7, Lemma 7.4.2]. Since $\mathcal{R}_t$ is not ε-regular, we conclude that

$$q(\mathcal{R}_{t+1}) - q(\mathcal{R}_t) \ge \frac{\varepsilon^4}{n^2} \sum_{(i,j)\in\mathbb{I}_\varepsilon(\mathcal{R}_t)} |V_i|\,|V_j| > \frac{\varepsilon^4}{n^2}\,\varepsilon\binom{n}{2} \ge \frac{\varepsilon^5}{4}. \tag{18}$$

We have that $q(\mathcal{P}) \le 1$ for any partition $\mathcal{P}$ (while trivially $q(\mathcal{P}) \ge 0$), see e.g. [7, Page 174]. By (18), we repeat the iteration procedure at most s times before we reach an ε-regular poset partition. As each part of $\mathcal{R}^t$ is split into at most $2^{|\mathcal{R}_t|-1}$ parts, we have that $|\mathcal{R}^{t+1}| \le |\mathcal{R}_t|\,2^{|\mathcal{R}_t|-1}$. Thus the final partition has at most M parts, as required. □

If we do not know $\prec$ but know an ε-regular partition $\mathcal{R} = (V_1, \dots, V_k)$ and the densities between all pairs of parts, then we can still derive various information about the poset P. For example, given two subsets $S, T \subseteq P$, one would expect to see approximately

$$e'(S,T) := \sum_{i<j} d(V_i, V_j)\,|V_i \cap S|\,|V_j \cap T|$$

directed arcs from S to T. Indeed, this is the case for posets.

Lemma 6.4. *Given the above assumptions, we have*

$$|e(S,T) - e'(S,T)| \le 3\varepsilon\binom{|P|}{2}. \tag{19}$$

Proof. Let $n := |P|$. Assuming the worst-case scenario, the edges inside a part or inside a non-ε-regular pair contribute at most $\varepsilon\binom{n}{2} + \varepsilon\binom{n}{2}$ to the left-hand side of (19). For every ε-regular pair (V_i, V_j) with $i < j$, we have

$$\Big|\, e(V_i \cap S, V_j \cap T) - d(V_i, V_j)\,|V_i \cap S|\,|V_j \cap T| \,\Big| \le \varepsilon |V_i|\,|V_j|.$$

Indeed, if $|S \cap V_i|\,|T \cap V_j| \le \varepsilon |V_i|\,|V_j|$, then we are trivially done; otherwise both S and T take more than ε-proportion of respectively V_i and V_j and the bound follows by the ε-regularity of (V_i, V_j). Thus the aggregate contribution of ε-regular pairs to (19) is at most $\varepsilon\binom{n}{2}$. □

7. An alternative proof of Theorem 1.7

Let $\{(P_n, \prec_n)\}_{n\in\mathbb{N}}$ be a convergent sequence of posets. We have to construct a kernel $([0,1], \mathcal{B}, \lambda, <, U)$ such that for every poset P we have

$$t(P, U) = \lim_{n\to\infty} t(P, P_n). \tag{20}$$

We construct U following closely the analogous construction of Lovász and Szegedy [13, Theorem 2.4] (see also [14, Theorem 5.1]). In brief, the proof proceeds by finding a $\frac{1}{k}$-regular partition $\mathcal{P}_{n,k}$ of P_n with the number of parts bounded by a function of k only. Then we construct a step-function $W_{n,k} : [0,1]^2 \to [0,1]$ that encodes the part ratios and densities of $\mathcal{P}_{k,n}$. Since the "complexity" of $W_{n,k}$ is bounded by a function of k, a diagonalisation process gives a subsequence $\{P_{n_i}\}_{i\in\mathbb{N}}$ such that, for every k, we have $W_{n,k} \to U_k$ a.e. for some $U_k : [0,1]^2 \to [0,1]$. Additionally, when we choose our partitions $\mathcal{P}_{k,n}$, we can assume that they are nested for each n. This allows us to write U_{k-1} as a conditional expectation of U_k and conclude that $\{U_k\}_{k\in\mathbb{N}}$ converges to some U a.e. Finally, we need to apply Lemma 3.1 to tranform an almost kernel U into a kernel.

Let us give more details. Let $m_1 = 1$ and inductively for $k = 2, 3, \dots$ let m_k be sufficiently large such that every poset partition with at most m_{k-1} parts admits a $\frac{1}{k}$-regular poset refinement with m_k parts. Such a number exists by Theorem 6.1. (Recall that we allow empty parts.) For each $n \in \mathbb{N}$, let $\mathcal{P}_{n,1} := (P_n)$ be the trivial partition and then inductively for $k = 2, 3, \dots$ let

$$\mathcal{P}_{n,k} = (V_{n,k,1}, \dots, V_{n,k,m_k}) \tag{21}$$

be a $\frac{1}{k}$-regular poset partition of $(P_n, \prec_n)$ that refines $\mathcal{P}_{n,k-1}$. This nestedness allows us for each n, to choose a total ordering $\prec'_n$ of $(P_n, \prec_n)$ which is *compatible* with every poset partition $\mathcal{P}_{n,k}$ (that is, $V_{n,k,i} \not\prec'_n V_{n,k,j}$ whenever $i > j$). By relabelling, let us assume that $P_n = \{1, \dots, |P_n|\}$ and $\prec'_n$ is the standard order.

Already at this point, it makes sense to start operating with functions. Let $W_n : [0,1]^2 \to \{0,1\}$ be the step-function that encodes the $\prec_n$-relation in the obvious way: W_n is constant on $[\frac{i-1}{v}, \frac{i}{v}) \times [\frac{j-1}{v}, \frac{j}{v})$, where $v := |P_n|$, and assumes value 1 there if and only if $i \prec_n j$. It is easy to see that

$$t(P, P_n) = t(P, W_n), \quad \text{for every poset } P$$

where we view W_n as a kernel on $([0,1], \mathcal{B}, \lambda, <)$.

Let $\mathcal{P}'_{n,k} = (V'_{n,k,1}, \dots, V'_{n,k,m_k})$ be the partition of $[0,1]$ into consecutive intervals corresponding to (21). (Thus, for example, $\lambda(V'_{n,k,i}) = |V_{n,k,i}|/|P_n|$.) Let $W_{n,k}$ be the step-function on $\mathcal{P}'_{n,k} \times \mathcal{P}'_{n,k}$, whose steps correspond to the parts of $\mathcal{P}_{n,k}$ and whose values correspond to densities between parts. We can write this more compactly as

$$W_{n,k} \sim \mathbb{E}(W_n | \sigma(\mathcal{P}'_{n,k})),$$

a conditional expectation of W_n with respect to the (finite) σ-algebra generated by $\mathcal{P}'_{n,k}$. Since $\sigma(\mathcal{P}'_{n,1}) \subseteq \sigma(\mathcal{P}'_{n,2}) \subseteq \dots$, we have

$$W_{n,k} \sim \mathbb{E}(W_{n,k+1} | \, \sigma(\mathcal{P}'_{n,k})), \quad k \geq 1,$$

which translates into the combinatorially obvious fact that the densities of $\mathcal{P}_{n,k}$ can be obtained by averaging over the densities in the finer partition $\mathcal{P}_{n,k+1}$.

Since each $W_{n,k}$ can be described by specifying part sizes and densities (which involves at most $m_k + \binom{m_k}{2}$ reals in $[0,1]$), the standard diagonalisation process gives a subsequence $\{n_i\}_{i\in\mathbb{N}}$ such that these parameters converge for every k. Thus $W_{n_i,k} \to U_k$ a.e. for some step-function U_k with m_k steps that are intervals and are ordered as $\mathcal{P}'_k = (V'_{k,1}, \dots, V'_{k,m_k})$. Since $\{P_n\}_{n\in\mathbb{N}}$ is convergent, passing to a subsequence does not affect (20); thus we can assume that $\{W_{n,k}\}_{n\in\mathbb{N}}$ itself a.e. converges to U_k. Clearly, $\sigma(\mathcal{P}'_1) \subseteq \sigma(\mathcal{P}'_2) \subseteq \dots$ a.e. and

$$U_k \sim \mathbb{E}(U_{k+1} | \, \sigma(\mathcal{P}'_k)).$$

Thus, by the Martingale Convergence Theorem (see e.g. [2, Theorem 10.3.3]), $U_k \to U$ a.e. for some $U : [0,1]^2 \to [0,1]$.

The obtained function U, as the a.e. pointwise limit of Borel functions, is Borel a.e. Clearly, the kernel axioms hold for $([0,1], \mathcal{B}, \lambda, U)$ for all inputs that do not require the evaluation of U on a point of

$$X := \big\{(x,y) \in [0,1]^2 : U_k(x,y) \not\to U(x,y) \text{ or } \exists\, k \; W_{n,k}(x,y) \not\to U_k(x,y)\big\},$$

the set where some convergence fails. Since X has measure zero, U is an almost kernel. By applying Lemma 3.1, we can assume that U is a kernel.

It remains to show that (20) holds. The *cut-norm* of a bounded measurable function $W : [0,1]^2 \to \mathbb{R}$ is defined by

$$\|W\|_\square = \sup_{S,T\in\mathcal{B}} \left| \int_{S\times T} W(x,y) \, \mathrm{d}\lambda(x) \, \mathrm{d}\lambda(y) \right|. \tag{22}$$

Claim 1. $\|W_n - W_{n,k}\|_\square \leq \frac{5}{2k}$ for any $k, n \in \mathbb{N}$.

Proof of Claim. Let $W := W_n - W_{n,k}$. Assume that $v := |P_n| > k$ for otherwise there is nothing do to as $W = 0$.

Observe that, up to an additive error $\frac{1}{v}$, it is enough to consider those S and T in (22) that are unions of intervals $V_i := [\frac{i-1}{v}, \frac{i}{v})$ for $i \in [v]$. Indeed, fix any $S, T \in \mathcal{B}$ with, say, $\int_{S\times T} W \geq 0$ and take $i \in [v]$ one by one. If we modify S and T inside V_i, then the integral of W over

$$\big((V_i \times V_i^c) \cup (V_i^c \times V_i)\big) \cap (S \times T)$$

is a linear function of $\lambda(V_i \cap S)$ and $\lambda(V_i \cap T)$. Thus we can make each of these to belong to $\{0, 1/v\}$ without decreasing the above contribution. Updating S and T accordingly, we

decrease $\int_{S\times T} W$ by at most $\int_{V_i\times V_i} |W| \le 1/v^2$. When we have iteratively processed all $i \in [v]$, both S and T have the desired form.

Thus, by (19), we obtain the required:

$$\|W\|_\square \le \frac{3}{k}\binom{v}{2}\frac{1}{v^2} + \frac{1}{v} \le \frac{3}{2k} + \frac{1}{k} = \frac{5}{2k}.$$

□

Now, we are ready to verify (20). Take any poset $(P, \prec)$ and $\varepsilon > 0$. Let $m := e(G_{P,\prec})$ be the number of pairs in $\prec$.

Since we deal with bounded measurable functions, all convergences also hold in the ℓ_1-space on $([0,1]^2, \mathcal{B}, \lambda)$ by [2, Theorem 2.2.3]. Thus there is $k \ge \frac{15m}{2\varepsilon}$ such that $\|U - U_k\|_1 \le \frac{\varepsilon}{3m}$ and, fixing this k, there is n_0 such that $\|U_k - W_{n,k}\|_1 \le \frac{\varepsilon}{3m}$ for all $n \ge n_0$. Clearly, $\|f\|_\square \le \|f\|_1$ for any integrable f. Thus, by the Triangle Inequality and Claim 1, we have that for all $n \ge n_0$

$$\begin{aligned}\|U - W_n\|_\square &\le \|U - U_k\|_\square + \|U_k - W_{n,k}\|_\square + \|W_{n,k} - W_n\|_\square \\ &\le \|U - U_k\|_1 + \|U_k - W_{n,k}\|_1 + \frac{5}{2k} \le \frac{\varepsilon}{m}.\end{aligned}$$

By [11, Lemma 6.4], we have that $|t(P,U) - t(P,W_n)| \le m\,\|U - W_n\|_\square \le \varepsilon$. Since ε and P were arbitrary, (20) follows.

Summarising, $([0,1], \mathcal{B}, \lambda, <, U)$ is a kernel that establishes Theorem 1.7.

Remark 7.1. An alternative way to proving that the densities of F in W_n and $W_{n,k}$ are close is to adopt the Counting Lemma (see e.g. [18, Theorem 5]) to our settings. We do not see any principal difficulties here but we expect that the error term would be larger.

Remark 7.2. In the above proof it is not generally true that $\|W_n - W_{n,k}\|_1$ is small for sufficiently large k: for example, $W_{n,k}$ may be strictly between 0 and 1 on a set of positive measure (while W_n is always $\{0,1\}$-valued).

8. Concluding remarks

There are two natural ways to extend the definition of convergence to the case when the poset orders do not tend to infinity. One is to just use (1). Another, adopted by Janson [11, Definition 3.2], is to say that $\{P_n\}_{n\in\mathbb{N}}$ with $|P_n| \not\to \infty$ is convergent if the sequence is eventually constant (up to isomorphism). The choice of which one to use (or none) is more a matter of convenience. For example, this choice may depend on whether we want the "limits" of $(P, P, \dots)$ and $(P^{(1)}, P^{(2)}, \dots)$ to be the same or not. Here the *blow-up* $P^{(k)}$ of P is obtained by cloning each vertex of P k times; obviously, $t(Q,P) = t(Q,P^{(k)})$ for every poset Q. Since all results stated in the Introduction can be trivially reduced to the case $|P_n| \to \infty$ by blowing posets up, we decided to use Definition 1.1.

Of course, the assumption that $(S, \mathcal{F}, \mu)$ is atomless is necessary in Theorems 1.9 and 1.10. This assumption can be removed if we are allowed to modify $([0,1], \mathcal{B}, \lambda, <)$ by shifting positive measure to a some countable subset $X \subseteq [0,1]$, where X depends on $(S, \mathcal{F}, \mu, \lhd)$. However, we believe that the versions presented in the Introduction are neater.

We cannot require in Theorems 1.9 and 1.10 that f preserves *every* relation (i.e. that the set in (7) is empty) as the following example demonstrates. Let $S := [0,1)$ with the Lebesgue measure λ on the Borel σ-algebra $\mathcal{B}$. Fix an irrational number τ. Let $T : S \to S$ map x to $x + \tau \pmod 1$. If we view S as a circle, then T is an aperiodic rotation. Define $x \lhd y$ if there is $k \in \mathbb{N}$ with $y = T^k(x)$. The constructed relation $\lhd$ is a Borel subset of S^2 (of measure zero). Let us suppose on the contrary that there is an inclusion $f : (S, \mathcal{B}, \lambda, \lhd) \to ([0,1], \mathcal{B}, \lambda, <)$ such that the set in (7) is empty. Let $A := f^{-1}([0, \frac{1}{2}])$. Since A is a down-closed set with respect to $<$, we have that $T^{-1}(A) \subseteq A$. Since T is measure preserving, we conclude that $T^{-1}(A) \sim A$. However, this contradicts the well-known fact (see e.g. [2, Example 10.9.9]) that T is ergodic. Alternatively, let $B := \cap_{k=1}^{\infty} T^{-k}(A)$. Then B is a measurable set such that $T^{-1}(B) = B$ (exactly) and $\mu(B) = 1/2$ (by σ-additivity). The same applies to B^c. By taking density points x and y of B and B^c respectively and a sequence of k such that $T^k(x) \to y$, one readily arrives at the desired contradiction.

Also, the assumption that W is strict in Theorem 1.9 is needed. For example, take $[0,1]^2$ with the Legesgue measure on the Borel sets and let $(x, y) \lhd (x', y')$ if $x < x'$. Let, for example, $W((x, y), (x', y'))$ be y' if $x' > x + 1/2$ and 0 otherwise. It is easy to see that every inclusion of this ordered probability space into the unit interval is a.e. equal to the projection onto the first coordinate. However, $W((x, y), (x', y'))$ is essentially non-constant on (x, x')-slices for $x' > 1/2 + x$ and thus cannot be equivalent to some pull-back.

References

1. T. Austin, *On exchangeable random variables and the statistics of large graphs and hypergraphs*, Probab. Surv. **5** (2008), 80–145.
2. V. I. Bogachev, *Measure theory*, Springer, Berlin, 2007.
3. C. Borgs, J. Chayes, and L. Lovász, *Moments of two-variable functions and the uniqueness of graph limits*, Geom. Func. Analysis **19** (2010), 1597–1619.
4. C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, *Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing*, Adv. Math. **219** (2008), 1801–1851.
5. G. Brightwell and N. Georgiou, *Continuum limits for classical sequential growth models*, Random Struct. Algorithms **36** (2010), 218–250.
6. P. Diaconis and S. Janson, *Graph limits and exchangeable random graphs*, Rendiconti di Matematica **28** (2008), 33–61.
7. R. Diestel, *Graph theory*, 4th ed., Springer, Berlin, 2010.
8. G. Elek and B. Szegedy, *A measure-theoretic approach to the theory of dense hypergraphs*, Adv. Math. **231** (2012), 1731–1772.

9. C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Ráth, and R. M. Sampaio, *Limits of permutation sequences*, E-Print `arXiv.org:1103.5844`, 2011.
10. C. Hoppen, Y. Kohayakawa, C. G. Moreira, and R. M. Sampaio, *Limits of permutation sequences through permutation regularity*, E-Print `arXiv.org:1106.1663`, 2011.
11. S. Janson, *Posets limits and exchangeable random posets*, Combinatorica **31** (2011), 529–563.
12. ———, *Limits of interval orders and semiorders*, J. Combin. **3** (2012), 163–184.
13. L. Lovász and B. Szegedy, *Limits of dense graph sequences*, J. Combin. Theory (B) **96** (2006), 933–957.
14. ———, *Szemerédi's Lemma for the analyst*, Geom. Func. Analysis **17** (2007), 252–270.
15. V. Patel, *Partitions of combinatorial structures*, PhD Thesis, London School of Economics, 2009.
16. H. J. Prömel, A. Steger, and A. Taraz, *Counting partial orders with a fixed number of comparable pairs*, Combin. Probab. Computing **10** (2001), 159–177.
17. A. Razborov, *Flag algebras*, J. Symb. Logic **72** (2007), 1239–1282.
18. M. Simonovits and V. T. Sós, *Szemerédi's partition and quasirandomness*, Random Struct. Algorithms **2** (1991), 1–10.
19. B. Szegedy, *Gowers norms, regularization and limits of functions on abelian groups*, E-Print `arxiv.org:1010.6211`, 2010.
20. E. Szemerédi, *Regular partitions of graphs*, Proc. Colloq. Int. CNRS, Paris, 1976, pp. 309–401.
21. T. Tao, *A correspondence principle between (hyper)graph theory and probability theory, and the (hyper)graph removal lemma*, J. Anal. Math. **103** (2007), 1–45.

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom

School of Mathematics, Birmingham University, Edgbaston, Birmingham B15 2TT, United Kingdom

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom

ON PATTERN-AVOIDING PERMUTONS

FREDERIK GARBE, JAN HLADKÝ, GÁBOR KUN, AND KRISTÝNA PEKÁRKOVÁ

Abstract. The theory of limits of permutations leads to limit objects called permutons, which are certain Borel measures on the unit square. We prove that permutons avoiding a given permutation of order k have a particularly simple structure. Namely, almost every fiber of the disintegration of the permuton (say, along the x-axis) consists only of atoms, at most $(k-1)$ many, and this bound is sharp. We use this to give a simple proof of the 'permutation removal lemma'.

1. Introduction

The main result of this paper concerns limit properties of pattern-avoiding permutations. Here, by a *permutation* we mean a bijection $\pi : [n] \to [n]$ for some $n \in \mathbb{N}$. We write $\mathbb{S}(n)$ for the set of all permutations on $[n]$. The fact that $[n]$ is equipped with the natural order allows us to define the notion of a pattern. Namely, for $k \in [n]$ and for a k-element set $K \in \binom{[n]}{k}$ we say that K *induces a pattern* $A \in \mathbb{S}(k)$ *in* π if for all $i, j \in [k]$ we have that $A(i) < A(j)$ if and only if the image of the ith smallest element of K under π is smaller than the image of the jth smallest element of K under π. The *density of* A *in* π, denoted by $t(A, \pi)$, is the proportion of k-tuples K that induce A. We say that π is A-*avoiding* if $t(A, \pi) = 0$. Pattern avoidance is one of the most vivid parts of the combinatorics of permutations, and indeed its treatise spans most of the standard books on permutations [3, 15]. A classical result that can be understood in this way is that of Erdős and Szekeres [10] on monotone patterns. Most of the questions in the area of pattern avoidance concern a setting in which the pattern is fixed and n is large. A very broad question is how restrictive the property of avoiding a specific pattern is. The most famous result in this direction is due to Marcus and Tardos [19], formerly known as the Stanley–Wilf conjecture, and says, that for any fixed pattern A there exists a constant c_A so that the number of A-avoiding permutations of any order n is at most c_A^n. The monographs of Bóna [3] and Kitaev [15] contain many further results regarding the number and structure of pattern-avoiding permutations.

Following the success of the theory of limits of dense graphs, Hoppen, Kohayakawa, Moreira, Ráth, and Sampaio [13] introduced a limit theory for finite permutations; the

JH: Research supported by Czech Science Foundation Project 21-21762X. FG, KP: This work has received funding from the MUNI Award in Science and Humanities (MUNI/I/1677/2018) of the Grant Agency of Masaryk University. GK: The work on the project leading to this application has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 741420), from the ÚNKP-20-5 New National Excellence Program of the Ministry of Innovation and Technology from the source of the National Research, Development and Innovation Fund, from Lendület grant no. 2022-58 and from the János Bolyai Scholarship of the Hungarian Academy of Sciences.

name and the following measure theoretic view on the limit object were introduced by Král' and Pikhurko [17]. Here, we recall only basics needed to state our main result and defer details to Section 2.1. A *permuton* Γ is a Borel probability measure on $[0,1]^2$ with uniform marginals, that is, $\Gamma(Z \times [0,1]) = \Gamma([0,1] \times Z)$ is equal to the Lebesgue measure of Z for any Borel set $Z \subset [0,1]$. The main feature of the theory of permutons is that they generalise permutations and form a compact metric space with the so-called *rectangular distance*. Furthermore, pattern densities are continuous with respect to this metric.

We now formally define pattern densities for permutons. Recall that for $A \in \mathbb{S}(k)$ and $\pi \in \mathbb{S}(n)$ with $k \leq n$ we have defined the density of A in π as the proportion of k-tuples of $[n]$ which induce the pattern A. To define pattern avoidance for permutons we will work with geometric representations of a pattern $A \in \mathbb{S}(k)$. Let $S_A \subset ([0,1]^2)^k$ be the collection of all sets of k points $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$ such that for all $1 \leq i < j \leq k$ we have $x_i < x_j$ and further that $y_i < y_j$ if and only if $A(i) < A(j)$. We call each element of S_A a *geometric representation* of A. Vice versa each collection $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$ of k points such that for every $i < j$ we have $x_i < x_j$ and $y_i \neq y_j$ *induces* a unique permutation $A \in \mathbb{S}(k)$ via $A(i) \coloneqq |\{j \in [k] \mid y_j \leq y_i\}|$. Note that each element of S_A induces A. We can now extend the notion of pattern density to permutons. The *density of A in a permuton Γ* is defined as $t(A, \Gamma) \coloneqq k! \cdot \Gamma^{\otimes k}(S_A)$, where $\Gamma^{\otimes k}$ denotes the k-dimensional product measure of Γ. A probabilistic interpretation of $t(A, \Gamma)$ is as follows. Suppose that we sample points $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$ independently at random from the measure Γ. Then $t(A, \Gamma)$ is the probability that when reading these points from left to right, their vertical positions are consistent with A. In particular, we say that Γ is *A-avoiding* if $t(A, \Gamma) = 0$.

Let $\mathfrak{M}$ be the space of all finite Borel measures on $[0,1]$. For a given $\gamma \in \mathfrak{M}$ and $\ell \in \mathbb{N}$ we say that γ is an *$\leq \ell$-molecule* if there exists a set $M \subset [0,1]$ with $|M| \leq \ell$ such that $\gamma([0,1] \setminus M) = 0$. We say that γ is an *ℓ-molecule* if it is an $\leq \ell$-molecule but not an $\leq (\ell - 1)$-molecule.

Our theorem below says that the support of the measure of any pattern-avoiding permuton has a particularly simple one-dimensional structure. This is phrased in terms of the disintegration of a measure, a classical concept which we briefly recall tailored to our setting and refer the reader to Section 10.6 of [1] for details. Suppose that $\{\Lambda_x \in \mathfrak{M}\}_{x \in [0,1]}$ is a collection of measures. We now add an integrability condition. A reader not very familiar with measure theory can understand the main point of the paper ignoring this condition just like one can understand quite a bit about real integration without knowing that not every function can be integrated. Suppose that $\{\Lambda_x \in \mathfrak{M}\}_{x \in [0,1]}$ satisfies that the map $x \mapsto \Lambda_x$ is measurable with respect to the Borel σ-algebra induced by the weak convergence topology on $\mathfrak{M}$. Then we can define a new Borel measure Λ on $[0,1]^2$ by

$$\Lambda(B) := \int_x \Lambda_x \left(\{y \in [0,1] \mid (x,y) \in B\}\right) \, \mathtt{d}\lambda(x) \quad \text{for each Borel } B \subset [0,1]^2 \,.$$

Then $\{\Lambda_x\}_{x\in[0,1]}$ is called a *disintegration* of Λ with respect to the Lebesgue measure λ. The Disintegration Theorem tells us that a disintegration always exists and is unique up to a nullset. We call the measures Λ_x *fibers*.

We give two examples which are particularly relevant for us. Firstly, the (unique, up to a nullset) disintegration of the 2-dimensional Lebesgue measure $\lambda^{\otimes 2}$ on $[0,1]^2$ is the collection of 1-dimensional Lebesgue measures, $\{\Lambda_x = \lambda\}_{x\in[0,1]}$. Secondly, consider the *diagonal permuton* D, that is, $D(Z) = \lambda(\{x \in [0,1] \mid (x,x) \in Z\})$ for every Borel $Z \subset [0,1]^2$. Then at each $x \in [0,1]$, the fiber of D at x is the Dirac measure Dirac_x. In particular, in the second example, all fibers are 1-molecules, whereas in the first example they are not $\leq \ell$-molecules for any ℓ.

We can now state our main result.

Theorem 1.1. *Suppose that A is a pattern of order k and Γ is an A-avoiding permuton. Let us fix a disintegration $\{\Gamma_x\}_{x\in[0,1]}$ of Γ. Then almost all fibers of Γ are $\leq (k-1)$-molecules.*

This bound is optimal: for every pattern A of order k there exists an A-free permuton such that almost all of its fibers are $(k-1)$-molecules.

Theorem 1.1 is a counterpart to a result of Cooper on finite permutations, [6, Theorem 3]. The language of permutons allows for a simpler formulation and proof. In particular, to prove Theorem 1.1 we use the Lebesgue density theorem while the proof in [6] is based on a regularity lemma for permutations (developed in that paper). This difference is yet another illustration how analytic statements combined with an appropriate theory of limits of discrete structures can substitute combinatorial tools such as regularity lemmas. One of the most fascinating demonstrations of this phenomenon is Elek's and Szegedy's approach to hypergraph limits [9].

Let us remark that Doležal, Máthé and the second author [8] proved a result in a similar spirit (but using completely different tools) for graphons: every graphon avoiding a specific graph must be countably-partite.

Our construction for the second half of the theorem will be piecewise linear. One may expect that pattern-free permutons are similarly nice. However, this need not be the case as we show in Section 5.1.

A reverse of Theorem 1.1 does not hold. That is, we can construct a permuton Γ_{all} such that every disintegration only consists of 1-molecule fibers, but Γ_{all} contains every pattern with positive density. For a given pattern A of size k we can construct a blow-up permuton Γ_A whose support consists of k increasing segments of equal length ordered according to A. Choosing an enumeration $\{A_1, A_2, \cdots\}$ of all countably many patterns we can scale each permuton Γ_{A_k} by the factor $\frac{1}{2^k}$ and form the direct sum $\Gamma_{all} = \bigoplus_{k\in\mathbb{N}} \frac{1}{2^k}\Gamma_{A_k}$ which just means that we place the scaled Γ_{A_k} along the diagonal. See Figure 1 for an illustration.

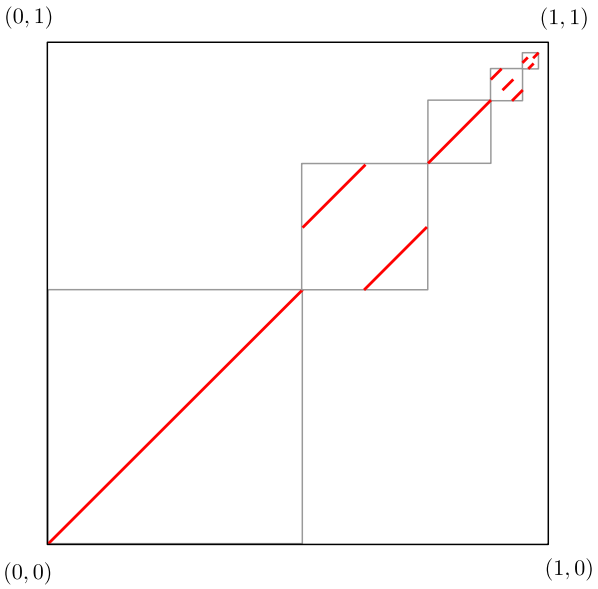


FIGURE 1. Permuton Γ_{all} for the sequence $A_1 = (12)$, $A_2 = (21)$, $A_3 = (123)$, $A_4 = (321)$, $A_5 = (213)$, ...

Formally, Γ_{all} is the unique permuton with the support

$$\bigcup_{k\in\mathbb{N},i\in[|A_k|]} \left\{ \left(\tfrac{i-1}{2^k|A_k|} + x + \sum_{j=1}^{k-1} \tfrac{1}{2^j}, \tfrac{A_k(i)-1}{2^k|A_k|} + x + \sum_{j=1}^{k-1} \tfrac{1}{2^j} \right) \mid 0 \le x \le \tfrac{1}{2^k|A_k|} \right\} .$$

Since every vertical or horizontal line intersects exactly one increasing segment in exactly one point every fiber of a disintegration is a 1-molecule. Furthermore, the occurrence of the scaled copy of Γ_{A_k} in Γ ensures that the each pattern A_k appears with positive density.

1.1. **A removal lemma as an application.** As an application, we provide a proof of the following removal lemma. This is a nice example how a result about limit objects can be used to show a result for finite permutations.

Theorem 1.2. *Let $k \in \mathbb{N}$ and $A \in \mathbb{S}(k)$. For every $\varepsilon > 0$ there exists $\delta > 0$ such that the following holds. Suppose that $\pi \in \mathbb{S}(n)$ is a permutation with $t(A, \pi) < \delta$. Then there exists a permutation $\tilde{\pi} \in \mathbb{S}(n)$ which is A-avoiding and for which we have*

$$\sum_{i=1}^{n} |\pi(i) - \tilde{\pi}(i)| < \varepsilon n^2 . \tag{1}$$

This 'permutation removal lemma' was first obtained by Klimošová and Král' [16] and reproved by Fox and Wei [11, 12]. The former proof gave an Ackermann-type dependency

between ε and δ.[a] The main upshot of the latter proof was in giving a polynomial dependency between ε and δ (with the degree of the polynomial being $2^{2^{O(k)}}$). Our proof is weaker in that it does not give a quantitative dependence between ε and δ. Its advantage, however, is that it is (with Theorem 1.1) almost computation-free, compared to the fairly involved proofs of Klimošová–Král' and Fox–Wei. Let us also note that the metric used to quantify (1) is called *Spearman's footrule distance*, and is within a factor of 2 within another, perhaps more common distance, the so-called *Kendall's tau distance* [7].

2. Notation and preliminaries

In this section we introduce some basic results from the theory of permutation limits and measure theory. This gives the context of Theorem 1.1, although its elementary proof will not make use of these additional results. However, we will use them in the proof of Theorem 1.2. In the following λ denotes the Lebesgue measure on $[0,1]$. For an arbitrary measure γ on a measure space X, we write $\gamma^{\otimes k}$ for its k-th power, the product measure on X^k.

2.1. **Limits of permutations.** We recall basic concepts of the theory of permutation limits as introduced in [13]. Earlier we already defined a *permuton* as a probability measure on the Borel σ-algebra on $[0,1]^2$ with the uniform marginal property. The relevant metric for the limit theory of permutations is the *rectangular distance* which is defined for two permutons Γ_1 and Γ_2 as

$$\mathrm{d}_{\square}(\Gamma_2,\Gamma_2) = \sup_{\substack{S,T\subseteq[0,1]\\ \text{intervals}}} |\Gamma_1(S\times T) - \Gamma_2(S\times T)| \,. \tag{2}$$

Remark 2.1. (2) *gives a metric for general measures (i.e., not necessarily permutons), as well.*

Note that each finite permutation $\pi\in\mathbb{S}(n)$ has a *permuton representation* Ψ_π, which is defined as follows. Take the union of rectangles $S \coloneqq \bigcup_{i=1}^{n}[\frac{i-1}{n},\frac{i}{n})\times[\frac{\pi(i)-1}{n},\frac{\pi(i)}{n})$. Now, for any Borel set $B\subset[0,1]^2$ we define $\Psi_\pi(B)\coloneqq n\cdot\lambda^{\otimes 2}(B\cap S)$. Using permuton representations we can extend the rectangular distance to all finite permutations. That is, given a permuton Γ and finite permutations $\pi_1\in\mathbb{S}(n_1)$ and $\pi_2\in\mathbb{S}(n_2)$ we write $\mathrm{d}_{\square}(\pi_1,\Gamma)\coloneqq\mathrm{d}_{\square}(\Psi_{\pi_1},\Gamma)$ and $\mathrm{d}_{\square}(\pi_1,\pi_2)\coloneqq\mathrm{d}_{\square}(\Psi_{\pi_1},\Psi_{\pi_2})$. The rectangular distance measures differences in global distributions of mass rather than differences in individual images, which corresponds to Spearman's footrule distance (see discussion below 1). A good illustration of this is the case when n is large and π_1 and π_2 are taken uniformly at random in $\mathbb{S}(n)$. Then, with

[a]Note that [16] claims a doubly exponential dependency. This claim is wrong as was pointed out in Footnote 2 of [11].

high probability, $\mathrm{d}_\square(\pi_1, \pi_2) \approx 0$, even though the images of individual elements of π_1 and of π_2 differ substantially. In fact, the 2-dimensional Lebesgue measure is the correct limit object, that is, with high probability, $\mathrm{d}_\square(\pi_1, \lambda^{\otimes 2}) \approx \mathrm{d}_\square(\pi_2, \lambda^{\otimes 2}) \approx 0$.

The following compactness result is the key result of the limit theory for permutations. We will use it to pass from finite permutations in Theorem 1.2 to a limit permuton to which we can apply Theorem 1.1.

Theorem 2.2 ([13])**.** *The space of permutons is compact with respect to the distance $d_\square$.*

In fact, Theorem 2.2 follows immediately from the fundamental theorem of Prokhorov that the weak topology (recalled in Section 2.2) of probability measures on $[0,1]^2$ is compact, since the rectangular distance gives the weak topology, and the closed subset of probability measures with uniform marginals is exactly the set of permutons, hence it is compact.

Furthermore, we will need that the convergence in rectangular distance also induces the convergence in pattern densities.

Theorem 2.3 ([13])**.** *Suppose that $\pi_1, \pi_2, \ldots$ is a sequence of permutations of growing order that converges to a permuton Γ in the rectangular distance. Then for any $k \in \mathbb{N}$ and any $A \in \mathbb{S}(k)$ we have $\lim_{n\to\infty} t(A, \pi_n) = t(A, \Gamma)$.*

2.2. **The Lévy-Prokhorov metric.** We will make use of the Lévy-Prokhorov metric to compare probability measures on $[0,1]$.

Definition 2.4 (Lévy-Prokhorov metric)**.** *For $\varepsilon > 0$ and a set $A \subseteq [0,1]$ we define the* ε-neighbourhood of A *by*

$$A^{\uparrow\varepsilon} = [0,1] \cap \bigcup_{q\in A} (q-\varepsilon, q+\varepsilon)\ .$$

Let α and β be two Borel probability measures on $[0,1]$. The Lévy-Prokhorov distance *between α and β is defined by*

$$\mathit{dist}_{LP}(\alpha,\beta) = \inf\{\varepsilon > 0 \mid \forall A \in \mathcal{B}([0,1])\ :\ \alpha(A) \le \beta(A^{\uparrow\varepsilon}) + \varepsilon,\ \beta(A) \le \alpha(A^{\uparrow\varepsilon}) + \varepsilon\}\ .$$

Recall that a sequence of Borel probability measures $\alpha_1, \alpha_2, \ldots$ on $[0,1]$ *converges weakly* to a measure α if for every Borel set $A \subset [0,1]$ we have $\lim_n \alpha_n(A) = \alpha(A)$. It is well-known that the Lévy-Prokhorov metric is a metrization of the topology from weak convergence. It is also well-known that the Lévy-Prokhorov metric is separable, that is, there exists a countable set $\{m_1, m_2, \ldots\}$ of Borel probability measures on $[0,1]$ so that for every Borel probability measure γ on $[0,1]$ and every $\varepsilon > 0$ there exists an index $i \in \mathbb{N}$ so that $\mathrm{dist}_{LP}(m_i, \gamma) < \varepsilon$. Last, recall that for any metric space, the properties of being

separable, having the property of Lindelöf and being second-countable are all equivalent.[b] In particular, we will use that the Lévy-Prokhorov metric gives a second-countable space.

Lemma 2.5. *Let Γ be a probability measure on $[0,1]^2$. For all $\delta > 0$, there exists $h \in \mathbb{N}$ and a partition $[0,1] = J_1 \cup J_2 \cup \ldots \cup J_h$ of intervals such that for a disintegration $\{\Gamma_x\}_{x\in[0,1]}$ of Γ there exists $X \subset [0,1]$ with $\lambda(X) < \delta$ such that $\forall i \in [h], x, y \in J_i \setminus X$: $dist_{LP}(\Gamma_x, \Gamma_y) < \delta$.*

Proof. We shall use Lusin's Theorem in the following form, see [14, Theorem 17.12]. Suppose that M is a second-countable topological space. Suppose that $f : [0,1] \to M$ is measurable. Then for every $\delta > 0$ there exists an open set $X \subset [0,1]$ with $\lambda(X) < \delta$ so that $f_{\restriction [0,1]\setminus X}$ is continuous.

We are in this setting of Lusin's Theorem as the space of all Borel probability measures on $[0,1]$ equipped with the Lévy-Prokhorov metric is second-countable, and the map $f : x \mapsto \Gamma_x$ is Borel. We use it with the same parameter as in the above statement. Since $[0,1] \setminus X$ is compact and the function f is continuous on it, it is also uniformly continuous. That is, for the given δ there exists $\alpha > 0$ so that for any $x, y \in [0,1] \setminus X$ with $|x - y| \le \alpha$ we have $\mathrm{dist}_{LP}(\Gamma_x, \Gamma_y) < \delta$. The lemma follows by choosing $(J_i)_{i\in[h]}$ to be consecutive intervals of length α. □

3. Proof of Theorem 1.1

3.1. **Proof of the upper bound.** Suppose for the sake of contradiction that the assertion does not hold. Then there exists a pattern A of order k, an A-avoiding permuton Γ together with a disintegration $\{\Gamma_x\}_{x\in[0,1]}$ and a set $X \subset [0,1]$ with $\lambda(X) > 0$ such that for every $x \in X$ the fiber Γ_x is not a $\le (k-1)$-molecule.

Figure 2 depicts the setting we introduce below. For $t \in \mathbb{N}$ and $m \in [t-1]$ we define $J_{m,t} \coloneqq [(m-1)/t, m/t)$ and $J_{t,t} \coloneqq [(t-1)/t, 1]$. For $S \in \binom{[t]}{k}$ and $i \in [k]$, write S_i for the ith smallest element of S. For each $S \in \binom{[t]}{k}$, set $X_{S,t} \coloneqq \{x \in X \mid \Gamma_x(J_{S_i,t}) > 0 \ \forall i \in [k]\}$. Note that for $x \in X$, since Γ_x is not $\le (k-1)$-molecule, we have that there exist $t \in \mathbb{N}$ and $S \in \binom{[t]}{k}$ such that $x \in X_{S,t}$. Hence $X = \bigcup_{t\in\mathbb{N}, S\in\binom{[t]}{k}} X_{S,t}$. Recalling that $\lambda(X) > 0$, it follows from the countable subadditivity of measures that there exist $t \in \mathbb{N}$ and $S \in \binom{[t]}{k}$ such that $\lambda(X_{S,t}) > 0$. By Lebesgue's density theorem, we can find a density point x_0 of $X_{S,t}$ and choose $\varepsilon > 0$ small enough such that $\lambda([x_0 - \varepsilon, x_0 + \varepsilon] \cap X_{S,t}) > 2\varepsilon(k-1)/k$. Then for $i \in [k]$ define $P_i \coloneqq [x_0 - \varepsilon + 2\varepsilon\cdot(i-1)/k, x_0 - \varepsilon + 2\varepsilon\cdot i/k)$ and note that $\lambda(X_{S,t} \cap P_i) > 0$. Finally, define $B_{i,j} \coloneqq P_i \times J_{S_j,t}$, for $i, j \in [k]$, and observe that every $p \in \prod_{i\in[k]} B_{i,A(i)}$

[b] We do not give definitions since the only way we use them is that we feed them into another theorem.

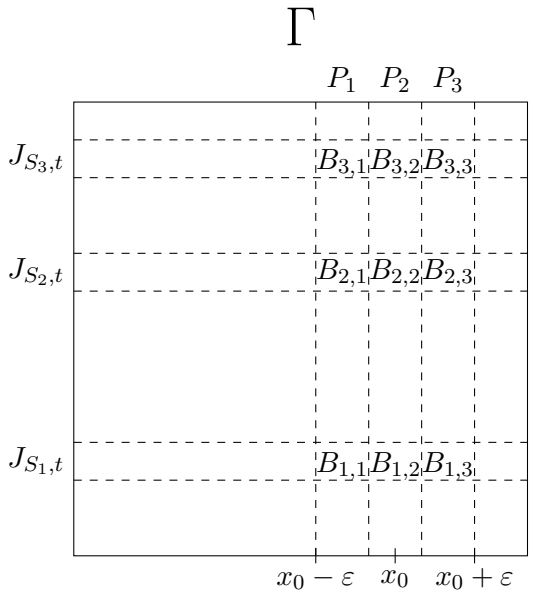


FIGURE 2. An example for the notation in Section 3 with $k = 3$. Each of the boxes $B_{i,j}$, $i, j \in [3]$, contains some positive mass of Γ, hence any 3×3-pattern occurs with positive density.

induces A and therefore

$$t(A, \Gamma) \geq \Gamma^{\otimes k}\left(\prod_{i\in[k]} B_{i,A(i)}\right) \geq \prod_{i\in[k]}\left(\int_{x\in P_i\cap X_{S,t}} \Gamma_x(J_{S_{A(i)},t})\right) > 0\,,$$

which is a contradiction.

3.2. **Optimality.** Given a pattern $A \in \mathbb{S}(k)$, we will construct an A-free permuton whose fibers with respect to a disintegration along the x-axis are almost all $(k-1)$-molecules. The role of the two coordinates is exchangeable, and it turns out that it is notationally simpler to construct an A^{-1}-free permuton whose horizontal fibers are $(k-1)$-molecules.

Set $\pi = A^{-1}$. We consider the piecewise linear function $L : [0,1] \to [0,1]$ consisting of $k-1$ pieces of slope $k-1$ or $1-k$ satisfying for $i = 1, \ldots, k-1$ that $L(\frac{i-1}{k-1}) = 0$ and $L(\frac{i}{k-1}) = 1$ if $\pi(i) > \pi(i+1)$, while $L(\frac{i-1}{k-1}) = 1$ and $L(\frac{i}{k-1}) = 0$ if $\pi(i) < \pi(i+1)$. (At the endpoints this might not be well-defined: in this case we choose the value defined on the interval on the left.) An example is given in Figure 3.

Putting an appropriate multiple of the 1-dimensional Lebesgue measure on the graph of the function L we obtain a permuton (that is, both uniform marginal properties are satisfied), which we call Γ. Firstly, almost all horizontal fibers of Γ are $(k-1)$-molecules. So, it remains to argue that $t(\pi, \Gamma) = 0$. To this end we prove the following.

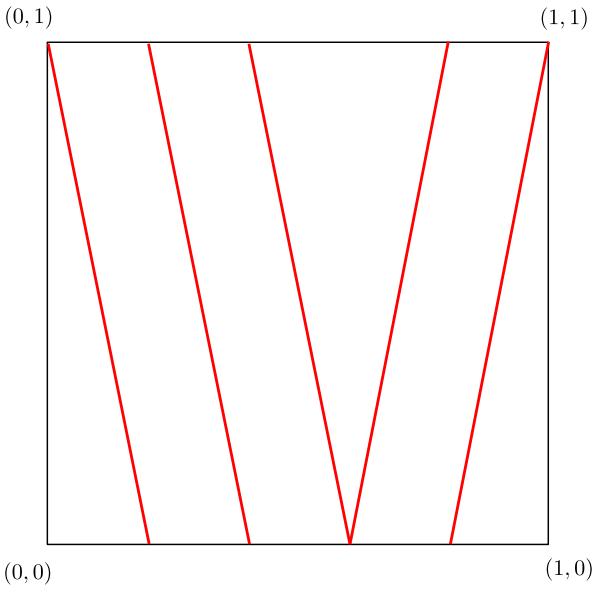


FIGURE 3. Construction of a piecewise linear function L in Section 3.2 for $\pi = (145632)$.

Claim. *Consider $\ell \in [k]$ and $0 \leq x_1 < \cdots < x_\ell \leq 1$ such that $x_i \notin \{\frac{j}{k-1} : j \in \mathbb{Z}\}$ for all $i \in [\ell]$. If the set of points $\{(x_1, L(x_1)) \ldots, (x_\ell, L(x_\ell))\}$ induces the pattern given by the first l entries of π then $x_\ell > \frac{\ell-1}{k-1}$.*

Observe that the above claim indeed proves that $t(\pi, \Gamma) = 0$, since it asserts that we cannot find geometric representations of π in the support of Γ, except for the nullset of k-tuples where at least one of the coordinates x_i lies in $\{\frac{j}{k-1} : j \in \mathbb{Z}\}$.

Let us now prove the claim by induction on ℓ. The base case $\ell = 1$ is obvious. Assume that the claim holds for $\ell-1$ and, in particular, for the $(\ell-1)$-tuple $(x_1, \ldots, x_{\ell-1})$. If $x_{\ell-1} > \frac{\ell-1}{k-1}$ then we are done, since $x_\ell > x_{\ell-1}$. It remains to handle the case $x_{\ell-1} \in (\frac{\ell-2}{k-1}, \frac{\ell-1}{k-1})$. Then $x_\ell \notin (\frac{\ell-2}{k-1}, \frac{\ell-1}{k-1})$, since $x_\ell > x_{\ell-1}$, and if $\pi(\ell-1) > \pi(\ell)$ then L is monotone increasing on this interval, while monotone decreasing otherwise. This completes the proof the theorem.

Note that while a.e. vertical fiber of the construction is a $(k-1)$-molecule, the horizontal fibers are 1-molecules, i.e., their support is a single atom of measure 1. It would be interesting to see, which lower bounds on the size of the support of a.e. horizontal and vertical fiber would guarantee that a given pattern has positive density.

4. Proof of Theorem 1.2

Let us first give an overview of the proof. By using the compactness of the space of permutons it suffices to show for a sequence $\pi_1, \pi_2, \ldots$ of not necessarily A-avoiding permutations converging to an A-free permuton Γ that for n large enough we can find similar but A-avoiding permutations $\tilde{\pi}_n$. Suppose that π_n is close to Γ in the rectangular distance. Fix a disintegration $\{\Gamma_x\}_{x\in[0,1]}$ of Γ. We generate an n-tuple of numbers uniformly at random in $[0,1]$, and read them in increasing order as $x_1 < x_2 < \cdots < x_n$. We set $y_i := (\pi_n(i) - 0.5)/n$. The set of points $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ is a geometric representation of π_n, and the uniform probability measure supported on this set is close to Γ in the rectangular distance. The key idea is that almost any n-tuple of points $(\tilde{x}_1, \tilde{y}_1)$, $\ldots$, $(\tilde{x}_n, \tilde{y}_n)$, whose x- and y-coordinates are pairwise distinct and where all the points lie in the support of the measure Γ, is a geometric representation of an A-free permutation. We shall use this with the choice $\tilde{x}_1 = x_1, \ldots, \tilde{x}_n = x_n$. We want to alter the y-coordinates of the points (x_i, y_i) such that the new $\tilde{y}_i$ lies in the support of the measure Γ_{x_i}, which is essentially equivalent to the point $(x_i, \tilde{y}_i)$ lying in the support of Γ.

Here is where Theorem 1.1 comes into play. We know that almost every Γ_{x_i} is a collection of at most $k-1$ atoms. Since π_n is close to Γ in the rectangular distance, we can also deduce that typically Γ_{x_i} must have positive mass around y_i. Therefore, there exists an atom $\tilde{y}_i$ of Γ_{x_i} which is close to y_i. The points $(x_i, \tilde{y}_i)$ then induce our new permutation $\tilde{\pi}_n$. See Figure 4.

We can now proceed with details. We shall prove Theorem 1.2 in the following form.

Theorem 4.1. *Let $k \in \mathbb{N}$ and $A \in \mathbb{S}(k)$. Suppose that we have a sequence of permutations $(\pi_n \in \mathbb{S}(n))_n$ converging to an A-free permuton Γ in the rectangular distance. Then for every $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for every $n > n_0$ there exists an A-free permutation $\tilde{\pi}_n \in \mathbb{S}(n)$ such that*

$$\sum_{i=1}^{n} |\pi_n(i) - \tilde{\pi}_n(i)| < \varepsilon n^2 \,. \tag{3}$$

Deducing Theorem 1.2 from Theorem 4.1. Suppose that Theorem 1.2 does not hold. That is, for a fixed pattern A, there exists a number $\varepsilon > 0$ and a sequence of permutations π_n with $t(A, \pi_n) \to 0$ for which we cannot find similar but A-free permutations. By compactness of the space of permutons (Theorem 2.2), there exists a subsequence (π_{n_ℓ}) which converges to a certain permuton Γ in the rectangular distance. By continuity of densities (Theorem 2.3), Γ is A-free. This setting contradicts Theorem 4.1, as was needed. □

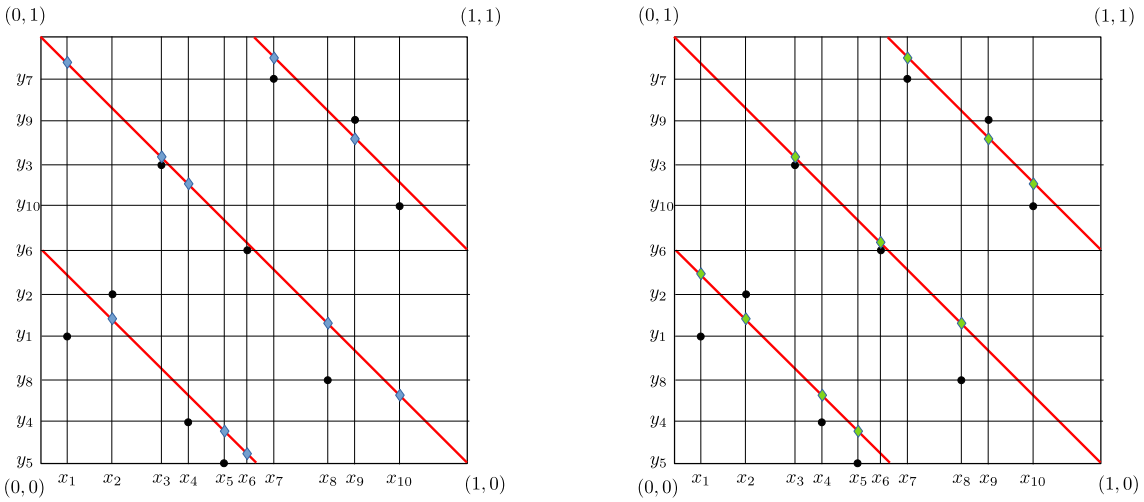


FIGURE 4. An example described in Section 1.1. In red, the support of a permuton Γ which is free of the pattern $A = (1234)$. Each fiber of the disintegration of Γ consists of two atoms, each of weight 0.5. Black dots show points (x_i, y_i) representing a permutation π_{10} which is close to Γ. *Left:* Grey diamonds show positions of points $(x_i, \tilde{y}_i)$ which represent an A-free permutation which might be an outcome of a random resnapping on the fibers Γ_{x_i}. The issue is that $|y_i - \tilde{y}_i|$ is big for $i = 1, 4, 6, 10$. *Right:* Green diamonds correspond to a resnapping to the closest atom on each fiber Γ_{x_i}.

Proof of Theorem 4.1. Given $\pi_n \overset{\mathrm{d}_\square}{\to} \Gamma$, where Γ is A-free, and $\varepsilon > 0$ we fix constants $\zeta, \delta > 0$ and $h, n_0 \in \mathbb{N}$ according to the following hierarchy

$$\tfrac{1}{n_0} \ll \zeta \ll \tfrac{1}{h} \ll \delta \ll \varepsilon \,.$$

First, we apply Lemma 2.5 to Γ and δ resulting in $h \in \mathbb{N}$, a partition $[0,1] = \bigcup_{i\in[h]} J_i$ into intervals, a set $X_1 \subseteq [0,1]$ with $\lambda(X_1) < \delta$, and a disintegration $\{\Gamma_x\}_{x\in[0,1]}$ of Γ such that for every $i \in [h]$ and $x, y \in J_i \setminus X_1$ we have

$$\mathrm{dist}_{LP}(\Gamma_x, \Gamma_y) < \delta \,. \tag{4}$$

Furthermore, since Γ is A-free and by Theorem 1.1 we can fix a null set $X_2 \subseteq [0,1]$ such that Γ_x is $\leq (k-1)$-molecule for every $x \in [0,1] \setminus X_2$. We then set $X := X_1 \cup X_2$.

Now, fix a permutation π_n with $n \geq n_0$. We want to show that there exists an A-free permutation $\tilde{\pi}_n \in \mathbb{S}(n)$ fulfilling (3). Note that by the constant hierarchy we can choose n_0 large enough such that we may assume that

$$\mathrm{d}_\square(\pi_n, \Gamma) < \zeta/2 \,. \tag{5}$$

Set $T \coloneqq \{i \in [n] \mid \lambda(((i-1)/n, i/n) \setminus X) > 0\}$. Note that $|T| > (1-\delta)n$, as $\lambda(X) < \delta$. We now generate a random n-tuple of points in the unit interval by choosing x_i uniformly at random from $((i-1)/n, i/n) \setminus X$ for every $i \in T$. For $i \in [n] \setminus T$ we choose x_i uniformly at random from $((i-1)/n, i/n) \setminus X_2$. Furthermore, we set

$$y_i \coloneqq (\pi_n(i) - 0.5)/n \tag{6}$$

and denote the measure which has mass $1/n$ on each point (x_i, y_i) by Δ,

$$\Delta = \frac{1}{n} \sum_{i \in [n]} \mathsf{Dirac}_{(x_i, y_i)} \ .$$

While Δ is not a permuton, it can be thought of as a perturbation of the permuton representation Ψ_{π_n} in which each rectangle $[\frac{i-1}{n}, \frac{i}{n}) \times [\frac{\pi_n(i)-1}{n}, \frac{\pi_n(i)}{n})$ is transformed to a single point of the same measure and within that rectangle. In particular $\mathrm{d}_\square(\Delta, \pi_n) \leq \frac{2}{n}$. From the triangle inequality (c.f. Remark 2.1) and (5) we get

$$\mathrm{d}_\square(\Delta, \Gamma) \leq \mathrm{d}_\square(\Delta, \pi_n) + \mathrm{d}_\square(\pi_n, \Gamma) \leq \tfrac{2}{n} + \zeta/2 < \zeta \ . \tag{7}$$

Claim 1. *With probability* 1 *with respect to the choice of the random points* $(x_i)_{i \in [n]}$ *the following holds.*

(i) If $(\tilde{y}_i \in [0,1])_n$ *is collection of pairwise distinct points such that for each* $i \in [n]$, $\tilde{y}_i$ *is an atom on* Γ_{x_i} *then the permutation induced by* $((x_i, \tilde{y}_i))_{i \in [n]}$ *is* A*-free.*

(ii) If $\Gamma_{x_i}(y) > 0$ *for some* $i \in [n]$ *and* $y \in [0,1]$*, then* $\Gamma_{x_j}(y) = 0$ *for every* $i \neq j$*.*

Proof of Claim. For proving (i) suppose for the sake of contradiction that there exists $Q \subseteq [0,1]^n$ with $\lambda^{\otimes n}(Q) > 0$ such that for every $x \in Q$ there exists $\tilde{y}(x) \in [0,1]^n$ such that $\prod_{i=1}^n \Gamma_{x_i}(\tilde{y}(x)_i) > 0$ and such that $(x_i, \tilde{y}(x)_i)_{i \in [n]}$ contains the pattern A. Furthermore, we can assume that the map $x \mapsto \tilde{y}(x)$ is measurable. Then the probability to sample the pattern A from Γ is at least

$$\int_Q \prod_{i=1}^n \Gamma_{x_i}(\tilde{y}(x)_i) \mathsf{d}\lambda^{\otimes n}(x) > 0 \ ,$$

contradicting that Γ is A-free.

For proving (ii) suppose for the sake of contradiction that there exists $y \in [0,1]$ such that for $Q_y \coloneqq \{x \in [0,1] \mid \Gamma_x(y) > 0\}$ holds that $\lambda(Q_y) > 0$. By continuity of the Lebesgue measure there exists $\beta > 0$ and $Q'_y \subseteq Q_y$ with $\lambda(Q'_y) > 0$ such that for every $x \in Q'_y$ we have $\Gamma_x(y) > \beta$. Then for an arbitrary $\alpha > 0$ we have

$$\Gamma([0,1] \times [y - \alpha, y + \alpha]) \geq \beta \lambda(Q'_y) \ ,$$

which for α small enough contradicts the uniform marginal property of Γ. □

We fix an outcome of our random selection of $x_1 < \cdots < x_n$ such that the above properties hold. Below, we abbreviate measures of intervals $[a,b]$ on fibers Γ_{x_i} as $\Gamma_{x_i}(a,b) := \Gamma_{x_i}([a,b])$.

Claim 2. *For all but at most $3\sqrt{\delta}n$ many $i \in [n]$ it holds that $\Gamma_{x_i}(y_i - 2\sqrt{\delta}, y_i + 2\sqrt{\delta}) > \delta$.*

Proof of Claim. We call a point (x_i, y_i) *bad* if $\Gamma_{x_i}(y_i - 2\sqrt{\delta}, y_i + 2\sqrt{\delta}) \leq \delta$ and let Y be the set of such bad points. We denote by $J(x)$ the interval J_j (where $j \in [h]$) such that $x \in J_j$. Consider a system of rectangles $\mathcal{C} = \{J(x_i) \times [y_i - \sqrt{\delta}, y_i + \sqrt{\delta}] \mid (x_i, y_i) \text{ bad}\}$ covering Y in $[0,1]^2$. Choose a minimal system $\mathcal{M} \subseteq \mathcal{C}$ which covers $\bigcup \mathcal{C}$. For $j \in [h]$, let $\mathcal{M}_j \subset \mathcal{M}$ be the collection of sets of the form $J_j \times U$. By the minimality, we have

$$|\mathcal{M}_j| \leq 2/(2\sqrt{\delta}) = 1/\sqrt{\delta}\ . \tag{8}$$

Therefore, $\bigcup \mathcal{C}$ can be written as the disjoint union of at most $h/\sqrt{\delta}$ many rectangles. By using (7) we get

$$|\Gamma(\cup\mathcal{C}) - \Delta(\cup\mathcal{C})| < \zeta h/\sqrt{\delta}\ . \tag{9}$$

For every rectangle $J(x_i) \times [y_i - \sqrt{\delta}, y_i + \sqrt{\delta}] = J_j \times B = M \in \mathcal{M}_j$ we have that

$$\Gamma(M \setminus (X \times [0,1])) = \int_{J_j \setminus X} \Gamma_x(B) \mathsf{d}\lambda(x) \overset{(4)}{\leq} \lambda(J_j) \cdot (\Gamma_{x_i}(B^{\uparrow\delta}) + \delta)\ .$$

In particular, since (x_i, y_i) is bad we get

$$\Gamma(M \setminus (X \times [0,1])) \leq 2\delta\lambda(J_j)\ . \tag{10}$$

Thus, the triangle inequality implies that

$$\begin{aligned}
\Delta(\cup\mathcal{C}) &\leq \Gamma(\cup\mathcal{C} \setminus (X \times [0,1])) + \Gamma(X \times [0,1]) + |\Gamma(\cup\mathcal{C}) - \Delta(\cup\mathcal{C})| \\
&\overset{(9)}{\leq} \sum_{j \in [h]} \sum_{M \in \mathcal{M}_j} \Gamma(M \setminus (X \times [0,1])) + \delta + \zeta h/\sqrt{\delta} \\
&\overset{(8),(10)}{\leq} \sum_{j \in [h]} \tfrac{1}{\sqrt{\delta}} \cdot 2\delta\lambda(J_j) + \sqrt{\delta} \\
&\leq 3\sqrt{\delta}\ ,
\end{aligned}$$

and therefore $|Y| \leq \Delta(\cup\mathcal{C})n \leq 3\sqrt{\delta}n$. □

Now set $S \coloneqq \{i \in T \mid \Gamma_{x_i}(y_i - 2\sqrt{\delta}, y_i + 2\sqrt{\delta}) > \delta\}$. Observe that by Claim 2 and the fact that $|T| \geq (1-\delta)n$ we have $|S| \geq n - 3\sqrt{\delta}n - \delta n \geq (1 - 4\sqrt{\delta})n$. For every $i \in S$, since Γ_{x_i} is $\leq (k-1)$-molecule there exists $\tilde{y}_i \in [y_i - 2\sqrt{\delta}, y_i + 2\sqrt{\delta}]$ such that $\Gamma_{x_i}(\tilde{y}_i) > 0$. For every $i \in [n] \setminus S$ choose an arbitrary atom $\tilde{y}_i$ of Γ_{x_i} which is possible, as $x_i \notin X_2$ for every $i \in [n]$. By Claim 1(ii) we have that the points $(x_i, \tilde{y}_i)_{i \in [n]}$ induce a permutation which

we denote by $\tilde{\pi}_n$. Furthermore, by Claim 1(i) $\tilde{\pi}_n$ is A-free. It remains to establish the key property (3). We split the summands $\sum_{i=1}^{n} |\pi_n(i) - \tilde{\pi}_n(i)|$ according to whether $i \in S$ or $i \notin S$. In the latter case, it is enough to use the trivial bound $|\pi_n(i) - \tilde{\pi}_n(i)| \leq n$. So, we need to have a good bound for $|\pi_n(i) - \tilde{\pi}_n(i)|$ when $i \in S$. Recall (6) which tells us that the spacing in the y-coordinates of the geometric representation of π_n was exactly $\frac{1}{n}$. Thus, if we change each point $(y_j)_{j\in S}$ by at most $2\sqrt{\delta}$, the point y_i changes its relative position[c] with at most $4\sqrt{\delta}n$ other points $(y_j)_{j\in S}$. Taking into account also elements which are not in S, we get $|\pi_n(i) - \tilde{\pi}_n(i)| \leq 4\sqrt{\delta}n + |[n] \setminus S|$. Putting all this together,

$$\sum_{i=1}^{n} |\pi_n(i) - \tilde{\pi}_n(i)| \leq \sum_{i\in S}(4\sqrt{\delta}n + |[n] \setminus S|) + \sum_{i\in[n]\setminus S} n \leq (4\sqrt{\delta} + 4\sqrt{\delta})n^2 + 4\sqrt{\delta}n^2 < \varepsilon n^2 \,.$$

□

5. Discussion and further questions

5.1. **Nondifferentiable pattern-avoiding permuton.** Theorem 1.1 together with Lusin's theorem tells us that we can think of the support of a pattern-avoiding permuton as a union of graphs of partial functions that are continuous on a set whose complement is of arbitrary small positive measure. The following example shows that continuity cannot be strengthened to differentiability. More precisely, there exists a function $f : [0, 1] \to [0, 1]$ whose restriction to any subset of $[0, 1]$ of positive measure is not differentiable and the permuton supported on the graph of this function is avoiding the pattern (3142).

Consider the quaternary expansion (with digits $0, 1, 2$ and 3) of the numbers in $[0, 1]$. This is well-defined for all, but countably many numbers, the rational numbers whose denominator is a power of two. We will ignore this countable set. Let f be the function that swaps 1 and 2 for every coordinate. This function is injective (up to a nullset) and measure preserving, hence its graph is the support of a permuton.

First we argue that the graph of f is free of (3142). Suppose for a contradiction that there are four numbers $x_1 < x_2 < x_3 < x_4$ such that $f(x_2) < f(x_4) < f(x_1) < f(x_3)$. Consider the first digit where x_1 and x_2 differ. Clearly this digit should be 1 for x_1 and 2 for x_2. The number x_4 cannot differ in an earlier digit, since $f(x_2) < f(x_4) < f(x_1)$. Hence x_3 being less than x_4 can also not differ in an earlier digit. We will get a contradiction when checking the critical digit: $f(x_3) > f(x_1), x_3 > x_2$ implies that this digit of x_3 should be 3, but then x_4 also has digit 3 here contradicting $f(x_1) > f(x_4)$.

It remains to show that the restriction of f to any set of positive measure cannot be differentiable: if there was such a set then on a subset of positive measure for every pair the inequality $|\frac{f(x)-f(y)}{x-y} - f'(x)| < \frac{1}{2}$ would hold. However, given a natural number n,

[c]that is 'bigger than/smaller than'

$i \in \{-3,-2,-1,1,2,3\}$ and real numbers $x, x+i4^{-n}$ agreeing until the first $(n-1)$ digits, note that $\frac{f(x)-f(x+i4^{-n})}{i4^{-n}} \in \{-2,-1,-\frac{1}{2},\frac{1}{2},1,2\}$, and this value depends on i and the nth digit of x (but not on n), and for every x and n we get three different values (depending on i). By the Lebesgue density theorem any set of positive measure contains more than three quarter of a small interval, so we can find four numbers in the set agreeing in all, but one digit. Hence the restriction of f to any set of positive measure cannot be differentiable.

5.2. **Lower bounds on the Stanley–Wilf constant.** For a permutation A, let $\mathbb{S}(A;n) \subset \mathbb{S}(n)$ be the set of A-avoiding permutations of order n. The Stanley–Wilf constant is defined as $c_A := \lim_n \sqrt[n]{|\mathbb{S}(A;n)|}$. The exact value of the Stanley–Wilf constant is only known for some permutations and permutation classes (see [3, Chapter 4]): trivially for $A \in \mathbb{S}(2)$ we have $c_A = 1$, for all $A \in \mathbb{S}(3)$ we have $c_A = 4$, for each $k \in \mathbb{N}$ we have $c_{\mathsf{identity}_k} = (k-1)^2$, and then the Stanley–Wilf constant is known for several sporadic examples, including $c_{(1342)} = 8$, $c_{(12453)} = 9 + 4\sqrt{2}$.

Consider an A-free permuton Γ. We can use Γ to generate many different A-free permutations of order n. Let us illustrate this with the (1234)-avoiding permuton Γ from Figure 4. Fix the numbers $x_1 < \ldots < x_n$ to be equidistant in $[0,1]$. Each fiber Γ_{x_i} has exactly two atoms. So, we choose y_i to be one of these two atoms. We have 2^n options in total of generating a geometric configuration. It can be shown that most of these choices represent different permutations, hence we obtain $c_{(1234)} \geq 2$.

5.3. **Permuton limits of typical pattern-avoiding permutations.** Consider the sequence of independent random permutations $\pi_1, \pi_2, \ldots$, where π_i is taken from $\mathbb{S}(i)$ uniformly at random. It is a well-known fact that such a sequence converges in the rectangular distance almost surely to the Lebesgue measure on $[0,1]^2$. What if we condition π_i to be free of a given pattern A?

Question 5.1. *Consider a pattern $A \in \mathbb{S}(k)$. Suppose that $\pi_1, \pi_2, \ldots$ is a sequence of independent random permutations, where π_i is taken from $\mathbb{S}(A;i)$ uniformly at random. Does this sequence converge in the rectangular distance almost surely? What is the limit permuton Δ_A?*

Theorem 1.1 asserts that the hypothetical permuton Δ_A is supported on the union of the graphs of finitely many functions.

Joshua Cooper [5] asked for two patterns A and B the expected density of B in a permutation chosen uniformly at random from $\mathbb{S}(A;n)$. If the answer to Question 5.1 is positive then it is easy to show that this expected density converges to $t(B, \Delta_A)$.

In a recent work Borga, Das, Mukherjee and Winkler [4] have introduced a Gibbs permutation model that might help to approach our question. This model does not include

the uniform distribution on permutations avoiding a fixed pattern A, but it allows to give an exponentially small probability w.r.t. the density of A to a permutation.

Question 5.1 is trivial for $A \in \mathbb{S}(2)$. For $A = (132)$, Section 5 of [2] tells us that for a permutation $\pi \in \mathbb{S}(A; n)$ we have $t(12, \pi) = o(1)$ asymptotically almost surely. This means that $\Delta_{(132)}$ is the antidiagonal permuton. Known bijections between $\mathbb{S}(123; n)$, $\mathbb{S}(132; n)$, $\mathbb{S}(213; n)$, $\mathbb{S}(231; n)$, $\mathbb{S}(312; n)$, and $\mathbb{S}(321; n)$ tell us that $\Delta_{(123)} = \Delta_{(213)}$ is also the antidiagonal permuton, whereas $\Delta_{(231)} = \Delta_{(312)} = \Delta_{(321)}$ is the diagonal permuton. For monotone increasing patterns it follows from a more general large deviations result [18] that $\Delta_{\mathsf{identity}_k}$ is the antidiagonal permuton. We do not know of other limit permutons of typical A-free permutations.

All the above examples lead to the diagonal or the antidiagonal limit permuton. Therefore, it is of particular interest to find limit permutons of a different shape. For the permutation $\sigma = (3142)$, suggested to us by Miklós Bóna, Δ_σ can neither be the diagonal nor the antidiagonal (if it exists). Indeed, suppose that Δ_σ is the diagonal. Then $\Delta_{\sigma^{-1}}$ must also be the diagonal. On the other hand, we can define the complement of a permutation $\pi \in \mathbb{S}(n)$ via $C(\pi)(i) = n + 1 - \pi(i)$ and observe that $\Delta_{C(\sigma)}$ then must be the antidiagonal contradicting that $C(\sigma) = (2413) = \sigma^{-1}$. One can argue similarly that Δ_σ cannot be the antidiagonal.

6. Acknowledgments

This work was initiated during the workshop *Interfaces of the Theory of Combinatorial Limits* at the Erdős center of the Rényi Institute. We thank the organisers for creating a very productive atmosphere. We thank Miklós Bóna, Joshua Cooper, Martin Doležal, Lara Pudwell and the two anonymous referees for comments and discussions.

References

[1] V. I. Bogachev. *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007.

[2] M. Bóna. The absence of a pattern and the occurrences of another. *Discrete Math. Theor. Comput. Sci.*, 12(2):89–102, 2010.

[3] M. Bóna. *Combinatorics of permutations.* Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL, second edition, 2012. With a foreword by Richard Stanley.

[4] J. Borga, S. Das, S. Mukherjee, and P. Winkler. Large Deviation Principle for Random Permutations. *International Mathematics Research Notices*, page rnad096, 05 2023.

[5] J. Cooper. Combinatorial Problems I like. Internet resource, https://people.math.sc.edu/cooper/combprob.html.

[6] J. N. Cooper. A permutation regularity lemma. *Electron. J. Combin.*, 13(1):Research Paper 22, 20, 2006.

[7] Persi Diaconis and Ronald L Graham. Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 39(2):262–268, 1977.

[8] M. Doležal, J. Hladký, and A. Máthé. Cliques in dense inhomogeneous random graphs. *Random Structures Algorithms*, 51(2):275–314, 2017.

[9] G. Elek and B. Szegedy. A measure-theoretic approach to the theory of dense hypergraphs. *Adv. Math.*, 231(3-4):1731–1772, 2012.

[10] P. Erdős and G. Szekeres. A combinatorial problem in geometry. *Compositio mathematica*, 2:463–470, 1935.

[11] J. Fox and F. Wei. Permutation property testing under different metrics with low query complexity. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1618–1637. SIAM, Philadelphia, PA, 2017.

[12] J. Fox and F. Wei. Fast property testing and metrics for permutations. *Combin. Probab. Comput.*, 27(4):539–579, 2018.

[13] C. Hoppen, Y. Kohayakawa, C. G. T. de A. Moreira, B. Ráth, and R. M. Sampaio. Limits of permutation sequences. *J. Comb. Theory, Ser. B*, 103(1):93–113, 2013.

[14] A. S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[15] S. Kitaev. *Patterns in permutations and words*. Monographs in Theoretical Computer Science. An EATCS Series. Springer, Heidelberg, 2011. With a foreword by Jeffrey B. Remmel.

[16] T. Klimošová and D. Král'. Hereditary properties of permutations are strongly testable. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1164–1173. ACM, New York, 2014.

[17] D. Král' and O. Pikhurko. Quasirandom permutations are characterized by 4-point densities. *Geom. Funct. Anal*, 23(2):570–579, 2013.

[18] N. Madras and L. Pehlivan. Large deviations for permutations avoiding monotone patterns. *Electron. J. Combin.*, 23(4):Paper 4.36, 20, 2016.

[19] A. Marcus and G. Tardos. Excluded permutation matrices and the Stanley–Wilf conjecture. *J. Comb. Theory, Series A*, 107(1):153–160, 2004.

Universität Heidelberg, Institut für Informatik, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany. Previous affiliation: Faculty of Informatics, Masaryk University, Botanická 68A, 602 00 Brno, Czech Republic.

Email address: garbe@informatik.uni-heidelberg.de

Institute of Computer Science of the Czech Academy of Sciences, Pod Vodárenskou věží 2, 182 00, Praha 8, Czechia.

Email address: hladky@cs.cas.cz

HUN-REN Alfréd Rényi Institute of Mathematics, Budapest, Hungary and Institute of Mathematics, Eötvös Loránd University, Budapest, Hungary.

Email address: kungabor@renyi.hu

Faculty of Informatics, Masaryk University, Brno, Czech Republic.

Email address: kristyna.pekarkova@mail.muni.cz