

Inhomogeneous random 2-SAT

Jan Hladký, Petr Savický

Institute of Computer Science of the Czech Academy of Sciences*

Abstract

We introduce an inhomogeneous variant of random 2-SAT. Each variable $v_1, \dots, v_n$ is assigned a type from a state space Λ , independently at random. Clause inclusion is governed by a symmetric measurable kernel W on $(\Lambda \times \{+, -\})^2$, in analogy with the inhomogeneous random graph model of Bollobás, Janson, and Riordan: given literals $\ell_i \in \{v_i, \neg v_i\}$ and $\ell_j \in \{v_j, \neg v_j\}$, the clause $\{\ell_i, \ell_j\}$ appears with probability $W(\text{type}(\ell_i), \text{type}(\ell_j))/(2n)$. In particular, for a variable v_i of type $x \in \Lambda$, the slices $W((+, x), \cdot)$ and $W((- , x), \cdot)$ describe how v_i and $\neg v_i$ interact with other literals.

We identify a parameter $\rho^*(W)$, defined as the spectral radius of an integral operator derived from W , and show that $\rho^*(W) < 1$ and $\rho^*(W) > 1$ correspond to asymptotically almost surely satisfiable and unsatisfiable instances, respectively. The satisfiability threshold of homogeneous random 2-SAT is long known to appear at the ratio 1 between the number of the clauses and the number of the variables which corresponds to $W \equiv 1$ and clause density $1/(2n)$. Our result extends this classical result to a broad class of models controlled by types of variables.

*Research supported by Czech Science Foundation Project 21-21762X. With institutional support RVO:67985807.

Contents

1	Introduction	3
1.1	Our model $\text{TwoSAT}(n, W)$	4
1.2	The statement of the result	5
2	Solving 2-SAT using the implication digraph	11
2.1	A convenient modification of $\text{TwoSAT}(n, W)$	12
2.2	Random digraph $\mathbb{G}(n, \vec{W})$	12
2.3	Basic idea behind the proof of the main theorem	13
2.4	Comparison to previous work on inhomogeneous random (di)graphs	14
3	Preliminaries	15
3.1	Measure theory	15
3.2	Digraphons	16
3.3	Banach space theory	16
3.4	More on digraphons	21
3.5	Proof of Proposition 1.12	21
4	Proof of Theorem 1.9(i)	22
4.1	Proposition 4.1 implies Theorem 1.9(i)	22
4.2	Proof of Proposition 4.1	22
5	Proof of Theorem 1.9(ii)	24
5.1	Proposition 5.1 implies Theorem 1.9(ii)	25
5.2	Proof of Proposition 5.1	26
6	Proof of Proposition 1.14	35
7	Further directions	36
7.1	Graphons with $\rho^*(W) = 1$	36
7.2	Asymptotic rate of convergence in Theorem 1.9	37
7.3	Higher clause sizes	37
	Index	37
	Bibliography	38

1 Introduction

For $k \in \{2, 3, \dots\}$, *random k -SAT* is one of the most studied problems in the intersection of combinatorics, probability theory, and theoretical computer science. It is a natural probabilistic model in the area of satisfiability, where the goal is to determine whether a given Boolean formula can be satisfied—i.e., whether there exists an assignment of truth values to variables that makes the entire formula evaluate to true.

In the k -SAT problem, the formula is expressed in conjunctive normal form and consists of variables $v_1, \dots, v_n$ and m clauses, where each clause is a disjunction of k literals. A literal is either a variable or its negation. A random k -SAT instance is generated by selecting each clause uniformly at random from the set of all possible clauses involving k distinct literals.

One of the key aspects of interest in random k -SAT is the behavior of the problem as the ratio of the number of clauses m to the number of variables n , denoted by $\alpha = \frac{m}{n}$, varies. As α increases, it is believed that the probability of the formula being satisfiable undergoes a sharp transition, commonly referred to as the *satisfiability threshold* α_k . Conjecturally, for values of α below α_k , a random k -SAT for α is satisfiable asymptotically almost surely (as $n \rightarrow \infty$), while for α above it, it becomes unsatisfiable asymptotically almost surely. As an alternative to parametrizing the model by n and m , we can parametrize the model by n and α . In this latter model, we form a formula by including each possible clause (there are $2^k \binom{n}{k}$ such clauses) with probability $\alpha n 2^{-k} \binom{n}{k}^{-1}$. So, the difference between the ‘ (n, m) -model’ and the ‘ (n, α) -model’ of random k -SAT is the same as the difference between the uniform model and the binomial model of Erdős–Rényi random graphs, and is insignificant for our purposes.

It is well-known that the k -SAT is in the computational complexity class P for $k = 2$ whereas it is NP-complete for every $k > 2$. In fact, it is known that each 2-SAT formula can be represented by an ‘implication digraph’ and that satisfiability of the formula then corresponds to a simple and computationally tractable problem of existence of certain ‘contradictory cycles’. There is a similar jump in the arduousness of the analysis of random k -SAT. In particular, in 1992, Goerdt [24] and independently Chvátal and Reed [13] proved the above ‘satisfiability conjecture’, for $k = 2$ and further determined that $\alpha_2 = 1$. Using the connection above, the bulk of their proofs is in fact about random digraphs. On the other hand, for every $k \geq 3$, even the existence of satisfiability threshold α_k was unknown (albeit Friedgut [20] got close). This changed only in 2014 with a preprint version of a *tour de force* paper [18], which used methods of statistical physics to determine α_k for all k sufficiently large. Among a huge body of work which looks at various aspects of random k -SAT, we chose [6, 14, 3, 2, 9] as representative references. We also refer to a slightly outdated survey [1].

While random k -SAT is an elegant mathematical model, many real-world scenarios that can be effectively represented by random models often involve constraints that are not uniformly distributed. A prominent example in this direction is random 2-SAT with prescribed literal degrees studied in [16]. For each n , this model is parametrized by integers $d_1, \bar{d}_1, d_2, \bar{d}_2, \dots, d_n, \bar{d}_n \in \mathbb{N}_0$ with $\sum_i (d_i + \bar{d}_i)$ even. Further, it is assumed that $\max_i \{d_i, \bar{d}_i\} \leq n^{1/11}$. A random formula is then taken uniformly from the set of all simple 2-SAT formulae¹ with exactly d_i occurrences of literal v_i and exactly $\bar{d}_i$ occurrences of literal $\neg v_i$. The

¹a 2-SAT formula is *simple* if it does not contain a repetition of any clauses, and also the literals of each clause arise from different variables

main result of [16] asserts that depending on whether $2 \sum_i d_i \cdot \bar{d}_i < (1 - \varepsilon) \sum_i (d_i + \bar{d}_i)$ or $2 \sum_i d_i \cdot \bar{d}_i > (1 + \varepsilon) \sum_i (d_i + \bar{d}_i)$, a random 2-SAT formula with prescribed literal degrees is asymptotically almost surely satisfiable or asymptotically almost surely unsatisfiable, respectively.

1.1 Our model $\text{TwoSAT}(n, W)$

Whenever we refer to a subset of a measure space, we implicitly assume it is measurable. Suppose that B is a probability space with measure β . Suppose that $p \in [1, \infty]$. Nonnegative functions in $L^p(B^2)$ are called L^p -*digraphons*. L^∞ -digraphons are simply called *digraphons*. L^p -digraphons which are symmetric with respect to the swap of the coordinates are called L^p -*graphons*. Again, L^∞ -graphons are simply called *graphons*. Note that compared to other literature, the values in graphons or digraphons in this paper are not necessarily bounded from above by 1. The combinatorial interpretation of the traditional bound of 1 is that each pair of vertices is connected by at most 1 edge. In this paper, however, we use graphons and digraphons mostly as sources of sparsified random graphs, that is, for a digraphon W , the edge inclusion probabilities are encoded in $\frac{W}{n}$, where $n \in \mathbb{N}$ is large.

Throughout the paper, Λ is an arbitrary Polish space and λ is a Borel probability measure on it. Let $\mathfrak{S} := \{+, -\}$ for the two-element probability space equipped with the uniform measure μ^{+-} . The space $\mathfrak{S}$ represents the positive and negative signs we equip the logical variables in our formula with. Let us consider the probability space $\mathbf{K} = \Lambda \times \mathfrak{S}$ with the corresponding product measure $\kappa = \lambda \times \mu^{+-}$. Define the negation map $\neg : \mathbf{K} \rightarrow \mathbf{K}$ by $\neg(x, +) = (x, -)$ and $\neg(x, -) = (x, +)$ for every $x \in \Lambda$. Let W be an L^1 -graphon on $\mathbf{K}$. For $n \in \mathbb{N}$, we define random 2-SAT formula $\phi \sim \text{TwoSAT}(n, W)$ on variables $\{v_1, \dots, v_n\}$ and let us denote the set of literals $\text{Lit}_n = \{v_1, \dots, v_n, \neg v_1, \dots, \neg v_n\}$. For this purpose, we sample elements $x_1, \dots, x_n \in \Lambda$ independently with distribution λ . For each pair $\{i, j\} \in \binom{[n]}{2}$ we insert clauses $\{v_i, v_j\}, \{v_i, \neg v_j\}, \{\neg v_i, v_j\}, \{\neg v_i, \neg v_j\}$ into ϕ independently at random with probabilities specified as follows. For $\mathbf{q}, \mathbf{s} \in \mathfrak{S}$, the clause $\{qv_i, sv_j\}$ is inserted with probability

$$\min \left\{ 1, \frac{W((x_i, \mathbf{q}), (x_j, \mathbf{s}))}{2n} \right\} \quad (1)$$

where we identify $+v$ with v and $-v$ with $\neg v$ for every logical variable v . Also, note that we view the clauses as unordered pairs, so that $\{qv_i, sv_j\} = \{sv_j, qv_i\}$. Therefore, the fact that W is symmetric is used in that (1) does not depend on the order of the literals.

1.1.1 Stochastic block model

While the number of clauses in $\text{TwoSAT}(n, W)$ is random, it is easy to see that it is concentrated (as $n \rightarrow \infty$) at $(1 \pm o(1))n\|W\|_1$. Note that the original (n, α) -model of random 2-SAT corresponds to $\text{TwoSAT}(n, \alpha)$, where α is the constant- α function. A class of models between the original random 2-SAT and the full generality of $\text{TwoSAT}(n, W)$ could be called ‘stochastic block model’.² This model is parametrized by the number of types $t \in \mathbb{N}$, the proportions $\gamma_1, \dots, \gamma_t > 0$, $\sum \gamma_i = 1$, and connection parameters $C_{i,j}^{\mathbf{q}, \mathbf{s}} \geq 0$, $\mathbf{q}, \mathbf{s} \in \mathfrak{S}$, $i, j \in [t]$ subject to symmetry $C_{i,j}^{\mathbf{q}, \mathbf{s}} = C_{j,i}^{\mathbf{s}, \mathbf{q}}$. For a given n , we take numbers $n_1, \dots, n_t$, where

²This term is borrowed from a related model of random graphs that first appeared in relation to a problem in sociology [28] and has been widely used since.

$n = n_1 + \dots + n_t$, $\frac{n_i}{n} \approx \gamma_i$ for each $i \in [t]$ (subject to some approximation condition) and Boolean variables $\{v_{i,k}\}_{i \in [t], k \in [n_i]}$. Each clause $\{qv_{i,k}, sv_{j,\ell}\}$ on distinct variables is included with probability $C_{i,j}^{q,s}/(2n)$. This, for many practical purposes, corresponds to $\text{TwoSAT}(n, W)$ if W is a graphon defined on $\Lambda \times \mathfrak{S}$, where $\Lambda = \Lambda_1 \sqcup \dots \sqcup \Lambda_t$, $\lambda(\Lambda_i) = \gamma_i$ and $W|_{(\Lambda_i \times \{q\}) \times (\Lambda_j \times \{s\})} \equiv C_{i,j}^{q,s}$. To see this, consider the stage of sampling elements $x_1, \dots, x_n \in \Lambda$ in the procedure of generating $\text{TwoSAT}(n, W)$. By the Law of Large Numbers, the number of indices $\ell \in [n]$ for which $x_\ell \in \Lambda_i$ (for a given $i \in [t]$) satisfies with high probability that $\frac{n_i}{n} \approx \gamma_i$. Given this event, the individual clauses are inserted with the same probability as in the stochastic block model.

1.2 The statement of the result

Our main result, Theorem 1.9, demonstrates that the threshold phenomenon known for the homogeneous random 2-SAT also extends to $\text{TwoSAT}(n, W)$. While the model $\text{TwoSAT}(n, W)$ is sensible for any nonnegative symmetric measurable function W on $\mathbf{K}^2$, in our main theorem we will impose mild additional integrability and operator-boundedness conditions. The critical parameter is characterized by the spectral properties of a specific operator on the Banach space $L^1(\mathbf{K})$ derived from the L^1 -graphon W . We introduce several concepts to state the result. In Definition 1.1 we introduce the implication L^p -digraphon $\vec{W}$. In Definition 1.2 we introduce a restriction of an L^p -digraphon to a set. In Definition 1.3 we introduce strong components of an L^1 -digraphon and in Theorem 1.4 we state the existence of decomposition of an L^1 -digraphon into its strong components, paralleling classical results on digraphs. In Definition 1.5 we define contradictory sets. Finally, in Definition 1.6 we recall the notions of eigenvalues and spectral radius of L^p -digraphons. So, while the central theme of the paper is random 2-SAT, some of the tools we develop have broader applications and are separated for clarity into [27]: the decomposition of digraphons into strong components and the study of their spectral properties.

In Section 1 we mentioned that a 2-SAT formula can be turned into an ‘implication digraph’. We recall this well-known transformation in Section 2. This transformation inspires a transformation of a graphon W (which parametrizes $\text{TwoSAT}(n, W)$) into an ‘implication digraphon’ $\vec{W}$. See Figure 1 for an illustration.

Definition 1.1 (implication digraphon). *Suppose that $p \geq 1$ is given. Given an L^p -graphon W on $\mathbf{K} = \Lambda \times \mathfrak{S}$, its implication L^p -digraphon $\vec{W}$ is an L^p -digraphon on $\mathbf{K}$ defined for $x, y \in \mathbf{K}$ by $\vec{W}(x, y) := W(\neg x, y)$.*

Note that the property of symmetry is in general lost in the implication digraphon. By using Definition 1.1, symmetry of W , and then Definition 1.1 again, we have

$$\vec{W}(x, y) := W(\neg x, y) = W(y, \neg x) = \vec{W}(\neg y, \neg x) \quad (2)$$

In the case of graphs, a corresponding property is called skew-symmetry.

Next, we define restriction of a digraphon to a subset of its ground set.

Definition 1.2 (restriction). *Suppose that $p \geq 1$ is given. For an L^p -digraphon U on a probability space (Ω, μ) , and for $A \subseteq \Omega$, let $U[A]$ be an L^p -digraphon on Ω defined by*

$$U[A](x, y) = \begin{cases} U(x, y) & \text{if } x, y \in A \\ 0 & \text{otherwise.} \end{cases}$$

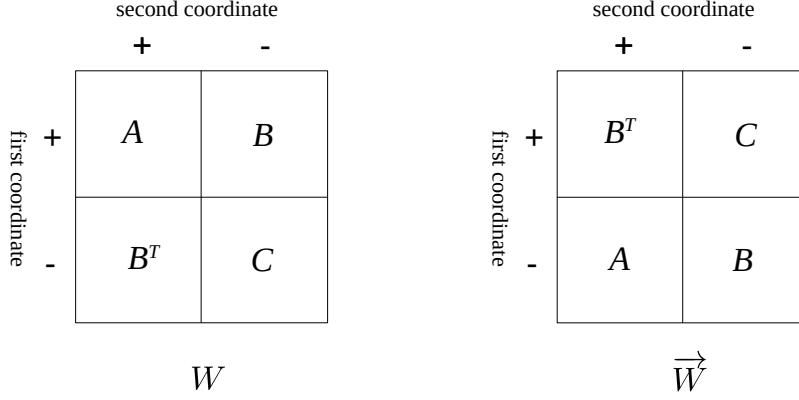


Figure 1: A visualization of the transformation of a graphon W into its implication digraphon $\vec{W}$. The graphon W consists of four parts: a symmetric part $A \in L^1((\Lambda \times \{+\}) \times (\Lambda \times \{+\}))$, a symmetric part $C \in L^1((\Lambda \times \{-\}) \times (\Lambda \times \{-\}))$ and a pair of mutually transposed parts B and B^T , where $B \in L^1((\Lambda \times \{+\}) \times (\Lambda \times \{-\}))$, $B^T \in L^1((\Lambda \times \{-\}) \times (\Lambda \times \{+\}))$. (The orientation of the plane here and elsewhere follows the matrix convention, that is, the main diagonal is in the $\searrow$ direction.)

In Definition 1.3 below, we introduce connectivity notions for digraphons taken from [27]. To motivate them by finite graphs, recall that a nonempty set of vertices C in digraph is *strongly connected* if for every partition $A \sqcup B = C$ into two nonempty sets, there is at least one directed edge going from A to B . Maximal strongly connected sets are called *strong components*. These notions have almost straightforward counterparts for digraphons except they are uniquely defined only modulo nullsets. The last notion in Definition 1.3 is of ‘fragmented sets’. This concept does not make any sense in finite digraphs. That is, one way to decompose a digraph into strong components is to start with an initial one-cell partition of the entire vertex set. If there is at any moment a cell violating the above condition on strong connectedness, we subdivide that cell accordingly. In a finite graph, this process must eventually terminate (since single vertices cannot be subdivided), resulting in the unique decomposition into connected components. In a digraphon, the regions where this process continues to split sets of positive measure into ever smaller ones constitute its fragmented sets.

Definition 1.3 (strongly connected set, strong component, fragmented set). *Suppose that Γ is an L^1 -digraphon on a probability space (Ω, μ) , and $X \subseteq \Omega$ is a set of positive measure.*

- (i) *We say that X is strongly connected in Γ if for every partition $A \sqcup B = X$ with $\mu(A), \mu(B) > 0$ we have $\int_{A \times B} \Gamma > 0$.*
- (ii) *We say that X is a strong component in Γ if X is strongly connected and for every $Y \subset \Omega$ with $\mu(X \cap Y) > 0$ and $\mu(Y \setminus X) > 0$ we have that Y is not strongly connected.*
- (iii) *We say that X is fragmented in Γ if every subset $Y \subset X$ of positive measure is not strongly connected.*

Recall that each digraph can be decomposed in a unique way into maximal strong components. Theo-

rem 1.4 below, which is one of the main results of [27], is a digraphon counterpart.³

Theorem 1.4 (Theorem 2.2 in [27]). *Suppose that Γ is an L^1 -digraphon on Ω . Then there exists a finite or a countable set I not containing 0 and a decomposition $\Omega = X_0 \sqcup \bigsqcup_{i \in I} X_i$ so that X_0 is either an empty set or is fragmented in Γ and each X_i is a strong component.*

Further, this decomposition is unique in the sense that if partitions $\{X_i\}_{i \in I \cup \{0\}}$ and $\{X'_i\}_{i \in I' \cup \{0\}}$ are two decompositions of Γ into strong components as above, then there exists a bijection $\pi : I \rightarrow I'$ such that $X'_{\pi(i)}$ equals X_i modulo a nullset for each $i \in I$, and X_0 equals to X'_0 modulo a nullset.

In Section 1 we mentioned that the analysis of random 2-SAT goes via translating a 2-SAT formula into its implication digraph, and its satisfiability boils down to the existence of contradictory cycles. We will introduce these concepts in Section 2. In our main theorem, Theorem 1.9, we determine which models $\text{TwoSAT}(n, W)$ yield almost surely satisfiable or almost surely unsatisfiable formulae. To this end, we decompose the implication digraphon $\vec{W}$ into strong connected components and disregard those components that cannot generate contradictory cycles. To this end, we use the concept of contradictory sets below.

Definition 1.5 (contradictory set). *A set $X \subseteq \mathbf{K}$ is contradictory if $\lambda(\{x \in \Lambda : (x, +), (x, -) \in X\}) > 0$.*

Spectral properties of digraphons are essential for formulating and proving our main result. While the spectral theory for graphons is well developed (see, e.g., Section 11.6 in [30]), its extension to digraphons poses nontrivial challenges. The main reason why spectral theory does not transfer directly is that graphons correspond to self-adjoint operators and thus admit a standard spectral decomposition, whereas digraphons generally do not. At the outset of this project, the spectral theory for digraphons was largely undeveloped. Addressing this gap — and, more broadly, developing basic tools related to spectral analysis and connectivity — led us to write a separate, self-contained paper [27]. In the meantime, Grebík, Král', Liu, Pikhurko, and Slipantschuk [26] posted a preprint which studies spectral properties of digraphons in the context of generating random oriented graphs, including an expression for the homomorphism density of oriented cycle in terms of the point spectrum. For now, we introduce only the essential definitions necessary to state Theorem 1.9, and recall the spectral tools from [27] in Section 3.3.

We fix $p \in [1, \infty)$. We work in the complex Banach space $L^p(\Omega)$, where Ω is a measure space (with an implicit sigma-algebra) equipped with a probability measure μ . For every L^1 -digraphon Γ on Ω , we may consider integral kernel operator $T_\Gamma : L^p(\Omega) \rightarrow L^p(\Omega)$, where for $f \in L^p(\Omega)$ the function $g := T_\Gamma(f)$ is defined by

$$g(x) := \int_{y \in \Omega} f(y) \Gamma(y, x) d\mu(y) \text{ for every } x \in \Omega.$$

Note that given $p \in [1, \infty)$, it is not automatic that $g \in L^p(\Omega)$, or even that the integral to define $g(x)$ is finite. Only when it is, we can view T_Γ as an integral kernel operator on $L^p(\Omega)$, and write $\Gamma \in \mathcal{B}(L^p(\Omega))$. Henceforth, any reference to a spectral concept of an L^1 -digraphon Γ will mean the corresponding concept for its associated integral kernel operator T_Γ (viewed on a Banach space $L^p(\Omega)$ for $p \in [1, \infty)$ which will be specified). We also write Γ instead of T_Γ .

³We state Theorem 1.4 in the generality of L^1 -digraphons whereas the original statement in [27] is only about digraphons. Given a general L^1 -digraphon Γ , we can however consider its *indicator digraphon* $\Gamma'(x, y) = \mathbb{1}_{\Gamma(x, y) > 0}$. It is easy to see that the notions of strong components and fragmented sets for Γ and for Γ' are equivalent, and thus we have reduced the general case of L^1 -digraphons to that of digraphons.

Definition 1.6 (eigenvalues, eigenfunctions, point spectrum, spectral radius). *Suppose that $p \in [1, \infty)$. Suppose that Γ is an L^1 -digraphon on Ω with the property that $\Gamma \in \mathcal{K}(L^p(\Omega))$. A complex number τ and a nonzero function $f \in L^p(\Omega)$ are called eigenvalue and eigenfunction of Γ with respect to the Banach space $L^p(\Omega)$, respectively, if $\Gamma(f) = \tau f$. The collection of all eigenvalues and 0 if Γ is not invertible is called the point spectrum of Γ with respect to the Banach space $L^p(\Omega)$, and denoted $\text{Spec}_p(\Gamma)$. The spectral radius of Γ , denoted $\rho_p(\Gamma)$, is the maximum modulus over the eigenvalues. For the most important choice $p = 2$, we write $\rho(\cdot) := \rho_2(\cdot)$.*

The definitions and the properties implicitly implied in Definition 1.6 are standard, see P7.3-5 in [29].

Remark 1.7. *Quite often, $\rho_p(\Gamma)$ does not depend on p , for a wide range of values of p . As a prominent example, it follows from Lemma 3.10 and Lemma 3.8(ii) that if $\Gamma \in L^p(\Omega^2)$ for some $p \in [2, \infty)$, then we have $\Gamma \in \bigcap_{q \in [2, p]} \mathcal{K}(L^q(\Omega))$ and for all $q \in [2, p]$ we have $\rho_p(\Gamma) = \rho_q(\Gamma)$. As another example which follows from Lemma 3.4, when Γ is a digraphon, then we have $\rho(\Gamma) = \rho_p(\Gamma)$ for all $p \in [1, \infty)$.*

Our main result, Theorem 1.9 below, asserts that the asymptotic almost sure (un)satisfiability is determined by a certain spectral parameter $\rho_p^*(W)$. We define this parameter here.

Definition 1.8. *Let W be an L^1 -graphon on $\mathbf{K} = \Lambda \times \mathfrak{S}$. Let $\mathbf{K} = \Omega_0 \sqcup \bigsqcup_{i \in I} \Omega_i$ be a decomposition of $\vec{W}$ into strong components. Let $I^* \subseteq I$ be the indices of the contradictory components Ω_i . For any $p \in [1, \infty)$ for which $\vec{W} \llbracket \bigcup_{i \in I^*} \Omega_i \rrbracket \in \mathcal{K}(L^p(\Omega))$, define $\rho_p^*(W) := \sup_{i \in I^*} \rho_p(\vec{W} \llbracket \Omega_i \rrbracket)$. For the most important choice $p = 2$, we write $\rho^*(\cdot) := \rho_2^*(\cdot)$.*

We can now state the main result, Theorem 1.9. The operator-theoretic assumption on the L^1 -graphon W involved in Part (ii) may look a bit technical. As we noted in Remark 1.7, it satisfied is when $W \in L^2(\mathbf{K}^2)$. In particular, when $W \in L^3(\mathbf{K}^2)$, both the satisfiability and the unsatisfiability part are applicable.

Theorem 1.9. *Let W be an L^1 -graphon on $\mathbf{K} = \Lambda \times \mathfrak{S}$.*

- (i) *If $W \in L^3(\mathbf{K}^2)$ and $\rho_3^*(W) < 1$, then as $n \rightarrow \infty$, $\text{TwoSAT}(n, W)$ is asymptotically almost surely satisfiable.*
- (ii) *Suppose that there exists $p \in [1, \infty)$ for which $W \in \mathcal{K}(L^p(\Omega))$, and $\rho_p^*(W) > 1$. Then, as $n \rightarrow \infty$, $\text{TwoSAT}(n, W)$ is asymptotically almost surely unsatisfiable.*

We prove Theorem 1.9(i) in Section 4 and Theorem 1.9(ii) in Section 5. In Section 2.3 we give a basic idea behind the proof.

Remark 1.10. *In many applications, including the stochastic block model from Section 1.1.1, the L^1 -graphon W is bounded in L^∞ . In that case, $\rho_p^*(W)$ in Definition 1.8 is defined for every $p \in [1, \infty)$ and does not depend on the choice of p . Other prominent examples where $\rho_p^*(W)$ stays constant for many choices of p were given in Remark 1.7.*

Most prominent applications of unbounded graphons involve random scale-free formulae. These are discussed in Section 1.2.1.

We believe that the integrability assumption $W \in L^3(\mathbf{K}^2)$ in Theorem 1.9(i) is an artifact of our proof and that it can be relaxed (at least) to $W \in L^2(\mathbf{K}^2)$.

Remark 1.11. *Theorem 1.9 does not cover the case $\rho^*(W) = 1$. In such a case, since the parameter $\rho^*(W)$ is easily seen to be multiplicative⁴, Theorem 1.9 tells us that for every $\varepsilon \in (0, 1)$, $\text{TwoSAT}(n, (1 - \varepsilon)W)$ and $\text{TwoSAT}(n, (1 + \varepsilon)W)$ are asymptotically almost surely satisfiable and asymptotically almost surely unsatisfiable, respectively. An open question concerning the asymptotic satisfiability of $\text{TwoSAT}(n, W)$ is given in Section 7.1.*

We complete this section with two propositions which may be useful to combine with Theorem 1.9. Proposition 1.12 asserts that each strong component Ω_i of $\vec{W}$ which is contradictory is in fact as contradictory as it can be. Namely, we have that $\Omega_i = Z \times \mathfrak{S}$ for some $Z \subseteq \Lambda$. This is a counterpart to a well-known and easy property of the implication digraph of a 2-CNF formula⁵ (see Section 2 for definition), namely that for each strong component of the implication digraph, we have that it either contains no pair of complementary literals or every literal comes with its complement.

Proposition 1.13 gives an alternative way to express the quantity $\rho^*(W)$ by considering not each contradictory component separately but rather all of them combined.

Proposition 1.12. *Let W be an L^1 -graphon on $\mathbf{K} = \Lambda \times \mathfrak{S}$. Let $\mathbf{K} = \Omega_0 \sqcup \bigsqcup_{i \in I} \Omega_i$ be the decomposition of $\vec{W}$ into strong components. Let $I^* \subseteq I$ be the indices of the contradictory components Ω_i . Then for each $i \in I^*$ we have that $\Omega_i =_0 Z_i \times \mathfrak{S}$ for some $Z_i \subseteq \Lambda$.*

Proposition 1.12 is proven in Section 3.5.

Proposition 1.13 (Proposition 5.1 and Remark 5.2 in [27]). *In the setting of Theorem 1.9, we have*

$$\sup_{i \in I^*} \rho(\vec{W}[\Omega_i]) = \max_{i \in I^*} \rho(\vec{W}[\Omega_i]) = \rho\left(\vec{W}\left[\bigcup_{i \in I^*} \Omega_i\right]\right) = \rho\left(\vec{W}\left[\Omega_0 \cup \bigcup_{i \in I^*} \Omega_i\right]\right).$$

Moreover, the spectral radii $\rho(\vec{W}[\Omega_i])$ and $\rho(\vec{W}[\bigcup_{i \in I^*} \Omega_i])$ are realized by nonnegative real eigenvalues, so the reference to the modulus in the definition of the spectral radius (Definition 1.6) is unnecessary in this case.

1.2.1 Theorem 1.9 and scale-free random formulae

Starting in the random graph community around 1999, researchers initiated study of random models in which vertices have heavy-tailed degree distributions. The most famous is arguably the dynamical Barabási–Albert model, [5]. For us, the static models of Norros–Reittu [33], and Chung–Lu [12] are more relevant. These models became known as ‘scale-free models’, and seem to capture many real-world scenarios. Section 16.4 of [10] shows that these models are roughly equivalent to the main model $\mathbb{H}(n, U)$ studied [10] when U is rank-1. We describe $\mathbb{H}(n, U)$ in Section 2.4. Recently, the scale-free phenomenon was introduced also to the setting of random k -SAT. Papers [21, 22] deal with scale-free random k -SAT for general k , and [4] deals with 2-SAT. Our Theorem 1.9 allows to study a number of scale-free models of random 2-SAT. Translating the main idea from Section 16.4 of [10] to the setting of random 2-SAT, we could take $\Lambda = (0, 1)$, and use kernels

⁴That is, $\rho^*(cW) = c\rho^*(W)$ for $c \geq 0$.

⁵2-CNF is an abbreviation for 2-Conjunctive Normal Form. These are boolean formula in conjunctive normal form with two literals per clause. So, each instance of random 2-SAT is a 2-CNF formula.

of the form $x^{-a}y^{-a}$ (for $a \in (0, 1)$) as building bricks. To give a particular example, take $W \in L^1(\mathbf{K}^2)$ defined as

$$\begin{aligned} W((x, +), (y, +)) &= x^{-\alpha}y^{-\alpha}, \\ W((x, +), (y, -)) &= W((y, -), (x, +)) = x^{-\gamma}y^{-\delta}, \text{ and} \\ W((x, -), (y, -)) &= (1-x)^{-\beta}(1-y)^{-\beta}. \end{aligned}$$

The purpose of this specific L^1 -graphon is that it illustrates the limitations on the exponents in Theorem 1.9 (which we discussed in Remark 1.10).

We must have $\alpha, \beta, \gamma, \delta < 1$ in order to satisfy $W \in L^1(\mathbf{K}^2)$. As for the additional requirement in Theorem 1.9(i), we must have $\alpha, \beta, \gamma, \delta < \frac{1}{3}$ in order to satisfy $W \in L^3(\mathbf{K}^2)$. As for the additional requirement in Theorem 1.9(ii), if we have $\alpha, \beta, \gamma, \delta < \frac{1}{2}$, then $\vec{W} \llbracket \bigcup_{i \in I^*} \Omega_i \rrbracket \in L^2(\mathbf{K}^2)$ and thus by Lemma 3.10, we have $\vec{W} \llbracket \bigcup_{i \in I^*} \Omega_i \rrbracket \in \mathcal{K}(L^2(\Omega))$.

1.2.2 Alternative scalings and graphons with $\rho^*(W) = 0$

In (1) we defined the probability of the inclusion of any one given clause to be of order $\frac{1}{n}$, resulting in $(1 \pm o(1))n\|W\|_1$ clauses in a typical instance of $\text{TwoSAT}(n, W)$. Some applications might call for other scalings, that is either for sparser models such as $\text{TwoSAT}(n, (\log n)^{-1} \cdot W)$ or $\text{TwoSAT}(n, n^{-0.5} \cdot W)$ or for denser models such as $\text{TwoSAT}(n, \log n \cdot W)$, $\text{TwoSAT}(n, \sqrt{n} \cdot W)$, or even $\text{TwoSAT}(n, n \cdot W)$. Note that $\text{TwoSAT}(n, f(n) \cdot W)$ defines the probability of the clauses using the scaled graphon $\frac{f(n)}{2n}W$. Theorem 1.9 together with multiplicativity from Footnote 4 has consequences on these other scalings as well. Indeed, first consider the sparser regime $\text{TwoSAT}(n, f(n) \cdot W)$, where $f(n) \rightarrow 0$. Clearly, there is n_0 such that $f(n) < 1/(2 \max\{1, \rho^*(W)\})$ for all $n \geq n_0$. We see that the random formula $\text{TwoSAT}(n, f(n) \cdot W)$ for $n \geq n_0$ is stochastically dominated as a set of clauses by $\text{TwoSAT}(n, U)$, where $U := \frac{W}{2 \max\{1, \rho^*(W)\}}$. Since $\rho^*(U) \leq \frac{1}{2}$, Theorem 1.9 tells us that $\text{TwoSAT}(n, U)$ is asymptotically almost surely satisfiable, and thus, so is $\text{TwoSAT}(n, f(n) \cdot W)$. Similarly, in the regime $f(n) \rightarrow \infty$, we see that $\text{TwoSAT}(n, f(n) \cdot W)$ eventually dominates $\text{TwoSAT}(n, U)$, where $U := \frac{2W}{\rho^*(W)}$, and thus is asymptotically almost surely unsatisfiable. This argument, however, does not work when $\rho^*(W) = 0$. The next proposition resolves this case entirely.

Proposition 1.14. *Suppose that W is an L^2 -graphon. The following statements are equivalent.*

- (i) *We have $\rho^*(W) = 0$.*
- (ii) *The L^2 -digraphon $\vec{W}$ contains no strong contradictory component.*
- (iii) *There exists a function $f : \mathbb{N} \rightarrow \mathbb{R}_+$ with $f(n) \rightarrow \infty$ so that $\text{TwoSAT}(n, f(n) \cdot W)$ is not asymptotically almost surely satisfiable.*
- (iv) *For every function $f : \mathbb{N} \rightarrow \mathbb{R}_+$ with $f(n) \rightarrow \infty$ we have that $\text{TwoSAT}(n, f(n) \cdot W)$ is asymptotically almost surely satisfiable.*

As a concrete — and somewhat surprising — example, if $\text{TwoSAT}(n, \log(\log n) \cdot W)$ is asymptotically almost surely satisfiable, then even the considerably denser $\text{TwoSAT}(n, n \cdot W)$ is as well. We give a proof of Proposition 1.14 in Section 6.

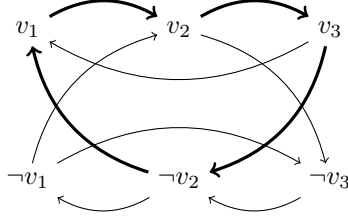


Figure 2: A formula $\phi = (\neg v_1 \vee v_2) \wedge (\neg v_2 \vee v_3) \wedge (\neg v_3 \vee v_1) \wedge (v_1 \vee v_2) \wedge (\neg v_3 \vee \neg v_2)$, and its implication digraph $D(\phi)$. One contradictory cycle is highlighted.

2 Solving 2-SAT using the implication digraph

The concept of the implication digraph of a 2-CNF formula translates the question of satisfiability into the language of graph theory. Suppose that ϕ is a 2-CNF formula on variables $\{v_1, \dots, v_n\}$. We create a digraph $D(\phi)$, called the *implication digraph* of ϕ , on the literals Lit_n as vertices as follows. For each clause $\{l_1, l_2\}$ we insert directed edges $(\neg l_1, l_2)$ and $(\neg l_2, l_1)$ representing the logically equivalent implications $\neg l_1 \rightarrow l_2$ and $\neg l_2 \rightarrow l_1$. As the edges are inserted in pairs, we see that for each pair l_i, l_j of literals on different variables, $D(\phi)$ contains either both directed edges (l_i, l_j) and $(\neg l_j, \neg l_i)$ or none of them. See Figure 2 for an example.

A directed closed walk or cycle⁶ C in a digraph D on vertex set Lit_n is *contradictory* if there is $i \in [n]$ such that $v_i, \neg v_i \in V(C)$. It is well-known and easy to verify that a 2-CNF formula ϕ is satisfiable if and only if $D(\phi)$ contains no contradictory closed walk. We will use the following stronger statement.

Proposition 2.1 ([25]). *Suppose that ϕ is a 2-CNF formula. Then ϕ is satisfiable if and only if $D(\phi)$ contains no contradictory cycle.*

Proposition 2.1 offers a plausible approach to proving Theorem 1.9. Namely, we need to prove that the implication digraph $D(\text{TwoSAT}(n, W))$ lacks or contains a contradictory cycle in the two respective regimes of Theorem 1.9, asymptotically almost surely. This idea drives the proofs in [13, 25] that the satisfiability threshold is $\alpha_2 = 1$. While the execution works straightforward for some parts of the proof, it turns out that the fact that a contradictory cycle may contain several pairs of complementary literals causes complications in others. To overcome this, [13] introduces a notion of a bicycle. We use a slightly modified form of a bicycle as follows. Suppose that we have integers $k, a, b, k \geq 2, 2 \leq a \leq k$, and $1 \leq b \leq k - 1$. Suppose that D is a digraph on Lit_n . A sequence $u_1, \dots, u_k$ of vertices of D is a *basis of a (k, a, b) -bicycle* if the literals $u_1, \dots, u_k$ are on pairwise different variables, $u_1 \cdots u_k$ is a directed path from u_1 to u_k , and further there is an edge from $\neg u_a$ to u_1 and an edge from u_k to $\neg u_b$ in D . We have the following.

Proposition 2.2. *Suppose that ϕ is an unsatisfiable 2-CNF formula. Then $D(\phi)$ contains a bicycle which is a subgraph of a contradictory cycle.*

Proof. By Proposition 2.1, $D(\phi)$ contains a contradictory cycle C . Let $P \subseteq C$ be a maximum path in C containing literals on pairwise different variables. P cannot contain all vertices of the cycle, since C contains

⁶in the context of directed graphs, we consider every closed walk or cycle directed

complementary literals. P will be the basis of our bicycle. Then, extend P by one edge at each end-point. The result is a bicycle, since C contains no literal twice. $\square$

2.1 A convenient modification of $\text{TwoSAT}(n, W)$

We introduce a convenient modification of the model $\text{TwoSAT}(n, W)$. Here, we recall that W is an L^1 -graphon on $\mathbf{K}$. First, sample $\phi \in \text{TwoSAT}(n, W)$. We create a formula $\phi^\dagger \sim \text{TwoSAT}^\dagger(n, W)$ by replacing some of the variables in ϕ by their negation. Each variable is replaced by its negation with probability $1/2$ and independently from all other variables. Note that ϕ and $\phi^\dagger$ are equisatisfiable, which means that either both are satisfiable or both are unsatisfiable. Indeed, we can take a satisfying assignment (if it exists) to one of these formulae, negate the values of the flipped variables, keep the values of the remaining variables, and we get a satisfying assignment of the other formula. The purpose of transforming $\text{TwoSAT}(n, W)$ to $\text{TwoSAT}^\dagger(n, W)$ is to simplify analysis of the random formula. The transformation introduces more symmetry into the distribution without affecting the satisfiability we are interested in.

The construction of $\text{TwoSAT}^\dagger(n, W)$ can be done in a single step as follows, and we assume this single-step construction in the rest of the paper. We sample elements $x_1, \dots, x_n \in \mathbf{K}$ independently with distribution κ and define a map $\tau : \text{Lit}_n \rightarrow \mathbf{K}$ so that for every i , $\tau(v_i) = x_i$ and $\tau(\neg v_i) = \neg \tau(v_i)$. Then, for every two literals $l_1, l_2 \in \text{Lit}_n$ on different variables, the clause $\{l_1, l_2\}$ is included in $\text{TwoSAT}^\dagger(n, W)$ with probability

$$\min \left\{ 1, \frac{W(\tau(l_1), \tau(l_2))}{2n} \right\}. \quad (3)$$

It is easy to verify that the two presented constructions of $\text{TwoSAT}^\dagger(n, W)$ are equivalent.

2.2 Random digraph $\mathbb{G}(n, \vec{W})$

By Proposition 2.1 we want to get a lower and an upper bound on the probability that the random digraph $D(\text{TwoSAT}^\dagger(n, W))$ contains a contradictory cycle. By construction, the edges of $D(\text{TwoSAT}^\dagger(n, W))$ are not independent, since they are included in pairs of the form (l_1, l_2) and $(\neg l_2, \neg l_1)$.⁷ In order to overcome the complications caused by these dependencies in the proof of Theorem 1.9(i) we introduce the following random digraph model $\mathbb{G}(n, U)$ where U is an L^1 -digraphon. The vertex set is Lit_n . Generate a random map $\tau : \text{Lit}_n \rightarrow \Omega$ as in Section 2.1 and for each ordered pair of literals $(l_1, l_2) \in \text{Lit}_n^2$ on different variables insert a directed edge (l_1, l_2) with probability

$$\min \left\{ 1, \frac{U(\tau(l_1), \tau(l_2))}{2n} \right\}, \quad (4)$$

independently of other choices.

Although the distributions of $\mathbb{G}(n, \vec{W})$ and of $D(\text{TwoSAT}^\dagger(n, W))$ are closely related, they are not the same. Indeed, $\mathbb{G}(n, \vec{W})$ can with positive probability contain a single edge, whereas $D(\text{TwoSAT}^\dagger(n, W))$ always has an even number of directed edges as explained above. However, restrictions of both these random digraphs on edge sets which do not contain pairs of equivalent edges yield the same marginal distributions. This is true, in particular, for bicycles.

⁷Recall the definition of the implication digraph in the first paragraph of Section 2.

Lemma 2.3. *Suppose that W is a graphon and $n \in \mathbb{N}$. Let $F \subseteq \text{Lit}_n^2$ be a set with the property that for every pair of literals $(l_1, l_2) \in \text{Lit}_n^2$ on different variables, we have $|F \cap \{(l_1, l_2), (\neg l_2, \neg l_1)\}| \leq 1$. Then $\mathbb{G}(n, \vec{W})|_F$ and $\text{D}(\text{TwoSAT}^\dagger(n, W))|_F$ have the same distribution. That is, for every $F' \subseteq F$, we have*

$$\mathbf{P} \left[\mathbb{G}(n, \vec{W}) \cap F = F' \right] = \mathbf{P} [\text{D}(\text{TwoSAT}^\dagger(n, W)) \cap F = F'] .$$

Proof. Recall that both models $\mathbb{G}(n, \vec{W})$ and $\text{TwoSAT}^\dagger(n, W)$ start with a random map $\tau : \text{Lit}_n \rightarrow \mathbf{K}$. Let $(l_1, l_2) \in F$ be arbitrary. The probability $\min \left\{ 1, \frac{\vec{W}(\tau(l_1), \tau(l_2))}{2n} \right\}$ in (4) for inclusion of the directed edge (l_1, l_2) is the same as the probability $\min \left\{ 1, \frac{W(\tau(\neg l_1), \tau(l_2))}{2n} \right\}$ for the inclusion of the clause $\{\neg l_1, l_2\}$ in (3). This is because of the way $\vec{W}$ is defined in Definition 1.1. The inclusion of the clause $\{\neg l_1, l_2\}$ is sufficient and necessary for the appearance of directed edge (l_1, l_2) in $\text{D}(\text{TwoSAT}^\dagger(n, W))$.

We conclude that (l_1, l_2) appears as a directed edge in $\mathbb{G}(n, \vec{W})$ and in $\text{D}(\text{TwoSAT}^\dagger(n, W))$ with the same probability. The occurrences of all possible edges in $\mathbb{G}(n, \vec{W})$, in particular, the edges of F , are independent conditionally on τ . Since the occurrences of edges of F in $\text{D}(\text{TwoSAT}^\dagger(n, W))$ are also independent conditionally on τ , the claim follows. $\square$

2.3 Basic idea behind the proof of the main theorem

We present the main idea behind the proof of Theorem 1.9. We rely on Proposition 2.1. That is, depending on whether either $\rho^*(W) < 1$ or $\rho^*(W) > 1$, we want to argue that the implication digraph of $\text{TwoSAT}^\dagger(n, W)$ asymptotically almost surely either does not or does contain a contradictory cycle, respectively. Since any contradictory cycle C itself is a strongly connected directed graph, we have (almost surely) that if $C \subset \text{D}(\text{TwoSAT}^\dagger(n, W))$ then all the vertices of C were sampled from one strong component Ω_i of $\vec{W}$, that is, (using the notation from Section 2.1) $\tau(l) \in \Omega_i$ for every $l \in V(C)$. Further, since C itself contains a pair of complementary literals, we have that Ω_i is (almost surely) a contradictory set.

We have therefore reduced the question of asymptotic almost sure nonexistence or existence of a contradictory cycle in $\mathbb{G}(n, \vec{W})$ to the same question in $\mathbb{G}(n, \vec{W}[\![\Omega_i]\!])$ for a single contradictory strong component Ω_i . It is our task to prove that a contradictory cycle asymptotically almost surely does not exist if $\rho(\vec{W}[\![\Omega_i]\!]) < 1$ and that it does exist if $\rho(\vec{W}[\![\Omega_i]\!]) > 1$. For this reason, set $\Gamma := \vec{W}[\![\Omega_i]\!]$ and $\rho := \rho(\Gamma)$.

Take $a \in \mathbb{N}$ and $b \in \mathbb{N}$. It turns out, that the regime when $a, b = \Theta(\log n)$ is the most relevant one. We first want to get bounds on the probability p that

$$v_n, v_1, v_2, \dots, v_{a-1}, v_a, \neg v_n, v_{a+1}, v_{a+2}, \dots, v_{a+b} \quad (5)$$

forms a contradictory cycle. We have (ignoring that (4) contains the term $\min\{1, \cdot\}$) that

$$p = \int_{x_n} \int_{x_1} \dots \int_{x_{a+b}} \frac{\Gamma(x_n, x_1)}{2n} \cdot \prod_{s=2}^a \frac{\Gamma(x_{s-1}, x_s)}{2n} \cdot \frac{\Gamma(x_a, \neg x_1)}{2n} \cdot \frac{\Gamma(\neg x_1, x_{a+1})}{2n} \cdot \prod_{s=a+2}^{a+b} \frac{\Gamma(x_{s-1}, x_s)}{2n} \cdot \frac{\Gamma(x_{a+b}, x_1)}{2n} d\kappa^{1+a+b} .$$

Such products correspond to powers of the digraphon Γ (as defined in Definition 3.1, very much in analogy with powers of square matrices). That is, we have

$$p = (2n)^{-(a+b+2)} \int_{x_n} \Gamma^{a+1}(x_n, \neg x_n) \cdot \Gamma^{b+1}(\neg x_n, x_n) d\kappa .$$

At this point, we use spectral theory. Since a reader might not be familiar with spectral theory of operators on a Hilbert space (and we certainly were not when we started this project), we rather draw an analogy with finite matrices.⁸ So, if we could treat Γ as a finite matrix with nonnegative entries, the Perron–Frobenius theorem would apply. It says that (under the assumptions of irreducibility and aperiodicity, the former one corresponds to Ω_i being a strong component, and the latter one is swept under the rug in this sketch) it tells us that $\Gamma^{a+1}(x_n, \neg x_n) = \Theta(\rho^{a+1})$ and $\Gamma^{b+1}(\neg x_n, x_n) = \Theta(\rho^{b+1})$. Note this sketched usage of the Perron–Frobenius theorem relies on the fact that Ω_i is contradictory. That is, the Perron–Frobenius theorem can only be used in the strongly connected setting, meaning, when there are paths in both directions between x_n and $\neg x_n$. The fact that Ω_i is contradictory provides this setting for a positive measure of x_n ’s.

To summarize, we have

$$p = \Theta(1) \cdot (2n)^{-(a+b+2)} \rho^{a+b+2}.$$

The number of rooted contradictory cycles of length $a + b + 2$ with the property that the only pair of complementary literals is the pair 1 and $(a + 2)$ -nd is equal to $2^{a+b+1}(n)_{a+b+1}$, where $(n)_{a+b+1}$ is a falling factorial. Indeed, we can think of this by replacing each of $v_n, v_1, v_2, \dots, v_{a+b}$ in (5) by arbitrary but distinct variables and additionally equipping that variable either with the positive or with the negative sign. Hence, the expected number of such all cycles is

$$p \cdot 2^{a+b+1}(n)_{a+b+1} = \Theta(1) \cdot (2n)^{-(a+b+2)} \rho^{a+b+2} \cdot 2^{a+b+1}(n)_{a+b+1} = \Theta(1) \cdot \frac{1}{n} \cdot \rho^{a+b+2}. \quad (6)$$

When $\rho < 1$ then this quantity is $o(1)$, even when summed over all a and b . We conclude that with high probability there are no contradictory cycles. (The above calculation is somewhat simplified since some contradictory cycle might use a several pairs of complementary literals. This is not captured in this sketch.)

When $\rho > 1$, then by choosing $a, b = \Theta(\log n)$ large enough, we have $\rho^{a+b+2} \gg n$. Hence the expectation in (6) goes to infinity. A second moment argument would be used to show that $\mathbb{G}(n, \Gamma)$ contains a contradictory cycle asymptotically almost surely.

In the actual proof, it is not contradictory cycles that are counted, but different structures: these are bicycles (for the proof of Theorem 1.9(i)) and what is later called ‘snakes’ (for the proof of Theorem 1.9(ii)).

2.4 Comparison to previous work on inhomogeneous random (di)graphs

One of the central results in the theory of random graphs is that the Erdős–Rényi random graph of order n and edge probability $\frac{c}{n}$ undergoes a phase transition at $c = 1$ with respect to the emergence of a giant component. Specifically, the largest component in such a graph is, with high probability, of order $o(n)$ if $c < 1$, and of order $\Theta(n)$ if $c > 1$.

Bollobás, Janson, and Riordan [10] studied a generalization of this phenomenon for a sparse inhomogeneous random graph model $\mathbb{H}(n, U)$, parametrized by an L^1 -graphon U defined on a space Λ . Let us recall the simplest version of their model.⁹ The vertex set of $\mathbb{H}(n, U)$ is $[n]$. To generate the edges, we first sample

⁸We introduce spectral theory of operators in Section 3.3.

⁹The full model of Bollobás, Janson, and Riordan is more general, informally allowing for different random perturbations for each n .

$x_1, \dots, x_n \in \Lambda$ independently according to the distribution λ . Each edge ij is then included independently with probability $\min \left\{ 1, \frac{U(x_i, x_j)}{n} \right\}$. This model has been a subject of extensive subsequent research.

Our model $\mathbb{G}(n, \overrightarrow{W})$ of random digraphs, introduced in Section 2.2, is very similar to this model—or rather, to its directed counterpart. This directed version was first explored by Bloznelis, Götze, and Jaworski [7], and later by Cao and Olvera-Cravioto [11]. We now explain these works and their relevance to our project.

A main result of [10] is a characterization of the existence versus absence of a giant component in $\mathbb{H}(n, U)$ for an L^1 -graphon U . Specifically, if $\rho(U) \leq 1$, then with high probability $\mathbb{H}(n, U)$ contains no giant component as $n \rightarrow \infty$, while if $\rho(U) > 1$, a giant component appears. The proof proceeds in two steps: first, a correspondence is established between the local neighborhood of a vertex in $\mathbb{H}(n, U)$ and an inhomogeneous Galton–Watson branching process $\mathcal{S}_U$; second, it is shown that $\rho(U) \leq 1$ corresponds to almost sure extinction of $\mathcal{S}_U$, whereas $\rho(U) > 1$ yields a positive survival probability. These arguments (included in Section 5 of [10]) are self-contained. Bloznelis, Götze, and Jaworski [7] provide a similar correspondence between branching processes and the size of the largest strong component for inhomogeneous random digraphs in the restricted case of a stochastic block model, but without connecting to spectral properties. Cao and Olvera-Cravioto [11] extended the analysis to a broader setting. They studied a variant of our model $\mathbb{G}(n, \Gamma)$ (extended similarly as described in Footnote 9), when Γ is continuous. The continuity condition is used heavily in the proof in [11], but we believe that the result is true even without it.¹⁰ More specifically the continuity assumption allows to reduce to the setting of the stochastic block model. In turn, the functional analytic/spectral side of the argument in [32] is rather rudimentary.

As explained in Section 2.3, our focus is not on the existence of a giant component as in [?, 11] but rather on the presence or absence of special (i.e., contradictory) cycles. However, the two concepts are clearly related: it is well-known, for instance, that sublinear components in Erdős–Rényi graphs contain only a small number of cycles, while the giant component contains exponentially many. Thus, the main novelty of our work is not in shifting attention from the giant component to the existence of particular cycles, but in employing advanced spectral techniques that enable us to analyze arbitrary digraphons Γ — without going through the approximation by stochastic block models.

3 Preliminaries

3.1 Measure theory

While parts of the paper specific to 2-SAT require to work in a the product space $\mathbf{K} = \Lambda \times \mathfrak{S}$, it is convenient to formulate some more basic parts with respect to a general probability space. We use a measure space Ω equipped with a probability measure μ to this end. That is, the measure μ always implicitly underlies the space Ω .

For measurable subsets A and B in a measure space, we write $A =_0 B$ and $A \subseteq_0 B$ for equality and containment modulo a nullset, respectively. We write $\text{ess inf } f$ and $\text{ess sup } f$ for the essential infimum and

¹⁰Similar continuity assumptions (probably also unnecessary) were used in other work on inhomogeneous random graphs, for example, regarding their clique number [32] or their connectivity [17].

essential supremum of a function f on a measure space.

3.2 Digraphons

Suppose that $p \in [1, \infty)$, Γ is an L^p -digraphon on Ω and D is an oriented graph on vertex set $[n]$. The *homomorphism density of D in Γ* is defined as

$$t(D, \Gamma) = \int_{x_1} \int_{x_2} \dots \int_{x_n} \prod_{(i,j) \in E(D)} \Gamma(x_i, x_j) d\mu^V. \quad (7)$$

This quantity may be infinite, but is certainly finite when Γ is a digraphon.

Suppose that $k \in \mathbb{N}$. Write P_k for the directed path $1, 2, 3, \dots, k, k+1$, and $P_k^{\bullet\bullet}$ for the directed path rooted at its terminal vertices. Write for the directed cycle $1, 2, 3, \dots, k$ and $C_k^\bullet$ for C_k rooted at vertex 1. We define $t_{x_1, x_{k+1}}^{\bullet\bullet}(P_k^{\bullet\bullet}, \Gamma)$ as a function of $x_1, x_{k+1} \in \Omega$ by disintegrating $t(P_k, \Gamma)$ with respect to x_1 and x_{k+1} . Likewise, we define $t_{x_1}^\bullet(C_k^\bullet, \Gamma)$ as a function of $x_1 \in \Omega$ by disintegrating $t(C_k, \Gamma)$ with respect to x_1 . That is,

$$\begin{aligned} t_{x_1, x_{k+1}}^{\bullet\bullet}(P_k^{\bullet\bullet}, \Gamma) &= \int_{x_2} \int_{x_3} \dots \int_{x_k} \prod_{i=1}^k \Gamma(x_i, x_{i+1}) d\mu^{k-1}, \\ t_{x_1}^\bullet(C_k^\bullet, \Gamma) &= \int_{x_2} \int_{x_3} \dots \int_{x_k} \prod_{i=1}^{k-1} \Gamma(x_i, x_{i+1}) \cdot \Gamma(x_k, x_1) d\mu^{k-1}. \end{aligned} \quad (8)$$

Next, we introduce the power of a digraphon. This definition is similar to the definition of matrix powers, and is also used in connection with integral kernel operators.

Definition 3.1 (Power of a digraphon). *Suppose that Γ is a digraphon on Ω , and let $k \in \mathbb{N}$. Define Γ^k as digraphon on Ω by $\Gamma^k(x, y) := t_{x,y}^{\bullet\bullet}(P_k^{\bullet\bullet}, \Gamma)$.*

It is straightforward to verify that the definition is consistent with operator powers, that is, for every $f \in L^2(\Omega)$ and every $k \in \mathbb{N}$ we have

$$\Gamma^k f = \underbrace{\Gamma(\Gamma(\dots(\Gamma(f))\dots))}_{k \text{ times}}.$$

Remark 3.2. *We will frequently use the concept of a power in the sense of Definition 3.1, as well as the notion of taking a power of the value of a digraphon. To distinguish between them, we write $\Gamma^k(x, y)$ for the former and $\Gamma(x, y)^k$ for the latter. For example, if Γ is a digraphon defined on the unit square, $\Gamma(x, y) := \mathbb{1}_{y \leq 0.5}$, then $\Gamma^k(0.1, 0.2) = 0.5^{k-1}$ and $\Gamma(0.1, 0.2)^k = 1$ for all $k \in \mathbb{N}$.*

3.3 Banach space theory

Our proof of Theorem 1.9 uses a fair amount of functional analysis. We summarize the tools we need in a way which should be accessible to discrete mathematicians.

As we said earlier, all our Hilbert/Banach spaces will be complex. We write $\mathbf{i}$ for the imaginary unit. Specifically, we will work with Banach spaces $L^p(\Omega)$ (for $p \in [1, \infty)$) and Hilbert space $L^2(\Omega)$. For a Banach space X , we write $\mathcal{B}(X)$ and $\mathcal{K}(X)$ for the set of bounded operators $X \rightarrow X$ and the set of compact operators $X \rightarrow X$, respectively.

Definition 3.3. Suppose that $p \in [1, \infty)$ is given. For an operator $T \in \mathcal{B}(L^p(\Omega))$, define its operator norm $\|T\|_{\text{op}(p)} = \sup_{f \in L^p(\Omega), \|f\|_p=1} \|Tf\|_p$.

Observe that for every $f \in L^p(\Omega)$,

$$\|Tf\|_p^p \leq \|T\|_{\text{op}(p)}^p \cdot \|f\|_p^p. \quad (9)$$

In Definition 1.6, we defined eigenvalues and eigenfunctions of an L^2 -digraphon Γ . These eigenvalues and eigenfunctions are sometimes also called *right*, *left eigenvalues* and *left eigenfunctions* of Γ are (right) eigenvalues and eigenfunctions of transposed L^2 -digraphon $\Gamma^\top$, $\Gamma^\top(x, y) := \Gamma(y, x)$.

The lemma below is standard and concerns digraphons as integral kernel operators.

Lemma 3.4. Suppose that Γ is a digraphon on Ω . Then for arbitrary $p \in [1, \infty)$, we have $\Gamma \in \mathcal{B}(L^p(\Omega))$. Further, consider an arbitrary eigenvalue $\gamma \neq 0$ and corresponding eigenfunction f , when Γ is viewed as an operator on $L^p(\Omega)$. Then $\|f\|_\infty \leq \frac{\|\Gamma\|_\infty \|f\|_1}{|\gamma|}$. Consequently, $\text{Spec}_p(\Gamma)$ and $\rho_p(\Gamma)$ does not depend on the choice of $p \in [1, \infty)$.

Proof. First, we show that $\Gamma \in \mathcal{B}(L^p(\Omega))$. Consider an arbitrary $h \in L^p(\Omega)$. We have $\|\Gamma h\|_p^p = \int_z \int_x |\Gamma(z, x)h(x)|^p \leq \|\Gamma\|_\infty^p \int_z |h(z)|^p < \infty$. Hence, $\Gamma h \in L^p(\Omega)$.

We now turn to L^∞ -boundedness of the eigenvalues. We have $\gamma f = \Gamma f$. Fix an arbitrary $x \in \Omega$. Thus, $|\gamma f(x)| = |\int_y f(y)\Gamma(y, x)| \leq \|\Gamma\|_\infty \cdot \|f\|_1$. $\square$

3.3.1 Kreĭn–Rutman theorem and beyond

Suppose that Γ is an digraphon on Ω and τ is its eigenvalue. We say that τ is *simple* if for every $f_1, f_2 \in L^2(\Omega)$ with $\Gamma f_i = \tau f_i$ we have that f_1 is a scalar multiple of f_2 .

We now introduce a version of the Kreĭn–Rutman Theorem. Let us give some background first. The Perron–Frobenius Theorem asserts that a real square matrix with positive entries has a unique nonnegative eigenvector, and that this eigenvector corresponds to the eigenvalue of the largest eigenvalue in absolute value. The Kreĭn–Rutman Theorem is often considered a counterpart of the Perron–Frobenius Theorem for nonnegative operators. While there are many version, we reproduce a version tailored to our language of digraphons, included as Theorem 3.16 in [27].

Theorem 3.5. Suppose that Γ is a strongly connected digraphon on Ω with $\rho(\Gamma) > 0$. Then for some $d \in \mathbb{N}$, the set of eigenvalues of maximal modulus is $\{\exp(-2\pi i k/d)\rho(\Gamma) : k = 0, \dots, d-1\}$. Further, there are right and left eigenfunctions v_{right} and v_{left} for the eigenvalue $\rho(\Gamma)$. Both these eigenfunctions are strictly positive and we have $v_{\text{right}}, v_{\text{left}} \in L^\infty(\Omega)$.

We call the number d from Theorem 3.5 the *peripheral multiplicity* of Γ .

We use the famous Gelfand’s formula (see P7.5-5 in [29]). Recall that the operator norm $\|R\|_{\text{op}}$ of an operator R on a Banach space $(X, \|\cdot\|)$ is defined as $\|R\|_{\text{op}} := \sup_{x \in X: \|x\| \leq 1} \|Rx\|$.

Proposition 3.6. Suppose that T is a bounded operator on a complex Banach space. Then we have $\rho(T) = \lim_{k \rightarrow \infty} \sqrt[k]{\|T^k\|_{\text{op}}}$.

Gelfand's formula holds for a variety of other norms. We recall another version, taken from [27] (though, it is very likely to be known). This version concerns the Hilbert–Schmidt norm, $\|\cdot\|_{\text{HS}}$. As we are concerned with L^2 -digraphons, we shall need the Hilbert–Schmidt norm only in the case of integral kernel operators, and thus we introduce it only in this particular setting. It is well-known that if T_K is an integral kernel operator on $L^2(\Omega)$ with kernel $K \in L^2(\Omega^2)$ then $\|T_K\|_{\text{HS}} = \|K\|_2$.

Proposition 3.7 (Proposition 3.11 in [27]). *Suppose that $\Gamma \in L^2(\Omega^2)$ is an L^2 -digraphon. Then we have $\rho(\Gamma) = \lim_{k \rightarrow \infty} \sqrt[k]{\|\Gamma^k\|_2}$.*

Gelfand's formula (Proposition 3.6) asserts that $\rho_p(T) = \lim_{k \rightarrow \infty} \left(\|T^k\|_{\text{op}(p)} \right)^{1/k}$. No additional assumptions on T are needed. The following lemma says that the spectral radius of a nonnegative operator is independent of the Banach space we work with in many scenarios. This is well known and we recall its proof for completeness. The statement and the proof of the lemma uses some basic notions (nonnegativity, irreducibility) from the theory of Banach lattices. The reader can find these notions explained in our language in Section 3.3.5 of [27]. What is important to us is that the nonnegativity of digraphons implies that they are nonnegative on each Banach lattice $L^q(\Omega)$.

Lemma 3.8. *Suppose that $1 \leq q \leq p < \infty$ are two numbers and T is a bounded operator, both as $T : L^q(\Omega) \rightarrow L^q(\Omega)$ and as $T : L^p(\Omega) \rightarrow L^p(\Omega)$.*

(i) *We have $\rho_q(T) \geq \rho_p(T)$.*

(ii) *If T is nonnegative (in the sense of Banach lattices) and $T \in \mathcal{K}(L^p(\Omega)) \cap \mathcal{K}(L^q(\Omega))$, then $\rho_p(T) = \rho_q(T)$.*

Proof. Part (i) is simple. Each eigenfunction of $T : L^p(\Omega) \rightarrow L^p(\Omega)$ is also an eigenfunction of $T : L^q(\Omega) \rightarrow L^q(\Omega)$ (with the same eigenvalue).

Let us turn to Part (ii). Without loss of generality, let's assume T is irreducible (if it were not, we would break it into irreducible parts). Let f be the eigenfunction corresponding to the eigenvalue of the maximum modulus when T is viewed as an operator on $L^p(\Omega)$, and let g be the eigenfunction corresponding to the eigenvalue of the maximum modulus when T is viewed as an operator on $L^q(\Omega)$. We will prove that f and g are the same (up to a multiple), which will prove the statement. The Kreĭn–Rutman theorem (applied twice, once for operators on $L^p(\Omega)$ and once for operators on $L^q(\Omega)$) tells us that f and g are in the positive cone (in other words, they are nonnegative function). Also, there is a uniqueness part of the Kreĭn–Rutman theorem, which asserts that (for irreducible operators), there is only one nonnegative eigenfunction. But in the space $L^p(\Omega)$ we seem to have two nonnegative eigenfunctions, namely f and g . So, this is only possible if f and g are the same (up to a multiple). $\square$

3.3.2 Approximating an integral kernel operator

In Section 5, we prove Theorem 1.9(ii). The calculation there do not work in the general case $W \in L^1(\Omega^2)$. So, as a first step in Section 5, we find a digraphon $V \in L^\infty(\mathbf{K}^2)$ with $V \leq W$ and $\rho^*(V) > 1$. This assumption of L^∞ -boundedness then allows us to compute that indeed $\text{TwoSAT}(n, V)$ is asymptotically almost surely unsatisfiable. Since $V \leq W$, the same is true for $\text{TwoSAT}(n, W)$. The following result is used for said approximation from below.

Proposition 3.9. *Suppose that we have $p \in [1, \infty)$ and W an L^1 -graphon on $\mathbf{K}$ with $W \in \mathcal{K}(L^p(\Omega))$. Then for every $\varepsilon > 0$, there exists a graphon V on $\mathbf{K}$ with the properties that $V \leq W$, and $\rho^*(V) \geq \rho_p^*(W) - \varepsilon$.*

Proof. We define a sequence of graphons $V_1 \leq V_2 \leq \dots \leq W$, where V_n is the pointwise minimum of W and the constant- n , $V_n := \min(W, n)$. Obviously, V_n is symmetric and bounded, and thus a graphon. Also, the decomposition of $\vec{V}_n$ into strong components is the same as of $\vec{W}$. So, the sets of contradictory components involved in the definition of $\rho_p^*(V_n)$ and of $\rho_p^*(W)$ are the same. Thus, we have $\lim_{n \rightarrow \infty} \rho_p^*(V_n) = \rho_p^*(W)$ by Theorem 2.4 in [34] (applied to the Banach space $L^p(\mathbf{K})$). We conclude that there exists a graphon $V \leq W$ with $\rho_p^*(V) \geq \rho_p^*(W) - \varepsilon$. Since V is a graphon, Lemma 3.4 applies. It tells us that the notion of eigenvalues and eigenfunctions of Γ does not depend on the choice of the space $L^q(\Omega)$ (over all $q \in [1, \infty)$). In particular, we have $\rho_p^*(V) = \rho^*(V)$. $\square$

3.3.3 L^p -digraphons as integral kernel operators on $L^p(\Omega)$

Suppose that W is an L^p -digraphon (in this section, we work with $p \in [2, \infty)$). In this section, we deduce that W as an integral kernel operator on $L^p(\Omega)$ is bounded and compact (Lemma 3.10) and express the spectral radius using a Gelfand-like formula involving the L^p -norm. While we were not able to find this result, we believe it could be known.

The first result is similar to Exercise 7 on page 177 in [15].

Lemma 3.10. *Suppose that $W \in L^p(\Omega^2)$ for some $p \geq 2$. Then W as an integral kernel operator on $L^p(\Omega)$ is bounded with $\|W\|_{\text{op}(p)} \leq \|W\|_p$. Furthermore, W is compact on $L^p(\Omega)$.*

Proof. Suppose that $f \in L^p(\Omega)$. First, we prove that Wf is well-defined as a function at almost every $x \in \Omega$. Let q be the Hölder conjugate to p , $\frac{1}{p} + \frac{1}{q} = 1$. We have that $q \leq 2 \leq p$. In particular, $\|f\|_q \leq \|f\|_p$. Also, notice that when $\|W\|_p < \infty$, then by Fubini's theorem, at almost every $x \in \Omega$ we must have $\|W(x, \cdot)\|_p < \infty$. Hence, using Hölder's inequality (HI),

$$|(Wf)(x)| = \left| \int W(x, y)f(y)d\mu(y) \right| \stackrel{\text{(HI)}}{\leq} \|W(x, \cdot)\|_p \cdot \|f\|_q \leq \|W(x, \cdot)\|_p \cdot \|f\|_p < \infty.$$

Let us now get a bound on the L^p -norm of Wf . We use the previous calculation

$$\|Wf\|_p^p = \int |(Wf)(x)|^p d\mu(x) \leq \int \|W(x, \cdot)\|_p^p \cdot \|f\|_p^p d\mu(x) \stackrel{\text{Fubini}}{=} \|W\|_p^p \cdot \|f\|_p^p.$$

Equivalently, $\|Wf\|_p \leq \|W\|_p \cdot \|f\|_p$. This shows that W as an integral kernel operator on $L^p(\Omega)$ is bounded with $\|W\|_{\text{op}(p)} \leq \|W\|_p$.

We now turn to proving compactness of W . Recall that the product sigma-algebra on Ω^2 is generated by sets of the form $S \times T$, where $S, T \subset \Omega$ are measurable. That is, it is known that for an arbitrary $\varepsilon > 0$, we can find a finite sequence $S_1, T_1, S_2, T_2, \dots, S_\ell, T_\ell \subset \Omega$ and coefficients $c_1, c_2, \dots, c_\ell \in \mathbb{R}$ so that for the function $W_\varepsilon := \sum_{i=1}^\ell c_i \mathbb{1}_{S_i \times T_i}$ we have $\|W - W_\varepsilon\|_p < \varepsilon$. By the previous (applied to $U = W - W_\varepsilon$), we have $\|W - W_\varepsilon\|_{\text{op}(p)} \leq \|W - W_\varepsilon\|_p < \varepsilon$. Also, the rank of W_ε is at most ℓ . That is, up to arbitrary precision, we are able to approximate W by a finite-rank operator in the operator norm. By a well-known fact (see e.g. Theorem 4.4 in [15]), it follows that W is compact. $\square$

There are many versions of Gelfand's formula (Proposition 3.6). We will need a different version, involving L^p -norms of kernels of a nonnegative L^p -digraphon.

Proposition 3.11. *Let $p \in [2, \infty)$. Suppose that W is an L^p -digraphon on Ω . Then we have $\rho_p(W) = \lim_{k \rightarrow \infty} \left(\|W^k\|_p \right)^{1/k}$.*

The next lemma is an important ingredient for the proof of Proposition 3.11. For it, we work with *composition of kernels*. That is, if $A, B \in L^p(\Omega^2)$, then we define $C := A * B$ as a function on Ω^2 , $C(x, y) := \int_z A(x, z)B(z, y)$ (provided, that the integral is defined). The next lemma tells us that under mild conditions, $A * B$ is indeed defined almost everywhere, and we have $A * B \in L^p(\Omega^2)$.

Lemma 3.12. *Let $p \in [1, \infty)$ be arbitrary. Let $A, B \in L^p(\Omega^2)$ be two kernels. Let T_A be the integral kernel operator associated with A , and assume that $T_A \in \mathcal{B}(L^p(\Omega))$. Set $C = A * B$. Then C is defined almost everywhere, $C \in L^p(\Omega^2)$, and $\|C\|_p^p \leq \|T_A\|_{\text{op}(p)}^p \cdot \|B\|_p^p$.*

Proof. For $x \in \Omega$, write C_x for the ‘‘column slice of C at x ’’. That is, C_x is a one-variable function, $C_x(y) := C(x, y)$. Fubini's Theorem gives

$$\|C\|_p^p = \int_x \|C_x\|_p^p d\mu(x).$$

With a similar slice convention for B , we have

$$\|B\|_p^p = \int_x \|B_x\|_p^p d\mu(x).$$

We will prove the lemma slice-wise, that is, we will prove that for every $x \in \Omega$, $\|C_x\|_p^p \leq \|T_A\|_{\text{op}(p)}^p \cdot \|B_x\|_p^p$. As $C = A * B$, we have $C_x = T_A(B_x)$. So, the inequality we need is just (9) applied to the function B_x and operator T_A . $\square$

Proof of Proposition 3.11. To get one inequality of the proposition, we use Gelfand's formula for the Hilbert–Schmidt norm (Proposition 3.7) and then use that $\|\cdot\|_2 \leq \|\cdot\|_p$ and $\rho(\cdot) = \rho_2(\cdot) \geq \rho_p(\cdot)$ (see Lemma 3.8(i)).

$$\rho_p(W) \leq \rho(W) = \lim_{k \rightarrow \infty} (\|W^k\|_2)^{1/k} \leq \liminf_{k \rightarrow \infty} \left(\|W^k\|_p \right)^{1/k}.$$

We now turn to the other inequality. Lemma 3.12 with $A = W^{k-1}$ and $B = W$ gives

$$\begin{aligned} \limsup_{k \rightarrow \infty} \left(\|W^k\|_p \right)^{1/k} &= \limsup_{k \rightarrow \infty} \left(\|W^k\|_p^p \right)^{1/(pk)} \leq \limsup_{k \rightarrow \infty} \left(\|T_W^{k-1}\|_{\text{op}(p)}^p \cdot \|W\|_p^p \right)^{1/(pk)} \\ &= \limsup_{k \rightarrow \infty} \left(\|T_W^{k-1}\|_{\text{op}(p)} \right)^{1/k} \cdot \lim_{k \rightarrow \infty} \left(\|W\|_p^p \right)^{1/(pk)}. \end{aligned}$$

The first term goes to the spectral radius $\rho_p(W)$ by Gelfand's formula (Proposition 3.6). The second term goes to 1. Hence, $\limsup_{k \rightarrow \infty} \left(\|W^k\|_p \right)^{1/k} \leq \rho_p(W) \cdot 1$, as was needed. $\square$

3.4 More on digraphons

3.4.1 Peripheral multiplicity and graphic periodicity of digraphons

In a directed graph, the *period* is defined as the greatest common divisor of the lengths of all its directed cycles. It is well-known that when this period is greater than 1, the vertex set can be partitioned into equivalence classes called *cyclic sets*. Edges in the digraph move vertices from one cyclic set to the next in a fixed cyclic order modulo the period. If the period is 1, the digraph is called *aperiodic*, and no nontrivial cyclic partition exists. Similar notion exists in the theory of Markov chains. In [27], a counterpart for digraphons was introduced.

Definition 3.13. *Suppose that Γ is a digraphon on Ω . For $d \in \mathbb{N}$, we say that Γ is graphically d -periodic if there exists a partition $\Omega = X_0 \sqcup X_1 \sqcup \dots \sqcup X_{d-1}$ such that (using the cyclic notation $X_d = X_0$) for every $j = 0, \dots, d-1$ we have $\Gamma|_{X_j \times (\Omega \setminus X_{j+1})} = 0$.*

The following was proven in [27].

Theorem 3.14 (Theorem 6.2 in [27]). *Suppose that Γ is a strongly connected digraphon on Ω . Suppose that the peripheral multiplicity of Γ is D . Then Γ is graphically D -periodic.*

3.4.2 Key asymptotics

The following is one of the main results of [27]. It asserts that high powers of a digraphon can be asymptotically expressed using the spectral radius, the left principal eigenfunction and the right principal eigenfunction.

Proposition 3.15 (Theorem 7.1 in [27]). *Suppose that Γ is a strongly connected digraphon on ground set Ω . We assume that there are left and right real eigenfunctions v_L, v_R for the eigenvalue $\rho(\Gamma)$ satisfying $\langle v_L, v_R \rangle = 1$.*

Let the peripheral multiplicity of Γ be D . Suppose that $\Omega = X_0 \sqcup X_1 \sqcup \dots \sqcup X_{D-1}$ is a decomposition as in Definition 3.13 provided by Theorem 3.14.

Let $\rho := \rho(\Gamma)$. There exists a number $\alpha \in (0, \rho)$ with the following property. For every $i, j \in \{0, \dots, D-1\}$ and every $x \in X_i$ and $y \in X_j$ we have

$$\Gamma^\ell(x, y) = \begin{cases} \rho^\ell v_R(x) v_L(y) + O(\alpha^\ell) & \text{if } \ell \equiv j - i \pmod{D}, \text{ or} \\ 0 & \text{otherwise,} \end{cases}$$

as $\ell \rightarrow \infty$. The term in $O(\cdot)$ does not depend on x and y .

3.5 Proof of Proposition 1.12

Suppose that $X \subset \mathbf{K}$ is a strong component of $\vec{W}$. Since X is a contradictory component, the set $P := \{z \in \mathbf{K} : z, \neg z \in X\}$ has positive measure. Define $R := X \setminus P$. Define $\bar{R} := \{\neg z : z \in R\}$. We shall prove that

$$P \sqcup R \sqcup \bar{R} \text{ is strongly connected.} \tag{10}$$

From that it will follow, using the fact that X is a strong component and Definition 1.3(ii), that $\bar{R}$ is null, which in turn yields that R is null. This will prove the statement.

To prove (10), consider a partition $A \sqcup B = P \cup R \cup \overline{R}$ into two arbitrary sets A and B of positive measure, as in Definition 1.3(i). We have $\int_{A \times B} \vec{W} \geq \int_{(A \cap X) \times (B \cap X)} \vec{W}$. The last term is positive by the fact that X is a strong component and Definition 1.3(i), provided that we prove that $A \cap X$ and $B \cap X$ have positive measure. So, it only remains to deal with the cases that $A \cap X$ or $B \cap X$ is null. Suppose for example that $A \cap X$ is null, the other case being analogous. Then $A \subset_0 \overline{R}$. In particular, for the set $\overline{A} := \{\neg z : z \in A\}$ we have $\overline{A} \subset R$. Also, the set $\overline{B} := \{\neg z : z \in B\}$ contains P . That means that $(\overline{B} \cap X) \sqcup \overline{A}$ is a partition of X into two sets of positive measures. By (2), we have $\int_{(A \cap X) \times (B \cap X)} \vec{W} = \int_{(\overline{B} \cap X) \times \overline{A}} \vec{W} > 0$, where the last inequality follows from Definition 1.3(i).

4 Proof of Theorem 1.9(i)

The bulk of the proof will deal with the following setting.

Proposition 4.1. *Suppose that U is an L^3 -digraphon on $\mathbf{K}$. By Lemma 3.10, we have that $\rho_3(U)$ is defined. If $\rho_3(U) < 1$, then for the random variable N_n counting the number of bicycles in $\mathbb{G}(n, U)$, we have $\mathbf{E}[N_n] \rightarrow 0$ as $n \rightarrow \infty$.*

4.1 Proposition 4.1 implies Theorem 1.9(i)

Let $U := \vec{W} \llbracket \bigcup_{i \in I^*} \Omega_i \rrbracket$. We have $U \in L^3(\mathbf{K}^2)$ and the associated integral kernel operator U is bounded and compact as an operator $U : L^3(\mathbf{K}) \rightarrow L^3(\mathbf{K})$. Restricting our attention to $\bigcup_{i \in I^*} \Omega_i$ is sensible as contradictory cycles cannot appear elsewhere. This is formally stated in the next lemma.

Lemma 4.2. *Suppose that $\phi \sim \text{TwoSAT}^\dagger(n, W)$ where in the sampling, we generated $\tau : \text{Lit}_n \rightarrow \Omega$. Then almost surely for each contradictory cycle $u_1, \dots, u_r$ in $D(\phi)$ we have that $\tau(u_1), \dots, \tau(u_r) \in \bigcup_{i \in I^*} \Omega_i$.*

Proof. This is just Proposition 4.10 in [27] translated to our setting. $\square$

Let M_n be the number of bicycles contained in some contradictory cycle in $D(\text{TwoSAT}^\dagger(n, W))$. Let N_n be the number of bicycles in $\mathbb{G}(n, U)$. Note that the edge set F of each potential bicycle satisfies the conditions of Lemma 2.3. From Lemma 2.3 and Lemma 4.2, we get

$$\mathbf{P}[N_n = 0] \leq \mathbf{P}[M_n = 0]. \quad (11)$$

We apply Proposition 4.1 and get that $\mathbf{E}[N_n] \rightarrow 0$, as $n \rightarrow \infty$. Markov's inequality and (11) gives that $M_n = 0$ asymptotically almost surely, as $n \rightarrow \infty$. By Proposition 2.2, $\text{TwoSAT}^\dagger(n, W)$ is asymptotically almost surely satisfiable.

4.2 Proof of Proposition 4.1

We set up constants α and C in a way which does not depend on n . Let $\alpha > 0$ be such that $\rho_3(U) + \alpha < 1$. Next, we set up C . By Proposition 3.11, we have $\rho_3(U) = \lim_{\ell \rightarrow \infty} (\|U^\ell\|_3)^{1/\ell}$. In particular, we can fix a constant $C > 0$ such that for every $\ell \in \mathbb{N}_0$, we have

$$\|U^\ell\|_3 \leq C(\rho_3(U) + \alpha)^\ell. \quad (12)$$

For each $k \geq 2$, $2 \leq a \leq k$, and $1 \leq b \leq k-1$, let the random variable $N_n^{k,a,b}$ count the number of (k, a, b) -bicycles in $\mathbb{G}(n, U)$. Let us first focus on the path $u_1 u_2 \dots u_k$ of any such bicycle. It is part of the definition of a bicycle that the variables in the literals u_i are pairwise distinct. By symmetry, the probability that any such sequence forms a basis of a (k, a, b) -bicycle is the same as for the sequence of positive literals $v_1, \dots, v_k$. That is, we have

$$\begin{aligned} \mathbf{E} [N_n^{k,a,b}] &= \binom{n}{k} k! \cdot 2^k \cdot \mathbf{P}_{\mathbb{G}(n,U)} [v_1, \dots, v_k \text{ forms a basis of a } (k, a, b)\text{-bicycle}] \\ &\leq n^k \cdot 2^k \cdot \mathbf{P}_{\mathbb{G}(n,U)} [v_1, \dots, v_k \text{ forms a basis of a } (k, a, b)\text{-bicycle}] . \end{aligned} \quad (13)$$

In order for $v_1, \dots, v_k$ to form a basis of a (k, a, b) -bicycle, the sequence $v_1, \dots, v_k$ has to form a path, and additionally, two extra edges, namely $(\neg v_a, v_1)$ and $(v_k, \neg v_b)$ have to be present.

We distinguish cases $a \geq b$ and $a < b$.

Case $a \geq b$.

We have

$$\begin{aligned} &\mathbf{P}_{\mathbb{G}(n,U)} [v_1, \dots, v_k \text{ forms a basis of a } (k, a, b)\text{-bicycle}] \\ &\leq \frac{1}{(2n)^{k+1}} \cdot \int_{x_a, x_b} U^b(\neg x_a, x_b) \cdot U^{a-b}(x_b, x_a) \cdot U^{k-a+1}(x_a, \neg x_b) . \end{aligned} \quad (14)$$

We use Hölder's inequality (HI) with exponents $\frac{3}{2}$ and 3, and the Cauchy-Schwarz Inequality (CSI),

$$\begin{aligned} &\int_{x_a, x_b} U^b(\neg x_a, x_b) \cdot U^{a-b}(x_b, x_a) \cdot U^{k-a+1}(x_a, \neg x_b) \\ &\stackrel{(\text{HI})}{\leq} \left(\int_{x_a, x_b} (U^b(\neg x_a, x_b))^{3/2} \cdot (U^{a-b}(x_b, x_a))^{3/2} \right)^{2/3} \cdot \left(\int_{x_a, x_b} (U^{k-a+1}(x_a, \neg x_b))^3 \right)^{1/3} \\ &\stackrel{(\text{CSI})}{\leq} \left(\int_{x_a, x_b} (U^b(\neg x_a, x_b))^3 \right)^{1/3} \cdot \left(\int_{x_a, x_b} (U^{a-b}(x_b, x_a))^3 \right)^{1/3} \cdot \left(\int_{x_a, x_b} (U^{k-a+1}(x_a, \neg x_b))^3 \right)^{1/3} \\ &= \|U^b\|_3 \cdot \|U^{a-b}\|_3 \cdot \|U^{k-a+1}\|_3 \\ &\stackrel{(12)}{\leq} C^3 (\rho_3(U) + \alpha)^{k+1} . \end{aligned}$$

We substitute this into (14),

$$\mathbf{P}_{\mathbb{G}(n,U)} [v_1, \dots, v_k \text{ forms a basis of a } (k, a, b)\text{-bicycle}] \leq C^3 \left(\frac{\rho_3(U) + \alpha}{2n} \right)^{k+1} ,$$

which can in turn be substituted into (13),

$$\mathbf{E} [N_n^{k,a,b}] \leq \frac{C^3}{2n} \cdot (\rho_3(U) + \alpha)^{k+1} . \quad (15)$$

Case $a < b$.

The calculations are similar. We have

$$\begin{aligned} &\mathbf{P}_{\mathbb{G}(n,U)} [v_1, \dots, v_k \text{ forms a basis of a } (k, a, b)\text{-bicycle}] \\ &\leq \frac{1}{(2n)^{k+1}} \cdot \int_{x_1, x_a, x_b, x_k} U(\neg x_a, x_1) U^{a-1}(x_1, x_a) \cdot U^{b-a}(x_a, x_b) \cdot U^{k-b}(x_b, x_k) U(x_k, \neg x_b) . \end{aligned} \quad (16)$$

We use Hölder's inequality (HI) with exponents $\frac{3}{2}$ and 3,

$$\begin{aligned}
& \int_{x_1, x_a, x_b, x_k} U(\neg x_a, x_1) U^{a-1}(x_1, x_b) \cdot U^{k-b}(x_b, x_k) U(x_k, \neg x_b) \cdot U^{b-a}(x_a, x_b) \\
& \stackrel{\text{(HI)}}{\leq} \left(\int_{x_1, x_a, x_b, x_k} (U(\neg x_a, x_1) U^{a-1}(x_1, x_a))^{3/2} \cdot (U^{k-b}(x_b, x_k) U(x_k, \neg x_b))^{3/2} \right)^{2/3} \cdot \left(\int_{x_a, x_b} (U^{b-a}(x_a, x_b))^3 \right)^{1/3} \\
& = \left(\int_{x_1, x_a} (U(\neg x_a, x_1) U^{a-1}(x_1, x_a))^{3/2} \right)^{2/3} \cdot \left(\int_{x_1, x_a} (U^{k-b}(x_b, x_k) U(x_k, \neg x_b))^{3/2} \right)^{2/3} \cdot \|U^{b-a}\|_3.
\end{aligned}$$

We can now use the Cauchy–Schwarz inequality on the first term,

$$\begin{aligned}
\left(\int_{x_1, x_a} (U(\neg x_a, x_1) U^{a-1}(x_1, x_a))^{3/2} \right)^{2/3} & \leq \left(\int_{x_1, x_a} (U(\neg x_a, x_1))^3 \right)^{1/3} \cdot \left(\int_{x_1, x_a} (U^{a-1}(x_1, x_a))^3 \right)^{1/3} \\
& = \|U\|_3 \cdot \|U^{a-1}\|_3.
\end{aligned}$$

We can now use the Cauchy–Schwarz inequality on the second term in the same way. Combined, we conclude that

$$\begin{aligned}
& \int_{x_1, x_a, x_b, x_k} U(\neg x_a, x_1) U^{a-1}(x_1, x_b) \cdot U^{k-b}(x_b, x_k) U(x_k, \neg x_b) \cdot U^{b-a}(x_a, x_b) \\
& \leq \|U\|_3 \cdot \|U^{a-1}\|_3 \cdot \|U\|_3 \cdot \|U^{k-b}\|_3 \cdot \|U^{b-a}\|_3 \\
& \stackrel{(12)}{\leq} C^5 (\rho_3(U) + \alpha)^{k+1}.
\end{aligned}$$

This allows us to get a counterpart to (15),

$$\mathbf{E} [N_n^{k,a,b}] \leq \frac{C^5}{2n} \cdot (\rho_3(U) + \alpha)^{k+1}. \quad (17)$$

Putting it together

We have

$$\mathbf{E}[N_n] \leq \sum_{k=2}^{\infty} \sum_{a=2}^k \sum_{b=1}^{k-1} \mathbf{E} [N_n^{k,a,b}] \stackrel{\text{by (15),(17)}}{\leq} \frac{1}{2n} (C^3 + C^5) \sum_{k=2}^{\infty} k^2 (\rho_3(U) + \alpha)^{k+1}.$$

Since $\rho_3(U) + \alpha < 1$, we obtain $\mathbf{E}[N_n] = O(\frac{1}{n})$. This proves Proposition 4.1.

5 Proof of Theorem 1.9(ii)

First, we use Proposition 3.9 to find a graphon $V \leq W$ with $\rho^*(V) > 1$.

Let Ω_i be an arbitrary contradictory component with $\eta := \rho(\vec{V}[\llbracket \Omega_i \rrbracket]) > 1$. Let $\mathbf{K}^* := \Omega_i$, and let κ^* be a measure on $\mathbf{K}$ naturally rescaled so that it becomes a probability measure, $\kappa^*(A) := \frac{\kappa(A)}{\kappa(\Omega_i)}$ for every $A \subset \Omega_i$. By Proposition 1.12, we have $\mathbf{K}^* =_0 \Lambda^* \times \mathfrak{S}$ and $\kappa^* = \lambda^* \times \mu^{+-}$ for a certain subspace $\Lambda^* \subset \Lambda$ and similarly rescaled probability measure λ^* on Λ^* . The negation map $\neg : \mathbf{K} \rightarrow \mathbf{K}$ from Section 1.1 naturally restricts to $\neg : \mathbf{K}^* \rightarrow \mathbf{K}^*$.

We define versions of V and $\vec{V}$ zoomed in to $\mathbf{K}^*$,

$$U \in L^\infty((\mathbf{K}^*)^2) \quad U(x, y) := V(x, y), \quad (18)$$

$$\Gamma \in L^\infty((\mathbf{K}^*)^2) \quad \Gamma(x, y) := \vec{V}(x, y). \quad (19)$$

In the remainder of this section, we will work (unless otherwise stated) within the space $(\mathbf{K}^*, \kappa^*)$. In particular, note that $\rho(\Gamma) = \frac{\eta}{\kappa(\Omega_i)}$.

Of course, we expect that there is a reason for unsatisfiability of $\text{TwoSAT}(n, W)$ within $\mathbf{K}^*$. This is expressed in Proposition 5.1 below. Let us do some preparations to state. For integers $a, b \geq 2$, an (a, b) -snake is any formula of the form

$$F = (\neg f \vee l_1) \wedge (\neg l_1 \vee l_2) \wedge \dots \wedge (\neg l_{a-1} \vee \neg f) \wedge (f \vee l_a) \wedge (\neg l_a \vee l_{a+1}) \wedge \dots \wedge (\neg l_{a+b-2} \vee f) \quad (20)$$

such that the variables of the literals $f, l_1, \dots, l_{a+b-2}$ are all different. We emphasize that equality of two snakes can include reordering of the clauses, for example

$$\begin{aligned} & (\neg v_{99} \vee v_1) \wedge (\neg v_1 \vee v_2) \wedge (\neg v_2 \vee \neg v_{99}) \wedge (v_{99} \vee v_3) \wedge (\neg v_3 \vee v_{99}) \\ &= (\neg v_{99} \vee \neg v_2) \wedge (\neg(\neg v_2) \vee \neg v_1) \wedge (\neg(\neg v_1) \vee \neg v_{99}) \wedge (v_{99} \vee v_3) \wedge (\neg v_3 \vee v_{99}) \\ &= (\neg v_{99} \vee v_1) \wedge (\neg v_1 \vee v_2) \wedge (\neg v_2 \vee \neg v_{99}) \wedge (v_{99} \vee \neg v_3) \wedge (\neg(\neg v_3) \vee \neg v_{99}) \\ &= (\neg v_{99} \vee \neg v_2) \wedge (\neg(\neg v_2) \vee \neg v_1) \wedge (\neg(\neg v_1) \vee \neg v_{99}) \wedge (v_{99} \vee \neg v_3) \wedge (\neg(\neg v_3) \vee \neg v_{99}) \end{aligned} \quad (21)$$

are 4 different choices of the literals $f, l_1, \dots, l_{a+b-2}$ which lead to the same $(3, 2)$ -snake, and no more choices exist. Later, we will take numbers a and b growing with n such that $a \neq b$ (see (25)). Then in general, there are exactly 4 different choices of the literals $f, l_1, \dots, l_{a+b-2}$ as in (20) that lead to the same (a, b) -snake.

For an (a, b) -snake F let $\text{Lit}(F)$ be the set of literals that have an occurrence in F . Note that $|\text{Lit}(F)| = 2(a + b) - 2$, since $\text{Lit}(F)$ is closed under negation, that is, if it contains a literal ℓ , it contains also $\neg\ell$.

When a and b are not specified, we call F simply a *snake*. The definition of snakes appeared previously (with minor modifications) in literature concerning 2-SAT (e.g. [23]). The importance of this definition is that any formula containing a snake as a subformula is obviously unsatisfiable.

Proposition 5.1. *Let U and η be as above. Let $\beta \in (\frac{\kappa(\Omega_i)}{\eta}, \infty)$ be arbitrary. As $n \rightarrow \infty$, $\text{TwoSAT}(\lfloor \beta n \rfloor, U)$ asymptotically almost surely contains a snake as a subformula.*

5.1 Proposition 5.1 implies Theorem 1.9(ii)

Below, we shall argue that the model $\text{TwoSAT}(n, V)$ is asymptotically almost surely unsatisfiable. Observe that this will imply the desired asymptotic almost sure unsatisfiability of $\text{TwoSAT}(n, W)$. Indeed, as $V \leq W$, the random clauses of $\text{TwoSAT}(n, V)$ are stochastically dominated by the random clauses of $\text{TwoSAT}(n, W)$.

Set $\beta := \frac{2\kappa(\Omega_i)}{1+\eta}$. The point of this choice is that we have $\beta \in (\frac{\kappa(\Omega_i)}{\eta}, \infty)$ (and so Proposition 5.1 applies) and $\beta < \kappa(\Omega_i)$ (which will be important when we apply the law of large numbers below). We take the asymptotics $n \rightarrow \infty$. Consider the way a random formula $\text{TwoSAT}(n, V)$ was generated in Section 1.1. Let $I \subset [n]$ be the set of indices i such that $x_i \in \Lambda^*$. For every $J \subset [n]$ of size $\lfloor \beta n \rfloor$, let $\mathcal{E}_J$ be the event that

$J \subset I$, and that $|I \cap \{1, 2, \dots, \max(J)\}| = |J|$. That is, $\mathcal{E}_J$ is the event that at least $\lfloor \beta n \rfloor$ many elements x_i were sampled from Λ^* , and that those $\lfloor \beta n \rfloor$ many one with the smallest indices i form the set J . Define the event $\mathcal{O}$ defined by $|I| < \lfloor \beta n \rfloor$. We have that the events $\mathcal{O}$ and $\{\mathcal{E}_J\}_J$ partition our probability space. By the law of large numbers, asymptotically almost surely, $|I| = (1 + o(1))\lambda(\Lambda^*)n = (1 + o(1))\kappa(\Omega_i)n$. In particular, $\mathbf{P}[\mathcal{O}] = o(1)$.

For any J as above, the formula $\text{TwoSAT}(n, W)$ in the conditional space $\mathcal{E}_J$ contains (up to renaming the variables) a formula $\text{TwoSAT}(\lfloor \beta n \rfloor, U)$. By Proposition 5.1, the latter formula contains a snake asymptotically almost surely, and thus is unsatisfiable asymptotically almost surely. We have

$$\begin{aligned} \mathbf{P}[\text{TwoSAT}(n, W) \text{ is satisfiable}] &\leq \mathbf{P}[\mathcal{O}] + \sum_{J \in \binom{[n]}{\lfloor \beta n \rfloor}} \mathbf{P}[\mathcal{E}_J] \cdot \mathbf{P}[\text{TwoSAT}(n, W) \text{ is satisfiable} \mid \mathcal{E}_J] \\ &\leq \mathbf{P}[\mathcal{O}] + \sum_J \mathbf{P}[\mathcal{E}_J] \cdot \mathbf{P}[\text{TwoSAT}(\lfloor \beta n \rfloor, U) \text{ is satisfiable}] \\ &\leq \mathbf{P}[\mathcal{O}] + \left(\sum_J \mathbf{P}[\mathcal{E}_J] \right) \cdot \mathbf{P}[\text{TwoSAT}(\lfloor \beta n \rfloor, U) \text{ is satisfiable}] \leq o(1) + 1 \cdot o(1). \end{aligned}$$

5.2 Proof of Proposition 5.1

In the proof, we treat the number $N := \lfloor \beta n \rfloor$ as $N = \beta n$. Let $\rho := \frac{\eta}{\kappa(\Omega_i)} = \rho(\Gamma)$.

Let D be the peripheral multiplicity of Γ . Fix a partition $\mathbf{K}^* = X_0 \sqcup X_1 \sqcup \dots \sqcup X_{D-1}$ as in Definition 3.13 provided by Theorem 3.14. Fix two indices $\iota_{\text{source}}, \iota_{\text{sink}} \in \{0, 1, \dots, D-1\}$ such that

$$\lambda^*(\{y \in \Lambda^* : (y, +) \in X_{\iota_{\text{source}}} \text{ and } (y, -) \in X_{\iota_{\text{sink}}}\}) > 0. \quad (22)$$

5.2.1 Setup for counting snakes

We first introduce an abstract version of our setup. That is, first we work in a general probability space. Suppose that $\{Z_s\}_{s \in \mathcal{S}}$ are indicator random variables indexed by a finite set $\mathcal{S}$. Let $Z = \sum_{s \in \mathcal{S}} Z_s$ be their sum. Let $\mathcal{T} \subset \mathcal{S}^2$ be an arbitrary superset of the set $\{(s_1, s_2) \in \mathcal{S}^2 : s_1 \neq s_2, \mathbf{E}[Z_{s_1} Z_{s_2}] \neq \mathbf{E}[Z_{s_1}] \mathbf{E}[Z_{s_2}]\}$. The Paley–Zygmund inequality tells us that

$$\begin{aligned} \mathbf{P}[Z = 0] &\leq \frac{\mathbf{E}[Z^2] - \mathbf{E}[Z]^2}{\mathbf{E}[Z^2]} = \frac{\sum_{(s_1, s_2) \in \mathcal{S}^2} (\mathbf{E}[Z_{s_1} Z_{s_2}] - \mathbf{E}[Z_{s_1}] \mathbf{E}[Z_{s_2}])}{\mathbf{E}[Z^2]} \\ &\stackrel{\boxed{\text{we have } Z_s^2 = Z_s}}{=} \frac{\sum_{s \in \mathcal{S}} (\mathbf{E}[Z_s] - \mathbf{E}[Z_s]^2) + \sum_{(s_1, s_2) \in \mathcal{T}} (\mathbf{E}[Z_{s_1} Z_{s_2}] - \mathbf{E}[Z_{s_1}] \mathbf{E}[Z_{s_2}])}{\mathbf{E}[Z^2]} \\ &\leq \frac{\sum_{s \in \mathcal{S}} \mathbf{E}[Z_s] + \sum_{(s_1, s_2) \in \mathcal{T}} \mathbf{E}[Z_{s_1} Z_{s_2}]}{\mathbf{E}[Z^2]} \\ &= \frac{\mathbf{E}[Z] + \sum_{(s_1, s_2) \in \mathcal{T}} \mathbf{E}[Z_{s_1} Z_{s_2}]}{\mathbf{E}[Z^2]}. \end{aligned} \quad (23)$$

We shall use the above setup for counting (a, b) -snakes on the set of variables $\{v_1, \dots, v_N\}$, where

$$a := 2D \cdot \lceil \log_\rho N \rceil + \iota_{\text{sink}} - \iota_{\text{source}} \quad \text{and} \quad b := 3D \cdot \lceil \log_\rho N \rceil + \iota_{\text{source}} - \iota_{\text{sink}}. \quad (24)$$

Note that

$$a \neq b \quad (25)$$

for N large enough.

That is, let $\mathcal{S}$ be the set of snakes on variables $\{v_1, \dots, v_N\}$. Taking into account the discussion around (21), we have $|\mathcal{S}| = 2^{a+b-3}(N)_{a+b-1}$, where $(N)_{a+b-1}$ is a falling factorial. If $a + b = o(\sqrt{N})$, then this gives

$$|\mathcal{S}| = \Theta((2N)^{a+b-1}). \quad (26)$$

For every $F \in \mathcal{S}$, let Z_F be an indicator variable representing the occurrence of F in our random formula $\text{TwoSAT}(N, U)$ obtained in the equivalent model $\text{TwoSAT}^\dagger(N, U)$ as described in Section 2.1. Let $Z = \sum_{F \in \mathcal{S}} Z_F$. For a formula F , write $\text{Vars}(F)$ for the set of all variables it contains. We define

$$\mathcal{T} := \{(F, G) \in \mathcal{S}^2 : F \neq G \text{ and } \text{Vars}(F) \cap \text{Vars}(G) \neq \emptyset\}. \quad (27)$$

Obviously, this $\mathcal{T}$ satisfies the assumptions above.

An (a, b) -snake F can be represented as a cycle of implications in several ways. Namely, F is equivalent to each of the following 4 different cycles

$$\begin{aligned} f &\rightarrow l_1 \rightarrow \dots \rightarrow l_{a-1} \rightarrow \neg f \rightarrow l_a \rightarrow \dots \rightarrow l_{a+b-2} \rightarrow f \\ f &\rightarrow \neg l_{a-1} \rightarrow \dots \rightarrow \neg l_1 \rightarrow \neg f \rightarrow l_a \rightarrow \dots \rightarrow l_{a+b-2} \rightarrow f \\ f &\rightarrow l_1 \rightarrow \dots \rightarrow l_{a-1} \rightarrow \neg f \rightarrow \neg l_{a+b-2} \rightarrow \dots \rightarrow \neg l_a \rightarrow f \\ f &\rightarrow \neg l_{a-1} \rightarrow \dots \rightarrow \neg l_1 \rightarrow \neg f \rightarrow \neg l_{a+b-2} \rightarrow \dots \rightarrow \neg l_a \rightarrow f \end{aligned} \quad (28)$$

If C is a rooted cycle in any of the forms (28) we call it an (a, b) -*serpent*. We let $\text{clauses}(C)$ denote the snake F . Moreover, let $\text{impl}(F)$ be the set of all implications equivalent to some of the clauses of F . Hence, $\text{impl}(F)$ is the union of the four serpents (28) (this union is generated by the first and the fourth serpent of (28), and also by the second and the third serpent of (28)).

5.2.2 First and second moment for snakes

The following lemma handles the probability of containment of a particular snake. Recall that U is bounded (see (18)).

Lemma 5.2. *The probability that any given (a, b) -snake F is contained in $\text{TwoSAT}(N, U)$ is*

$$\mathbf{P}[Z_F] = \frac{1}{(2N)^{a+b}} \int_{x \in \mathbf{K}^*} \Gamma^a(x, \neg x) \Gamma^b(\neg x, x) d\kappa^*(x), \quad (29)$$

when $N \geq 2\|U\|_\infty$.

Proof. The quantity $\mathbf{P}[Z_F]$ can be expressed using the way we generate random formulas in Section 2.1. That is, we integrate over the representatives $x, y_1, \dots, y_{a+b-2}$ of the literals of the snake, and use (3) on every clause of the snake (see (20)). This results in a $(a + b)$ -fold product of terms coming from (3) in the integral,

$$\begin{aligned} \mathbf{P}[Z_F] = \frac{1}{(2N)^{a+b}} \int_{x, y_1, \dots, y_{a+b-2} \in \mathbf{K}^*} & U(\neg x, y_1) U(\neg y_1, y_2) \dots U(\neg y_{a-1}, \neg x) \cdot \\ & \cdot U(x, y_a) U(\neg y_a, y_{a+1}) \dots U(\neg y_{a+b-2}, x) d\kappa^*(x) d\kappa^*(y_1) \dots d\kappa^*(y_{a+b-2}). \end{aligned}$$

Using the definition of Γ , this can be rewritten as

$$\mathbf{P}[Z_F] = \frac{1}{(2N)^{a+b}} \int_{x, y_1, \dots, y_{a+b-2} \in \mathbf{K}^*} \Gamma(x, y_1) \Gamma(y_1, y_2) \dots \Gamma(y_{a-1}, \neg x) \cdot \\ \cdot \Gamma(\neg x, y_a) \Gamma(y_a, y_{a+1}) \dots \Gamma(y_{a+b-2}, x) d\kappa^*(x) d\kappa^*(y_1) \dots d\kappa^*(y_{a+b-2}) .$$

Note that each of the variables $y_1, \dots, y_{a-1}$ has an occurrence in exactly two terms in the integrand and these terms have the form $\Gamma(z_1, y_i) \Gamma(y_i, z_2)$ with some other variables z_1 and z_2 . Since no other term contains y_i , the integration over y_i produces $\Gamma^2(z_1, z_2)$. Using this by induction, we can integrate over all the variables $y_1, \dots, y_{a-1}$ and replace the product of the terms containing them by $\Gamma^a(x, \neg x)$. Similarly, the terms containing $y_a, \dots, y_{a+b-2}$ can be replaced by $\Gamma^b(\neg x, x)$. $\square$

Observe that Proposition 3.15 and Lemma 3.4 tell us that there exists a constant $c \geq 0$ such that for every $\ell \in \mathbb{N}$,

$$\|\Gamma^\ell\|_\infty \leq c\rho^\ell . \quad (30)$$

The next lemma uses Lemma 5.2 to get the order of magnitude of $\mathbf{P}[Z_F]$.

Lemma 5.3. *For each (a, b) -snake F we have*

$$\mathbf{P}[Z_F] = \Theta \left(\left(\frac{\rho}{2N} \right)^{a+b} \right) . \quad (31)$$

Proof. For the upper bound, we use (30) and see that the integrand in (29) is at most $c^2 \rho^{a+b}$. Thus, $\mathbf{P}[Z_F] \leq c^2 \left(\frac{\rho}{2N} \right)^{a+b}$.

For the lower bound, we again invoke (29), but this time together with a more careful estimate directly from Proposition 3.15. Let $\alpha \in (0, \rho)$ be given by Proposition 3.15 for the digraphon Γ . Let v_L and v_R be the left and the right eigenvector for the eigenvalue $\rho(\Gamma)$ of Γ . By Theorem 3.5, the functions $v_L, v_R \in L^2(\mathbf{K}^*)$ are positive almost everywhere. Combining with (22), we see that there exists $\delta > 0$ such that for the set

$$R := \{y \in \Lambda^* : (y, +) \in X_{\iota_{\text{source}}}, v_L((y, +)) \geq \delta \text{ and } (y, -) \in X_{\iota_{\text{sink}}}, v_R((y, -)) \geq \delta\} ,$$

we have $\lambda^*(R) \geq \delta$. To obtain a lower bound on the integral in (29), we proceed as follows. We have

$$\int_{x \in \mathbf{K}^*} \Gamma^a(x, \neg x) \Gamma^b(\neg x, x) d\kappa^*(x) \geq \frac{1}{2} \int_{y \in R} \Gamma^a((y, +), (y, -)) \Gamma^b((y, -), (y, +)) d\lambda^*(y) . \quad (32)$$

For $y \in R$, we have $(y, +) \in X_{\iota_{\text{source}}}$ and $(y, -) \in X_{\iota_{\text{sink}}}$. By (24), we have that $a \equiv \iota_{\text{sink}} - \iota_{\text{source}} \pmod{D}$ and $b \equiv \iota_{\text{source}} - \iota_{\text{sink}} \pmod{D}$. Thus, we can continue with (32) with the help of the point estimate from Proposition 3.15,

$$\begin{aligned} & \int_{x \in \mathbf{K}^*} \Gamma^a(x, \neg x) \cdot \Gamma^b(\neg x, x) d\kappa^*(x) \\ & \geq \frac{1}{2} \int_{y \in R} (\rho^a v_R((y, +)) v_L((y, -)) + O(\alpha^a)) \cdot (\rho^b v_R((y, -)) v_L((y, +)) + O(\alpha^b)) d\lambda^*(y) \\ & \stackrel{\boxed{\text{def of } R}}{\geq} \frac{\delta}{2} (\rho^a \delta^2 + O(\alpha^a)) \cdot (\rho^b \delta^2 + O(\alpha^b)) = \Theta(\rho^{a+b}) . \end{aligned}$$

It suffices to plug this bound into (29). $\square$

Lemma 5.4. *For the sum $Z := \sum_{F \in \mathcal{S}} Z_F$ we have $\mathbf{E}[Z] = \Theta\left(\frac{\rho^{a+b}}{N}\right)$. In particular, by the choice of a and b in (24), we have $\mathbf{E}[Z] \rightarrow \infty$.*

Proof. We combine (26) and Lemma 5.3. □

To use the bound (23), we define $\Delta := \sum_{(F,G) \in \mathcal{T}} \mathbf{E}[Z_F Z_G]$. To prove Proposition 5.1, we will show that $\Delta = o(\mathbf{E}[Z^2])$. By symmetry, we can fix an arbitrary $F_0 \in \mathcal{S}$. Below, we write

$$F_0 = (\neg f \vee l_1) \wedge (\neg l_1 \vee l_2) \wedge \dots \wedge (\neg l_{a-1} \vee \neg f) \wedge (f \vee l_a) \wedge (\neg l_a \vee l_{a+1}) \wedge \dots \wedge (\neg l_{a+b-2} \vee f) .$$

We then have

$$\Delta = |\mathcal{S}| \sum_{G \in \mathcal{S}: (F_0, G) \in \mathcal{T}} \mathbf{E}[Z_{F_0} Z_G] . \quad (33)$$

Every term $Z_{F_0} Z_G$ in (33) is the indicator of the event the concatenation of the formulas F_0 and G appears in $\text{TwoSAT}^\dagger(N, U)$. We denote $Z_{F_0 \cup G} := Z_{F_0} Z_G$.

In order to get an upper bound on $\mathbf{E}[Z_{F_0 \cup G}]$, we use a generalization of the method used to prove Lemma 5.2. Consider an arbitrary mapping $\psi : \text{Vars}(F_0 \cup G) \rightarrow \mathbf{K}^*$. Extend it into a mapping $\psi : \text{Lit}(F_0 \cup G) \rightarrow \mathbf{K}^*$ by defining $\psi(\neg v) = \neg \psi(v)$. Let $t(F_0 \cup G, \Gamma, \psi)$ be the product of the weights of all clauses in $F_0 \cup G$, where a clause $(h \vee h')$ has weight

$$\frac{U(\psi(h), \psi(h'))}{2N} = \frac{\Gamma(\psi(\neg h), \psi(h'))}{2N} = \frac{\Gamma(\psi(\neg h'), \psi(h))}{2N} . \quad (34)$$

Moreover, we consider $\psi(v)$ for every $v \in \text{Vars}(F_0 \cup G)$ as a variable whose range is $\mathbf{K}^*$ and $t(F_0 \cup G, \Gamma, \psi)$ as a function of these variables. The expectation $\mathbf{E}[X_{F_0 \cup G}]$ will be expressed as an integral of $t(F_0 \cup G, \Gamma, \psi)$ over the variables in V , so we write

$$\mathbf{E}[X_{F_0 \cup G}] = \int_{\psi} t(F_0 \cup G, \Gamma, \psi) d((\mathbf{K}^*)^{\text{Vars}(F_0 \cup G)}) . \quad (35)$$

Given any serpent $C = (g \rightarrow l'_1 \rightarrow \dots \rightarrow l'_{a-1} \rightarrow \neg g \rightarrow l'_a \rightarrow \dots \rightarrow l'_{a+b-2} \rightarrow g)$, we say that C is *original* if we have $\text{clauses}(C) \neq F_0$. We say that C is *overlapping* if $(F_0, \text{clauses}(C)) \in \mathcal{T}$. Note that by (21), serpents are in a 4-to-1 correspondence to all (a, b) -snakes. That is, we can rewrite (33) as

$$\Delta = \frac{1}{4} |\mathcal{S}| \sum_{C \text{ original overlapping serpent}} \mathbf{E}[Z_{F_0 \cup C}] . \quad (36)$$

(Here, $Z_{F_0 \cup C}$ is a short for $Z_{F_0 \cup \text{clauses}(C)}$. Similarly, later on, $\text{Vars}(C)$ will be a short for $\text{Vars}(\text{clauses}(C))$.)

To get an upper bound on Δ , we classify serpents, and obtain bounds on $\mathbf{E}[Z_{F_0 \cup C}]$ depending on the particular classification of serpent C . This is done in the following definition. See also Figure 3 for an illustration.

Definition 5.5. *Suppose that C is an original serpent rooted at a literal g . A literal in C is called *free*, if it is not in $\text{Lit}(F_0) \cup \{g, \neg g\}$. A directed edge of C is an *intersection edge* if it belongs to $C \cap \text{impl}(F_0)$, and a *non-intersection edge* otherwise. A nonempty sequence of consecutive (i.e., increasing along C) non-intersection edges of C , such that the terminal vertices of the sequence are non-free literals and there is*

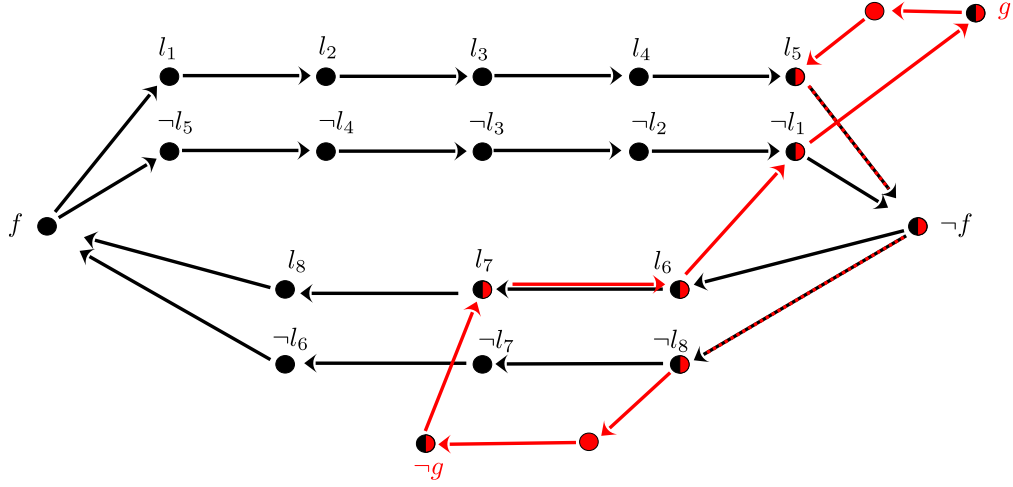


Figure 3: An example of a serpent when $a = 6$, $b = 4$. The serpent is depicted in red. Its free literals are depicted in full red circles, its non-free literals are depicted in half-red circles. Its intersection edges are depicted in dotted. The good non-intersection sequences written in the format $(start, end, length)$ are: $(g, l_5, 2)$, $(\neg l_8, \neg g, 2)$, $(\neg g, l_7, 1)$.

no other non-free literal in the sequence, is called a good non-intersection sequence. The length of a good non-intersection sequences is 1 plus the number of free literals in it.

Let $r(C)$ be the number of free literals in C . Let $s(C) = |C \setminus \text{impl}(F_0)|$.

We say that C is of Type I if $g \notin \text{Lit}(F_0)$. We say that C is of Type II if $g \in \text{Lit}(F_0) \setminus \{f, \neg f\}$. We say that C is of Type III if $g \in \{f, \neg f\}$.

Definition 5.6. Suppose that C is a serpent. We define the intersection pattern of C as the rooted cycle coming from C (with the same root) where we replace the free literals by positive literals on new variables $z_1, z_2, \dots, z_{r(C)}$ (not appearing in Lit_N) in the order defined by the numbering of the literals in C . Hence, the intersection pattern is uniquely determined by C .

For a given serpent C , let $[C]$ be the equivalence class of all serpents having the same intersection pattern, and let $\mathfrak{C}$ be the set of all these equivalence classes.

Note that all serpents C within one intersection pattern have the same type (Type I/Type II/Type III) and the same parameters $r(C)$ and $s(C)$. Thus we can define $r(\mathcal{E})$ and $s(\mathcal{E})$ for $\mathcal{E} \in \mathfrak{C}$.

Lemma 5.7. Suppose that C is an original serpent. The number of good non-intersection sequences of C is $s(C) - r(C) \geq 1$. The sum of the lengths of all good non-intersection sequences of C is equal to $s(C)$

Proof. Let g be the root of C . Since g and $\neg g$ are non-free literals, every maximal sequence of consecutive non-intersection edges and free literals has non-free terminal vertices, so it is a good non-intersection sequence. The number of edges in a good non-intersection sequence is 1 plus the number of free literals in it. Hence, it

contributes 1 to $s(C) - r(C)$. Moreover, there is at least one non-intersection edge, since $F_0 \neq \text{clauses}(C)$. This implies that there is at least one good non-intersection sequence, so we have $s(C) - r(C) \geq 1$.

The claim about the sum of the lengths is obvious. $\square$

Note that a serpent has no free literals if it goes through the set of literals $\text{Lit}(F_0)$ in a way different from F_0 .

Let us derive an upper bound on $\mathbf{E}[X_{F_0 \cup C}]$ in terms of the parameters $r(C)$ and $s(C)$.

Lemma 5.8. *Let C be a serpent. We have*

$$\mathbf{E}[Z_{F_0 \cup C}] \leq \Theta \left(\left(\frac{\rho}{2N} \right)^{a+b} c^{s(C)-r(C)} \left(\frac{\rho}{2N} \right)^{s(C)} \right) \quad (37)$$

Proof. Let g be the root of C . Consider the following partition of the set of variables $\text{Vars}(F_0 \cup C) = V_1 \sqcup V_2 \sqcup V_3$, where

$$\begin{aligned} V_1 &:= \text{Vars}(F_0) , \\ V_2 &:= \text{Vars}(\{g\}) \setminus \text{Vars}(F_0) , \\ V_3 &:= \text{Vars}(C) \setminus (\text{Vars}(\{g\}) \cup \text{Vars}(F_0)) . \end{aligned}$$

Note that V_2 is either empty or a singleton and V_3 are the variables from V corresponding to the free literals of C . The integral (35) becomes

$$\mathbf{E}[Z_{F_0 \cup C}] = \int_{\psi_1 \in (\mathbf{K}^*)^{V_1}} \int_{\psi_2 \in (\mathbf{K}^*)^{V_2}} \int_{\psi_3 \in (\mathbf{K}^*)^{V_3}} t(F_0 \cup C, \Gamma, \psi_1 \oplus \psi_2 \oplus \psi_3) d((\kappa^*)^{V_3}) d((\kappa^*)^{V_2}) d((\kappa^*)^{V_1}) . \quad (38)$$

First, we prove that

$$\mathbf{E}[Z_{F_0 \cup C}] \leq c^{s(C)-r(C)} \left(\frac{\rho}{2N} \right)^{s(C)} \cdot \int_{\psi_1 \in (\mathbf{K}^*)^{V_1}} t(F_0, \Gamma, \psi_1) d((\kappa^*)^{V_1}) . \quad (39)$$

Indeed, the integrand in (39) is the product of the weights of the edges of F_0 . This is the part of $t(F_0 \cup C, \Gamma, \psi_1 \oplus \psi_2 \oplus \psi_3)$ that is not affected by integration over V_3 and V_2 . We now need to incorporate the contribution of the non-intersection edges to $t(F_0 \cup C, \Gamma, \psi_1 \oplus \psi_2 \oplus \psi_3)$. We do it by looking at the good non-intersection sequences one-by-one. Consider each good non-intersection sequence, say consisting of ℓ edges, and use (30) and (34). After integrating over the $\ell - 1$ free variables in its free literals, we see from this that such a sequence contributes at most $c \cdot \left(\frac{\rho}{2N} \right)^\ell$.

Finally, we multiply the above upper bounds over all good non-intersection sequences. The number of these sequences is $s(C) - r(C)$ and the sum of their lengths is $s(C)$, so the product over all good sequences is the term

$$c^{s(C)-r(C)} \left(\frac{\rho}{2N} \right)^{s(C)}$$

in (39).

Lets now move from (39) to (37). To this end, it suffices to combine (39) with (31). $\square$

Proposition 5.9. *Let C be an arbitrary serpent. Then*

$$|\mathcal{S}| \sum_{C' \in [C]} \mathbf{E}[Z_{F_0 \cup C'}] \leq \Theta \left(\frac{1}{2N} \left(\frac{c}{2N} \right)^{s(C)-r(C)} \rho^{a+b+s(C)} \right) . \quad (40)$$

Proof. Each serpent $C' \in [C]$ is given by a choice of a replacement of the free literals of C . We have at most $(2N)^{r(C)}$ choices of the free literals. We combine this with (26) and (37) and get the desired bound. $\square$

For every combination of r, s which are admissible parameters $r(C)$ and $s(C)$ for serpents, let $M_{r,s,1}$ denote the number of intersection patterns of serpents of Type I such that the parameters r, s , have the given values. We define similarly $M_{r,s,2}$ and $M_{r,s,3}$ for serpents of Type II and Type III.

Lemma 5.10. *For any $r, s \in \mathbb{N}_0$, we have*

$$M_{r,s,1} \leq 4N(2(a+b))^{3(s-r)}, \quad (41)$$

$$M_{r,s,2} \leq 4(a+b)(2(a+b))^{3(s-r)}, \text{ and} \quad (42)$$

$$M_{r,s,3} \leq 4(2(a+b))^{3(s-r)}. \quad (43)$$

Proof. First, we look at (41). A specific intersection pattern of Type I and with parameters r, s is determined by the following choices

- one of at most $2N$ possible literals for the root g ,
- at most one of two choices ‘up’/‘down’ explained below,
- for each good non-intersection sequence, by a choice of its two terminal vertices and length. The terminal vertices are elements of $\text{Lit}(F_0) \cup \{g, \neg g\}$ which has size $2(a+b)$. The length of a good non-intersection sequence is an integer between 1 and $a+b$. By Lemma 5.7, there are $s-r$ good non-intersection sequences. Hence, we have at most $(2(a+b))^{2(s-r)}(a+b)^{s-r} \leq (2(a+b))^{3(s-r)}$ possibilities.

Let us verify that the set of intersection edges is uniquely determined by the above choices. These edges can be split into maximum sequences of consecutive edges in $\text{impl}(F_0)$. If g is contained as an interior point of some of these sequences, we split it further, so that g is an terminal vertex. The number of sequences of intersection edges is not important, so we can safely do this. After this, the terminal vertices of intersection sequences are determined by the root g and the terminal vertices of good non-intersection sequences (that is, by the end of one good non-intersection sequence and the beginning of the next one). There is one exception, an example of which is shown in Figure 4. Namely, Figure 4 shows two serpents, with two different intersection sequences, both both starting at l_6 and ending at l_1 . Such a situation can occur only for intersection sequences which entirely contain either one of the two branches of F_0 from f to $\neg f$ or one of the two branches of F_0 from $\neg f$ to f . Each serpent contains at most one such intersection sequence. Hence, this non-uniqueness can be encoded by two choices, say ‘up’/‘down’.

We now turn our attention to (42). The argument is as for intersection patterns of Type I, except that we can sharpen the bound on the number of choices of the root g from $2N$ to $2(a+b)-2$.

Last, we cover (43). The argument is again the same, with even a better bound on the number of choices of the root, namely $g \in \{f, \neg f\}$, i.e., 2 choices. $\square$

As we said, we prove Proposition 5.1 by using (23). For this, it remains to show that $\Delta = o(\mathbf{E}[Z^2])$. To this end, we use (36). We partition $\mathfrak{C} = \mathfrak{C}_1 \sqcup \mathfrak{C}_2 \sqcup \mathfrak{C}_3$ into intersection patterns whose serpents are of Type I, Type II, or Type III, respectively. That is, Proposition 5.1 will follow from the following three statements.

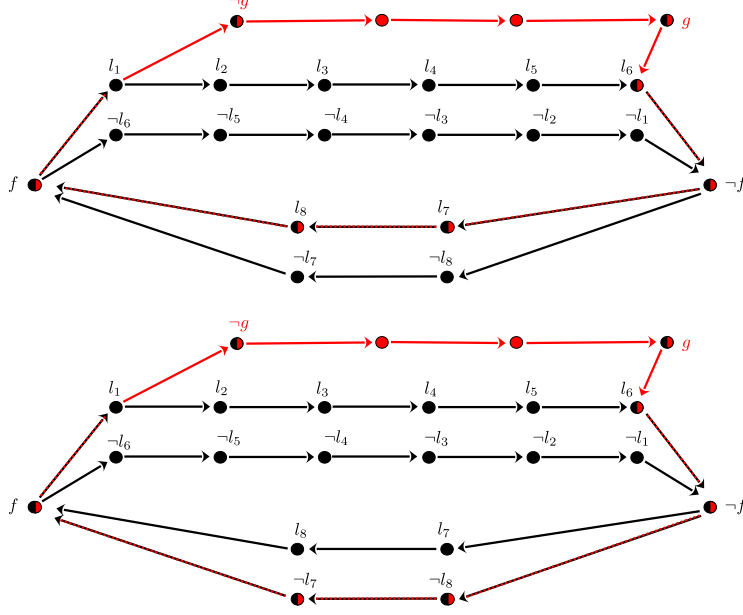


Figure 4: An example of two serpents which do not have the same intersection pattern but have the same set of good non-intersection sequences. Written in the format $(start, end, length)$, these are: $(g, l_6, 1)$, $(l_1, \neg g, 1)$, $(\neg g, g, 3)$. Here, $a = 7$, $b = 3$.

Lemma 5.11. *We have*

$$|\mathcal{S}| \sum_{\mathcal{E} \in \mathfrak{C}_1} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = o(\mathbf{E}[Z^2]) .$$

Lemma 5.12. *We have*

$$|\mathcal{S}| \sum_{\mathcal{E} \in \mathfrak{C}_2} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = o(\mathbf{E}[Z^2]) .$$

Lemma 5.13. *We have*

$$|\mathcal{S}| \sum_{\mathcal{E} \in \mathfrak{C}_3} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = o(\mathbf{E}[Z^2]) .$$

For the proofs of Lemmas 5.11, 5.12, 5.13, we partition $\mathfrak{C}_t$ (where $t = 1, 2, 3$) as $\mathfrak{C}_t = \sqcup_{r,s} \mathfrak{C}_{s,s,t}$, where r and s are the parameters of the intersection pattern as in Definition 5.6.

In fact, in the proofs of the three lemmas above, we shall prove that the left-hand sides are $o(\mathbf{E}[Z]^2)$, which is a slightly stronger bound. To this end, we shall use Lemma 5.4, which tells us that

$$\mathbf{E}[Z]^2 = \Theta \left(\frac{\rho^{2(a+b)}}{N^2} \right) . \quad (44)$$

5.2.3 Proof of Lemma 5.11

Let $r \in \mathbb{N}_0$, $s \in \mathbb{N}$. Multiplying (40) by (41) yields

$$|\mathcal{S}| \sum_{\mathcal{E} \in \mathfrak{C}_{r,s,1}} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = \Theta \left((2(a+b))^{3(s-r)} \left(\frac{c}{2N} \right)^{s-r} \rho^{a+b+s} \right) . \quad (45)$$

We claim that $\mathfrak{C}_{r,s,1} = \emptyset$ for $s - r < 3$. By Lemma 5.7, $s - r$ is the number of good non-intersection sequences. In order to prove that we have at least 3 such sequences, note that C_G contains at least 3 non-free literals: g , $\neg g$, and at least one further literal, say h from F_0 (indeed, otherwise we would have $\text{Vars}(F_0) \cap \text{Vars}(C) = \emptyset$ and C would not be overlapping). Each of the literals g and $\neg g$ is connected to two non-intersection edges in C . They contribute to $s - r$, but the contribution depends on whether they belong to different good sequences. Two of them may belong to the same sequence, if they are in the branch between g and $\neg g$ containing only free literals. However, one of the branches contains h , so this branch contributes at least 2 to $s - r$ and the other at least 1.

Now, fix $s \in \mathbb{N}$. The sum of (45) over decreasing $r \geq 1$ satisfying $s - r \geq 3$ is a finite geometric progression with the common ratio $(2(a+b))^3/(2N) = o(1)$. Hence,

$$|\mathcal{S}| \sum_{r \in \mathbb{N}: s-r \geq 3} \sum_{\mathcal{E} \in \mathfrak{C}_{r,s,1}} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = \Theta \left((2(a+b))^9 \left(\frac{c}{2N} \right)^3 \rho^{a+b+s} \right).$$

Since $\rho > 1$ is a constant, taking the sum over $1 \leq s \leq a+b$ yields

$$|\mathcal{S}| \sum_{\mathcal{E} \in \mathfrak{C}_1} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = \Theta \left((2(a+b))^9 \left(\frac{c}{2N} \right)^3 \rho^{2(a+b)} \right). \quad (46)$$

To finish the proof of the lemma, we use (44). That is, we need to show the right-hand side of (46) is $o\left(\frac{\rho^{2(a+b)}}{N^2}\right)$. By canceling $\rho^{2(a+b)}$ on both sides, we see that this is equivalent to showing that

$$(2(a+b))^9 \left(\frac{c}{2N} \right)^3 = o\left(\frac{1}{N^2}\right),$$

which is evident.

5.2.4 Proof of Lemma 5.12

Let $r \in \mathbb{N}_0$, $s \in \mathbb{N}$. Multiplying (40) by (42) yields

$$|\mathcal{S}| \sum_{\mathcal{E} \in \mathfrak{C}_{r,s,2}} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = \Theta \left(\frac{a+b}{N} (2(a+b))^{3(s-r)} \left(\frac{c}{2N} \right)^{s-r} \rho^{a+b+s} \right). \quad (47)$$

We claim that $\mathfrak{C}_{r,s,2} = \emptyset$ for $s - r < 2$. By Lemma 5.7, we only need to treat the case $s - r = 1$. So, suppose $s - r = 1$. Then one of the halfcycles of C , that is either the halfcycle starting at g and ending at $\neg g$ or the halfcycle starting at $\neg g$ and ending at g consists only of intersection edges and due to the structure of $\text{impl}(F_0)$ contains both literals f and $\neg f$. This is a contradiction, since C contains only one pair of contradictory literals.

Now, fix $s \in \mathbb{N}$. Similarly as in the previous case, taking the sum of (47) over decreasing $r \geq 1$ satisfying $s - r \geq 2$ yields

$$|\mathcal{S}| \sum_{r \in \mathbb{N}_0: s-r \geq 2} \sum_{\mathcal{E} \in \mathfrak{C}_{r,s,2}} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = \Theta \left(\frac{a+b}{N} (2(a+b))^6 \left(\frac{c}{2N} \right)^2 \rho^{a+b+s} \right)$$

and the sum over $1 \leq s \leq a+b$ is

$$|\mathcal{S}| \sum_{\mathcal{E} \in \mathfrak{C}_2} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = \Theta \left(\frac{a+b}{N} (2(a+b))^6 \left(\frac{c}{2N} \right)^2 \rho^{2(a+b)} \right). \quad (48)$$

To finish the proof of the lemma, we use (44). That is, we need to show the right-hand side of (48) is $o\left(\frac{\rho^{2(a+b)}}{N^2}\right)$. By canceling $\rho^{2(a+b)}$ on both sides, we see that this is equivalent to showing that

$$\frac{a+b}{N}(2(a+b))^6 \left(\frac{c}{2N}\right)^2 = o\left(\frac{1}{N^2}\right),$$

which is evident.

5.2.5 Proof of Lemma 5.12

Let $r \in \mathbb{N}_0$, $s \in \mathbb{N}$. Multiplying (40) by (43) yields

$$|\mathcal{S}| \sum_{\mathcal{E} \in \mathfrak{C}_{r,s,3}} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = \Theta \left(\frac{1}{N} (2(a+b))^{3(s-r)} \left(\frac{c}{2N}\right)^{s-r} \rho^{a+b+s} \right). \quad (49)$$

In order to get an upper bound, we distinguish two subcases, namely $s-r \geq 2$ and $s-r = 1$. We use the upper bound $M_{r,s,3}$ for each of these subcases, so we loose a factor of at most 2.

Now, fix $s \in \mathbb{N}$. Similarly as in the previous case, taking the sum of (49) yields

$$|\mathcal{S}| \sum_{r \in \mathbb{N}: s-r \geq 2} \sum_{\mathcal{E} \in \mathfrak{C}_{r,s,3}} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = \Theta \left(\frac{1}{N} (2(a+b))^6 \left(\frac{c}{2N}\right)^2 \rho^{a+b+s} \right).$$

and the sum over $1 \leq s \leq a+b$ is

$$|\mathcal{S}| \sum_{s \in \mathbb{N}} \sum_{r \in \mathbb{N}_0: s-r \geq 2} \sum_{\mathcal{E} \in \mathfrak{C}_{r,s,3}} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = \Theta \left(\frac{1}{N} (2(a+b))^6 \left(\frac{c}{2N}\right)^2 \rho^{2(a+b)} \right).$$

We immediately see, that this term is negligible compared to (44). It remains to get the same conclusion about the terms with $s-r = 1$. If $s-r = 1$, then (49) gives

$$|\mathcal{S}| \sum_{r \in \mathbb{N}: s-r=1} \sum_{\mathcal{E} \in \mathfrak{C}_{r,s,3}} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = \Theta \left(\frac{1}{N} (2(a+b))^3 \left(\frac{c}{2N}\right) \rho^{a+b+s} \right).$$

By Lemma 5.7, the above formula counts the contribution of intersection patterns with a single non-intersection sequence. One can verify that such intersection patterns satisfy $s \leq \max(a, b) = b$ (recall (24)). Taking the sum over $1 \leq s \leq b$ yields

$$|\mathcal{S}| \sum_{r \in \mathbb{N}: s-r=1} \sum_{\mathcal{E} \in \mathfrak{C}_{r,s,3}} \sum_{C \in \mathcal{E}} \mathbf{E}[Z_{F_0 \cup C}] = \Theta \left(\frac{(\log N)^3}{N^2} \rho^{a+2b} \right) \stackrel{(44)}{=} O \left(\frac{(\log N)^3}{\rho^a} \cdot \mathbf{E}[Z]^2 \right), \quad (50)$$

as was needed.

6 Proof of Proposition 1.14

Equivalence $(i) \Leftrightarrow (ii)$. This is Proposition 5.1(ii) in [27].

Implication $\neg(i) \Rightarrow \neg(iii)$. This is explained in Section 1.2.2, we just repeat the argument. Assume that $\rho^*(W) > 0$. Let f be an arbitrary function tending to infinity. We see that random clauses $\text{TwoSAT}(n, f(n) \cdot W)$ eventually stochastically dominate random clauses of $\text{TwoSAT}(n, U)$, where $U := \frac{2W}{\rho^*(W)}$. Since we have $\rho^*(U) = \frac{2\rho^*(W)}{\rho^*(W)} > 1$, Theorem 1.9 tells us that $\text{TwoSAT}(n, U)$ is asymptotically almost surely unsatisfiable. We conclude that $\text{TwoSAT}(n, f(n) \cdot W)$ is asymptotically almost surely unsatisfiable, too.

Implication $(iv) \Rightarrow (iii)$. This is obvious.

Implication $(ii) \Rightarrow (iv)$. We will introduce another model of random 2-SAT, $\text{DensestTwoSAT}(n, W)$. We will prove that assuming (ii),

(D1) for every $n \in \mathbb{N}$ and for every $c > 0$, random clauses of $\text{DensestTwoSAT}(n, W)$ stochastically dominate random clauses of $\text{TwoSAT}(n, c \cdot W)$, and

(D2) $\text{DensestTwoSAT}(n, W)$ is asymptotically almost surely satisfiable.

This will obviously prove that $\text{TwoSAT}(n, f(n) \cdot W)$ is asymptotically almost surely satisfiable for every function f . The model $\text{DensestTwoSAT}(n, W)$ is defined exactly like $\text{TwoSAT}(n, W)$ except that the insertion probability in (1) is replaced by

$$\begin{cases} 0 & \text{if } W((x_i, \mathfrak{q}), (x_j, \mathfrak{s})) = 0, \\ 1 & \text{if } W((x_i, \mathfrak{q}), (x_j, \mathfrak{s})) > 0. \end{cases}$$

It is obvious that (D1) is satisfied. Let us now turn to (D2). Let $\mathbf{K} = \Omega_0 \sqcup \bigsqcup_{i \in I} \Omega_i$ be the decomposition of $\vec{W}$ into strong components. By Proposition 4.10 in [27], we almost surely have the property that elements $y_{j_1}, y_{j_2}, \dots, y_{j_k} \in \mathbf{K}$ representing an arbitrary cycle, say $C = v_{j_1} v_{j_2} \cdots v_{j_k}$, in the implication digraph $D(\text{DensestTwoSAT}(n, W))$ lie within one strong component, say Ω_i . But as Ω_i is not a contradictory component, we have that C is not a contradictory cycle, almost surely. We conclude that $D(\text{DensestTwoSAT}(n, W))$ contains no contradictory cycles almost surely. Thus, $\text{DensestTwoSAT}(n, W)$ is satisfiable almost surely by Proposition 2.1.

7 Further directions

7.1 Graphons with $\rho^*(W) = 1$

If we take $W = \mathbb{1}$, then the result of [9] implies that the probability of satisfiability of $\text{TwoSAT}(n, \mathbb{1})$ lies in the interval $(c, 1 - c)$ for some $c > 0$. It is in fact reasonable to conjecture that the limit of $\mathbf{P}[\text{TwoSAT}(n, \mathbb{1}) \text{ is satisfiable}]$ exists. More generally, it could be, that for every graphon W with $\rho^*(W) = 1$, we have that the limit of $\mathbf{P}[\text{TwoSAT}(n, W) \text{ is satisfiable}]$ exists and lies in $(0, 1)$. The main focus of [9] is on the ‘scaling window’ of 2-SAT. This a phenomenon first described in the setting of the giant component of random graphs, [8, 31]. The main result of [9] asserts that

$$0 < \liminf_n \mathbf{P}[\text{TwoSAT}(n, (1 + Cn^{-1/3})\mathbb{1}) \text{ is satisf.}] \leq \limsup_n \mathbf{P}[\text{TwoSAT}(n, (1 + Cn^{-1/3})\mathbb{1}) \text{ is satisf.}] < 1$$

for every $C \in \mathbb{R}$, and

$$\liminf_n \mathbf{P}[\text{TwoSAT}(n, (1 - f(n))\mathbb{1}) \text{ is satisf.}] = 1 \quad \text{and} \quad \limsup_n \mathbf{P}[\text{TwoSAT}(n, (1 + f(n))\mathbb{1}) \text{ is satisf.}] = 0,$$

whenever $f : \mathbb{N} \rightarrow \mathbb{R}_+$ grows much faster than $n^{-1/3}$. The same question makes sense for any graphon with $\rho^*(W) = 1$. Specifically, it is natural to ask whether the critical scaling is always of order $\Theta(n^{-1/3})$. This phenomenon — often referred to as belonging to a ‘universality class’ — has been observed in many random discrete structures.

7.2 Asymptotic rate of convergence in Theorem 1.9

One might want to obtain the rate at which the probability of the satisfiability $\text{TwoSAT}(n, W)$ approaches 0 or 1 in Theorem 1.9. The main result of [9] asserts that if $W \equiv C$ is a constant graphon, then the probability of satisfiability is exponentially small when $C > 1$, and the probability of unsatisfiability is $\Theta(n^{-1})$ when $C \in (0, 1)$. We believe that if W is a graphon with $\rho^*(W) > 1$ then $\mathbf{P}[\text{TwoSAT}(n, W) \text{ is satisf.}] < \exp(-\varepsilon_W n)$ for some $\varepsilon_W > 0$ that depends on W . The case $\rho^*(W) < 1$ seems more delicate, allowing for different rates of the probability of satisfiability, depending on the structure of W .

7.3 Higher clause sizes

One could study inhomogeneous variants of random k -SAT for $k > 2$. In the simplest version, the model would be given by a symmetric bounded measurable function $W : \mathbf{K}^k \rightarrow [0, \infty)$. Here, *symmetric* means that $W(x_1, \dots, x_k) = W(x_{\pi(1)}, \dots, x_{\pi(k)})$ for every permutation π . For $n \in \mathbb{N}$, we define random k -SAT formula $\phi \sim \text{RandomkSAT}(n, W)$ on variables Lit_n by sampling elements $x_1, \dots, x_n \in \Lambda$ independently with distribution λ . For each tuple $\{i_1, \dots, i_k\} \in \binom{n}{k}$ and each tuple of signs $\mathfrak{s}_1, \dots, \mathfrak{s}_k \in \mathfrak{S}$, we insert the clause $\{\mathfrak{s}_1 v_{i_1}, \dots, \mathfrak{s}_k v_{i_k}\}$ into ϕ independently from other choices with probability

$$\min \left\{ 1, \frac{W((x_{i_1}, \mathfrak{s}_1), \dots, (x_{i_k}, \mathfrak{s}_k))}{n^{k-1}} \right\}.$$

The scaling $\frac{1}{n^{k-1}}$ is chosen so that the resulting formula typically has $\Theta(n)$ clauses, which is the order of magnitude relevant for the homogeneous random k -SAT.

The theory of hypergraph limits (as established in [19, 35]) offers even more complicated models. In those, the parameterizing function would not be k -dimensional, but rather $(2^k - 2)$ -dimensional, $U : \mathbf{K}^k \times \Lambda^{2^k - (k+2)} \rightarrow [0, \infty)$ (subject to certain symmetries). The right way to think about the power $2^k - (k+2)$ is that it represents all subsets of $[k]$ of sized $2, 3, \dots, k-1$. To create a random formula $\phi \sim \text{RandomkSAT}(n, U)$, we generate a random collection $\{x_S \in \Lambda\}_{S \subset [k], |S|=2,3,\dots,k-1}$. The clause $\{\mathfrak{s}_1 v_{i_1}, \dots, \mathfrak{s}_k v_{i_k}\}$ is inserted into ϕ independently from other choices with probability

$$\min \left\{ 1, \frac{U((x_{\{i_1\}}, \mathfrak{s}_1), \dots, (x_{\{i_k\}}, \mathfrak{s}_k), (x_S)_{S \subset [k], |S|=2,3,\dots,k-1})}{n^{k-1}} \right\}.$$

Given the difficulty of the homogeneous case of random k -SAT for $k \geq 3$ ([18]), it is hopeless to obtain a complete picture of the inhomogeneous model. But there are some intermediate goals. For example, for a given U , it is plausible that there exists $c_U \in (0, +\infty]$ such that for $c < c_U$, $\text{RandomkSAT}(n, cU)$ is asymptotically almost surely satisfiable and for $c > c_U$, $\text{RandomkSAT}(n, cU)$ is asymptotically almost surely unsatisfiable.¹¹

¹¹The case $c_U = \infty$ corresponds to U s for which $\text{RandomkSAT}(n, cU)$ is almost asymptotically surely satisfiable for every $c > 0$. This is a higher-dimensional counterpart to Proposition 1.14.

Index

(Λ, λ) , 4	Lit_n , 4
(Ω, μ) , 15	$\text{TwoSAT}(n, W)$, 4
$(\mathbf{K}, \kappa)$, 4	$\text{Spec}_p(\Gamma)$, 8
$(\mathfrak{S}, \mu^{+-})$, 4	$\text{ess inf } f, \text{ess sup } f$, 15
$=_0, \subseteq_0$, 15	$\mathfrak{C}$, 30
$C_k, C_k^\bullet$, 16	$\text{clauses}(C)$, 27
$P_k, P_k^{\bullet\bullet}$, 16	$\neg$, 4
$U[A]$, 5	$\vec{W}$, 5
$[C]$, 30	$\rho(\cdot)$, 8
$\mathcal{B}(L^p(\Omega))$, 7	$\rho^*(\cdot)$, 8
$\mathcal{B}(X)$, 16	$\rho_p^*(\cdot)$, 8
$\mathcal{K}(X)$, 16	$\rho_p(\cdot)$, 8
$\mathbb{G}(n, U)$, 12	$r(\mathcal{C}), s(\mathcal{C})$, 30
Γ^k , 16	$r(\mathcal{E}), s(\mathcal{E})$, 30
$\mathbf{i}$, 16	$t(D, \Gamma)$, 16
$\text{D}(\phi)$, 11	$t_{x,y}^{\bullet\bullet}(P_k^{\bullet\bullet}, \Gamma)$, 16
$\text{Lit}(F)$, 25	$t_x^\bullet(C_k^\bullet, \Gamma)$, 16

References

- [1] ACHLIOPTAS, D. Random satisfiability. In *Handbook of Satisfiability*, A. Biere, M. Heule, H. van Maaren, and T. Walsh, Eds., vol. 185 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, 2009, pp. 245–270.
- [2] ACHLIOPTAS, D., COJA-OGHLAN, A., HAHN-KLIMROTH, M., LEE, J., MÜLLER, N., PENSCHUCK, M., AND ZHOU, G. The number of satisfying assignments of random 2-SAT formulas. *Random Structures Algorithms* 58, 4 (2021), 609–647.
- [3] ACHLIOPTAS, D., AND PERES, Y. The threshold for random k -SAT is $2^k \log 2 - O(k)$. *J. Amer. Math. Soc.* 17, 4 (2004), 947–973.
- [4] ANSÓTEGUI, C., BONET, M. L., AND LEVY, J. Scale-free random sat instances. *Algorithms* 15, 6 (2022).
- [5] BARABÁSI, A.-L., AND ALBERT, R. Emergence of scaling in random networks. *Science* 286, 5439 (1999), 509–512.
- [6] BAYATI, M., GAMARNIK, D., AND TETALI, P. Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. *Ann. Probab.* 41, 6 (2013), 4080–4115.

- [7] BLOZNELIS, M., GÖTZE, F., AND JAWORSKI, J. Birth of a strongly connected giant in an inhomogeneous random digraph. *J. Appl. Probab.* 49, 3 (2012), 601–611.
- [8] BOLLOBÁS, B. The evolution of random graphs. *Trans. Amer. Math. Soc.* 286, 1 (1984), 257–274.
- [9] BOLLOBÁS, B., BORGS, C., CHAYES, J. T., KIM, J. H., AND WILSON, D. B. The scaling window of the 2-SAT transition. *Random Structures Algorithms* 18, 3 (2001), 201–256.
- [10] BOLLOBÁS, B., JANSON, S., AND RIORDAN, O. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms* 31, 1 (2007), 3–122.
- [11] CAO, J., AND OLVERA-CRAVIOTO, M. Connectivity of a general class of inhomogeneous random digraphs. *Random Structures & Algorithms* 56, 3 (2020), 722–774.
- [12] CHUNG, F., AND LU, L. The average distance in a random graph with given expected degrees. *Internet Math.* 1, 1 (2003), 91–113.
- [13] CHVÁTAL, V., AND REED, B. A. Mick gets some (the odds are on his side). In *33rd Annual Symposium on Foundations of Computer Science, Pittsburgh, Pennsylvania, USA, 24-27 October 1992* (1992), IEEE Computer Society, pp. 620–627.
- [14] COJA-OGHLAN, A., AND PANAGIOTOU, K. The asymptotic k -SAT threshold. *Adv. Math.* 288 (2016), 985–1068.
- [15] CONWAY, J. B. *A course in functional analysis*, second ed., vol. 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1990.
- [16] COOPER, C., FRIEZE, A., AND SORKIN, G. B. Random 2-SAT with prescribed literal degrees. *Algorithmica* 48, 3 (2007), 249–265.
- [17] DEVROYE, L., AND FRAIMAN, N. Connectivity of inhomogeneous random graphs. *Random Structures Algorithms* 45, 3 (2014), 408–420.
- [18] DING, J., SLY, A., AND SUN, N. Proof of the satisfiability conjecture for large k . *Ann. of Math. (2)* 196, 1 (2022), 1–388.
- [19] ELEK, G., AND SZEGEDY, B. A measure-theoretic approach to the theory of dense hypergraphs. *Advances in Mathematics* 231, 3 (2012), 1731–1772.
- [20] FRIEDGUT, E. Sharp thresholds of graph properties, and the k -sat problem. *J. Amer. Math. Soc.* 12, 4 (1999), 1017–1054. With an appendix by Jean Bourgain.
- [21] FRIEDRICH, T., KROHMER, A., ROTHENBERGER, R., SAUERWALD, T., AND SUTTON, A. M. Bounds on the satisfiability threshold for power law distributed random SAT. In *European Symposium on Algorithms (ESA)* (sep 2017), vol. 87 of *LIPIcs*, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, pp. 37:1–37:15.

- [22] FRIEDRICH, T., KROHMER, A., ROTHENBERGER, R., AND SUTTON, A. M. Phase transitions for scale-free SAT formulas. In *Conference on Artificial Intelligence (AAAI)* (2017), AAAI Press, pp. 3893–3899.
- [23] FRIEDRICH, T., AND ROTHENBERGER, R. The Satisfiability Threshold for Non-Uniform Random 2-SAT. In *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)* (Dagstuhl, Germany, 2019), C. Baier, I. Chatzigiannakis, P. Flocchini, and S. Leonardi, Eds., vol. 132 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, pp. 61:1–61:14.
- [24] GOERDT, A. A threshold for unsatisfiability. In *Mathematical Foundations of Computer Science 1992, 17th International Symposium, MFCS’92, Prague, Czechoslovakia, August 24–28, 1992, Proceedings* (1992), I. M. Havel and V. Koubek, Eds., vol. 629 of *Lecture Notes in Computer Science*, Springer, pp. 264–274.
- [25] GOERDT, A. A threshold for unsatisfiability. *Journal of Computer and System Sciences* 53, 3 (1996), 469–486.
- [26] GREBÍK, J., KRÁL’, D., LIU, X., PIKHURKO, O., AND SLIPANTSCHUK, J. Convergence of spectra of digraph limits. arXiv:2506.04426, 2025.
- [27] HLADKÝ, J., AND SAVICKÝ, P. Digraphons: connectivity and spectral aspects. preprint.
- [28] HOLLAND, P. W., LASKEY, K. B., AND LEINHARDT, S. Stochastic blockmodels: first steps. *Social Networks* 5, 2 (1983), 109–137.
- [29] KREYSZIG, E. *Introductory functional analysis with applications*. John Wiley & Sons, New York-London-Sydney, 1978.
- [30] LOVÁSZ, L. *Large networks and graph limits*, vol. 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.
- [31] LUCZAK, T. Component behavior near the critical point of the random graph process. *Random Structures Algorithms* 1, 3 (1990), 287–310.
- [32] MCKINLEY, G. Superlogarithmic cliques in dense inhomogeneous random graphs. *SIAM J. Discrete Math.* 33, 3 (2019), 1772–1800.
- [33] NORROS, I., AND REITTU, H. On a conditionally Poissonian graph process. *Adv. in Appl. Probab.* 38, 1 (2006), 59–75.
- [34] SCHEP, A. R. Positive diagonal and triangular operators. *J. Operator Theory* 3, 2 (1980), 165–178.
- [35] ZHAO, Y. Hypergraph limits: A regularity approach. *Random Structures & Algorithms* 47, 2 (2015), 205–226.

Contact details

Postal address: Institute of Computer Science of the Czech Academy of Sciences
Pod Vodárenskou věží 2
182 00, Prague
Czechia

Email: `hladky@cs.cas.cz`, `savicky@cs.cas.cz`