THE UNIVERSITY OF WARWICK

AUTHOR: **Jan Hladký** DEGREE: **Ph.D.**

TITLE: **Graph containment problems**

DATE OF DEPOSIT: ....

I agree that this thesis shall be available in accordance with the regulations governing the University of Warwick theses.

I agree that the summary of this thesis may be submitted for publication.

I **agree** that the thesis may be photocopied (single copies for study purposes only).

Theses with no restriction on photocopying will also be made available to the British Library for microfilming. The British Library may supply copies to individuals or libraries. subject to a statement from them that the copy is supplied for non-publishing purposes. All copies supplied by the British Library will carry the following statement:

> "Attention is drawn to the fact that the copyright of this thesis rests with its author. This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's written consent."

AUTHOR'S SIGNATURE: ....

USER'S DECLARATION

1. I undertake not to quote or make use of any information from this thesis without making acknowledgement to the author.
2. I further undertake to allow no-one else to use this thesis while it is in my care.

DATE SIGNATURE ADDRESS

....

....

....

....

....

Graph containment problems

by

Jan Hladký

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Computer Science

May 2011

THE UNIVERSITY OF WARWICK

Contents

Acknowledgments

Much of the thesis stems from our work with my dearest collaborators Peter, Julia, and Diana.

The funding I received during my studies was mostly due to efforts of Artur Czumaj, Dan Král', and Anusch Taraz. This support was essential and my possibilities of doing research would be limited otherwise. I hope I was a good investment. I spent one year at TU Munich in Anusch's group and then almost two years at Warwick where Artur was my triple boss (as the head of Department of Computer Science and DIMAP, and as my thesis advisor), and much enjoyed both places.

I am grateful for the training I received during my undergraduate studies at Charles University in Prague. The care of Dan Král' and Jarik Nešetřil was exceptional.

Thanks to Anna and Michal Adamaszek, Demetres Christofides, Gábor Elek, Codruţ Grosu, Peter Heinig, Michael Krivelevich, Fiachra Knox, Jan Kynčl, András Máthé, Laci Lovász, Sergey Norin, Miki Simonovits, Balász Szegedy, Endre Szemerédi, and Andreas Würfl for mathematical discussions.

Most of my funding during the work on this thesis came from EPSRC (EP/D063191/1) through the Centre of Discrete Mathematics and Its Applications (DIMAP) at the University of Warwick. Further, I was supported by the Grant Agency of Charles University (GAUK 202-10/258009), and I was a holder of the DAAD and the BAYHOST fellowships.

Last let me thank the anonymous referees of [52, 3, 6], and Amin Coja-Oghlan and Deryk Osthus who served as examiners for this thesis. Their valuable comments are projected in the text.

Declarations

This thesis consists of six chapters.

(a) Chapter 1 contains overview of the thesis, motivation, background, and notation. It also contains some preliminary results – some unoriginal, and some obtained in collaboration with the collaborators below.

(b) Results from Chapter 2 were obtained in collaboration with Codruţ Grosu and the corresponding paper [52] is accepted for publication in *European Journal of Combinatorics*. In the same chapter we also mention a recent result obtained with Peter Allen, Julia Böttcher, and Diana Piguet [4] an extended abstract of which was accepted to proceedings of the EuroComb 2011 Conference.

(c) Results from Chapter 3 were obtained in collaboration with Peter Allen, and Julia Böttcher and the corresponding paper [3] is accepted for publication in *Journal of the London Mathematical Society*. This result was announced in the PhD thesis of Julia Böttcher ([21, p. 177-178]).

(d) Results from Chapter 4 were obtained in collaboration with Peter Allen, Julia Böttcher, and Diana Piguet and the corresponding paper [6] is accepted for publication in *Random Structures & Algorithms*.

(e) Results from Chapter 5 were obtained in collaboration with Peter Allen, Julia Böttcher, and Diana Piguet and the corresponding paper [5] will be submitted for publication.

(f) Results from Chapter 6 were obtained in collaboration with Demetres Christofides, and András Máthé and the corresponding paper [23] is submitted for publication. An extended abstract of [23] was accepted to proceedings of the EuroComb 2011 Conference.

With the exception described under point (c), none of these results appeared in any other thesis.

The collaborators above have agreed with the inclusion of our joint work into this thesis.

Abstract

In the thesis we study various graph containment problems. Our motivation comes from Turán's Theorem which determines the threshold – denoted by $\mathrm{ex}(n, K_r)$ – for the maximum number of edges of an n-vertex graph which does not contain a copy of the clique K_r of order r. We study the threshold $\mathrm{ex}(n, \ell \times H)$ defined as the maximum number of edges in an n-vertex graph which does not contain ℓ vertex-disjoint copies of a graph H in the case when H is bipartite. In a similar spirit, we study the minimum-degree condition for containment of a distance-square of a path and of a cycle of a given length, respectively.

Turán's Theorem gives the bound on the number of edges of any graph in which every r-tuple of vertices is forbidden to induce a K_r. What happens if the cliques K_r are forbidden only on certain locations? We introduce a notion of *Turánnical hypergraphs.* These are r-uniform hypergraphs $\mathcal{H}$ with the property that no graph on the same vertex set and with no r-clique on a hyperedge of $\mathcal{H}$ has more edges than the Turán bound. Besides an explicit construction of Turánnical hypergraphs we explore Turánnical hypergraphs from the probabilistic point of view.

Lovász asked whether each connected vertex-transitive graph G contains a Hamilton path. We answer Lovász' question in positive under an additional assumption that G is sufficiently dense. In fact, we show that such graphs contain a Hamilton cycle and moreover we provide a polynomial time algorithm for finding such a cycle.

Chapter 1

Introduction, notation, and preliminaries

In this thesis we investigate conditions on the host graph which guarantee containment of a specific subgraph. This is one of the loci of extremal graph theory. Our motivation comes from the fundamental result of Turán [97] from 1941 – often cited as the starting point of extremal graph theory itself – which determines the maximum edge density of graphs not containing a copy of the clique K_r; see Section 1.5 for further information on Turán's Theorem. Even though Turán's proof (as well as many subsequent proofs of the same result) is simple and elementary the result has led to an immense volume of consequent development in graph theory, and – even more importantly – to development of methods, such as the Regularity Method, the Probabilistic Method, and Flag Algebras. Graph theory is often (unjustly, we believe) regarded as somewhat isolated from mainstream mathematics, but this was never the case with extremal graph theory. Interaction with other fields was crucial from the beginnings – most notably with probability, algebra, and algebraic geometry. In the thesis we rely on the Regularity Method, and some of results in Chapter 4 are of probabilistic nature. Further, we believe that algebra, and representation theory in particular will be needed to answer our Question 4.10 below (we try to indicate some links in Section 4.4.1).

A *Turán-type problem* asks whether a certain density condition (usually parametrised by the density of edges, or the minimum degree) in the host structure (which is typically a graph, digraph, or a hypergraph) guarantees the existence of a specific substructure. Despite bounty of the area, the only two existing surveys – a slightly outdated Füredi's [46] and a very recent Keevash's [56] – focus on hypergraphs. In this thesis we restrict ourselves to Turán-type problems for graphs, even though digraphs and hypergraphs will occasionally emerge as auxiliary objects.

1.1 Motivation for our results

We use some elementary graph-theoretic notation throughout this section, and the reader may need to consult it with Section 1.2.

1.1.1 Motivation for Chapters 2 and 3

Turán's Theorem gives a sharp threshold for the maximum number of edges of an n-vertex graph with no copy of K_r. Erdős and Stone [34] generalised this result to avoiding a fixed r-vertex graph H. Even though Turán's Theorem applies to any pair of values n and r, the interesting instances are rather those when n is large compared to r. The Erdős-Stone Theorem on the other hand does need the assumption that $r \ll n$. On the other end of the research in extremal graph theory are results about containment of a spanning subgraphs. The most classic of these is the Dirac Theorem [32] which determines the minimum degree of a host graph which guarantees containment of a Hamilton cycle. Note that the density parameter occurring in a reasonable statement about containment of a spanning subgraph H must be the minimum degree (rather than the average degree). Indeed, unless H contains isolated vertices there exist very dense graphs (a clique and one isolated vertex, for example), which do not contain H.

In Chapters 2–3 we establish several Turán-type results for "subgraphs of intermediate size". These results are in between results about small subgraphs and results about spanning subgraphs. Typical example of such a problem is the following question: *Does the presence of 73% of the edges in a graph G guarantee existence of a path covering 20% of the vertices of G?* This question was actually answered by Erdős and Gallai in 1959 [36]. More precisely, Erdős and Gallai gave a tight bounds on the size of the maximum matching, the length of the longest path, and the length of the longest cycle in a graph of given order and a given number of edges. Other results about intermediate-sized subgraphs include the Hajnal-Szemerédi Theorem [54] which determines the minimum-degree which guarantees covering of a fixed proportion of the host by copies of K_r (see Section 2.1.1). In the same vein as the Erdős-Stone Theorem generalizes Turán's Theorem, the Hajnal-Szemerédi Theorem was later generalised by Komlós [59]. The main result of Chapter 2 provides a density condition which guarantees tiling of a fixed proportion of the host graph by an arbitrary fixed bipartite graph. In Section 2.4 we mention our recent work on the same question for triangle tilings. Last, in Chapter 3 we determine the minimum degree condition which guarantees containment of a distance-square of a path or a cycle of specified length. This result is an analogue of the Erdős-Gallai Theorem when paths or cycles are replaced by their distance-squares.

1.1.2 Motivation for Chapters 4 and 5

Turán's Theorem provides an upper bound on the number of edges of a graph in which no r-tuple induces a copy of the complete graph K_r. Can the same bound be guaranteed if K_r's are forbidden only on certain r-tuples? We call an r-uniform hypergraph $\mathcal{H}$ on a vertex set V *Turánnical* if every graph G on vertex set V which contains no cliques on edges of $\mathcal{H}$ has at most the number of edges given by the Turán bound. In Chapter 4 we thoroughly investigate Turánnical hypergraphs from a probabilistic point of view. In particular, we investigate for what edge densities is a typical random hypergraph Turánnical. It turns out that even very sparse random hypergraphs are typically Turánnical.

In Chapter 5 we then provide with an explicit construction of Turánnical hypergraphs and provide a related extension of Turán's Theorem.

Our result could be put into a more general framework which – to the best of our knowledge had not been explored prior to our work. Indeed, questions in extremal combinatorics usually fit the pattern *Maximize a certain parameter f over a set of combinatorial structures $\mathcal{S}$ satisfying certain restrictions $\mathcal{R}$*. A common modification is that one works with random substructures of $\mathcal{S}$ rather than with $\mathcal{S}$ itself. Here, we therefore propose another model of randomization of problems in extremal combinatorics; instead one randomizing $\mathcal{S}$, we consider just a random subset of the restriction set $\mathcal{R}$. More details and examples are given in Section 4.8.

1.1.3 Motivation for Chapter 6

In Chapter 6 we provide a result on a problem coming from algebraic graph theory. A famous conjecture of Lovász [75] states that every connected vertex-transitive graph contains a Hamilton path. We confirm the conjecture in the case that the graph is dense and sufficiently large. In fact, we show that such graphs contain a Hamilton cycle and moreover we provide a polynomial time algorithm for finding such a cycle. We use tools from the Extremal Graph Theory, and the Regularity Method in particular. Even though these are rather standard techniques to the best of our knowledge this is the first time they were used in algebraic graph theory.

1.2 Notation

1.2.1 General notation

Our notation is standard and we draw attention only to symbols which may possibly cause confusion.

The difference of sets A and B is denoted by $A-B$, the symmetric difference by $A\triangle B$. For a set X and a positive integer r we write $\binom{X}{r}$ for the set of all subsets of X of size r. We write $\mathbb{N},\mathbb{Z},\mathbb{R}$ for the sets of natural numbers (the smallest of them being one), integers, and reals. Given a positive integer m we will often denote the set $\{1,\ldots,m\}$ of the first m positive integers by $[m]$. Given a (finite) set X and a function $f:X\to\mathbb{R}$ we will write $\|f\|_1$ for the sum $\sum_{x\in X}|f(x)|$.

When we say that a statement $\mathsf{S}(\varepsilon,\varepsilon')$ holds for positive real numbers $\varepsilon\gg\varepsilon'>0$, then we mean that, given an arbitrary $\varepsilon>0$, we can find an $\varepsilon''>0$ such that $\mathsf{S}(\varepsilon,\varepsilon')$ holds for all $\varepsilon'\in(0,\varepsilon'']$.

Finally, to avoid unnecessarily complicated calculations, we will sometimes omit floor and ceiling signs and treat large numbers as if they were integers.

1.2.2 Graph theory

All graphs are finite. Loops and multiple edges are not allowed. Given a graph $G=(V,E)$, we write $V(G):=V$ and $E(G):=E$ for the vertex set, and edge set of G, respectively. As usual we write $xy\in E(G)$ instead the correct $\{x,y\}\in E(G)$ to denote that the pair $\{x,y\}$ forms an edge of G. We define the *order* of G as $v(G):=|V|$, and further, $e(G):=|E|$. (The number $e(G)$ is often called the *size* of G in literature, however we do not use this terminology.)

Given two graphs G and H, we say that a map $\psi:V(G)\to V(H)$ is an *isomorphism* if ψ is a bijection preserving edges and non-edges, i.e., we have $xy\in E(G)\Leftrightarrow\psi(x)\psi(y)\in E(H)$. If at least one isomorphism exists, we say that G and H are *isomorphic*, and write $G\simeq H$. A map $\phi:V(G)\to V(H)$ is a *homomorphism* if it preserves the edges of G. We write $\phi:G\to H$ in this case.

G is a *subgraph* of a graph H if there exists a graph $G'\simeq G$ such that $V(G')\subseteq V(H)$ and $E(G')\subseteq E(H)$. We write $G\subseteq H$ in this case. G is called *spanning subgraph* of H if $G\subseteq H$ and $v(G)=v(H)$.

A set $U\subseteq V(G)$ is called *independent* if there is no edge of G with both endvertices in U.

G is *r-partite* if there exists a partition $V(G)=V_1\dot\cup\cdots\dot\cup V_r$ such that each set V_i is an independent set in G. For $r=2$ we call G *bipartite*. We refer to the sets $V_1,\ldots,V_r$ as *colour classes* of G. Observe that colour classes need not be unique.

If $X \subseteq V(G)$ then $G[X]$ is the graph *induced graph* by X, that is $V(G[X]) = X$, and an edge $xy \in \binom{X}{2}$ is present in $G[X]$ if and only of it is present in G. Similarly, for two disjoint sets $X, Y \subseteq V(G)$, then $G[X,Y] \subseteq G$ is the bipartite graph with colour classes X and Y, and edge set inherited from G.

Given a vertex $v \in V(G)$, its *neighbourhood* is defined as $\mathrm{N}(v) := \{u \in V(G) : uv \in E(G)\}$. For a set $U \subseteq V(G)$, we write $\mathrm{N}_U(v)$ for the restricted neighbourhood $\mathrm{N}_U(v) := \mathrm{N}(v) \cap U$. We denote the sizes of $\mathrm{N}(v)$ and $\mathrm{N}_U(v)$ by $\deg(v)$ and $\deg(v, U)$, respectively. For $U \subseteq V(G)$, the symbol $\mathrm{N}(U)$ is the *united neighbourhood*, $\mathrm{N}(U) := \bigcup_{u \in U} \mathrm{N}(u)$. The *common neighbourhood* on the other hand is defined by $\mathrm{N}^\wedge(U) := \bigcap_{u \in U} \mathrm{N}(u)$.

The maximum and minimum degree in a graph G are defined by $\deg^{\max}(G) := \max\{\deg(v) : v \in V(G)\}$, and $\deg^{\min}(G) := \min\{\deg(v) : v \in V(G)\}$. For a set $E' \subseteq E(G)$ we write $\deg^{\max}(E')$ for the maximum degree of the subgraph induced by E'. For for two disjoint sets $A, B \subseteq V(G)$ we write ${\deg^{\max}}_G(A, B)$ for the maximum degree of the bipartite graph $G[A, B]$. For two sets $X, Y \subseteq V(G)$ we define ${\deg^{\min}}_Y(X) := \min\{\deg(x, Y) : x \in X\}$ and ${\deg^{\min}}_G(X) := {\deg^{\min}}_{V(G)}(X)$.

If every vertex of a graph G has the same degree k then we say that G has *valency* k, and write $\deg(G) = k$. It is perhaps more standard to call G *k-regular* in this case, however our usage of the term "regular" is reserved for the context of Szemerédi's Regularity Lemma (see Section 1.4).

An *automorphism* of a graph G is an isomorphism from G to itself. The automorphisms of G form a group, denoted by $\mathrm{Aut}(G)$. The unit element of $\mathrm{Aut}(G)$ is the identity map, and multiplication is defined as composition of the corresponding automorphisms. A graph is called *vertex-transitive* if $\mathrm{Aut}(G)$ acts transitively on $V(G)$, i.e., for each pair of vertices $x, y \in V(G)$ there exists $g \in \mathrm{Aut}(G)$ such that $y = g(x)$. A graph G is a *Cayley graph* if there exists a group Γ and a set $X \subseteq \Gamma$ such that $V(G) = \Gamma$, and $xy \in E(G)$ if and only if $xy^{-1} \in X \cup X^{-1}$. Note that each Cayley graph is vertex-transitive, and that each vertex-transitive graph has valency k for some k.

A *colouring* of a graph G is any function $f : V(G) \to [\ell]$ such that $f(x) \neq f(y)$ whenever $xy \in E(G)$. If at least one such colouring exists for a given ℓ, we say that G is *ℓ-colourable*, or *ℓ-chromatic*. The least ℓ for which G is ℓ-colourable is called the *chromatic number* of G, and denoted by $\chi(G)$.

A *Hamilton path* is a spanning path, and a *Hamilton cycle* is a spanning cycle. Graphs which have a Hamilton cycle are called *hamiltonian*.

A graph G is *connected* if for each two its vertices u_1, u_2 there exists a path in G from u_1 to u_2. G graph is *ℓ-connected* (more precisely, ℓ-vertex-connected) if removal of each $j \leqslant \ell - 1$ vertices of G results in a connected graph.

Some special graphs

Suppose that $n \in \mathbb{N}$. Then K_n is the *complete graph* (or also *clique*) of order n, i.e., $V(K_n) = [n]$ with all possible edges present, $E(K_n) = \binom{[n]}{2}$. For $a, b \in \mathbb{N}$, $K_{a,b}$ is the *complete bipartite graph* with colour-classes of sizes a and b, i.e., $V(K_{a,b}) = [a+b]$, $E(K_{a,b}) = \{xy : x \in [a], y \in [a+b] - [a]\}$.

P_n is the *path* of length $n-1$, i.e., $V(P_n) = [n]$, and vertices i and j are adjacent if and only if $|i-j| = 1$. Finally for $n \geqslant 3$, C_n is the *cycle* of length n, i.e., C_n can be constructed from P_n by adding the edge $1n$.

1.2.3 Hypergraphs

In order to distinguish hypergraphs from graphs, we will use a calligraphy font to denote them, i.e., we write $\mathcal{H} = (V, \mathcal{E})$ for a hypergraph on the vertex set V and hyperedges $\mathcal{E}$. We shall call hyperedges simply *edges* when no confusion can arise. Recall that $\mathcal{H}$ is r*-uniform* if $|e| = r$ for each $e \in \mathcal{E}$.

1.3 Basic probability theory and random (hyper)graphs

All our probability spaces are finite, with their σ-algebras generated by elementary events. In a probability space Ω and for an event $A \subseteq \Omega$ we write $\mathbb{P}(A)$ for the probability of A. If $Y : \Omega \to \mathbb{R}$ is a random variable, then we write $\mathbb{E}(Y)$ for its expectation.

The notion of random graphs was introduced in a seminar paper of Erdős and Rényi [37]. The so-called *Erdős-Rényi model* $G(n,p)$ (for $n \in \mathbb{N}$ and $p \in [0,1]$) assigns each graph G on the vertex set $[n]$ probability $p^{e(G)}(1-p)^{\binom{n}{2}-e(G)}$. Therefore the model corresponds to inserting each possible edge with probability p independently of other edges. Most of the research on random graphs concerns *asymptotic properties* of the model. Let S be a graph predicate. Suppose that $(p_n)_{n=1}^{\infty}$ is a sequence of probabilities. We then say that S *holds asymptotically almost surely*[1] *for* $G(n, p_n)$ if

$$\lim_{n \to \infty} \mathbb{P}\big(\mathsf{S}(G(n, p_n))\big) = 1 .$$

The theory of random hypergraphs has witnessed some enormous development in the last 50 years. We refer the reader to books [55, 18].

[1] abbreviated by *a.a.s.*

In an analogy to random graphs, the Erdős-Rényi model of random hypergraphs $\mathcal{R}^{(r)}(n,p)$ assigns probability $p^{e(\mathcal{H})}(1-p)^{\binom{n}{r}-e(\mathcal{H})}$ to any r-uniform hypergraph $\mathcal{H}$ on the vertex set $[n]$. There is a corresponding notion of a hypergraph property being satisfied a.a.s.

1.4 The Regularity Lemma

Results from Chapter 2, 3, and 6 in our thesis rely on the Szemerédi Regularity Lemma. Roughly speaking the lemma asserts that each graph can be decomposed into random-looking parts. First seeds of the result can be seen in Szemerédi's resolution of a conjecture of Erdős and Turán on arithmetic progressions in dense subsets of the integers [95]. The Regularity Lemma has found numerous applications in number theory, graph theory, and property testing since then. Here, we use a form of the lemma from 1978, also due to Szemerédi [96]. (Even though some further strengthenings exist, this form is suitable for most applications in graph theory.) We refer the reader to surveys [65, 64, 70] on the Regularity Method and its applications in graph theory. After introducing the crucial notion of ε-regular pairs and stating the lemma we further give tools which often accompany the Regularity Lemma.

Let $G=(V,E)$ be a graph and $\varepsilon, d \in (0,1]$. For disjoint nonempty $U, W \subseteq V$ the *density* of the pair (U,W) is $d(U,W) := e(U,W)/|U||W|$. A pair (U,W) is *ε-regular* if $|d(U',W')-d(U,W)| < \varepsilon$ for all $U' \subseteq U$ and $W' \subseteq W$ with $|U'| \geqslant \varepsilon|U|$ and $|W'| \geqslant \varepsilon|W|$. The pair (U,W) is called *(ε,d)-super-regular* if it is ε-regular, and further $\mathrm{deg}^{\min}{}_W(U) \geqslant d|W|$ and $\mathrm{deg}^{\min}{}_U(W) \geqslant d|U|$. An *$\varepsilon$-regular partition* of G is a partition $V_0 \dot\cup V_1 \dot\cup \ldots \dot\cup V_k$ of V with $|V_0| \leqslant \varepsilon|V|$, $|V_i| = |V_j|$ for all $i,j \in [k]$, and such that for all but at most εk^2 pairs $(i,j) \in [k]^2$, the pair (V_i, V_j) is ε-regular.

Given some $0 < d < 1$ and a pair of disjoint vertex sets (V_i, V_j) in a graph G, we say that (V_i, V_j) is *(ε,d)-regular* if it is ε-regular and has density at least d. We say that an ε-regular partition $V_0 \dot\cup V_1 \dot\cup \ldots \dot\cup V_k$ of a graph G is an *(ε,d)-regular partition* if the following is true. For every $1 \leqslant i \leqslant k$, and every vertex $v \in V_i$, there are at most $(\varepsilon+d)n$ edges incident to v which are not contained in (ε,d)-regular pairs of the partition.

Given an (ε,d)-regular partition $V_0 \dot\cup V_1 \dot\cup \ldots \dot\cup V_k$ of a graph G, we define a graph R, called the *reduced graph* of the partition of G, where $R=(V(R),E(R))$ has $V(R) = \{V_1,\ldots,V_k\}$ and $V_iV_j \in E(R)$ whenever (V_i,V_j) is an (ε,d)-regular pair. We will usually omit the partition, and simply say that G has *(ε,d)-reduced graph R*. We refer to the spanning subgraph $G' \subseteq G$ formed by edges $xy \in E(G)$, $x \in V_i, y \in V_j$, with $V_iV_j \in E(R)$, as the graph *corresponding to the reduced graph*

R. Observe that the property of (ε, d)-regular partition above translates as

$$\deg_{G'}(v) \geqslant \deg_G(v) - (\varepsilon + d)n \;, \tag{1.1}$$

for each $v \in \bigcup_{i=1}^{k} V_i$.

We call the partition classes V_i with $i \in [k]$ *clusters* of G. Observe that our definition of the reduced graph R implies that for $T \subseteq V(R)$ we can for example refer to the set $\bigcup T$, which is a subset of $V(G)$.

Suppose that $U_1 \dot\cup \ldots \dot\cup U_\ell$ and $V_0 \dot\cup V_1 \dot\cup \ldots \dot\cup V_k$ are two partitions of the set V. We then say that the partition $V_0 \dot\cup V_1 \dot\cup \ldots \dot\cup V_k$ *refines* $U_1 \dot\cup \ldots \dot\cup U_\ell$, if for every $j \in [k]$ there exists $i \in [\ell]$ such that $V_j \subseteq U_i$. Note that this is weaker than the usual notion of refinement as we do not require V_0 to be contained in any U_i.

The celebrated Szemerédi Regularity Lemma [96] states that every large graph has an ε-regular partition with a bounded number of clusters. Further, we may require to refine any other partition with a bounded number of parts. Here we state the so-called degree form of this lemma (see, e.g., [65, Theorem 1.10]).

Lemma 1.1 (Regularity Lemma, degree form)**.** *For every $\varepsilon > 0$ and every integer N', there is $N := N(\varepsilon, N')$ such that for every $d \in [0,1]$ every graph $G = (V, E)$ on $n \geqslant N$ vertices has an (ε, d)-reduced graph R on m vertices with $N' \leqslant m \leqslant N$.*

Furthermore, if any partition $V(G) = U_1 \dot\cup U_2 \dot\cup \ldots \dot\cup U_{N'}$ is given, then we may require the clusters of R to refine $U_1 \dot\cup U_2 \dot\cup \ldots \dot\cup U_{N'}$.

Remark 1.2. In the "furthermore" part of Lemma 1.1 we could have as well required the clusters of R to refine any partition $V(G) = U_1 \dot\cup U_2 \dot\cup \ldots \dot\cup U_\ell$, with $\ell \leqslant N'$. This is not really a stronger assertion as one can obtain it from the original version by introducing auxiliary sets $U_{\ell+1} = U_{\ell+2} = \ldots = U_{N'} := \emptyset$.

Remark 1.3. It turns out that for the proofs of Theorems 6.20 and 6.21 we need to work with two threshold densities $d_1 < d_2$ of the reduced graph. The degree form of the Regularity Lemma can be adapted in order to accommodate this need. In particular we can get a partition $V_0, V_1, \ldots, V_k$ of the vertex set of G and spanning subgraphs G_1, G_2 of G such that properties of the Regularity Lemma hold for both G_1 and G_2 with the corresponding densities d_1 and d_2.

For our work in Chapter 3 it is more convenient to work with even a different version of the regularity lemma, which takes into account that we are dealing with graphs of high minimum degree. This lemma is an easy corollary of Lemma 1.1. A proof can be found, e.g., in [72, Proposition 9].

Lemma 1.4 (Regularity Lemma, minimum degree form)**.** *For all ε, d, γ with $0 < \varepsilon < d < \gamma < 1$ and for every m_0, there is m_1 such that every graph G on $n > m_1$ vertices with $\deg^{\min}(G) \geqslant \gamma n$ has an (ε, d)-reduced graph R on m vertices with $m_0 \leqslant m \leqslant m_1$ and $\deg^{\min}(R) \geqslant (\gamma - d - \varepsilon)m$.*

This lemma asserts that the reduced graph R of G inherits the high minimum degree of G.

1.4.1 Properties of regular pairs

First we recall that regularity of a pair is inherited even to subpairs of substantial size. Lemma 1.5 below has a standard proof which we include only for completeness as we were unable to find any in surveys on the topic.

Lemma 1.5. *Let (A, B) be an ε-regular pair with density d, and let $A' \subseteq A, |A'| \geqslant \alpha|A|, B' \subseteq B, |B'| \geqslant \alpha|B|$, $\alpha \geqslant \varepsilon$. Then (A', B') is an ε'-regular pair with $\varepsilon' := \max\{\varepsilon/\alpha, 2\varepsilon\}$, and for its density d' we have $|d' - d| < \varepsilon$.*

Proof. The fact that $|d' - d| < \varepsilon$ follows immediately from the fact that (A, B) is ε-regular.

To verify that (A', B') is ε'-regular, consider two arbitrary sets $A'' \subseteq A'$, $B'' \subseteq B'$ such that $|A''| \geqslant \varepsilon'|A'|$ and $|B''| \geqslant \varepsilon'|B'|$. We have $|A''| \geqslant \varepsilon|A|$, and $|B''| \geqslant \varepsilon|B|$. By the regularity of (A, B), we therefore have $|d(A'', B'') - d| < \varepsilon$. By the triangle inequality, we therefore have $|d' - d(A'', B'')| < 2\varepsilon \leqslant \varepsilon'$. □

Given any bounded degree subgraph H of the reduced graph R we can make the pairs corresponding to its edges super-regular by removing a small fraction of the vertices of each cluster to the exceptional set. We will only need this fact in the case that H is a matching.

Lemma 1.6. *Suppose $0 < 4\varepsilon < d \leqslant 1$ and let $V_0, V_1, \ldots, V_k$ be an (ε, d)-regular partition of a graph G. Let m be the size of any (nonexceptional) cluster. Let R be the reduced graph with respect to this partition and the parameters ε and d. Let M be a matching in R. Then we can move exactly εm vertices from each cluster V_i $(i > 0)$ into V_0 such that each pair of clusters corresponding to an edge of M is $(2\varepsilon, d/2)$-super-regular while each pair of clusters corresponding to an edge of R is $(2\varepsilon, d/2)$-regular.*

Proof. We include a proof for the sake of completeness even though it is standard. For a cluster V_i $(i > 0)$ its *partner* is the cluster which is matched to V_i by M. We do not define partners of clusters not covered by M.

For each cluster V_i covered by M we let W_i be the set of exactly εm vertices with the least degrees in the partner of V_i. For clusters V_i not covered by M let $W_i \subseteq V_i$ be an arbitrary set of size εm. We move the vertices of the sets W_i into V_0. Let $V_i' := V_i - W_i$ be the modified clusters. We claim, that the modified clusters satisfy the assertion of the lemma.

Let $V_iV_j \in E(R)$. By the regularity of the pair (V_i, V_j) we have $d(V_i', V_j') = d(V_i, V_j) \pm \varepsilon \geqslant d/2$. Further, for any sets $A \subseteq V_i'$ and $B \subseteq V_j'$ with $|A| \geqslant 2\varepsilon|V_i'|$, $|B| \geqslant 2\varepsilon|V_j'|$ we have $d(A, B) = d(V_i, V_j) \pm \varepsilon$. In particular, $d(A, B) = d(V_i', V_j') \pm 2\varepsilon$, proving that (V_i', V_j') is $(2\varepsilon, d/2)$-regular.

It remains to prove that for any vertex $x \in V_i'$ we have $\deg(x, V_j') \geqslant d|V_j'|/2$, where V_j is the partner of V_i. Suppose not. Then in particular, $\deg(x, V_j) \leqslant d|V_j'|/2 + |W_j| < 3d|V_j|/4$. By the choice of the set W_i we therefore have $\deg(y, V_j) < 3d|V_j|/4$ for each $y \in W_i$. Consequently, the subpair (W_i, V_j) is a substantial pair of density less than $3d/4$, a contradiction to the regularity of (V_i, V_j). □

Given an (ε, d)-super-regular pair (A, B), we will need to isolate a small sub-pair that maintains super-regularity in any sub-pair that contains it. To this end we introduce the following definition. For $A^* \subseteq A$ and $B^* \subseteq B$ we say that (A^*, B^*) is an (ε^*, d^*)-*ideal for* (A, B) if for any $A^* \subseteq A' \subseteq A$ and $B^* \subseteq B' \subseteq B$ the pair (A', B') is (ε^*, d^*)-super-regular. The following lemma shows that ideals exist.

Lemma 1.7 ([25, Lemma 15])**.** *Suppose* $0 < \varepsilon \ll \theta, d < 1/2$, *and let* (A, B) *be an* (ε, d)*-super-regular pair with* $|A| = |B| = m$, *where* m *is sufficiently large. Then there exists subsets* $A^* \subseteq A$ *and* $B^* \subseteq B$ *of sizes* θm *such that* (A^*, B^*) *is an* $(\varepsilon/\theta, \theta d/4)$*-ideal for* (A, B).

The proof of the above lemma given in [25] is probabilistic. (It proves that random subsets of sizes θm have the required property with high probability.) We use Lemma 1.7 in Chapter 6 to prove a certain existential result. More specifically, this lemma is used to prove that a certain class of graphs contains a Hamilton cycle (Theorem 6.2). However, we also aim in that chapter to provide an algorithmic counterpart of the result, i.e., to find a Hamilton cycle efficiently in any graph from this class (Theorem 6.22). Therefore, we will also need a 'constructive' proof of this lemma. We proceed to give such a proof.

Proof of Lemma 1.7. By Lemma 1.5, it is enough to construct subsets $A^* \subseteq A$ and $B^* \subseteq B$ of sizes θm such that every vertex $a \in A$ has $\deg(a, B^*) \geqslant \theta dm/4$ and every vertex $b \in B$ has $\deg(b, A^*) \geqslant \theta dm/4$. By symmetry, it is enough to show how to construct a subset $A^* \subseteq A$ of size θm such that very vertex $b \in B$ has

$\deg(b, A^*) \geqslant \theta dn/4$. We will construct this set A^* by adding to it one vertex at every step. At each step we will say that a vertex b of B is *unhappy* if it has $k < \theta dm/4$ neighbours in A^*. If a vertex b is unhappy we will define its unhappiness $u(b)$ to be $u(b) := \sum_{r=k+1}^{\theta dm/4} 2^{-r}$. Otherwise we define its unhappiness $u(b)$ to be equal to 0. We also denote by U the total unhappiness $U := \sum_{b \in B} u(b)$ of vertices of B. Observe that if in the next step we add to A^* a neighbour of b then the unhappiness of b is reduced by at least $u(b)/2$. Note also that if a vertex b is unhappy, then it has at least $dm - \theta dm/4 \geqslant dm/2$ neighbours outside of A^*. We now give to every edge joining b to a vertex of $A - A^*$ a weight equal to $u(b)/2$. Then the total weight on these edges is at least $\sum_{b \in B} u(b) dm/4 = U dm/4$. In particular there is a vertex $a \in A - A^*$ where the total weight on its incident edges is at least $Ud/4$. Adding this vertex to A^* we get that the new total unhappiness is at most $(1 - d/4)U$. Initially the total unhappiness was at most m. So after θm steps the total unhappiness is at most $(1 - d/4)^{\theta m} m \leqslant m e^{-\theta md/4} < 2^{-\theta dm/4}$, when m is sufficiently large. But no unhappy vertex can have unhappiness less than $2^{-\theta dm/4}$. It follows that after θm steps there is no unhappy vertex in B, as required. $\square$

1.4.2 The Blow-up Lemma

Many applications of the Regularity Lemma in graph theory rely on the so-called Blow-up Lemma of Komlós, Sarközy and Szemerédi [61]. Roughly, this lemma asserts that a bounded degree graph H can be embedded into a graph G with reduced graph R if there is a homomorphism from H to R which does not overfill any of the clusters in R.

Here, we recall the most general version of the Blow-up Lemma [61, Theorem 1, Remark 13], which we then tailor to the current need in each chapter separately.

Lemma 1.8 (Komlós, Sarközy & Szemerédi [61])**.** *Given a graph R on the vertex set $\{1, \ldots, r\}$ and $d > 0$ and $\Delta \in \mathbb{N}$, there exists $\varepsilon = \varepsilon(\delta, \Delta, r)$ such that the following holds. Let $n_1, \ldots, n_r$ be arbitrary positive integers and let us replace the vertices $1, \ldots, r$ by disjoint sets $V_1, \ldots, V_r$ of sizes $n_1, \ldots, n_r$. Replace each edge ij of R by an arbitrary (ε, d)-super-regular pair between V_i and V_j, thus obtaining a graph G.*

Suppose that a graph H with $\deg^{\max}(H) \leqslant \Delta$ is given together with a homomorphism $\phi : H \to R$. Suppose further that $|\phi^{-1}(i)| \leqslant n_i$ for each $i \in [r]$. Then $H \subseteq G$, and more specifically we can find a copy of H in G which corresponds to ϕ.

Furthermore, we can strengthen the assertion as follows. Suppose that a set $\{u_1, \ldots, u_\ell\} \subseteq V(H)$ is given, $\ell < \varepsilon \min_i\{n_i\}$. Let $U_1, \ldots, U_\ell$ be sets with $U_i \subseteq$

$V_{\phi(u_i)}$, $|U_i| > dn_{\phi(u_i)}$. *Then we can find a copy of H in G such that each vertex u_i lies in the set U_i.*

1.5 Turán's Theorem

This thesis is centered around Turán's Theorem. In particular, Chapters 2–5 can be viewed as its extensions. In this section we give concise information on Turán's result, and historical development it initiated.

Let us start with stating the theorem. Let $\mathrm{T}_r(n)$ denote the complete *balanced* $(r-1)$-partite graph on n vertices (i.e., the part sizes of $\mathrm{T}_r(n)$ are as equal as possible) and $t_r(n)$ the number of its edges. We have

$$t_r(n) = \binom{r-1}{2}\left\lfloor\frac{n}{r-1}\right\rfloor + \left(n-(r-1)\left\lfloor\frac{n}{r-1}\right\rfloor\right)\left(n-1-\left\lfloor\frac{n}{r-1}\right\rfloor\right).$$

Theorem 1.9 (Turán [97]). *Suppose that G is an n-vertex graph with no copy of the complete graph K_r. Then G has at most $t_r(n)$ edges.*

Turán's paper [97] appeared in 1941. However, a particular case of forbidding the triangle K_3 was proven more than thirty years before that.

Theorem 1.10 (Mantel [78]). *Suppose that G is an n-vertex graph with no copy of the triangle K_3. Then G has at most $t_3(n) = \lfloor\frac{n}{2}\rfloor\lceil\frac{n}{2}\rceil$ edges.*

The bound in Turán's Theorem is obviously best possible because of the graph $\mathrm{T}_r(n)$ which is K_r-free. Theorem 1.9 can be strengthened to give *uniqueness* as well: each K_r-free graph n-vertex graph with $t_r(n)$ edges is isomorphic to $\mathrm{T}_r(n)$.

There are several elementary (and reasonably simple) proofs of Theorem 1.9. We refer to [1, Chapter 36] for a collection of five different proofs. Let us provide some intuition for the problem (which can be actually converted to a proper proof quite easily). Suppose that G is an n-vertex graph with more than $t_r(n)$ edges; our task is to find a copy of K_r in G. In other words, the average degree of G is more than $\frac{(r-2)n}{r-1}$ (let us neglect a small rounding error we have just made). Let us make an assumption that the degree of *each* vertex of G is exactly the average degree. Then a copy of the clique K_r may be constructed sequentially: suppose we already have a copy K_p in G, with $p < r$. The vertices of this copy have more than $n - p\frac{(r-2)n}{r-1} \geqslant 0$ common neighbors in G. Taking any of these common neighbors gives a copy of K_{p+1}, giving eventually a copy of K_r. One application of the Cauchy-Schwarz inequality can be used to reduce the general case to the case with all vertices having the same degree.

Turán's Theorem is often stated using the notion of the *extremal number.* Given a graph H and an integer $n \in \mathbb{N}$ we define $\mathrm{ex}(n, H)$ as the maximum number of edges an n-vertex graph not containing a copy of H can have. Then Turán's Theorem asserts that $\mathrm{ex}(n, K_r) \leqslant t_r(n)$.

Last, we note that even thought Turán's Theorem gives a tight bound for any pair r and n, the result gets interesting rather when n is large compared to r. Most of the extensions of Turán's Theorem apply only to this enviroment (cf. Section 1.5.1).

1.5.1 Extensions of Turán's Theorem

Let us highlight some extensions of Turán's Theorem. Erdős and Stone [34] determined the asymptotic behaviour of the function $\mathrm{ex}(n, H)$ (as a function of n) for a fixed graph H. They discovered that $\mathrm{ex}(n, H)$ is essentially governed by the chromatic number $\chi(H)$.

Theorem 1.11 (Erdős & Stone [34])**.** *Suppose that a graph H is given, $\chi(H) =: r$. For any $\varepsilon > 0$ there exists $n_0 = n_0(\varepsilon, H)$ such that for each $n > n_0$ we have $\mathrm{ex}(n, H) \leqslant t_r(n) + \varepsilon n^2$.*

Theorem 1.11 is again asymptotically tight because of the graphs $\mathrm{T}_r(n)$. Note that when H is bipartite then we have $t_{\chi(H)}(n) = 0$. Consequently, the error term εn^2 in Theorem 1.11 dominates the bound on $\mathrm{ex}(n, H)$. Thus Theorem 1.11 only asserts that $\mathrm{ex}(n, H)$ is a subquadratic function in n in this case. To get a more precise bound turned out to be a very challenging question, known as the *Zarankiewicz problem.* Let us recall that when H has colour classes of sizes s and t, $s \leqslant t$, then the Kövari-Sós-Turán Theorem [66] asserts that

$$\mathrm{ex}(n, H) \leqslant \mathcal{O}(n^{2-1/s}) = o(n^2)\ . \tag{1.2}$$

On the other hand, a standard random graph argument[2] gives that $\mathrm{ex}(n, K_{s,t}) \geqslant \Omega(n^{2-(s+t-2)/(st-1)})$.

The second direction of extending Turán's Theorem we want to discuss – as it is related to our work in Chapter 4 – is the recent resolution of the Kohayakawa-Łuczak-Rödl Conjecture[3] [57]. To motivate the conjecture, we need yet another

[2]This bound was recently refined in [17] for some bipartite graphs using another randomised construction.

[3]The paper [57] contains also another conjecture which would imply Conjecture 1.12. Often, it is this stronger conjecture which is referred to as the Kohayakawa-Łuczak-Rödl Conjecture. While Conjecture 1.12 was recently solved (cf. Theorem 4.6), the stronger Kohayakawa-Łuczak-Rödl Conjecture remains open.

reformulation of Turán's Theorem: Each K_r-free subgraph of the complete graph K_n has at most (roughly) $\frac{r-2}{r-1}$ fraction of the edges of K_n. Using simple averaging it can be shown that each graph G contains a K_r-free subgraph with at least $\frac{r-2}{r-1}e(G)$ edges. Thus, it is natural to ask for which graphs G we cannot find a K_r-free subgraph with more than (roughly) $\frac{r-2}{r-1}e(G)$ edges. Kohayakawa, Łuczak and Rödl conjectured that even a fairly sparse typical random graph $G(n,p)$ has this property.

Conjecture 1.12 (Kohayakawa, Łuczak & Rödl [57])**.** *Given $\varepsilon > 0$ and r there exists a constant C such that the following is true. For $q \geqslant Cn^{-2/(r+1)}$, a.a.s. $G = G(n,q)$ has the property that every subgraph of G with at least $(1+\varepsilon)\frac{r-2}{r-1}e(G)$ edges contains a copy of K_r.*

It is not difficult to see that the lower-bound on the probability $Cn^{-2/(r+1)}$ is best possible. This also follows from our (more general) result, the 0-statement of Theorem 4.8.

Conjecture 1.12 was recently proven by Schacht [89] (see also [45]), and independently by Conlon and Gowers [28] using different methods. See Theorem 4.6. Schacht uses elementary (but complicated) probabilistic calculations while Conlon and Gowers give a proof relying on functional analysis.

Chapter 2

Partial tilings with bipartite graphs

2.1 Introduction

It is natural to extend the *existential questions* in Extremal Graph Theory (such as Turán's Theorem, Theorem 1.9) to *tiling questions*. In such a setting one asks for the maximum number of edges of an n-vertex graph which does not contain ℓ vertex-disjoint copies of a graph H. This quantity is denoted by $\mathrm{ex}(n, \ell \times H)$. Erdős and Gallai [36] in 1959 gave a complete solution to the problem in the case when $H = K_2$.

Theorem 2.1 (Erdős & Gallai [36]). *Suppose that $\ell \leqslant n/2$. Then*

$$\mathrm{ex}(n, \ell \times K_2) = \max\left\{(\ell-1)(n-\ell+1) + \binom{\ell-1}{2}, \binom{2\ell-1}{2}\right\} .$$

Given $n, x \in \mathbb{N}$, $x \leqslant n$, we define two graphs $M_{n,x}$ and $L_{n,x}$ as follows. The graph $M_{n,x}$ is an n-vertex graph whose vertex set is split into sets A and B, $|A| = x, |B| = n - x$, A induces a clique, B induces an independent set, and $M_{n,x}[A, B] \simeq K_{x,n-x}$. The graph $L_{n,x}$ is the complement of $M_{n,n-x}$, i.e., it is an n-vertex graph whose edges induce a clique of order x. Obviously, $e(M_{n,\ell-1}) = (\ell-1)(n-\ell+1) + \binom{\ell-1}{2}$, and $e(L_{n,2\ell-1}) = \binom{2\ell-1}{2}$. Moreover, it is easy to check that there are no ℓ vertex-disjoint edges in either of the graphs $M_{n,\ell-1}$, $L_{n,2\ell-1}$. Therefore, when $\ell < \frac{2}{5}n + O(1)$, the graph $M_{n,\ell-1}$ is (the unique) graph showing that $\mathrm{ex}(n, \ell \times K_2) \geqslant (\ell-1)(n-\ell+1) + \binom{\ell-1}{2}$. The graph $L_{n,2\ell-1}$ is the unique extremal graph for the problem otherwise.

Moon [79] started the investigation of $\mathrm{ex}(n, \ell \times K_r)$. Allen, Böttcher, Hladký, and Piguet [4] only recently determined the behaviour of $\mathrm{ex}(n, \ell \times K_3)$ for the whole range of ℓ; we briefly describe the result in Section 2.4. Simonovits [93] determined the value $\mathrm{ex}(n, \ell \times H)$ for a non-bipartite graph H, fixed value of ℓ and large n.

2.1.1 The Hajnal-Szemerédi Theorem as a tiling result

An equally important density parameter which can be considered in the context of tiling questions is the minimum degree of the host graph. That is, we ask what is the largest possible minimum degree of an n-vertex graph which does not contain ℓ vertex-disjoint copies of H. In the case $H = K_r$, the precise answer is given by the Hajnal-Szemerédi Theorem[1] [54]. In its original formulation, the Hajnal-Szemerédi Theorem asserts that an n-vertex graph G with minimum-degree at least $\frac{r-1}{r}n$ contains a K_r-tiling missing at most $r-1$ vertices of G, thus giving an answer only to the question of almost perfect tilings. The at most $r-1$ exceptional vertices are necessary as n need not be a multiple of r. When the minimum-degree of G is lower, we can however add auxiliary vertices which are complete to G and obtain an n'-vertex graph G' such that the Hajnal-Szemerédi Theorem applies to G'. The restriction of the almost perfect K_r-tiling of G' to G gives a K_r-tiling which is optimal in the worst case; the extremal graphs for the problem are complete r-partite graphs with $r-1$ parts of equal size while the r-th part is smaller.

An asymptotic threshold for a general fixed graph H was determined by Komlós [59]. In this case, the threshold depends on a parameter which Komlós calls the *critical chromatic number*. The critical chromatic number of H is a real between $\chi(H)-1$ and $\chi(H)$. Roughly speaking, graphs H which possess a coloring with $\chi(H)$ colors with one of the color classes small, have the critical chromatic number close to $\chi(H)-1$. On the other hand, graphs H which have only approximately balanced $\chi(H)$-colorings have the critical chromatic number close to $\chi(H)$. There is a natural way how to state our main result, Theorem 2.2, using the critical chromatic number. However, we chose not to as in the bipartite setting of Theorem 2.2 it is possible to give a self-contained formula for the problem. Let us also note that Komlós' result [59] gives an asymptotic min-degree threshold even in the case when H is bipartite. In this case the near-extremal graphs for the problem are complete bipartite graphs.

[1]the special cases of $r = 2$ (tiling with edges) and $r = 3$ (tiling with triangles) are covered by the Dirac Theorem [32], and the Corrádi-Hajnal Theorem [29], respectively

2.1.2 The result

In this chapter we use a variation of the technique developed by Komlós to determine the asymptotic behaviour of the function $\mathrm{ex}(n, \ell \times H)$ for a fixed bipartite graph H. Let H be an arbitrary bipartite graph. Suppose that $b : V(H) \to [2]$ is a proper coloring of H which minimizes $|b^{-1}(1)|$. We define quantities $s(H) := |b^{-1}(1)|$, $t(H) := |b^{-1}(2)|$. Obviously, $s(H) \leqslant t(H)$, and $s(H) + t(H) = v(H)$. Furthermore, we define $V_1(H) := b^{-1}(1)$ and $V_2(H) := b^{-1}(2)$. The sets $V_1(H)$ and $V_2(H)$ are uniquely defined provided that H does not contain a balanced bipartite graph as one of its components; in this other case we fix a coloring b satisfying the above conditions and use it to define uniquely $V_1(H)$ and $V_2(H)$.

Given $s, t \in \mathbb{N}$, we define a function $T_{s,t} : (0,1) \to (0,1)$ by setting

$$T_{s,t}(\alpha) := \max \left\{ \frac{2s\alpha}{s+t} \left(1 - \frac{s\alpha}{2(s+t)} \right), \alpha^2 \right\} , \tag{2.1}$$

for $\alpha \in (0,1)$. Note that $T_{s',t'} = T_{s,t}$ when $s' = ks$ and $t' = kt$. Also, note that

$$T_{s,s}(\alpha) \binom{n}{2} = \mathrm{ex}\left(n, \frac{\alpha n}{2} \times K_2 \right) + o(n^2) , \tag{2.2}$$

and, in general, for $s \leqslant t$, the number $T_{s,t}(\alpha)\binom{n}{2}$ is asymptotically the maximum between the number of edges of $M_{n, \frac{\alpha s}{s+t} n}$ and $L_{n,\alpha n}$.

Our main result is the following.

Theorem 2.2. *Suppose that H is a bipartite graph with no isolated vertices, $s := s(H), t := t(H)$. Let $\alpha \in (0,1)$ and $\varepsilon > 0$. Then there exists an $n_0 = n_0(s,t,\alpha,\varepsilon)$ such that for any $n \geqslant n_0$, any graph G with n vertices and at least $T_{s,t}(\alpha)\binom{n}{2}$ edges contains more than $(1-\varepsilon)\frac{\alpha}{s+t} n$ vertex-disjoint copies of the graph H.*

Let H, s and t be as in the hypothesis of the theorem, $\varepsilon' > 0$ and $\beta \in (0,1)$. Then we may find an $\alpha > \beta(s+t)$ and an $\varepsilon < \varepsilon'$ sufficiently small, such that for n large enough, by Theorem 2.2, any graph G with n vertices and at least $T_{s,t}(\alpha)\binom{n}{2} < (T_{s,t}(\beta(s+t)) + \varepsilon')\binom{n}{2}$ edges contains at least βn vertex-disjoint copies of H. Hence $\mathrm{ex}(n, \beta n \times H) \leqslant T_{s,t}(\beta(s+t))\binom{n}{2} + \varepsilon' n^2$. This asymptotically matches the lower bound which comes — as in Theorem 2.1 — from graphs $M_{n,\beta s n - 1}$ and $L_{n,\beta(s+t)n-1}$. Indeed, neither of these graphs contains βn vertex-disjoint copies of H, as any such copy would require at least s vertices in the clique subgraph of $M_{n,\beta s n-1}$, and at least $s + t = v(H)$ non-isolated vertices in $L_{n,\beta(s+t)n-1}$, respectively. Note however that for most graphs H, the graphs $M_{n,\beta s n-1}$ and $L_{n,\beta(s+t)n-1}$ are not extremal for the problem. For example, we can replace the independent set in the graph $L_{n,\beta(s+t)n-1}$ by any H-free graph. This links us to the Zarankiewicz problem, and suggests that an exact result is not within the reach of current techniques.

The assumption on H to contain no isolated vertices in Theorem 2.2 is made just for the sake of compactness of the statement. Indeed, let H' be obtained from H by removing all the isolated vertices. Then there is a simple relation between the sizes of optimal coverings by vertex disjoint copies of H and H' in an n-vertex graph G. Let x and x' be the number of vertices covered by a maximum family of vertex-disjoint copies of H and H' in G, respectively. We have that

$$x = \min\left\{ v(H)\left\lfloor \frac{n}{v(H)} \right\rfloor, \frac{x'v(H)}{v(H')} \right\} .$$

One can attempt to obtain an analogue of Theorem 2.2 for graphs with higher chromatic number. This however appears to be substantially more difficult. To indicate the difficulty, let us recall that there are two types ($M_{n,x}$ and $L_{n,x}$) of extremal graphs for the H-tiling problem for bipartite H. The graphs $M_{n,x}$ and $L_{n,x}$ have a *block structure*, i.e., their vertex set can be partitioned into *blocks* (two, in this case), such that any two vertices from the same block have almost the same neighborhoods. These two graphs appear even in the simplest case of $H = K_2$ (cf. Theorem 2.1). However, when H is not balanced, if we let α go from 0 to 1, the transition between the two extremal structures which determine the threshold function occurs at a different time in the evolution. On the other hand, there are five types of extremal graphs for the problem of determining $\mathrm{ex}(n, \ell \times K_3)$ as shown in [4]. All the five types have a block structure. It is plausible that when H is a general 3-colorable graph, the same five types of extremal graphs determine the threshold function for H-tilings. However, the transitions between them occur at different times and the block sizes depend on various structural properties of H. In particular, we have indications that the critical chromatic number alone does not determine $\mathrm{ex}(n, \alpha n \times H)$ in this situation.

If $\mathcal{F}$ is a family of graphs, and G is a graph, an $\mathcal{F}$*-tiling* in G is a set of vertex-disjoint subgraphs of G, each of them isomorphic to a graph in $\mathcal{F}$. If $\mathcal{F} = \{H\}$ then we simply say H-tiling. $V(F)$ denotes the vertices of G covered by an $\mathcal{F}$-tiling F, and $|F| = |V(F)|$ is the *size* of the tiling F. If F is a collection of bipartite graphs, we let $V_1(F) = \bigcup_{H\in F} V_1(H)$ and $V_2(F) = \bigcup_{H\in F} V_2(H)$. For $n \in \mathbb{N}$, we write $[n]$ to denote the set $\{1, 2, \ldots, n\}$.

2.2 Tools for the proof of the main result

Given four positive numbers a, b, x, y we say that the pair a, b *dominates* the pair x, y, if $\max\{x, y\}/\min\{x, y\} \geqslant \max\{a, b\}/\min\{a, b\}$. The following easy lemma states that $K_{a,b}$ has an almost perfect $K_{s,t}$-tiling provided that a, b dominates s, t.

Lemma 2.3. *For any $s, t \in \mathbb{N}$ there exists a constant C such that the following holds. Suppose that the pair $a, b \in \mathbb{N}$ dominates s, t. Then the graph $K_{a,b}$ contains a $K_{s,t}$-tiling containing all but at most C vertices of $K_{a,b}$.*

Proof. If $s = t$ then necessarily $a = b$. There obviously exists a $K_{s,t}$-tiling containing all but at most $C := 2(s-1)$ vertices of $K_{a,b}$.

With no loss of generality, we may suppose that $a \leqslant b$ and $s < t$. Then $as \leqslant bt$ and $bs \leqslant at$. A tiling with $\lfloor (bt-as)/(t^2-s^2)\rfloor$ copies of $K_{s,t}$ with the s-part of the $K_{s,t}$ placed in the a-part of the $K_{a,b}$ and $\lfloor (at-bs)/(t^2-s^2)\rfloor$ copies placed the other way misses at most $C := 2(s+t-1)$ vertices of $K_{a,b}$. □

The next two lemmas are easy consequences of Lemma 1.8.

Lemma 2.4. *For every $d > 0, \gamma \in (0, 1)$ and any two graphs R and H, there is an $\varepsilon = \varepsilon(H, d, \gamma) > 0$ such that the following holds for all positive integers s. Let R_s be the graph obtained from R by replacing every vertex of R by s vertices, and every edge of R by a complete bipartite graph between the corresponding s-sets. Let G be any graph obtained similarly from R by replacing every vertex of R by s vertices, and every edge of R with an ε-regular pair of density at least d. If R_s contains an H-tiling of size at least $\gamma v(R_s)$ then so does G.*

Lemma 2.5. *For every bipartite graph H and every $\gamma, d > 0$ there exists an $\varepsilon = \varepsilon(H, d, \gamma) > 0$ such that the following holds. Suppose that there is an H-tiling in $K_{a,b}$ of size x. Let (A, B) be an arbitrary ε-regular pair with density at least d, $|A| = a$, $|B| = b$. Then the pair (A, B) contains an H-tiling of size at least $x - \gamma(a + b)$.*

Finally, let us state a straightforward corollary of the König Matching Theorem (see for example [31, Theorem 2.1.1]).

Fact 2.6. *Let $G = (A \dot\cup B, E)$ be a bipartite graph with color classes A and B. If G has no matching with $l + 1$ edges, then $e(G) \leqslant l \max\{|A|, |B|\}$.*

2.3 The proof

In this section, we first state and prove the main technical result, Lemma 2.7. Then, we show how it implies Theorem 2.2.

For $s,t \in \mathbb{N}$, we set

$$\mathcal{F}_1 := \{K_{s,t}, K_{s,t-1}, K_2\} \quad \text{and} \quad \mathcal{F}_2 := \{K_{st,t^2}, K_{st-1,(t-1)t}, K_{st,(t-1)t}, K_2\} .$$

Let us note that when $s < t$, the sizes of the two color classes of any graph from $\mathcal{F}^* := \mathcal{F}_1 \cup \mathcal{F}_2$ dominate s and t.

Let F be a $K_{s,t}$-tiling in a graph G, $s < t$. Suppose E_0 and E_1 are matchings in $G[V(G)-V(F), V_1(F)]$ and $G[V_2(F)]$, respectively, such that each copy K of $K_{s,t}$ in F has at most one vertex matched by E_0 and at most one vertex matched by E_1. If any $K \in F$ which has a vertex matched by E_0, also has a vertex matched by E_1, then we call the pair (E_0, E_1) an *F-augmentation.* Note that in this case E_0 and E_1 are vertex disjoint, as $V_1(F) \cap V_2(F) = \emptyset$.

The main step in our proof of Theorem 2.2 is the following lemma.

Lemma 2.7. *Let $t > s \geqslant 1, \alpha \in (0,1)$ and $\varepsilon > 0$. Then there exists an $\varepsilon' = \varepsilon'(s,t,\alpha,\varepsilon) > 0$ and an $h = h(s,t,\alpha,\varepsilon) > 0$ such that the following holds. Suppose G is an n-vertex graph with $n \geqslant h$ and $e(G) \geqslant T_{s,t}(\alpha)\binom{n}{2}$, and F is a $K_{s,t}$-tiling in G of maximum size with $|F| \leqslant (1-\varepsilon)\alpha n$. Then one of the following is true:*

(i) there exists an $\mathcal{F}_1$-tiling F' in G with $|F'| \geqslant |F| + \varepsilon' n$, or

(ii) there exists an F-augmentation (E_0, E_1) such that E_0 contains at least $\varepsilon' n$ edges.

Proof. Set

$$\varepsilon' := \frac{1}{4}\min\left\{\frac{\varepsilon\alpha^2}{3t+1}, \frac{\varepsilon s\alpha}{(3t+1)(s+t)}\right\} ,$$

and let h be sufficiently large.

Suppose for a contradiction that the assertions of the lemma are not true.

Set $L := V(G) - V(F)$ and $m := |L|$. Let $\mathcal{C} := \{V_1(K) : K \in F\}, \mathcal{D} := \{V_2(K) : K \in F\}$ and $C := \bigcup \mathcal{C}, D := \bigcup \mathcal{D}$. We call members of $\mathcal{C}$ *lilliputs* while members of $\mathcal{D}$ are *giants.* We say that giant $V_2(K)$ ($K \in F$) is *coupled* with lilliput $V_1(K)$.

As F is a maximum size $K_{s,t}$-tiling in G, by (1.2) we have that

$$e(G[L]) = o(n^2) . \tag{2.3}$$

Let r be the number of copies of $K_{s,t}$ in F. Then $r \leqslant (1-\varepsilon)\alpha n/(s+t)$. Moreover, we have

$$m = n - (s+t)r . \tag{2.4}$$

Let us define an auxiliary graph $H = (V', E')$ as follows. The vertex-set of H is $V' := \mathcal{C} \cup \mathcal{D} \cup L$. For any $x \in L$ and $K \in F$ the edge $xV_1(K)$ belongs to E' iff $\mathrm{N}_G(x) \cap V_1(K) \neq \emptyset$. Similarly, the edge $xV_2(K)$ belongs to E' iff $\mathrm{N}_G(x) \cap V_2(K) \neq \emptyset$. Finally, for any distinct $K, K' \in F$ the edge $V_2(K)V_2(K')$ belongs to E' iff $E_G(V_2(K), V_2(K')) \neq \emptyset$. The vertices L and the vertices $\mathcal{C}$ induce two independent sets in H.

As (i) does not hold, $H[L, \mathcal{D}]$ does not contain a matching with at least $\varepsilon' n$ edges. It follows from Fact 2.6 that

$$e_G(L, D) \leqslant \varepsilon' nt \max\{m, r\} \leqslant t\varepsilon' n^2 \ . \tag{2.5}$$

Let M be a maximum matching in $H[L, \mathcal{C}]$ with l edges. Obviously, $l \leqslant r$. By Fact 2.6, we have that

$$e_G(L, C) \leqslant ls \max\{m, r\} \ . \tag{2.6}$$

Let $\mathcal{C}' \subseteq \mathcal{C}$ be the lilliputs matched by M. We write $\mathcal{D}' \subseteq \mathcal{D}$ for the giants coupled with $\mathcal{C}'$. Set $D' = \bigcup \mathcal{D}'$.

Suppose for a moment that $H[\mathcal{D}'] \cup H[\mathcal{D}', \mathcal{D} - \mathcal{D}']$ contains a matching T with at least $\varepsilon' n$ edges. Let $\mathcal{D}''$ be the giants in $\mathcal{D}'$ matched by T and M' the set of edges in M matching the lilliputs coupled with $\mathcal{D}''$. Then M' and T give rise to an F-augmentation (E_0, E_1) in G with $|E_0| = |M'| \geqslant |T| \geqslant \varepsilon' n$, contradicting our assumption that (ii) does not hold.

So $H[\mathcal{D}'] \cup H[\mathcal{D}', \mathcal{D} - \mathcal{D}']$ does not contain a matching with at least $\varepsilon' n$ edges. Applying Theorem 2.1 and passing to the graph G, we get

$$e(G[D'] \cup G[D', D - D']) \leqslant t^2 \mathrm{ex}(r, \varepsilon' n \times K_2) + r\binom{t}{2} \leqslant 2t^2 \varepsilon' nr + r\binom{t}{2} \ .$$

Therefore,

$$\begin{aligned} e(G[C \cup D]) &= e(G[D'] \cup G[D', D - D']) + e(G[D - D']) + e(G[C]) + e_G(C, D) \\ &\leqslant 2t^2 \varepsilon' nr + r\binom{t}{2} + \binom{(r-l)t}{2} + \binom{rs}{2} + r^2 st. \end{aligned} \tag{2.7}$$

Summing up the bounds (2.3), (2.5), (2.6), and (2.7) we get:

$$\begin{aligned} e(G) &= e(G[L]) + e_G(L, D) + e_G(L, C) + e(G[C \cup D]) \\ &\leqslant o(n^2) + t\varepsilon' n^2 + ls \max\{m, r\} + 2\varepsilon' nrt^2 + r\binom{t}{2} + \binom{(r-l)t}{2} + \binom{rs}{2} + r^2 st. \end{aligned}$$

Using the convexity of $f(l) := ls\max\{m,r\} + \binom{(r-l)t}{2}$ on $[0,r]$, and the fact that $rt \leqslant n$, we get:

$$e(G) \leqslant o(n^2) + 3t\varepsilon' n^2 + r\binom{t}{2} + r^2st + \binom{rs}{2} + \max\left\{\binom{rt}{2}, rs\max\{m,r\}\right\}.$$

However, $r^2s \leqslant \binom{rt}{2} + o(n^2)$, and hence from (2.4) we get:

$$\begin{aligned}e(G) &\leqslant o(n^2) + 3t\varepsilon' n^2 + r\binom{t}{2} + r^2st + \binom{rs}{2} + \max\left\{\binom{rt}{2}, rs(n-(s+t)r)\right\}\\ &< \max\left\{\binom{(s+t)r}{2}, \binom{rs}{2} + rs(n-rs)\right\} + (3t+1)\varepsilon' n^2,\end{aligned}$$

where in the last inequality we have majorized the term $r\binom{t}{2} + o(n^2)$ by $\varepsilon' n^2$. But

$$\binom{(s+t)r}{2} + (3t+1)\varepsilon' n^2 \leqslant \binom{(1-\varepsilon)\alpha n}{2} + \frac{\varepsilon\alpha^2 n^2}{4} < \left(1-\frac{\varepsilon}{2}\right)\frac{\alpha^2 n^2}{2},$$

and

$$\begin{aligned}\binom{rs}{2} + rs(n-rs) + (3t+1)\varepsilon' n^2 &< rsn - \frac{r^2s^2}{2} + \frac{\varepsilon s\alpha n^2}{4(s+t)}\\ &\leqslant \frac{2s\alpha}{s+t}\left(1 - \frac{\alpha s}{2(s+t)} + \frac{\varepsilon(2-\varepsilon)\alpha s}{2(s+t)} - \frac{3\varepsilon}{4}\right)\frac{n^2}{2}\\ &< \frac{2s\alpha}{s+t}\left(1 - \frac{\alpha s}{2(s+t)} - \frac{\varepsilon}{4}\right)\frac{n^2}{2}.\end{aligned}$$

Consequently for large enough n,

$$e(G) < T_{s,t}(\alpha)\binom{n}{2},$$

a contradiction. $\square$

Suppose $G = (V, E)$ is a graph and $r \in \mathbb{N}$. The r-*expansion* of G is the graph $G' = (V', E')$ defined as follows. The vertex set of G' is $V \times [r]$. For $a, b \in [r]$, an edge $((u,a),(v,b))$ belongs to E' iff uv belongs to E. Note that there is a natural projection $\pi_{G'} : V' \to V$ that maps every vertex (u,a) from G' to the vertex u in G. We are interested in the following property of r-expansions. Suppose that K is a copy of any graph from $\mathcal{F}^*$ in G. Then $\pi_{G'}^{-1}(V(K))$ contains a complete bipartite graph B with color classes of sizes $s(K)r$ and $t(K)r$. By Lemma 2.3 we can tile B almost perfectly with copies of $K_{s,t}$. If F is an $\mathcal{F}^*$-tiling in G, we can apply the above operation on each member $K \in F$ and obtain a new tiling F' — which we call *retiling* — in the graph G'.

We are now ready to prove Theorem 2.2.

Proof of Theorem 2.2. Note that it suffices to prove the theorem for $H \simeq K_{s,t}$.

We first deal with the particular case $t = s$. Set $\alpha' := (1 - \varepsilon/4)\alpha$. Let $\varepsilon_1 := \frac{1}{5}(T_{s,t}(\alpha) - T_{s,t}(\alpha'))$, and ε_2 be given by Lemma 2.5 for input parameters H, $d := \varepsilon_1$ and $\gamma := \alpha\varepsilon/8$. Suppose that k_0 is sufficiently large. Let M be the bound from Lemma 1.1 for precision $\varepsilon_R := \min\{\varepsilon_1, \varepsilon_2\}$ and minimal number of clusters k_0 (we shall not utilize the "furthermore" part of Lemma 1.1, and thus use the input $\ell = 1$). Let C be given by Lemma 2.3 for the input parameters s, t. Fix $n_0 \gg MC$. Suppose that G is an n-vertex graph, $n \geqslant n_0$, with at least $T_{s,t}(\alpha)\binom{n}{2}$ edges. We apply Lemma 1.1 on G to obtain an (ε_R, d)-reduced graph R with k clusters, $k_0 \leqslant k \leqslant M$. We have that

$$\begin{aligned} e(R) &\geqslant (T_{s,t}(\alpha) - d - 3\varepsilon_1)\binom{k}{2} = \left(T_{s,t}(\alpha') + \frac{1}{5}(T_{s,t}(\alpha) - T_{s,t}(\alpha'))\right)\binom{k}{2} \\ &\overset{(2.2)}{>} \mathrm{ex}\left(k, \frac{\alpha' k}{2} \times K_2\right) . \end{aligned}$$

Therefore, R contains at least $\frac{\alpha' k}{2}$ independent edges. These edges correspond to regular pairs in G which can be tiled almost perfectly with copies of $K_{s,t}$, by means of Lemma 2.3 and Lemma 2.5. Elementary calculations give that in this way we get a tiling of size at least $(1 - \varepsilon)\alpha n$.

Consequently we may suppose that $t > s$. We first define a handful of parameters. Set

$$\alpha' := \frac{6 - 4\varepsilon}{6 - 3\varepsilon}\alpha, \quad \gamma := (1 - \varepsilon/2)\alpha', \quad d := \frac{2}{5}(T_{s,t}(\alpha) - T_{s,t}(\alpha')) .$$

Note that $\gamma = (1 - 2\varepsilon/3)\alpha$.

Let ε_R be given by Lemma 2.4 for input graph $K_{s,t}$, density $d/2$ and approximation parameter γ. We may suppose that ε_R is sufficiently small such that $\gamma(1 - \varepsilon_R) > (1 - \varepsilon)\alpha$ and $\varepsilon_R < d/2$. Let C be given by Lemma 2.3 for input s, t. Further, let ε' and h be given by Lemma 2.7 for input parameters α' and $\varepsilon/4$. We may assume that $\varepsilon' < \varepsilon$. Set

$$p := t^2 \left\lceil \frac{4C}{\varepsilon'} \right\rceil , \quad q := \left\lceil \frac{2t}{\varepsilon'} \right\rceil$$

Let M be the upper bound on the number of clusters given by Lemma 1.1 for input parameters h (for the minimal number of clusters), $\ell = 1$ (for the complexity of prepartition), and $\varepsilon_R p^{-q}/2$ (for the precision). Let $n_0 > Mp^q$ be sufficiently large.

Suppose now that G is a graph with $n > n_0$ vertices and at least $T_{s,t}(\alpha)\binom{n}{2}$ edges. We first apply Lemma 1.1 to G with parameters $\varepsilon_R p^{-q}/2$ and h. In this way we obtain an $(\varepsilon_R p^{-q}/2, d)$-reduced graph R with at least h vertices.

Let us now define a sequence of graphs $R^{(i)}$ by setting $R^{(0)} = R$ and letting $R^{(i)}$ be the p-expansion of $R^{(i-1)}, i = 1, 2, \dots, q$. Note that $e(R^{(i)}) \geqslant T_{s,t}(\alpha')\binom{v(R^{(i)})}{2}$ for every $i \in \{0, 1, \dots, q\}$.

Let $F^{(i)}$ be a maximum size $K_{s,t}$-tiling in $R^{(i)}$ for $i = 0, 1, \dots, q$. We claim that

$$|F^{(i)}| \geqslant \min\left\{\frac{i\varepsilon' v(R^{(i)})}{2t}, \left(1 - \frac{\varepsilon}{2}\right)\alpha' v(R^{(i)})\right\} . \tag{2.8}$$

To this end it suffices to show that for any $i \geqslant 1$,

(C1) if $|F^{(i-1)}| > (1 - \varepsilon/4)\alpha' v(R^{(i-1)})$, then $\frac{|F^{(i)}|}{v(R^{(i)})} \geqslant \frac{|F^{(i-1)}|}{v(R^{(i-1)})} - \frac{\varepsilon\alpha'}{4}$, and

(C2) if $|F^{(i-1)}| \leqslant (1 - \varepsilon/4)\alpha' v(R^{(i-1)})$, then $\frac{|F^{(i)}|}{v(R^{(i)})} \geqslant \frac{|F^{(i-1)}|}{v(R^{(i-1)})} + \frac{\varepsilon'}{2t}$.

In the case (C1), according to Lemma 2.3, the retiling of $F^{(i-1)}$ in $R^{(i)}$ has size at least $|F^{(i-1)}|(p - C) > (1 - \varepsilon/2)\alpha' v(R^{(i)})$, thus proving the statement.

Consequently we may suppose that we are in case (C2). Apply Lemma 2.7 to the graph $R^{(i-1)}$ and the tiling $F^{(i-1)}$, with parameters α' and $\varepsilon/4$.

Suppose first that assertion (i) of the lemma holds. Then $R^{(i-1)}$ contains an $\mathcal{F}_1$-tiling F with $\frac{|F|}{v(R^{(i-1)})} \geqslant \frac{|F^{(i-1)}|}{v(R^{(i-1)})} + \varepsilon'$. By retiling F, we get a $K_{s,t}$-tiling in $R^{(i)}$ with size at least $|F|(p - C) > i\varepsilon' v(R^{(i)})/(2t)$, thus proving the statement.

Suppose now that assertion (ii) of Lemma 2.7 is true. Then $R^{(i-1)}$ contains an $F^{(i-1)}$-augmentation (E_0, E_1) with $|E_0| \geqslant \varepsilon' v(R^{(i-1)})$. Let $r = p/t$. We shall denote by T the t-expansion of $R^{(i-1)}$ and by T' the r-expansion of T. Note that T' is isomorphic to $R^{(i)}$.

Let us build an $\mathcal{F}_2$-tiling in T in the following way.

For every edge $e = (u, v) \in E_0$ with $u \in V(F^{(i-1)})$ we choose an edge $e' = (u', v')$ in T with $\pi_T(u') = u$ and $\pi_T(v') = v$. We shall denote by w_e the vertex u' corresponding to u.

For every edge $e = (u, v) \in E_1$ we choose a set S_e of t independent edges in $\pi_T^{-1}(e)$.

For every $K \in F^{(i-1)}$ we shall also choose a subgraph K' of T. We distinguish the following cases. If K has no vertex matched by E_0 or E_1, then we let $K' := T[\pi_T^{-1}(K)]$. If K has a vertex u matched by E_1 but no vertex matched by E_0, we let $K' := T[\pi_T^{-1}(K-u)]$. Then $K' \simeq K_{st,(t-1)t}$. Finally, if K has a vertex u matched by an edge $e \in E_0$ and a vertex v matched by an edge in E_1, we let $K' := T[\pi_T^{-1}(K-v)] - w_e$. Note that in this last case $K' \simeq K_{st-1,(t-1)t}$.

It is easy to see that

$$F := \{e' : e \in E_0\} \cup \{K' : K \in F^{(i-1)}\} \cup \left(\bigcup_{e \in E_1} S_e\right)$$

is an $\mathcal{F}_2$-tiling in T. Moreover, we have that $\frac{|F|}{v(T)} \geqslant \frac{|F^{(i-1)}|}{v(R^{(i-1)})} + \frac{\varepsilon'}{t}$. So the retiling of F in T' has size at least $|F|(r-C) \geqslant i\varepsilon' v(R^{(i)})/(2t)$. This proves (C2) and also (2.8).

Using Lemma 1.5, we may subdivide every cluster corresponding to a vertex of R into p^q equal-sized parts, by discarding some vertices if necessary. This gives us an $(\varepsilon_R, d/2)$-reduced graph R'. By construction $R' \simeq R^{(q)}$. By (2.8), there is a $K_{s,t}$-tiling F in R' with size at least $(1-\varepsilon/2)\alpha' v(R')$. Let G' be the subgraph of G induced by the clusters corresponding to the vertices of R'. By applying Lemma 2.4 to R', we see that G' has a $K_{s,t}$-tiling of size at least $\gamma v(G') \geqslant \gamma(1-\varepsilon_R)v(G) > (1-\varepsilon)\alpha v(G)$, and so does G.

This finishes the proof of Theorem 2.2. □

2.4 Tiling with triangles

In this section we describe a recent result of Allen, Böttcher, Piguet and the author which determines $\mathrm{ex}(n, \ell \times K_3)$ for large enough n. This section is brief and meant only to demonstrate connections between the problem and other parts of the thesis; the picture of the tiling problems concerning determination of $\mathrm{ex}(n, \ell \times H)$ is then summarised in Section 2.5. In particular, we omit the (lengthy) proof of the main result.

Mantel's Theorem 1.10 asserts that each n-vertex graph G with more than $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil$ edges contains a triangle[2]. What happens when the threshold $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil$ is exceeded? Can we quantify the presence of triangles in G?

One natural approach to this broad question is to determine how many triangles is G guaranteed to have, as a parameter of the edge density of G. The first instance of this problem is when G on n vertices contains $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil + 1$ edges. In this case Rademacher (unpublished, around 1960) proved that there must be at least $\lfloor \frac{n}{2} \rfloor$ triangles in G. Solving a long-standing open problem, Razborov [83] determined a

[2]This was later extended by Turán, see Section 1.5. However we restrict ourselves to investigating extremal problems concerning only triangles, and in that sense Mantel's theorem is more relevant.

tight bound $f(\alpha)$ such that each n-vertex graph with αn^2 edges contains at least $(f(\alpha) + o_{n\to\infty}(1))n^3$ triangles. Note that $f(\alpha) = 0$ for $\alpha \in [0, \frac{1}{4}]$, while by Mantel's theorem and the Supersaturation Theorem [38], $f(\alpha) > 0$ for $\alpha \in (\frac{1}{4}, \frac{1}{2})$. It is striking that the function $f(\alpha)$ exhibits very complicated behaviour.

In this section, we deal with a different measure of the presence of triangles. We ask what edge density in an n-vertex graph guarantees k *vertex-disjoint* triangles, i.e., we investigate the number $\mathrm{ex}(n, k \times K_3)$. Prior to our work this question was considered by Erdős [33] and by Moon [79]; the former proved the exact result when $n \geqslant 400k^2$, and the latter when $n \geqslant 9k/2 + 4$. Interestingly, although Moon states that his result 'almost certainly remains valid for somewhat smaller values of n also', in fact he almost reaches a natural barrier: the graph which Moon proved to be extremal (the first in Figure 2.1 below) is only extremal when $n \geqslant 9k/2+3$. We give a precise answer to the question for all values of k when n is greater than an absolute constant n_0 in Theorem 2.9 below.

One can deduce the following generalisation of the Corrádi-Hajnal Theorem as indicated in Section 2.1.1. Every n-vertex graph G with $\frac{n}{2} < \delta(G) < \frac{2n}{3}$ contains a triangle tiling with at least $2\delta(G) - n$ triangles. This bound is tight, as is shown by unbalanced complete tripartite graphs. Our main result, Theorem 2.9 below, is therefore a density version of the Corrádi-Hajnal theorem.

In Definition 2.8 we construct four graphs $E_1(n, k)$, $E_2(n, k)$, $E_3(n, k)$, $E_4(n, k)$ each on n vertices. Theorem 2.9 below, asserts that these graphs are extremal with respect to the number of edges subject to not containing $(k+1) \times K_3$. We say that an edge e (or more generally a set of vertices) *meets* a set of vertices X if e and X intersect. The edge e meets X in X' if $X' = X \cap e$.

Definition 2.8 (extremal graphs)**.** *Let n and k be non-negative integers with $k \leqslant \frac{n}{3}$. We define the following four graphs (see also Figure 2.1).*[3]

$E_1(n,k)$: *Let $X\dot\cup Y_1\dot\cup Y_2$ with $|X| = k$, $|Y_1| = \lceil\frac{n-k}{2}\rceil$, and $|Y_2| = \lfloor\frac{n-k}{2}\rfloor$ be the vertices of $E_1(n,k)$. Insert all edges intersecting X, and between Y_1 and Y_2.*

$E_2(n,k)$: *The second class of extremal graphs is defined only for $k < \frac{n-1}{4}$. Let $X\dot\cup Y_1\dot\cup Y_2$ with $|X| = 2k+1$, $|Y_1| = \lfloor\frac{n}{2}\rfloor$, and $|Y_2| = \lceil\frac{n}{2}\rceil - 2k - 1$ (or $|Y_1| = \lceil\frac{n}{2}\rceil$, and $|Y_2| = \lfloor\frac{n}{2}\rfloor - 2k - 1$) be the vertices of $E_2(n,k)$. Insert all edges within X, and between Y_1 and $X \cup Y_2$. If n is odd, this construction captures two graphs, if n is even just one.*

[3] The constructions for $E_2(n, k)$ and $E_4(n, k)$ do not give unique graphs. We collectively denote all graphs constructed in this way by $E_2(n, k)$ and $E_4(n, k)$, respectively.

$E_3(n,k)$: *Let $X\dot{\cup}Y_1$ with $|X| = 2k+1$ and $|Y_1| = n-2k-1$ be the vertices of $E_3(n,k)$. Insert all edges intersecting X.*

$E_4(n,k)$: *The fourth class of extremal graphs is defined only for $k \geqslant \frac{n}{6} - 2$. The vertex set is formed by five disjoint sets X, Y_1, Y_2, Y_3, and Y_4, with $|Y_1| = |Y_3|$, $|Y_2| = |Y_4|$, $|Y_1| + |Y_2| = n - 3k - 2$, and $|X| = 6k - n + 4$. Insert all edges in X, between X and $Y_1 \cup Y_2$, and between $Y_1 \cup Y_4$ and $Y_2 \cup Y_3$. Thus the choice of $|Y_1|$ determines a particular graph in the class $E_4(n,k)$. All graphs in $E_4(n,k)$ have the same number of edges.*

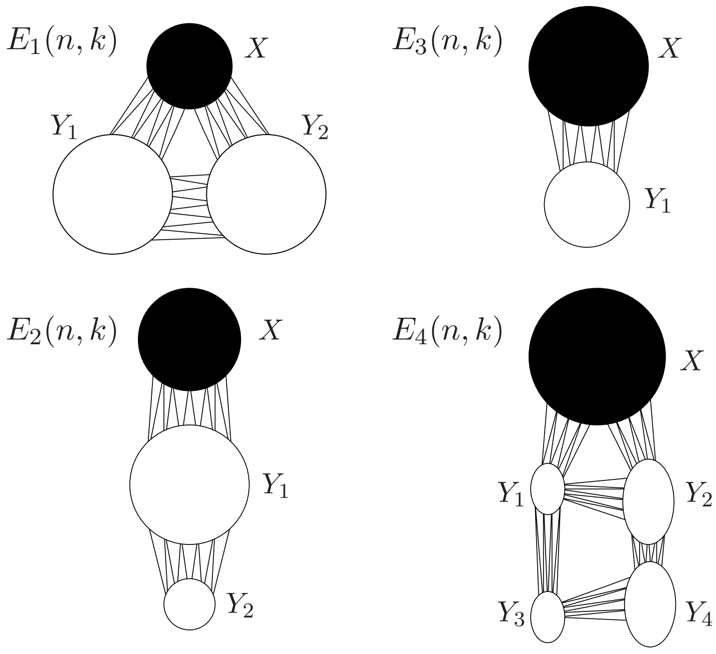


Figure 2.1: The extremal graphs.

Theorem 2.9. *There exists n_0 such that for each $n > n_0$ we have the following. Let $n \geqslant n_0$ and $k \leqslant \frac{n}{3}$ be positive integers. Let G be a $(k+1) \times K_3$-free graph on n vertices. Then*

$$e(G) \leqslant \max_{j \in [4]} e\big(E_j(n,k)\big) \ . \tag{2.9}$$

In the sequel we shall provide explicit formulas for the number of edges of $E_i(n,k)$ and then analyse for which values of k which of the extremal graphs dominates, i.e., attains the maximal number of edges in (2.9).

Clearly, the graphs $E_i(n,k)$ are edge-maximal subject to not containing $(k+1) \times K_3$, the only exception is $E_4(n,k)$ for $k \lessapprox \frac{n}{5}$. However $E_4(n,k)$ irrelevant for the statement of Theorem 2.9 in the range $0 \leqslant k \lessapprox \frac{n}{5}$; see the discussion below and

Table 2.1. The graphs $E_i(n,k)$ have the following numbers of edges (after an exact formula we identify the leading terms; to this end we use symbol $\approx$).

$$
\begin{aligned}
e\big(E_1(n,k)\big) &= \binom{k}{2} + k(n-k) + \left\lceil \frac{n-k}{2} \right\rceil \left\lfloor \frac{n-k}{2} \right\rfloor \approx \tfrac{1}{4}n^2 - \tfrac{1}{4}k^2 + \tfrac{1}{2}kn \ , \\
e\big(E_2(n,k)\big) &= \binom{2k+1}{2} + \left\lceil \frac{n}{2} \right\rceil \left\lfloor \frac{n}{2} \right\rfloor \approx \tfrac{1}{4}n^2 + 2k^2 \ , \\
e\big(E_3(n,k)\big) &= \binom{2k+1}{2} + (2k+1)(n-2k-1) \approx 2kn - 2k^2 \ , \\
e\big(E_4(n,k)\big) &= \binom{6k-n+4}{2} + (6k-n+4)(n-3k-2) + (n-3k-2)^2 \\
&\approx \frac{n^2}{2} + 9k^2 - 3kn \ .
\end{aligned}
$$

Comparing these edge numbers reveals that, as k grows from 1 to $n/3$, the extremal graphs dominate in the following order (for n sufficiently large). In the beginning $E_1(n,k)$ has most edges of these four graphs until $k \approx \frac{2n}{9}$, where it is surpassed by $E_2(n,k)$. At $k \approx \frac{n-1}{4}$ this extremal structure ceases to exist and is replaced by $E_3(n,k)$, until finally at $k \approx (5+\sqrt{3})n/22$ the graph $E_4(n,k)$ takes over. The exact thresholds are listed in Table 2.1. Further, the edge numbers of the graphs $E_i(n,k)$ are plotted on Figure 2.2.

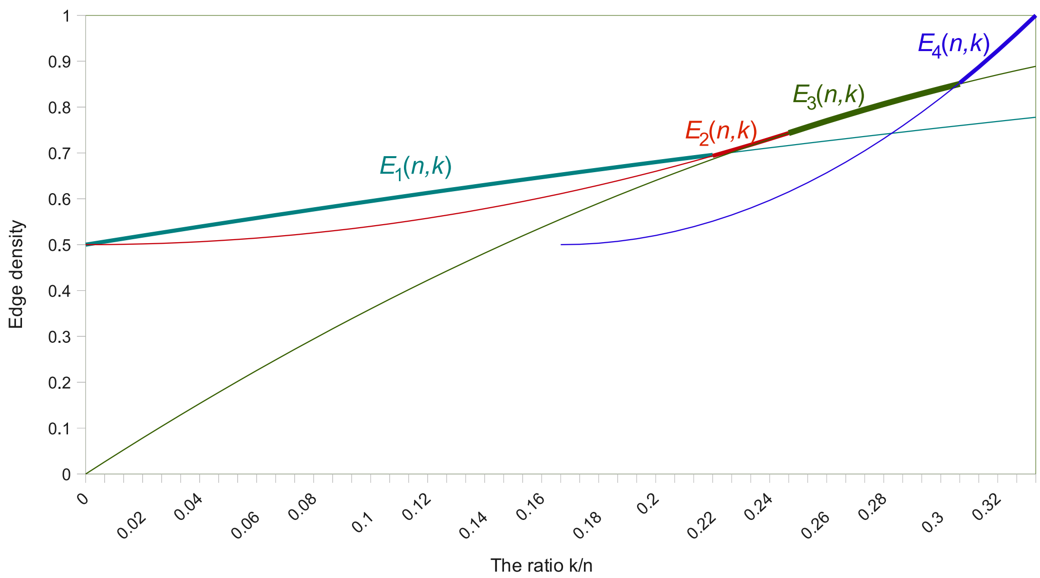


Figure 2.2: Edge densities of the graphs $E_i(n,k)$ where k ranges from 0 to $\frac{n}{3}$.

Observe that for fixed n, as k increases, the transitions from $E_1(n,k)$ to

$E_2(n,k)$ and from $E_3(n,k)$ to $E_4(n,k)$ are not continuous: $\Theta(n^2)$ edges must be edited to change from the former to the latter structure. The transition from $E_2(n,k)$ to $E_3(n,k)$ however is continuous.

graph	**extremal for**
$E_1(n,k)$	$1 \leqslant k \leqslant \frac{2n-6}{9}$
$E_2(n,k)$	$\frac{2n-6}{9} \leqslant k \leqslant \frac{n-1}{4}$
$E_3(n,k)$	$\frac{n-1}{4} \leqslant k \leqslant \frac{5n-12+\sqrt{3n^2-10n+12}}{22}$
$E_4(n,k)$	$\frac{5n-12+\sqrt{3n^2-10n+12}}{22} \leqslant k \leqslant \frac{n}{3}$

Table 2.1: Transitions between the extremal graphs.

Observe that Theorem 2.9 extends the previous partial result by Moon. Moon's result covers exactly the range when the graph $E_1(n,k)$ is extremal.

We believe that our requirement for n to be large in the statement of Theorem 2.9 is just an artefact of our proof which is caused by employing a stability-type argument (in the sense of Simonovits [92]). We use this argument however only when k is in the range for which $E_4(n,k)$ is extremal. Thus our theorem actually gives a tight bound *for all* n and for all values $k \lessapprox (5+\sqrt{3})n/22$.

2.4.1 Sketch of the proof of Theorem 2.9

The basic idea of the proof of Theorem 2.9 is the most natural one: For a $(k+1)\times K_3$-free graph G let $\mathcal{T}(G)$ denote any set of k vertex-disjoint triangles which maximises the cardinality of the set $\mathcal{M}(G)$ of vertex-disjoint edges outside $\mathcal{T}(G)$ (i.e., $\mathcal{M}(G)$ is a matching). Let $\mathcal{I}(G)$ be the remaining set of independent vertices. One would now aim to upper-bound the edge counts within the sets $\mathcal{T}(G)$, $\mathcal{M}(G)$, and $\mathcal{I}(G)$ and between them using the maximality of the choice of $\mathcal{T}(G)$ and $\mathcal{M}(G)$. It turns out that that decomposition of G is too rough and insufficient to get the overall upper bound $\max_{i\in[4]} e(E_i(n,k))$. We therefore partition $\mathcal{T}(G)$ into four sets depending on how the triangles are connected to $\mathcal{M}(G)$ and $\mathcal{I}(G)$. We are then able to provide with good bounds on the number of edges inside different parts of this refined decomposition. The proof itself however is unpleasantly technical and spans over thirty pages. Interestingly enough, we need Theorem 5.3 as an auxiliary result for proving Theorem 2.9. (The historical development was the other way round, though: our need for a result similar to what is now Theorem 5.3 during our work on triangle tilings led us to a much more general notion of *Turánnical hypergraphs*, thoroughly studied in Chapters 4 and 5).

2.5 Extremal theory of partial tilings

As discussed in Section 2.1.1, asymptotic problems on partial tilings by an arbitrary graph H which involve the minimum degree condition are completely answered by Komlós [59]. The min-degree bound depends on just two parameters: the chromatic number $\chi(H)$, and the critical chromatic number $\chi_{cr}(H)$ (introduced in [59]). Even though computing $\chi(H)$ and $\chi_{cr}(H)$ is an intractable problem from a computational point of view in general, both parameters have a simple combinatorial description. In Theorem 2.2 we answered asymptotically problems on partial tilings by an arbitrary bipartite graph H which involve the bound on the number of edges. The quantities in Theorem 2.2 again depended on a simple parameter of H (which can be actually reformulated in terms of $\chi_{cr}(H)$). Theorem 2.9 on the other hand showed that even for $H = K_3$ the picture is quite complex, with four types of extremal graphs. Further, in the same vein as Komlós' result extends the Hajnal-Szemerédi Theorem, and as our Theorem 2.2 extends the Erdős-Gallai Theorem 2.1 we could ask for a counterpart of Theorem 2.9 for tilings with a general three-colourable graph H. In that respect we think that the behaviour of the extremal graph will again undergo three transitions (as in Theorem 2.9), and that the extremal graphs will be based on the graphs $E_i(n,k)$ defined in Definition 2.8. However the actual extremal graphs seems to be certain "skewed" versions of $E_i(n,k)$, i.e., the ratio of the block sizes is changed compared to the original graphs depending on H. We were unable to puzzle out the relation between the structure of H and the change of the block sizes of the extremal graphs.

Even more ambitiously, one could ask for a graph parameter which describes the asymptotic behaviour of the function $\mathrm{ex}(n, k \times H)$ for a general graph H. It would be most interesting to be able to say something even about the simplest open case $H = K_4$.

Chapter 3

Between Turán's Theorem and Pósa's Conjecture

3.1 Introduction

As noted in Section 1.5, Turán's Theorem (as many other fundamental results in Extremal Graph Theory) deals with the problem of finding a *small subgraph* (such as the triangle) in a large graph. Dirac's theorem [32] – another classical result from the area – on the other hand considers *spanning* target graphs. Clearly, any average degree condition on the host graph (which parametrises Turán's Theorem, for example) that enforces a connected spanning subgraph must be trivial, and hence the average degree needs a suitable replacement in this setting. Here, the minimum degree is a natural candidate, and indeed, Dirac's theorem asserts that every graph G with minimum degree $\deg^{\min}(G) > \frac{1}{2}n$ has a Hamilton cycle. This implies in particular that G has a matching covering $2\lfloor n/2 \rfloor$ vertices.

A 3-chromatic version of this matching result follows from a theorem by Corrádi and Hajnal [29]: the minimum degree condition $\deg^{\min}(G) \geqslant 2\lfloor n/3 \rfloor$ implies the existence of a so-called *spanning triangle factor* in G, that is, a collection of $\lfloor n/3 \rfloor$ vertex disjoint triangles. A well-known conjecture of Pósa (see, e.g., [35]) asserts that roughly the same minimum degree actually guarantees the existence of a connected super-graph of a spanning triangle factor. It states that any graph G with $\deg^{\min}(G) \geqslant \frac{2}{3}n$ contains a spanning *squared cycle* C_n^2, where the square of a graph, F^2, is obtained from F by adding edges between all pairs of vertices with distance 2 in F. This can be seen as a 3-chromatic analogue of Dirac's theorem, which turned out to be much more difficult than its 2-chromatic cousin.

Fan and Kierstead [39] proved an approximate version of Pósa's conjecture for large n. In addition they determined a sufficient and best possible minimum degree condition for the case that the squared cycle in Pósa's conjecture is replaced by a *squared path* P_n^2, i.e., the square of a spanning path P_n.

Theorem 3.1 (Fan & Kierstead [40]). *If G is a graph on n vertices with minimum degree* $\deg^{\min}(G) \geqslant (2n-1)/3$*, then G contains a spanning squared path P_n^2.*

The Pósa Conjecture was verified for large values of n by Komlós, Sarközy, and Szemerédi [60]. The proof in [60] actually asserts the following stronger result, which guarantees not only spanning squared cycles but additionally squared cycles of all lengths between 3 and n that are divisible by 3.

Theorem 3.2 (Komlós, Sárközy & Szemerédi [60]). *There exists an integer n_0 such that for all integers $n > n_0$ any graph G of order n and minimum degree* $\deg^{\min}(G) \geqslant \frac{2}{3}n$ *contains all squared cycles $C_{3\ell}^2 \subseteq G$ with* $3 \leqslant 3\ell \leqslant n$*. If furthermore $K_4 \subseteq G$, then $C_\ell^2 \subseteq G$ for any* $3 \leqslant \ell \leqslant n$ *with* $\ell \neq 5$.

For squared cycles C_ℓ^2 with ℓ not divisible by 3 the additional condition $K_4 \subseteq G$ is necessary because these target graphs are not 3-colourable and hence a complete 3-partite graph shows that one cannot hope to force C_ℓ^2 unless $\deg^{\min}(G) \geqslant (2n+1)/3$. If $\deg^{\min}(G) \geqslant (2n+1)/3$, on the other hand, then Turán's Theorem asserts that G contains a copy of K_4 and hence Theorem 3.2 implies $C_\ell^2 \subseteq G$ for any $3 \leqslant \ell \leqslant n$ with $\ell \neq 5$. The case $\ell = 5$ has to be excluded because C_5^2 is the 5-chromatic K_5.

In this chapter we address the question of what happens between these two extrema of target graphs with constant order and order n. We are interested in essentially best possible minimum degree conditions that enforce subgraphs covering a certain percentage of the host graph.

Let us start with a simple example. It is easy to see that every graph G with minimum degree $\deg^{\min}(G) \geqslant \delta$ for $0 \leqslant \delta \leqslant \frac{1}{2}n$ has a matching covering at least 2δ vertices (see Proposition 3.11(a)). This gives a linear dependence between the forced size of a matching in the host graph and its minimum degree. A more general form of the result of Corrádi and Hajnal [29] mentioned earlier is a variant of this linear dependence for triangle factors.

Theorem 3.3 (Corrádi & Hajnal [29]). *Let G be a graph on n vertices with minimum degree* $\deg^{\min}(G) = \delta \in [\frac{1}{2}n, \frac{2}{3}n]$*. Then G contains $2\delta - n$ vertex disjoint triangles.*

(See Section 2.1.1 for a sketch of a reduction of the version of the Corrádi-Hajnal Theorem above to the original version.)

The main theorem of this chapter is a corresponding result mediating between Turán's theorem and Pósa's conjecture. More precisely, our aim is to provide exact minimum degree thresholds for the appearance of a squared path P_ℓ^2 and a squared cycle C_ℓ^2.

There are at least two reasonable guesses one might make as to what minimum degree $\deg^{\min}(G) =: \delta$ will guarantee which length $\ell := \ell(n, \delta)$ of squared path (or longest squared cycle). On the one hand, the degree threshold for a *spanning* squared path or cycle and for a spanning triangle factor are approximately the same. So perhaps this remains true for smaller ℓ: in light of Theorem 3.3 one could expect that $\ell(n, \delta)$ were roughly $3(2\delta - n)$. This turns out to be far too optimistic.

On the other hand, proofs of preceding results dealing with spanning subgraphs essentially combine greedy techniques with local changes. They simply start to construct the desired subgraph in (almost) any location, and in the event of getting stuck change only a few of the vertices embedded so far; at no time do they scrap an entire half-constructed object and start anew. It would not be unreasonable to believe that this technique also leads to best possible minimum degree conditions for large but not spanning subgraphs. Clearly, in the case of (unsquared) paths such a greedy strategy provides a path of length $\deg^{\min}(G) + 1$. As G might be disconnected, however, it cannot guarantee longer paths if $\deg^{\min}(G) < n/2$. For squared paths the following construction shows that with an arbitrary starting location one cannot hope for squared paths on more than $\frac{3}{2}(2\deg^{\min}(G) - n)$ vertices: If G contains disjoint cliques C and C' of orders $2\delta - n$ and $n - \delta$, and an independent set I of order $n - \delta$ such that all vertices of C and C' are connected to all vertices of I but not to other vertices of G, then it is not difficult to see that the longest squared path in G starting in an edge of C has length $\frac{3}{2}(2\deg^{\min}(G) - n)$. This could lead to the idea that $\ell(n, \delta)$ were approximately $\frac{3}{2}(2\delta - n)$. It is true that there are squared paths of this length in G—but this lower bound is almost always excessively pessimistic. In other words, it turns out that one has to carefully choose the 'region' of G to look for the desired squared path. Since spanning squared paths use all vertices of G this problem does not occur for these subgraphs.

For fixed n both guesses propose a linear dependence between δ and the length $\ell(n, \delta)$ of a forced squared path (or cycle) of any n-vertex graph with minimum degree δ. As we will see below $\ell(n, \delta)$ as a function of δ behaves very differently: it is piece-wise linear but jumps at certain points. (These jumps can be viewed as phase transitions for the appearance of squared paths or cycles.) To make this precise we introduce the following functions. Given two positive integers n and δ with $\delta \in (\frac{1}{2}n, n - 1]$, we define $r_p(n, \delta)$ to be the largest integer r such that

$n-\delta+\lfloor\delta/r\rfloor > \delta$ and $r_c(n,\delta)$ to be the largest integer r such that $n-\delta+\lceil\delta/r\rceil > \delta$. We then define

$$\begin{aligned} \mathrm{sp}(n,\delta) &:= \min\left\{ \left\lceil \tfrac{3}{2}\lceil \delta/r_p(n,\delta)\rceil + \tfrac{1}{2} \right\rceil, n \right\}, \quad \text{and} \\ \mathrm{sc}(n,\delta) &:= \min\left\{ \left\lfloor \tfrac{3}{2}\lceil \delta/r_c(n,\delta)\rceil \right\rfloor, n \right\}. \end{aligned} \tag{3.1}$$

Observe that $\mathrm{sc}(n,\delta) \leqslant \mathrm{sp}(n,\delta)$ and that for almost every $\alpha \in (0,1)$ we have $\lim_{n\to\infty} \mathrm{sc}(n,\alpha n)/n = \lim_{n\to\infty} \mathrm{sp}(n,\alpha n)/n$. The dependence between $\mathrm{sp}(n,\delta)$ and δ is illustrated in Figure 3.1.

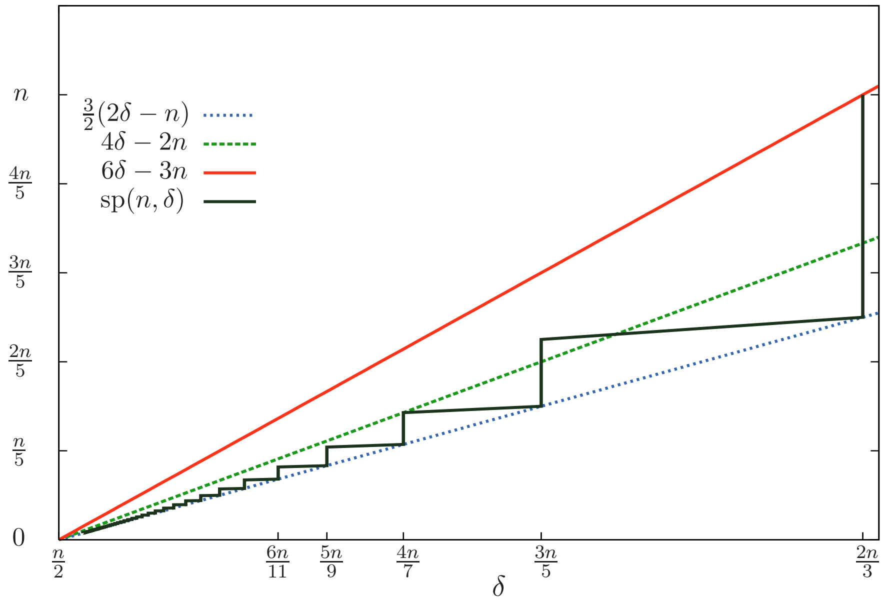


Figure 3.1: The behaviour of $\mathrm{sp}(n,\delta)$.

Our main theorem now states states that $\mathrm{sp}(n,\delta)$ and $\mathrm{sc}(n,\delta)$ are the maximal lengths of squared paths and cycles, respectively, forced in an n-vertex graph G with minimum degree δ. More generally, and in accordance with Theorem 3.2, we show that G also contains any shorter squared cycle with length divisible by 3 (see (i) of Theorem 3.4). We shall show below that these results are tight by explicitly constructing extremal graphs $G_p(n,\delta)$ and $G_c(n,\delta)$ for squared paths and cycles. While the extremal graphs of all previously discussed results are Turán graphs (complete r-partite graphs, where $r=3$ in the case of squared paths and cycles) the graphs $G_p(n,\delta)$ and $G_c(n,\delta)$ have a rather different structure. In fact they do contain squared cycles C_ℓ^2 for all $3 \leqslant \ell \leqslant \mathrm{sc}(n,\delta)$ with $\ell \neq 5$. If any one of these 'extra' squared cycles with chromatic number 4 is not present in the host graph G, then (*ii*) of Theorem 3.4 guarantees even much longer squared cycles C_ℓ^2 in G, where ℓ is a multiple of 3.

Theorem 3.4. *For any $\nu > 0$ there exists an integer n_0 such that for all integers $n > n_0$ and $\delta \in [(\frac{1}{2}+\nu)n, \frac{2}{3}n]$ the following holds for all n-vertex graphs G with minimum degree $\deg^{\min}(G) \geqslant \delta$.*

(*i*) *$P^2_{\mathrm{sp}(n,\delta)} \subseteq G$ and $C_\ell^2 \subseteq G$ for every $\ell \in \mathbb{N}$ with $3 \leqslant \ell \leqslant \mathrm{sc}(n,\delta)$ such that 3 divides ℓ.*

(*ii*) *Either $C_\ell^2 \subseteq G$ for every $\ell \in \mathbb{N}$ with $3 \leqslant \ell \leqslant \mathrm{sc}(n,\delta)$ and $\ell \neq 5$, or $C_\ell^2 \subseteq G$ for every $\ell \in \mathbb{N}$ with $3 \leqslant \ell \leqslant 6\delta - 3n - \nu n$ such that 3 divides ℓ.*

The proof of this result relies on Szemerédi's Regularity Lemma and is presented together with the main lemmas in Section 3.2. Theorem 3.4 cannot be extended to all values of $\deg^{\min}(G)$ with $\deg^{\min}(G) - \frac{1}{2}n = o(n)$ because for infinitely many values of m there are C_4-free graphs F on m vertices with $\deg^{\min}(F) \geqslant \frac{1}{2}\sqrt{m}$ (see [84]). Then, letting G be the n-vertex graph obtained from F by adding an independent set I on $m - \lfloor\frac{1}{2}\sqrt{m}\rfloor$ vertices and inserting all edges between F and I, it is easy to see that $\deg^{\min}(G) > \frac{1}{2}n + \frac{1}{5}\sqrt{n}$ but G does not contain a copy of C_6^2.

The following *extremal graphs* show that the bounds in (*i*) and (*ii*) of Theorem 3.4 are tight (see also Figure 3.2). For (*ii*) consider the complete tripartite graph $K_{n-\delta,n-\delta,2\delta-n}$. Clearly, this graph has minimum degree δ and does not contain C_ℓ^2 for any $\ell \geqslant 3$ not divisible by 3 or $\ell \geqslant 3(2\delta - n)$. For the *first part* of (*i*), let $G_p(n,\delta)$ be the n-vertex graph obtained from the disjoint union of an independent set Y on $n-\delta$ vertices and $r := r_p(n,\delta)$ cliques $X_1, \dots, X_r$ with $|X_1| \leqslant \dots \leqslant |X_r| \leqslant |X_1| + 1$ on a total of δ vertices, by inserting all edges between Y and X_i for each $i \in [r]$. It is easy to check that $\deg^{\min}(G_p(n,\delta)) = \delta$. Moreover any squared path $P_m^2 \subseteq G_p(n,\delta)$ contains vertices from at most one clique X_i. As Y is independent and P_m^2 has independence number $\lceil m/3 \rceil$ we have $\lfloor 2m/3 \rfloor \leqslant \lceil \delta / r_p(n,\delta) \rceil$ and thus $m \leqslant \lfloor \frac{1}{2}(3\lceil \delta/r_p(n,\delta)\rceil + 1) \rfloor = \mathrm{sp}(n,\delta)$. For the *second part* of (*i*), we construct the graph $G'_c(n,\delta)$ in the same way as $G_p(n,\delta)$ but with $r := r_c(n,\delta)$ and with $|X_i| = \lceil \delta/r \rceil$ for *all* $i \in [r]$. To obtain an n-vertex graph $G_c(n,\delta)$ from $G'_c(n,\delta)$ choose v_i in X_i arbitrarily for each $i \in [r]$ and identify all v_i with $i \leqslant r\lceil \delta/r \rceil - \delta$. Again $G_c(n,\delta)$ has minimum degree δ, any squared cycle C_m^2 in $G_c(n,\delta)$ touches only one of the X_i, and hence $m \leqslant \mathrm{sc}(n,\delta)$.

Before closing this introduction let us remark that similar phenomena to those described in Theorem 3.4 are observed with simple paths and cycles. Every graph with minimum degree δ contains a path of length $\lceil n/\lfloor n/(\delta+1)\rfloor \rceil$, and the

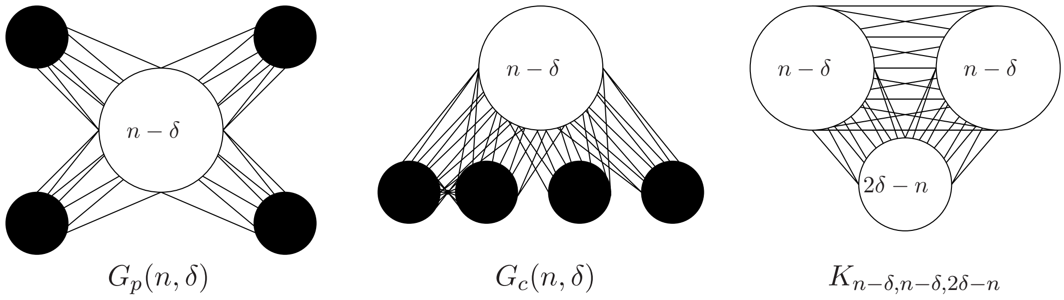


Figure 3.2: The extremal graphs, for the case $r_p(n, \delta) = r_c(n, \delta) = 4$.

extremal graph is a vertex disjoint union of cliques. This follows from an easy adjustment of the proof of Dirac's theorem. Improving on results of Nikiforov and Schelp [81], Allen proved the following theorem in [2]. The methods used for obtaining this result are quite different from those applied to prove Theorem 3.4. In particular they do not rely on the Regularity Lemma.

Theorem 3.5 (Allen [2]). *Given an integer $k \geqslant 2$ there is n_0 such that whenever $n \geqslant n_0$ and G is an n-vertex graph with minimum degree $\delta \geqslant n/k$, the following are true.*

(*i*) *G contains C_t for every even $4 \leqslant t \leqslant \lceil n/(k-1) \rceil$,*

(*ii*) *if G does not contain a cycle of every length from $\lfloor 2n/\delta \rfloor - 1$ to $\lceil n/(k-1) \rceil$ inclusive then G does contain C_t for every even $4 \leqslant t \leqslant 2\delta$.*

3.2 Main lemmas and the proof of Theorem 3.4

Our proof of Theorem 3.4 combines the Stability Method pioneered by Simonovits [92], the Regularity Method which pivots around the joint application of Szemerédi's Regularity Lemma (Lemma 1.4), and Blow-up Lemma (Lemma 1.8). The combination of these three methods has proved useful for a variety of exact embedding results and was applied for example in [60]. However, this well-established technique provides only a rather loose framework for proofs of this kind. For our application we will embellish this framework with a new concept, which we call the connected triangle components of a graph.

In this section we explain how we use connected triangle components in conjuction with the Regularity Method and the Blow-up technique, and the Stability Method. We first provide the necessary definitions, formulate our main lemmas (whose proofs are provided in the remaining sections of this chapter), and sketch how they work together in the proof of Theorem 3.4. The details of this proof are then presented at the end of this section.

3.2.1 Connected triangle components and triangle factors

Connected triangle components and connected triangle factors are the main protagonists in the proof of Theorem 3.4. Roughly speaking, in a connected triangle component we can start in an arbitrary triangle and reach each other triangle by "walking" through a sequence of triangles, and a connected triangle factor is a collection of vertex disjoint triangles each pair of which is connected in this way.

To make this precise, let $G = (V, E)$ be a graph. A *triangle walk* in G is a sequence of edges $e_1, \ldots, e_p$ in G such that e_i and e_{i+1} share a triangle of G for all $i \in [p-1]$. We say that e_1 and e_p are *triangle connected* in G. A *triangle component* of G is a maximal set of edges $C \subseteq E$ such that every pair of edges in C is triangle connected. Observe that this induces an equivalence relation on the edges of G, but a vertex may be part of many triangle components. In addition a triangle component does not need to form an induced subgraph of G in general. The *vertices of a triangle component* C_i are all vertices v such that some edge uv of G is contained in C_i. We define the *size* $|C|$ *of a triangle component* C to be the number of vertices of C.

A *triangle factor* T in a graph G is a collection of vertex disjoint triangles in G. T is a *connected triangle factor* if all edges of T are in the same triangle component of G. We define the *size* of T to be the number of vertices covered by T. We let $\mathrm{CTF}(G)$ denote the maximum size of a connected triangle factor in G. It is not difficult to check for example that any connected triangle factor in $G_p(n,\delta)$ contains only vertices of at most one of the cliques X_i (cf. the definition of $G_p(n,\delta)$ below Theorem 3.4) and of the independent set Y. Hence

$$\mathrm{CTF}\big(G_p(n,\delta)\big) = 3\left\lfloor \frac{\mathrm{sp}(n,\delta)}{3} \right\rfloor. \tag{3.2}$$

Suppose that a graph G contains a square-path of length ℓ. Then obviously, $\mathrm{CTF}(G) \geqslant 3\lfloor \ell/3 \rfloor$. Thus, (3.2) together with Theorem 3.4(i) says that $G_p(n,\delta)$ minimises CTF among all graphs of order n and minimum degree δ.

We will usually find that the number of vertices in a triangle component and the size of a maximum connected triangle factor in that component are quite different. As we will explain next, for the purposes of embedding squared paths and squared cycles, it is the size of a connected triangle factor that is important.

3.2.2 A blow-up type statement

As indicated in Section 1.4.2, regular pairs of positive density behave almost like complete bipartite graphs with respect to embedding a bounded degree graph. In order to prove Theorem 3.4 we shall use that in each triangle t of a connected triangle factor T of the reduced graph R (of the host graph G) we find a squared path in G that almost fills the clusters of G corresponding to t. By using the fact that T is triangle connected it is then possible to connect these squared paths into squared paths or cycles of the desired overall length. In addition, the Blow-up Lemma allows for some control about the start- and end-vertices of the path that is constructed in this way. This is stated in Lemma 3.6. A similar statement is implicit, e.g., in [60].

Lemma 3.6 (Embedding Lemma). *For all $d > 0$ there exists $\varepsilon_{\mathrm{EL}} > 0$ with the following property. Given $0 < \varepsilon < \varepsilon_{\mathrm{EL}}$, for every $m_{\mathrm{EL}} \in \mathbb{N}$ there exists $n_{\mathrm{EL}} \in \mathbb{N}$ such that the following hold for any graph G on $n \geqslant n_{\mathrm{EL}}$ vertices with (ε, d)-reduced graph R' on $m \leqslant m_{\mathrm{EL}}$ vertices.*

(i) $C^2_{3\ell} \subseteq G$ for every $\ell \in \mathbb{N}$ with $3\ell \leqslant (1-d)\,\mathrm{CTF}(R')\frac{n}{m}$.

(ii) If $K_4 \subseteq C$ for each triangle component C of R', then $C^2_\ell \subseteq G$ for every $\ell \in \mathbb{N} - \{5\}$ with $3 \leqslant \ell \leqslant (1-d)\,\mathrm{CTF}(R')\frac{n}{m}$.

Furthermore, let T be a connected triangle factor in a triangle component C of R with $K_4 \subseteq C$, let $u_1v_1, u_2v_2 \in E(G)$ be disjoint edges, and suppose that there are (not necessarily disjoint) edges $X_1Y_1, X_2Y_2 \in C$ such that the edge u_iv_i has at least $2d\frac{n}{m}$ common neighbours in each cluster X_i and Y_i for $i = 1, 2$. Then

(iii) $P^2_\ell \subseteq G$ for every $\ell \in \mathbb{N}$ with $6(m+2)^3 < \ell < (1-d)|T|\frac{n}{m}$, such that P^2_ℓ starts in u_1, v_1 and ends in u_2, v_2 (in those orders) and at most $(\varepsilon + d)n$ vertices of P^2_ℓ are not in $\bigcup T$.

The copies of K_4 that are required in this lemma play a crucial rôle when embedding squared cycles which are not 3-chromatic.

For the proof of Lemma 3.6 we apply the following version (which is a special case) of the Blow-up Lemma 1.8.

Lemma 3.7. *Given fixed $c, d > 0$, there exist $\varepsilon_0 > 0$ and n_{BL} such that for any $0 < \varepsilon < \varepsilon_0$ the following holds. Let H be any graph on at least n_{BL} vertices with $V(H) = V_1 \dot\cup V_2 \dot\cup V_3$ and $|V_i| \geqslant \frac{1}{6}|V(H)|$, in which each bipartite graph $H[V_i, V_j]$ is $(3\varepsilon, d)$-regular and furthermore $\deg^{\min}{}_{V_i}(V_j) \geqslant \frac{1}{2}d|V_i|$ for each $1 \leqslant i, j \leqslant 3$.*

Let F be any subgraph of the complete tripartite graph with parts V_1, V_2 and V_3 such that the maximum degree of F is at most four. Assume further, that at most *four vertices x_i ($i \in [4]$) of F are endowed with sets $C_{x_i} \subseteq V_j$ such that $x_i \in V_j$ and $|C_{x_i}| \geqslant c|V_j|$*

Then there is an embedding $\psi\colon V(F) \to V(H)$ of F into H with $\psi(x_i) \in C_{x_i}$ for $i \in [4]$.

We also say that the x_i in Lemma 3.7 are *image restricted* to C_{x_i}.

Proof of Lemma 3.6. Given d, we let $c := d^2/4$. Now Lemma 3.7 gives values $\varepsilon_0 > 0$ and n_{BL}. We choose $\varepsilon_{\mathrm{EL}} := \min(\varepsilon_0, d^2/24)$. Given $\varepsilon < \varepsilon_{\mathrm{EL}}$ and m_{EL} we choose

$$n_{\mathrm{EL}} := \max\left(2m_{\mathrm{EL}}n_{\mathrm{BL}}, \frac{6m^4}{\varepsilon}\right).$$

Let $n \geqslant n_{\mathrm{EL}}$, let G be an n-vertex graph, and let R' be an (ε, d)-reduced graph of G on $m \leqslant m_{\mathrm{EL}}$ vertices.

Fix a set $\mathcal{T}' = \{T'_1, \ldots, T'_{\mathrm{CTF}(R')/3}\}$ of vertex-disjoint triangles in a triangle component of R' covering $\mathrm{CTF}(R')$ vertices. For each triangle $T'_i = X'_{i,1}X'_{i,2}X'_{i,3}$ we may by regularity for each $j \in [3]$ remove at most $\varepsilon|X'_{i,j}|$ vertices from $X'_{i,j}$ to obtain a set $X_{i,j}$ such that each pair $(X_{i,j}, X_{i,k})$ is not only $(2\varepsilon, d)$-regular but also satisfies ${\deg^{\min}}_{X_{i,k}}(X_{i,j}) \geqslant (d - 3\varepsilon)|X_{i,k}|$. We let R be the $(2\varepsilon, d)$-reduced graph corresponding to the new vertex partition given by replacing each $X'_{i,j}$ with $X_{i,j}$; then every edge of R' carries over to R, and we let $\mathcal{T}$ be the set of $\mathrm{CTF}(R')/3$ vertex disjoint triangles in R corresponding to $\mathcal{T}'$. We set $r := \mathrm{CTF}(R')/3$.

Our strategy now is as follows. We shall first fix a collection of suitable triangle walks $W_1, \ldots, W_{r-1}$ and W' in R. Next, for each of these triangle walks $W = (E_1, E_2, \ldots)$ we do the following. Let $\overrightarrow{U_1V_1}$ be (a suitable) orientation of the first edge E_1 of W. We shall construct a sequence $Q(W, \overrightarrow{U_1V_1})$ of vertices of R whose first two vertices are U_1 and V_1, in that order, and which has the property that every vertex in the sequence is adjacent to the two preceding vertices (as is the case for a squared path). Then we use this sequence $Q(W, \overrightarrow{U_1V_1})$ to obtain a squared path in G following W, whose first two vertices are in U_1 and V_1. Finally, connecting suitable paths appropriately will lead to a proof of (*i*), (*ii*), and (*iii*).

We first construct the triangle walks $W_1, \ldots, W_{r-1}$ and W'. For each $1 \leqslant i \leqslant r-1$ let W_i be a fixed triangle walk in R whose first edge is in T_i and whose last is in T_{i+1}. We suppose (repeating edges in the triangle walk W_i if necessary) that each triangle walk W_i contains at least ten edges, that the first edge of W_{i+1} is not the same as the last edge of W_i, and such that each walk with more than ten edges is of minimal length. We have $|W_i| \leqslant \binom{m}{2}$ for each i. Let W' be the triangle walk obtained by concatenating $W_1, \ldots, W_{r-1}$.

Next, we describe how to construct the sequence $Q(W,\overrightarrow{A_1B_1})$ for any triangle walk $W=(E_1,E_2,\ldots)$ in R and any orientation $\overrightarrow{A_1B_1}$ of its first edge E_1. We construct $Q(W,\overrightarrow{A_1B_1})$ iteratively as follows. Let $Q_1:=(A_1,B_1)$. Now for each $2\leqslant i\leqslant |W|$ successively, we define Q_i as follows. The last two vertices A_{i-1},B_{i-1} of Q_{i-1} are an orientation of E_{i-1}. If $E_i=A_{i-1}B_i$ we create Q_i by appending (B_i,A_{i-1}) to Q_{i-1}; if $E_i=B_{i-1}B_i$ we append $(B_i,A_{i-1},B_{i-1},B_i)$ to Q_{i-1} to create Q_i. At each step the final two vertices of Q_i are an orientation of E_i. Furthermore every vertex of Q_i is adjacent in R to the two vertices preceding it in Q_i. Finally, we let $Q(W,\overrightarrow{A_1B_1}):=Q_{|W|}$.

We shall need the following observations concerning the lengths of sequences constructed in this way. It is easy to check by induction that for any triangle walk W with at least two edges whose first edge is U_1V_1, we have

$$|Q(W,\overrightarrow{U_1V_1})|+|Q(W,\overrightarrow{V_1U_1})|\equiv 1 \mod 3\,. \tag{3.3}$$

Now consider the concatenation W' of the walks W_i. Let U_1V_1 be the first edge of W_1. If we construct $Q(W',\overrightarrow{U_1V_1})$ then the first edge U_iV_i and the last edge $U_i'V_i'$ of each W_i obtains an orientation, say $\overrightarrow{U_iV_i}$ and $\overrightarrow{U_i'V_i'}$. Clearly, there are sequences $\tilde{Q}_i$ of vertices in T_i for $1<i<r$, such that $Q(W',\overrightarrow{V_1U_1})$ is the concatenation of

$$Q(W_1,\overrightarrow{V_1U_1}),\tilde{Q}_2,Q(W_2,\overrightarrow{V_2U_2}),\ldots,\tilde{Q}_{r-1},Q(W_{r-1},\overrightarrow{V_{r-1}U_{r-1}})\,.$$

Further we let $\tilde{Q}_1=T_1-U_1V_1$ and $\tilde{Q}_r=T_r-U'_{r-1}V'_{r-1}$. We define $f_i=|\tilde{Q}_i| \mod 3$ for $i\in[r]$. Together with (3.3) we obtain

$$|Q(W',\overrightarrow{U_1V_1})|+|Q(W_1,\overrightarrow{V_1U_1})|+\sum_{1<i<r}\big(|Q(W_i,\overrightarrow{V_iU_i})|+f_i\big)\equiv 1 \mod 3$$

and hence

$$|Q(W',\overrightarrow{U_1V_1})|+\sum_{i\in[r-1]}\big(|Q(W_i,\overrightarrow{V_iU_i})|+f_i\big)+f_r\equiv 0 \mod 3\,. \tag{3.4}$$

This will enable us to construct cycles of lengths divisible by three later.

In order to construct squared paths in G from short vertex sequences in R we use the following fact.

Claim 3.6.1. *Let X_1,X_2,X_3 be vertices of R (not necessarily distinct), and Z be any set of at most $2\varepsilon|X_1|$ vertices of G. Suppose that X_1X_2 and X_1X_3 are edges of R. Suppose furthermore that we have two vertices u and v of G such that u and v*

have at least $(d-2\varepsilon)^2|X_1|$ *common neighbours in* X_1*, and* v *has at least* $(d-2\varepsilon)|X_2|$ *neighbours in* X_2*.*

Then there is a vertex $w \in X_1 - Z$ *adjacent to* u *and* v *such that* v *and* w *have at least* $(d-2\varepsilon)^2|X_2|$ *common neighbours in* X_2 *and* w *has at least* $(d-2\varepsilon)|X_3|$ *neighbours in* X_3*.*

Proof of Claim 3.6.1. Let W be the set of common neighbours of u and v in X_1. Since $X_1X_2 \in E(R)$, at most $2\varepsilon|X_1|$ vertices of W have fewer than $(d-2\varepsilon)|\mathrm{N}(v) \cap X_2| \geqslant (d-2\varepsilon)^2|X_2|$ common neighbours with v in X_2. Since $X_1X_3 \in E(R)$ at most $2\varepsilon|X_1|$ vertices of W have fewer than $(d-2\varepsilon)$ neighbours in X_3. Finally since $6\varepsilon|X_1| < (d-2\varepsilon)^2|X_1|$ we can find a vertex of $W - Z$ satisfying the desired properties. □

With these building bricks at hand we can now turn to the proofs of (i), (ii), and (iii).

Proof of (i), i.e., G contains $C_{3\ell}^2$ for each $3\ell \leqslant (1-d)\,\mathrm{CTF}(R)n/m$: When $\ell \leqslant (1-d)n/m$ we have $C_{3\ell}^2 \subseteq K_{(1-d)n/m,(1-d)n/m,(1-d)n/m}$ and thus by Lemma 3.7 we can find $C_{3\ell}^2$ as a subgraph of G (whose vertices are in T_1, with no restrictions required). Otherwise we use the following strategy. Let UV be the first edge of the triangle walk W_1.

Our first goal will be to construct a squared path P' in G which 'connects' T_1 to T_2, T_2 to T_3, and so on. For this purpose we choose two adjacent vertices u and v of G in U and V respectively, such that u and v have $(d-2\varepsilon)^2 n/m$ common neighbours in both the third vertex of T_1 and the third vertex of $Q(W',\overrightarrow{UV})$, such that v has $(d-2\varepsilon)n/m$ neighbours in the fourth vertex of $Q(W',\overrightarrow{UV})$, and such that u has $(d-2\varepsilon)n/m$ neighbours in V. This is possible by the regularity of the various pairs. (Observe that the required sizes for the neighbourhoods and joint neighbourhoods are chosen large enough for an application of Lemma 3.7 in the triangle T_1.) Now we apply Claim 3.6.1 with the vertices u and v and the third, fourth and fifth vertices of $Q(W',\overrightarrow{UV})$ to obtain a third vertex v' in the third vertex of $Q(W',\overrightarrow{UV})$ such that u and v are adjacent to v'. By repeatedly applying Claim 3.6.1 we construct a sequence of vertices P' (starting with u, v, v'), where the ith vertex of P' is in the ith set of $Q(W',\overrightarrow{UV})$ and is adjacent to its two predecessors, and where the vertices are all distinct (noting that $3|W'| < \varepsilon n/m$). Thus P' is a squared path running from T_1 to T_{r-1} following all the triangle walks W_i.

In addition we construct similarly (and without re-using vertices) for each $1 \leqslant i \leqslant r-1$ a squared path P_i following the triangle walk W_i. However, this time we use the opposite orientation for the first edge: that is, instead of constructing P_1

from $Q(W_1, \overrightarrow{UV})$ we use $Q(W_1, \overrightarrow{VU})$, and similarly for each P_i we use the opposite orientation of the first edge of W_i to that used in P'. Again, for each P_i we insist that the first two vertices have suitable neighbourhoods in T_i, and the last two in T_{i+1}, for an application of Lemma 3.7 in these triangles. Again, this is possible by regularity.

We note that the total number of vertices on all of these squared paths is not more than $6m\binom{m}{2} < \varepsilon n/m$. Finally, we remove from T_1 all vertices of $P := P' \cup P_1 \cup \cdots \cup P_{r-1}$. Since at most $\varepsilon n/m$ vertices are removed, and each cluster of T_1 has size at least $(1-3\varepsilon)n/m$, even after removal all three pairs remain $(3\varepsilon, d)$-regular and each cluster still has size at least $(1-4\varepsilon)n/m$.

Thus the conditions of Lemma 3.7 are satisfied, and hence we may embed a squared path S_1 into T_1, with the four restrictions that its first vertex is a common neighbour of the first two vertices of P', its second a neighbour of the first vertex of P', its penultimate vertex a neighbour of the first vertex of P_1 and its final vertex a common neighbour of the first two vertices of P_1 (noting that by choice of the first two vertices of P' and of P_1 the sets to which these vertices are restricted are indeed of size cn/m because $c = d^2/4$). In this way we can construct a squared path on $3\ell_1 + f_1$ vertices for any integer $\ell_1 \in [10, (1-d)n/m]$ (since $3 \cdot 4\varepsilon < d$), where $f_1 \in \{0, 1, 2\}$ is as defined above (3.4). Similarly we may apply Lemma 3.7 to each T_i ($2 \leqslant i \leqslant r$), after removing P from T_i, to obtain squared paths S_i of length $3\ell_i + f_i$ for any integer $\ell_i \in [10, (1-d)n/m]$.

Finally $S := P' \cup S_1 \cup P_1 \cup \ldots \cup P_{r-1} \cup S_r$ forms a squared cycle in G. It follows from (3.4) that the length of S is divisible by three. We conclude that indeed $S = C_{3k}^2$, where we may choose any integer k with $3k \in [6m^3, (1-d)\,\mathrm{CTF}(R)n/m]$, as required.

Proof of (ii): When every triangle component of R contains K_4 we also want to obtain squared cycles whose lengths are not divisible by three. Observe that if $ABCD$ is a copy of K_4 in R, then the vertex sequences ABC, $ABCDABC$ and $ABCDABCDABC$ each start and end with the same pair. Hence, with the help of Claim 3.6.1, these sequences can be used to construct squared paths in G of length 3 (which is $0 \mod 3$), length 7 ($1 \mod 3$), and length 11 ($2 \mod 3$).

We construct C_ℓ^2 for $\ell \in [3, 20] - \{5\}$ within a copy of K_4 in R directly (by the above methods). To obtain C_ℓ^2 with $21 \leqslant \ell \leqslant 3(1-d)n/m$ we remove at most $2\varepsilon n/m$ vertices from each of A, B and C to obtain a triangle satisfying the conditions of Lemma 3.7, construct a short path in A, B, C, D following the appropriate vertex sequence for $\ell \mod 3$ and apply Lemma 3.7 to obtain C_ℓ^2. Finally, to obtain longer squared cycles we perform the same construction as above, with the exception that

W' is any triangle walk to and from a copy of K_4, and so $Q(W', \overrightarrow{UV})$ may be taken (using one of the three vertex sequences above) to have any desired number of vertices modulo three (and not more than $2m^2$ in total).

Proof of (iii): Lastly, when we are required to construct a squared path between two specified edges u_1v_1 (with $2dn/m$ common neighbours in both X_1 and Y_1) and u_2v_2 (with $2dn/m$ common neighbours in both X_2 and Y_2) using triangles T in R, we apply the identical strategy, noting that the conditions on u_1v_1 and u_2v_2 are already suitable for an application of Claim 3.6.1. □

3.2.3 The Stability Method

The strategy we just described leaves us with the task of finding a big connected triangle factor T in the reduced graph R of G. However, there is one problem with this approach: The proportion τ of R covered by T is roughly equal to the proportion of G covered by the squared path P that we obtain from the Embedding Lemma (Lemma 3.6). However, as explained above, the relative minimum degree $\gamma_R := \deg^{\min}(R)/|V(R)|$ of R is in general slightly smaller than $\gamma_G := \deg^{\min}(G)/|V(G)|$, but the extremal graphs for squared paths and connected triangle factors are the same. It follows that we cannot expect that τ is larger than the proportion a maximum squared path covers in a graph with relative minimum degree γ_R, and hence smaller than the proportion we would like to cover for relative minimum degree γ_G.

Consequently we need to be more ambitious and shoot for a bigger connected triangle factor in R than we can expect for this minimum degree (cf. Lemma 3.8 (S1) and (S2)). This will of course not always be possible, but it will only fail if R (and hence G) is 'very close' to the extremal graph $G_p(|V(R)|, \deg^{\min}(R))$ (and hence also to $G_c(|V(R)|, \deg^{\min}(R)))$ in which case we will say that R is *near-extremal* (cf. Lemma 3.8 (S3)).

This approach is called the Stability Method and the following lemma states that it is feasible for our purposes. This lemma additionally guarantees copies of K_4 as required by the Embedding Lemma. We formulate this lemma for graphs G, but use it on the reduced graph R later. Its proof does not rely on the Regularity Lemma and is given in Section 3.3.

Lemma 3.8 (Stability Lemma)**.** *Given $\mu > 0$, for any sufficiently small $\eta > 0$ there exists n_0 such that if G has $n > n_0$ vertices and* $\deg^{\min}(G) = \delta \in ((\frac{1}{2} + \mu)n, \frac{2n-1}{3})$, *then either*

(S1) $\mathrm{CTF}(G) \geqslant 3(2\delta - n)$, *or*

(S2) $\mathrm{CTF}(G) \geqslant \min(\mathrm{sp}(n, \delta+\eta n), \frac{11n}{20})$, *or*

(S3) *G has an independent set of size at least $n-\delta-11\eta n$ whose removal disconnects G into components, each of size at most $\frac{19}{10}(2\delta-n)$.*

Moreover, in cases (S2) *and* (S3) *each triangle component of G contains a K_4.*

By the discussion above, it remains to handle the graphs with near-extremal reduced graph. For these graphs we have a lot of structural information which enables us to show directly that they contain the squared paths and squared cycles we desire, as the following lemma documents. The proof of this lemma is provided in Section 3.4. In this proof we shall again make use of the embedding lemma, Lemma 3.6. Accordingly Lemma 3.9 inherits the upper bound m_{EL} on the number of clusters from Lemma 3.6.

Lemma 3.9 (Extremal Lemma)**.** *For every $\nu > 0$, given $0 < \eta, d < 10^{-8}\nu^4$ there exists $\varepsilon_0 > 0$ such that for every $0 < \varepsilon < \varepsilon_0$ and every m_{EL}, there exists N such that the following holds. Suppose that*

(*i*) *G is an n-vertex graph with $n > N$ and $\deg^{\min}(G) = \delta > \frac{n}{2} + \nu n$,*

(*ii*) *R is an (ε, d)-reduced graph of G of order $m \leqslant m_{\mathrm{EL}}$,*

(*iii*) *each triangle component of R contains a copy of K_4.*

(*iv*) *$V(R) = I\dot\cup B_1\dot\cup B_2\dot\cup\cdots\dot\cup B_k$ with $k \geqslant 2$,*

(*v*) *I is an independent set with $|I| \geqslant (n-\delta-11\eta n)m/n$,*

(*vi*) *for each $i \in [k]$ we have $0 < |B_i| \leqslant 19m(2\delta-n)/(10n)$, and for every $j \in [k]-\{i\}$ there are no edges between B_i and B_j in R.*

Then G contains $P^2_{\mathrm{sp}(n,\delta)}$ and C^2_ℓ for each $\ell \in [3, \mathrm{sc}(n,\delta)] - \{5\}$.

It is interesting to notice that, although the two functions $\mathrm{sp}(n,\delta)$ and $\mathrm{sc}(n,\delta)$ are different—their jumps as δ increases occur at slightly different values—they are similar enough that the Stability Lemma covers them both. We will only need to distinguish between squared paths and squared cycles when we examine the near-extremal graphs.

3.2.4 Proof of Theorem 3.4

With this we have all the ingredients for the proof of our main theorem, which uses the Regularity Lemma (in form of Lemma 1.4) to construct a regular partition with reduced graph R of the host graph G, the Stability Lemma (Lemma 3.8) to conclude that R either contains a big connected triangle factor or is near-extremal, the Embedding Lemma (Lemma 3.6) to find long squared paths and cycles in G in the first case, and the Extremal Lemma (Lemma 3.9) in the second case.

Proof of Theorem 3.4. We require our constants to satisfy

$$\nu \gg \mu \gg \eta \gg d \gg \varepsilon > 0\,,$$

which we choose, given ν, as follows. First, we choose $\mu := \nu/2$. We then choose $\eta > 0$ to be small enough for both Lemma 3.8 and Lemma 3.9. Now we set $d > 0$ to be small enough for Lemma 3.9 and such that $d \leqslant \nu/10$ and $d \leqslant \eta/10$. For this d Lemma 3.6 then produces a constant ε_{EL}. We choose $\varepsilon > 0$ to be smaller than $\min\{\varepsilon_{\text{EL}}, \nu/10\}$ and sufficiently small for Lemma 3.9. We choose m_0 to be sufficiently large to apply Lemma 3.8 to any graph with at least m_0 vertices. We then choose m_{EL} such that Lemma 1.4 guarantees the existence of an (ε, d)-regular partition with at least m_0 and at most m_{EL} parts. Finally we choose $n_0 > n_{\text{EL}}$ to be sufficiently large for both Lemma 3.6 and Lemma 3.9.

Let $n > n_0$ and $\delta \in (n/2 + \nu n, n-1]$. Let G be any n-vertex graph with $\deg^{\min}(G) \geqslant \delta$. Observe that it suffices to show that $P^2_{\text{sp}(n,\delta)} \subseteq G$ and that (ii) of Theorem 3.4 holds. We first apply Lemma 1.4 to G to obtain an (ε, d)-reduced graph R on $m_0 \leqslant m \leqslant m_{\text{EL}}$ vertices. Let $\delta' := \deg^{\min}(R) \geqslant (\delta/n - d - \varepsilon)m > m/2 + \mu m$. Then we apply Lemma 3.8 to R. There are three possibilities.

First, we could find that $\text{CTF}(R) \geqslant 3(2\delta' - m)$. In this case by Lemma 3.6 we are guaranteed that for every integer ℓ' with $3\ell' < (1-d)\,\text{CTF}(R)n/m$ we have $C^2_{3\ell'} \subseteq G$. By choice of d and ε we have $(1-d)\cdot 3(2\delta' - m)n/m > 6\delta - 3n - \nu n$. Noting that $P^2_\ell \subseteq C^2_\ell$ we conclude that $P^2_{\text{sp}(n,\delta)} \subseteq G$ and $C^2_\ell \subseteq G$ for each integer $\ell \leqslant 6\delta - 3n - \nu n$ such that 3 divides ℓ, i.e., the second case of Theorem 3.4(ii) holds.

Second, we could find that $\text{CTF}(R) \geqslant \min(\text{sp}(m, \delta' + \eta m), \frac{11m}{20})$ and that every triangle component of R contains a copy of K_4. By Lemma 3.6 we are guaranteed that for every $\ell \in [6, (1-d)\,\text{CTF}(R)n/m] - \{5\}$ we have $C^2_\ell \subseteq G$. Recall that we have $\text{sp}(n,\delta) \leqslant 0.51n$, and in particular $(1-d)\frac{11m}{20} \times n/m > \text{sp}(n,\delta)$. Therefore, and by choice of η and d we have $(1-d)\,\text{CTF}(R)n/m > \text{sp}(n,\delta) \geqslant \text{sc}(n,\delta)$, so we have $P^2_{\text{sp}(n,\delta)} \subseteq G$ and for each integer $\ell \in [3, \text{sc}(n,\delta)] - \{5\}$ we have $C^2_\ell \subseteq G$, i.e., the first case of Theorem 3.4(ii) holds.

Third, we could find that R is near-extremal. Then R contains an independent set on at least $m - \delta' - 11\eta m$ vertices whose removal disconnects R into components of size at most $\frac{19}{10}(2\delta' - m)$, and each triangle component of R contains a copy of K_4. But now G satisfies the conditions of Lemma 3.9. It follows that G contains $P^2_{\mathrm{sp}(n,\delta)}$ and for each $\ell \in [3, \mathrm{sc}(n,\delta)] - \{5\}$ the graph G contains C^2_ℓ, i.e., the first case of Theorem 3.4(*ii*) holds. $\square$

3.3 Triangle components and the proof of Lemma 3.8

In this section we provide a proof of our stability result for connected triangle factors, Lemma 3.8. Distinguishing different cases, we analyse the sizes and the structure of the triangle components in the graph G under study. Before we give more details about our strategy and a sketch of the proof, we introduce some additional definitions and provide a preparatory lemma (Lemma 3.10).

Let G be a graph with triangle components $C_1, \ldots, C_r$. The *interior* $\mathrm{int}(G)$ of G is the set of vertices of G which are in more than one of the triangle components. For a component C_i, the interior of C_i, written $\mathrm{int}(C_i)$, is the set of vertices of C_i which are in $\mathrm{int}(G)$. The remaining vertices of C_i are called the *exterior* $\partial(C_i)$. That is, $\partial(C_i)$ is formed by the set of vertices of C_i which are in no other triangle component of G. To give an example, by definition the graph $G_p(n,\delta)$ has $r_p(n,\delta)$ triangle components; its interior is the independent set Y (using the notation of the construction of $G_p(n,\delta)$ on page 35 in Section 3.1), with the component exteriors being the cliques $X_1, \ldots, X_r$.

The following lemma collects some observations about triangle components.

Lemma 3.10. *Let G be an n-vertex graph with* $\deg^{\min}(G) = \delta > n/2$. *Then*

(*a*) *each triangle component C of G satisfies $|C| > \delta$,*

(*b*) *for distinct triangle components C, C' we have $e(\partial(C), \partial(C')) = 0$,*

(*c*) *for each triangle component C, each vertex u of C, and $U := \{v \colon uv \in C\}$, the minimum degree in $G[U]$ is at least $2\delta - n$ and hence $|G[U]| \geqslant 2\delta - n + 1$.*

Proof. To see (*a*) let M be the vertices of a maximal clique in C (clearly $|M| \geqslant 3$). If u and v are in M, and x is a common neighbour of u and v, then x is also in C. Thus vertices of $G - C$ are adjacent to at most 1 vertex of M and vertices of C are adjacent to at most $|M| - 1$ vertices of M, by maximality of M. This gives the inequality

$$|M|\delta \leqslant \sum_{m \in M} \deg(m) \leqslant \sum_{x \in C} (|M| - 1) + \sum_{x \notin C} 1$$

and hence $|M|\delta - n \leqslant (|M|-2)|C|$. Since $n < 2\delta$ we have $|C| > \delta$ as required.

Since $\delta > n/2$, we have that $\mathrm{N}^\wedge(u,u') \neq \emptyset$ for any two vertices u and u'. Now, if $u \in \partial(C)$, $u' \in \partial(C')$, $x \in \mathrm{N}^\wedge(u,u')$, and uu' was an edge, then $uu'x$ would form a triangle. Then u and u' would be together in some triangle component C'', contradicting the fact that they are in the exterior. Therefore, the assertion (b) follows.

Moreover, for an edge uv of C we have $\mathrm{N}^\wedge(u,v) \subseteq C$ as C is a triangle component. Since $|\mathrm{N}^\wedge(u,v)| \geqslant 2\delta - n$ we get (c). $\square$

Now let us sketch the proof of Lemma 3.8. Lemma 3.10(a) states that triangle components cannot be too small. However, it is not solely the size of the triangle components we are interested in: we want to find a triangle component that contains many vertex disjoint triangles. At this point, Lemma 3.10(c) comes into play. It asserts that certain spots in a triangle component induce a graph with minimum degree $2\delta - n$. In the proof of Lemma 3.8 we shall usually (i.e., for many values of δ) use this fact in order to find a big matching M in such spots (Proposition 3.11(a) below asserts that this is possible). Clearly all edges in such a matching are triangle connected and hence it will remain to extend M to a set of vertex disjoint triangles. For this purpose we will analyse the size of the common neighbourhood $\mathrm{N}^\wedge(u,v)$ of an edge uv in M. We will usually find that $\mathrm{N}^\wedge(u,v)$ is so big that a simple greedy strategy allows us to construct the triangles. For estimating $\mathrm{N}^\wedge(u,v)$ we will often use the following technique: We find a large set X such that neither u nor v has neighbours in X. This implies $|\mathrm{N}^\wedge(u,v)| \geqslant 2\delta - (n - |X|)$. Observe that Lemma 3.10(b) implies that $\partial(C)$ can serve as X if both $u, v \in \partial(C')$ for some triangle components C and C'.

The strategy we just described works for most values of δ below $\frac{3}{5}n$ (we describe the exceptions below). For $\delta \geqslant \frac{3}{5}n$ however, the greedy type argument fails, the reason being that we usually bound the common neighbourhood of an edge used in the argument above by $4\delta - 2n$. But for $\delta \geqslant \frac{3}{5}n$ we might have $\mathrm{sp}(n,\delta) > 4\delta - 2n$ (see Figure 3.1). We solve this problem by using a different strategy in this range of δ. We will still start with a big connected matching M as before, but use a Hall-type argument to extend M to a triangle factor T. More precisely, we find M in the exterior of some triangle component and then consider for each edge uv of M all common neighbours of uv in $\mathrm{int}(G)$. The Hall-type argument then permits us to find distinct extensions for the edges of M. To make this argument work we use the fact that in this range of δ the set $\mathrm{int}(G)$ is an independent set.

We indicated earlier that there are some exceptional values of δ that require special treatment: namely δ close to $\frac{3}{5}n$ and $\frac{4}{7}n$. Observe that in both ranges the number of triangle components of $G_p(n,\delta)$ changes (from 2 to 3 for $\frac{3}{5}n$, and from 3 to 4 for $\frac{4}{7}n$) and thus the value $\mathrm{sp}(n,\delta)$ as a function in δ jumps (see Figure 3.1). Roughly speaking, the reason that these two ranges need to be treated separately is that again $\mathrm{sp}(n,\delta)$ is not substantially smaller than $4\delta-2n$ here, but we also do not know now that $\mathrm{int}(G)$ is an independent set. For dealing with these values of δ we will use a somewhat technical case analysis which we provide at the end of this section (as proof of Claim 3.8.5).

As explained above, we will apply the following simple observations about matchings in graphs of given minimum degree.

Proposition 3.11.

(*a*) *Let $G=(X,E)$ be a graph with minimum degree δ. Then G has a matching covering $2\min(\delta,\lfloor|X|/2\rfloor)$ vertices.*

(*b*) *Let $G=(A\dot\cup B,E)$ be a bipartite graph with parts A and B, such that every vertex in A has degree at least a and every vertex in B has degree at least b. Then G has a matching covering $2\min(a+b,|A|,|B|)$ vertices.*

Proof. We first prove (*a*). Let M be a maximum matching in G, and assume that M contains less than $\min(\delta,\lfloor|X|/2\rfloor)$ edges. In particular, there are two vertices $x,y\in X$ not covered by M. Clearly, all neighbours of x and y are covered by M.

We claim that there is an edge uv in M with $xu,yv\in E$. Indeed, suppose that this is not the case. Then $|e\cap \mathrm{N}(x)|+|e\cap \mathrm{N}(y)|\leqslant 2$ for each $e\in M$. We therefore have

$$\delta+\delta\leqslant|\mathrm{N}(x)|+|\mathrm{N}(y)|=\sum_{e\in M}(|e\cap \mathrm{N}(x)|+|e\cap \mathrm{N}(y)|)\leqslant 2|M|\,,$$

contradicting the fact that $\delta>|M|$.

Now, let $uv\in M$ be an edge as in the claim above. Since $xu,yu\in E$ we get that x,u,v,y is an M-augmenting path, a contradiction.

Next we prove (*b*). Let M be a maximum matching in G. We are done unless there are vertices $u\in A$ and $v\in B$ not contained in M. There cannot be an edge $xy\in M$ such that uy and xv are edges of G by maximality of M, since then u,y,x,v was an M-augmenting path. But u has at least a neighbours in $V(M)\cap B$, and v at least b neighbours in $V(M)\cap A$, so there must be at least $a+b$ edges in M. □

Before turning to the proof of Lemma 3.8 let us quickly collect some analytical data about $\mathrm{sp}(n,\delta)$ and $r_p(n,\delta) =: r$. It is not difficult to check that

$$\begin{aligned} \frac{(r+1)n-r}{2(r+1)-1} \leqslant \delta < \frac{rn-r+1}{2r-1} \quad \text{and} \\ \frac{n-\delta}{2\delta-n+1} \leqslant r < \frac{\delta+1}{2\delta-n+1}\,. \end{aligned} \tag{3.5}$$

For the proof of Lemma 3.8 it will be useful to note in addition that given $\mu > 0$, for every $0 < \eta < \eta_0 := \eta_0(\mu)$, there is $n_1 := n_1(\eta)$ such that the following holds for all $n \geqslant n_1$. For all $\delta, \delta' > \frac{n}{2} + \mu n$, where δ is such that $\mathrm{sp}(n, \delta+\eta n) \leqslant \frac{11}{20}n$, and where δ' is such that we have $r_p(n,\delta') \geqslant 3$ and either $r_p(n,\delta') \geqslant 5$ or $r_p(n,\delta') = r_p(n,\delta'+\eta n)$, we have

$$\mathrm{sp}(n,\delta+\eta n) \leqslant \frac{3}{2}\min\Big(\frac{\delta}{r_p(n,\delta+\eta n)-1} - 2,\ \frac{\delta+3\eta n}{r_p(n,\delta+\eta n)} - 2\Big), \tag{3.6}$$

$$\begin{aligned} \mathrm{sp}(n,\delta+\eta n) &\leqslant \tfrac{19}{20}\cdot 3(2\delta-n) - 2 \leqslant 6\delta - 3n - 100\eta n, \quad \text{and} \\ \mathrm{sp}(n,\delta'+\eta n) &\leqslant 4\delta' - 2n, \end{aligned} \tag{3.7}$$

which follows immediately from the definition of $\mathrm{sp}(n,\delta)$ in (3.1) (see also Figure 3.1).

Proof of Lemma 3.8. Given μ and any $0 < \eta < \min(\frac{1}{1000}, \eta_0(\mu), 2\mu^2/3)$, where $\eta_0(\mu)$ is as above (3.6), let $n_0 := \max(n_1(\eta), 2/\eta)$ with $n_1(\eta)$ as above (3.6). Let $n \geqslant n_0$. This in particular means that we may assume the inequalities (3.6) and (3.7) in what follows. Define $\gamma := \delta/n$, and $r := r_p(n,\delta)$ and $r' := r_p(n, \delta+\eta n)$.

If G has only one triangle component then Theorem 3.3 guarantees that $\mathrm{CTF}(G) \geqslant 6\delta - 3n$ and so we are in Case (S1). Thus we may assume in the following that G has at least two triangle components. Then Lemma 3.10(*a*) implies that $\mathrm{int}(C) \neq \emptyset$ for any triangle component C.

Suppose that C is a triangle component of G which does not contain a copy of K_4. Let u be a vertex of C, and $U := \{v\colon uv \in C\}$. By Lemma 3.10(*c*) we have $\deg^{\min}(G[U]) \geqslant 2\delta - n$. Because C contains no copy of K_4, U contains no triangle. By Turán's theorem we have $|U| \geqslant 2(2\delta - n)$, and so by Proposition 3.11(*a*) the set U contains a matching M with $2\delta - n$ edges. Finally we choose greedily for each $e \in M$ a distinct vertex $v \in V(G)$ such that ev is a triangle. Since U is triangle free all these vertices must lie outside U, and since $|\mathrm{N}^\wedge(e)| \geqslant 2\delta - n$ we cannot fail to find distinct vertices for each edge. This yields a set T of $2\delta - n$ vertex-disjoint triangles which are all in C. So $\mathrm{CTF}(G) \geqslant 6\delta - 3n$ and we are in case (S1). Henceforth we assume that every triangle component of G contains a copy of K_4.

We continue by considering the case $\frac{3n-2}{5} \leqslant \delta < \frac{2n-1}{3}$. The following observation readily implies the lemma in this range, as we will see in Claim 3.8.2.

Claim 3.8.1. *If* $\deg^{\min}(G) \geqslant (\frac{3}{5} - 2\eta)n$*,* G *has exactly* 2 *triangle components,* $\mathrm{int}(G)$ *is independent, and either* $|\,\mathrm{int}(G)| < n - \delta - 11\eta n$ *or the exterior* X *of the triangle component with most vertices satisfies* $|X| \geqslant \frac{19}{10}(2\delta - n)$*, then* $\mathrm{CTF}(G) \geqslant \min(\mathrm{sp}(n, \delta + \eta n), \frac{11}{20}n)$*.*

Proof of Claim 3.8.1. First, by Lemma 3.10(*b*) a vertex $x \in X$ cannot have neighbours in the exterior of the other triangle component, so $\mathrm{N}(x) \subseteq X \cup \mathrm{int}(G)$. Thus $\deg^{\min}(G[X]) \geqslant \delta - |\,\mathrm{int}(G)|$, which by Proposition 3.11(*a*) means that there is a matching M in $G[X]$ with

$$|M| = \min(\delta - |\,\mathrm{int}(G)|, \lfloor |X|/2 \rfloor) \tag{3.8}$$

edges.

We aim to pair off edges of M with vertices of $\mathrm{int}(G)$ to form a sufficiently large number of vertex-disjoint triangles. To see that a triangle factor resulting from this process will be connected, observe that all edges of M are in X, and since X is a triangle component exterior, the edges of M are triangle connected. To form triangles from edges of M and vertices of $\mathrm{int}(G)$, we introduce an auxiliary bipartite graph H with vertex set $M \dot{\cup} \mathrm{int}(G)$, where $uv \in M$ is adjacent in H to $w \in \mathrm{int}(G)$ iff uvw is a triangle of G. Every vertex of X has at least $\delta - |X|$ neighbours in $\mathrm{int}(G)$, and so every edge of M has at least $a := 2(\delta - |X|) - |\,\mathrm{int}(G)|$ common neighbours in $\mathrm{int}(G)$. At the same time, since $\mathrm{int}(G)$ is independent, every vertex of $\mathrm{int}(G)$ has at least $\delta - (n - |\,\mathrm{int}(G)| - |X|)$ neighbours in X, of which all but $|X| - 2|M|$ must be in M. So every vertex of $\mathrm{int}(G)$ must have at least

$$b := \delta - (n - |\,\mathrm{int}(G)| - |X|) - (|X| - 2|M|) - |M| = \delta - n + |\,\mathrm{int}(G)| + |M|$$

edges of M in its neighbourhood. By Proposition 3.11(*b*) there is a matching in H on at least $\min(a + b, |M|, |\,\mathrm{int}(G)|)$ edges, and hence a connected triangle factor in G with so many triangles. Observe that

$$\begin{aligned} a + b &= 2\delta - 2|X| - |\,\mathrm{int}(G)| + \delta - n + |\,\mathrm{int}(G)| + |M| \\ &= 3\delta - n - 2|X| + |M|\,. \end{aligned} \tag{3.9}$$

Since there are two triangle components in G, there is a vertex u in a triangle component exterior which is not X. Therefore u has no neighbour in X, so $|X| <$

$n-\delta$. Since $\delta \geqslant (\frac{3}{5}-2\eta)n$, by (3.9) we have

$$a+b > |M| - 10\eta n\,. \tag{3.10}$$

Furthermore,

$$\text{if} \quad |X| \leqslant (\tfrac{2}{5}-3\eta)n\,, \quad \text{then} \quad a+b \geqslant |M|\,. \tag{3.11}$$

By Lemma 3.10(a) we have $|\operatorname{int}(G)| \geqslant 2\delta - n \geqslant \frac{n}{5} - 4\eta n$. Since $\eta \leqslant \frac{1}{1000}$ we have

$$3|\operatorname{int}(G)| \geqslant \frac{3n}{5} - 12\eta n > \frac{11n}{20}\,.$$

Thus we have $\mathrm{CTF}(G) \geqslant \frac{11n}{20}$ if we find a matching in H covering $\operatorname{int}(G)$. It remains, then, to check that we have

$$3\min(a+b,|M|) \geqslant \min(\mathrm{sp}(n,\delta+\eta n), \frac{11}{20}n). \tag{3.12}$$

We distinguish two cases.

Case 1: $a+b<|M|$. By (3.11) this forces $|X| > (\frac{2}{5}-3\eta)n$. Since we have $|M| = \min(\delta - |\operatorname{int}(G)|, \lfloor |X|/2 \rfloor)$ by (3.8), there are two possibilities. If $|M| = \lfloor |X|/2 \rfloor$ then we have

$$a+b \overset{(3.10)}{\geqslant} \left\lfloor \frac{|X|}{2} \right\rfloor - 10\eta n > \frac{n}{5} - 12\eta n > \frac{11n}{60}\,,$$

which proves (3.12) in this subcase. If, on the other hand, $|M| = \delta - |\operatorname{int}(G)|$, then we use that $\operatorname{int}(G)$ is independent, which implies $\operatorname{int}(G) \leqslant n-\delta$ and thus

$$\begin{aligned} a+b &\overset{(3.10)}{\geqslant} |M| - 10\eta n = \delta - |\operatorname{int}(G)| - 10\eta n \geqslant 2\delta - n - 10\eta n \\ &\overset{(3.7)}{\geqslant} \tfrac{1}{3}\,\mathrm{sp}(n,\delta+\eta n)\,, \end{aligned}$$

which proves (3.12) in this subcase.

Case 2: $a+b \geqslant |M|$. In this case, H contains a matching of size $|M|$, so we have $\mathrm{CTF}(G) \geqslant 3|M| = 3\min(\delta - |\operatorname{int}(G)|, \lfloor |X|/2 \rfloor)$. Again there are two possibilities, depending on $|M|$. If $|M| = \delta - |\operatorname{int}(G)|$, we are done by (3.7) exactly as before. If, on the other hand, $|M| = \lfloor |X|/2 \rfloor$, then (3.12) holds (and hence we are done) unless

$$3\lfloor \tfrac{|X|}{2} \rfloor < \min\left(\mathrm{sp}(n,\delta+\eta n), \tfrac{11}{20}n\right). \tag{3.13}$$

We now assume (3.13) in order to derive a contradiction, and make a final subcase distinction.

First assume that $\mathrm{sp}(n, \delta+\eta n) < \frac{11}{20}n$. Then $r' \geqslant 2$ and hence (3.13) and (3.1) imply

$$|X| < \tfrac{1}{2}(\delta + \eta n) + 3 < \tfrac{51}{100}\delta < \tfrac{19}{10}(2\delta - n)\,,$$

because $\delta \geqslant (\frac{3}{5} - 2\eta)n$ and $\eta \leqslant \frac{1}{1000}$. Furthermore, since G has two triangle components whose exterior is of size at most X by assumption we have $|\operatorname{int}(G)| > n - 2|X| = n - \delta - \eta n - 6$, a contradiction to the the conditions of Claim 3.8.1.

Now assume that $\mathrm{sp}(n, \delta + \eta n) \geqslant \frac{11}{20}n$. Then we have $\delta > (\frac{2}{3} - 2\eta)n$. By Lemma 3.10($a$) we have $|X| \leqslant n - \delta < (\frac{1}{3} + 2\eta)n$ and so $|X| < \frac{19}{10}(2\delta - n)$. Further $|\operatorname{int}(G)| \geqslant n - 2|X| \geqslant 2\delta - n > \frac{n}{3} - 4\eta n > n - \delta - 11\eta n$, which again contradicts the conditions of Claim 3.8.1. $\square$

Claim 3.8.2. *Lemma 3.8 is true for* $\frac{3n-2}{5} \leqslant \delta < \frac{2n-1}{3}$.

Proof of Claim 3.8.2. Observe that in this range $r = 2$. Assume G has an edge uv in $\operatorname{int}(G)$, let x be a common neighbour of u and v and C be the triangle component containing ux and vx. Since $uv \in \operatorname{int}(G)$ there are edges uy and vz of G outside C. The sets $\mathrm{N}^\wedge(u, y)$, $\mathrm{N}^\wedge(v, z)$ and $\{u, v, x, y, z\}$ are pairwise disjoint, and x is not adjacent to $\mathrm{N}^\wedge(u, y) \cup \mathrm{N}^\wedge(v, z) \cup \{y, z\}$. So $\delta \leqslant \deg(x) \leqslant (n-1) - 2(2\delta - n) - 2$ which is only possible when $\delta \leqslant (3n-3)/5$, a contradiction. Thus $\operatorname{int}(G)$ is an independent set, which implies $|\operatorname{int}(G)| \leqslant n - \delta$. Hence, by Lemma 3.10($a$), G cannot have more than two triangle components. In particular, all vertices in $\operatorname{int}(G)$ lie in both triangle components of G. So if $|\operatorname{int}(G)| \geqslant n - \delta - 11\eta n$ then $\operatorname{int}(G)$ is the desired large independent set for Case (S3). If moreover all triangle component exteriors are of size $\frac{19}{10}(2\delta - n)$ at most we are in Case (S3). Otherwise (if $\operatorname{int}(G)$ is small or a triangle component exterior is large) Claim 3.8.1 gives $\mathrm{CTF}(G) \geqslant \min(\mathrm{sp}(n, \delta + \eta n), \frac{11}{20}n)$ which is Case (S2). $\square$

For the remainder of the proof, we suppose $\delta < \frac{3n-2}{5}$ and accordingly $r \geqslant 3$ and $r' \geqslant 2$. For dealing with this case we first establish two auxiliary facts. The first one captures the greedy technique for finding a large connected triangle factor that we sketched in the beginning of this section. We will use this technique throughout the rest of the proof.

Claim 3.8.3. *If there are two sets* $U_1, U_2 \subseteq V(G)$ *such that no vertex in* U_1 *has a neighbour in* U_2*, all edges in* $G[U_1]$ *are triangle connected and* $\deg^{\min}(G[U_1]) \geqslant \delta_1$ *then* $\mathrm{CTF}(G) \geqslant \min(3\lfloor |U_1|/2 \rfloor, 3\delta_1, 2\delta - n + |U_2|)$.

Proof of Claim 3.8.3. By Proposition 3.11(a) we can find a matching M' in U_1

covering

$$\min(2\lfloor |U_1|/2 \rfloor, 2\delta_1)$$

vertices. Let M be a subset of M' covering $\min(2\lfloor |U_1|/2\rfloor, 2\delta_1, (4\delta - 2n + 2|U_2|)/3)$ vertices. For each edge $e \in M$ in turn we pick greedily a common neighbour of e outside both M and the previously chosen common neighbours to obtain a set T of disjoint triangles. For any $x, y \in U_1$ we have $|\mathrm{N}^\wedge(x, y)| \geqslant 2\delta - (n - |U_2|)$. We claim that this implies that T can be constructed, covering all of M. Indeed, in each step of the greedy procedure we have strictly more than $2\delta-(n-|U_2|)-3|M| \geqslant 0$ common neighbours of $e \in M$ available. Hence T covers at least $\min(3\lfloor |U_1|/2\rfloor, 3\delta_1, 2\delta - n + |U_2|)$ vertices. Note further that T is a connected triangle factor because all edges in $G[U_1]$ are triangle connected. □

Below, our goal will be to show that $\mathrm{int}(G)$ is an independent set. The following fact prepares us for this step.

Claim 3.8.4. *Let uv be an edge in $\mathrm{int}(G)$. Unless $r' = 2$ at least one vertex, u or v, is contained in at most $r' - 1$ triangle components.*

Proof of Claim 3.8.4. Let C_1 be the triangle component containing $uv \in \mathrm{int}(G)$ along with the (non-empty) common neighbourhood $\mathrm{N}^\wedge(u, v)$ (and perhaps some other neighbours of u or v separately). Suppose that $C \neq C_1$, and u is a vertex of C. Then by Lemma 3.10(c), there are at least $2\delta - n + 1$ neighbours x of u such that the edge ux is in C. Now suppose that u lies in at least $r' - 1$ triangle components other than C_1. It follows that there is a set $U_u \subseteq \mathrm{N}(u)$ of vertices x such that ux is not in C_1, with $|U_u| \geqslant (r' - 1)(2\delta - n + 1)$, since no edge lies in two distinct triangle components. Suppose furthermore that v too lies in at least $r' - 1$ triangle components other than C_1. Then there exists an analogously defined set U_v. Since all vertices of $\mathrm{N}^\wedge(u, v)$ form triangles of C_1 with u and v, the three sets $\mathrm{N}^\wedge(u, v)$, U_u and U_v are pairwise disjoint, and thus $|U_u \cup U_v| \geqslant (2r' - 2)(2\delta - n + 1)$. Now given any $x \in \mathrm{N}^\wedge(u, v)$, since ux and vx are both in C_1, x cannot be adjacent to any vertex of $U_u \cup U_v$. But then $\delta \leqslant \deg(x) < n - (2r' - 2)(2\delta - n + 1)$ which is equivalent to $2r' - 2 < (n - \delta)/(2\delta - n + 1)$. By (3.5) the right-hand side is at most r and thus we get $2r' - 2 < r$. Since $r \leqslant r' + 1$ however this is a contradiction unless $r' \leqslant 2$. □

We assume from now on, that

$$\mathrm{CTF}(G) < \mathrm{sp}(n, \delta + \eta n)\,, \tag{3.14}$$

that is, we are not in Cases (S1) or (S2). Our aim is to conclude that then $(*)$ $\mathrm{int}(G)$ is an independent set and that its vertices are contained in at least r' triangle components. It turns out, however, that we need to consider the cases $r = r'+1 = 3$ and $r = r'+1 = 4$ (i.e., the cases when the minimum degree δ is just a little bit below $\frac{3}{5}n$ and $\frac{4}{7}n$, respectively) separately. Unfortunately these two cases, which are treated by Claim 3.8.5, require a somewhat technical case analysis, which we prefer to defer to the end of the section.

Claim 3.8.5. *If* $r = r'+1 = 3$ *or* $r = r'+1 = 4$ *then* $\mathrm{int}(G)$ *is an independent set all of whose vertices are contained in at least* r' *triangle components.*

Assuming this fact is true we can deduce $(*)$ for all values $r \geqslant 3$ as follows.

Claim 3.8.6. *The set* $\mathrm{int}(G)$ *is an independent set (and hence of size at most* $n-\delta$*) all of whose vertices are contained in at least* r' *triangle components.*

Proof of Claim 3.8.6. Recall that we have $r \geqslant 3$ at this point of the proof. Moreover, the cases $r = r'+1 = 3$ and $r = r'+1 = 4$ are handled by Claim 3.8.5. So we assume we are not in these cases; in particular, $r' \geqslant 3$. We will show that then each vertex of $\mathrm{int}(G)$ is contained in at least r' triangle components. Once we establish this, Claim 3.8.4 implies that there are no edges in $\mathrm{int}(G)$ and so $\mathrm{int}(G)$ is an independent set as desired.

To prove that each vertex of $\mathrm{int}(G)$ is contained in at least r' triangle components we assume the contrary and show that then $\mathrm{CTF}(G) \geqslant \mathrm{sp}(n, \delta+\eta n)$, a contradiction to (3.14). Indeed, let $w \in \mathrm{int}(G)$ and suppose that there are $k > 1$ triangle components $C_1, \dots, C_k$ containing w. For $i \in [k]$ let U_i be the set of neighbours u of w such that $uw \in C_i$. By Lemma 3.10(*c*) we have $\deg^{\min}(G[U_i]) \geqslant 2\delta - n$ and $|U_i| \geqslant 2\delta - n + 1$. Suppose that U_1 is the largest of the U_i. No vertex in U_1 has a neighbour in U_2, since the components are distinct. In addition, all edges in $G[U_1]$ are triangle connected, because $U_1 \subseteq \mathrm{N}^\wedge(w)$. Therefore Claim 3.8.3 implies that there is a connected triangle factor T in G covering $\min(3\lfloor |U_1|/2 \rfloor, 3(2\delta - n), 2\delta - n + |U_2|) \geqslant \min(3\lfloor |U_1|/2 \rfloor, 4\delta - 2n)$ vertices. If w lies only in $r'-1$ triangle components then $|U_1| \geqslant \delta/(r'-1)$ and therefore T covers at least $\min(3\lfloor \delta/(2r'-2) \rfloor, 4\delta - 2n)$ vertices. Now since (3.6) holds, we have $\frac{3}{2}\delta/(r'-1) - 2 \geqslant \mathrm{sp}(n, \delta + \eta n)$. Since $r \geqslant r' \geqslant 3$ and we have excluded the case $r = r'+1 = 4$, by (3.7) we have $4\delta - 2n \geqslant \mathrm{sp}(n, \delta + \eta n)$. It follows that T covers at least $\mathrm{sp}(n, \delta+\eta n)$ vertices, in contradiction to (3.14). □

Claim 3.8.7. *We are in Case* (S3).

Proof of Claim 3.8.7. Claim 3.8.6 tells us that $\mathrm{int}(G)$ is an independent set. By Lemma 3.10(a) and the fact that $\delta > n-\delta$ we have that every triangle component in G has an exterior, and by Lemma 3.10(b) that there are no edges between any triangle component exteriors. Hence, to show that we are in Case (S3), it is enough to prove that

$$|\,\mathrm{int}(G)| := \alpha \geqslant n-\delta-11\eta n \qquad \text{and} \qquad |X_1| \leqslant \frac{19}{10}(2\delta-n) \tag{3.15}$$

for the biggest triangle component exterior X_1 in G. Suppose for a contradiction that this is not the case. We first claim that this forces G to have exactly r' triangle components.

Indeed, assume G has $k \geqslant r'+1$ triangle components. Each of these components C has vertices in its exterior $\partial(C)$, and so by Lemma 3.10(b) the minimum degree of G implies $|\partial(C)| \geqslant \delta-\alpha+1 \geqslant 2\delta-n+1$. We let these triangle component exteriors be $X_1,\dots,X_k$, with X_1 being the biggest. Since $n = |X_1\dot\cup\dots\dot\cup X_k\dot\cup\,\mathrm{int}(G)|$, we have $(r'+1)(\delta-\alpha)+\alpha < n$. We distinguish two cases.

Case 1: (3.15) fails because $\alpha < n-\delta-11\eta n$. Then we obtain

$$\begin{aligned}(r'+1)\delta &< n+r'\alpha < n+r'(n-\delta-11\eta n)\\ &= (r'+1)n-(9r'-1)\eta n - r'\delta-(2r'+1)\eta n\,.\end{aligned}$$

Straightforward manipulation gives

$$\delta+\eta n < \frac{(r'+1)n-(9r'-1)\eta n}{2(r'+1)-1}\,.$$

Since $(9r'-1)\eta n \geqslant 9r'-1 \geqslant r'$ this contradicts (3.5) applied to $r' = r_p(n,\delta+\eta n)$.

Case 2: (3.15) fails because $|X_1| > \frac{19}{10}(2\delta-n)$. Let x be any vertex in X_2. Since x has at least δ neighbours, none of which are in $X_1\dot\cup X_3\dot\cup\dots\dot\cup X_k$, we have

$$\begin{aligned}1+\delta+\frac{19}{10}(2\delta-n)+(k-2)(2\delta-n+1) &\leqslant n\,, \text{ hence}\\ \frac{19}{10}(2\delta-n)+(r'-1)(2\delta-n) &< n-\delta\,.\end{aligned}$$

By (3.5) we have $r' \geqslant (n-\delta-\eta n)/(2\delta+2\eta n-n+1)$. Combined with the last inequality, this gives

$$\frac{9}{10}(2\delta-n)+\frac{n-\delta-\eta n}{2\delta-n+1+2\eta n}(2\delta-n) < n-\delta$$

Now provided that $\eta < 2\mu^2/3$, and since $2\delta - n \geqslant 2\mu n$, we have

$$\begin{aligned}(2\delta - n + 2\eta n + 1)(1-\mu) &< 2\delta - n + 3\eta n - \mu(2\delta - n)\\ &\leqslant 2\delta - n + 3\eta n - 2\mu^2 n < 2\delta - n\,,\end{aligned}$$

and we obtain $\frac{9}{5}\mu n + (1-\mu)(n-\delta-\eta n) < n - \delta$ which is a contradiction since $n - \delta < n/2$ and $\eta < \mu$.

Hence, if (3.15) fails, then G has indeed exactly r' triangle components.

Now we use this fact in order to derive a contradiction to (3.14). Observe that, if $r' = 2$, and accordingly $\delta \geqslant (\frac{3}{5} - 2\eta)n$, then Claim 3.8.1 implies that (3.15) holds, because according to (3.14) we have $\mathrm{CTF}(G) < \mathrm{sp}(n, \delta + \eta n)$. In the remainder we assume $r' \geqslant 3$.

Since every vertex in X_1 has neighbours only in X_1 and $\mathrm{int}(G)$, and $|\,\mathrm{int}(G)| \leqslant n - \delta$, we have $\delta(G[X_1]) \geqslant 2\delta - n$. Furthermore, since no vertex in X_1 has neighbours in either X_2 or X_3, and $|X_2 \dot\cup X_3| \geqslant 2(2\delta - n + 1)$, we can apply Claim 3.8.3 to obtain

$$\begin{aligned}\mathrm{CTF}(G) &\geqslant \min\big(3\lfloor |X_1|/2\rfloor, 3(2\delta - n), 2\delta - n + 2(2\delta - n + 1)\big)\\ &= \min\big(3\lfloor |X_1|/2\rfloor, 3(2\delta - n)\big)\,.\end{aligned}$$

Now by (3.7), $\mathrm{CTF}(G) \geqslant 3(2\delta - n)$ is a contradiction to (3.14), so to complete our proof it remains to show that if (3.15) fails, then $\mathrm{CTF}(G) \geqslant 3\lfloor |X_1|/2\rfloor$ is also a contradiction to (3.14). Again, we distinguish two cases.

Case 1: (3.15) fails because $\alpha < n - \delta - 11\eta n$. Since X_1 is the largest exterior, we have $|X_1| \geqslant (\delta + 11\eta n)/r'$. But we have by (3.6) that

$$\mathrm{sp}(n, \delta + \eta n) \leqslant \frac{3}{2}\frac{\delta + 3\eta n}{r'} - 2 < 3\Big\lfloor \frac{\delta + 11\eta n}{2r'}\Big\rfloor\,,$$

so that $\mathrm{CTF}(G) \geqslant 3\lfloor |X_1|/2\rfloor$ is indeed a contradiction to (3.14).

Case 2: (3.15) fails because $|X_1| > \frac{19}{10}(2\delta - n)$. Then $\mathrm{CTF}(G) \geqslant 3\lfloor |X_1|/2\rfloor \geqslant \frac{57}{20}(2\delta - n) - 2$, which by (3.7) is a contradiction to (3.14), as desired. □

This completes, modulo the proof of Claim 3.8.5, the proof of Lemma 3.8. □

It remains to show Claim 3.8.5. Note that we can use all facts from the proof of Lemma 3.8 that precede Claim 3.8.5. We will further assume that all constants and variables are set up as in this proof.

Proof of Claim 3.8.5. Recall that we assumed (3.14), i.e., $\mathrm{CTF}(G) < \mathrm{sp}(n, \delta + \eta n)$, in this part of the proof of Lemma 3.8. We distinguish two cases.

Case 1: $r = 3$ and $r' = 2$. In this case $\deg^{\min}(G) \in [(\frac{3}{5} - 2\eta)n, (\frac{3}{5} + \eta)n]$. Trivially each vertex of $\mathrm{int}(G)$ is contained in at least $r' = 2$ triangle components. Suppose for a contradiction that there is an edge uv in $\mathrm{int}(G)$. Let x be a common neighbour of u and v, and C be the triangle component containing the triangle uvx. Let $U_1 := \{y : uy \in C\}$ and $V_1 := \{y : vy \in C\}$ and let $U_2 := \mathrm{N}(u) - U_1$ and $V_2 := \mathrm{N}(v) - V_1$. Observe that $U_2 \cap V_2 = \emptyset$.

By definition x is not in, and has no neighbour in, $U_2 \dot{\cup} V_2$. It follows that $|U_2 \dot{\cup} V_2| < n - \delta \leqslant (\frac{2}{5} + 2\eta)n$. On the other hand, by Lemma 3.10(c), we have $|U_2|$, $|V_2| > 2\delta - n \geqslant \frac{1}{5}n - 4\eta n$, and thus

$$|U_2|, |V_2| \in \left[(\tfrac{1}{5} - 4\eta)n, (\tfrac{1}{5} + 6\eta)n\right].$$

Since $\deg(u) \geqslant \delta \geqslant (\frac{3}{5} - 2\eta)n$, we have $|U_1| \geqslant \delta - |U_2| \geqslant (\frac{2}{5} - 8\eta)n$. But no vertex in U_2 is adjacent to any vertex in U_1. This implies that every vertex in U_2 is adjacent to all but at most $n - \delta - |U_1| \leqslant 10\eta n$ vertices outside U_1. Since $\eta < \frac{1}{1000}$ we have $|U_2| > 20\eta n$, so $\deg^{\min}(G[U_2]) > |U_2|/2$, and by Proposition 3.11(a), U_2 contains a matching M_u with $\lfloor |U_2|/2 \rfloor$ edges. Since each vertex of U_2 has at most $10\eta n$ non-neighbours outside U_1, each pair of vertices has common neighbourhood covering all but at most $20\eta n$ vertices of $V(G) - U_1$. In particular, the common neighbourhood of each edge of M_u covers all but at most $20\eta n$ vertices of $V(G) - U_1$. Similarly, V_2 contains a matching M_v with $\lfloor |V_2|/2 \rfloor$ edges, and the common neighbourhood of each edge covers all but at most $20\eta n$ vertices of $V(G) - V_1$.

Since $20\eta n < |U_2|/4$ and $U_2 \cap V_1 = \emptyset$, the common neighbourhood of each edge of M_v contains more than half of the edges of M_u. By symmetry, the reverse is also true. Thus all edges in $M_u \dot{\cup} M_v$ are in the same triangle component of G. Finally, each edge of $M_u \dot{\cup} M_v$ has at least $\delta - 10\eta n - |U_2 \dot{\cup} V_2| \geqslant (\frac{1}{5} - 24\eta)n$ common neighbours outside $U_2 \dot{\cup} V_2$. Choosing greedily for each edge of $M_u \dot{\cup} M_v$ in succession distinct common neighbours outside $U_2 \dot{\cup} V_2$, we obtain a connected triangle factor with $\min(\lfloor |U_2|/2 \rfloor + \lfloor |V_2|/2 \rfloor, (\frac{1}{5} - 24\eta)n) = (\frac{1}{5} - 24\eta)n$ triangles. But then $\mathrm{CTF}(G) \geqslant (\frac{3}{5} - 72\eta)n > n/2 > \mathrm{sp}(n, \delta + \eta n)$, a contradiction to (3.14). This proves Claim 3.8.5 for the case $r = 3$ and $r' = 2$.

Case 2: $r = 4$ and $r' = 3$. This implies that $(\frac{4}{7} - 2\eta)n \leqslant \deg^{\min}(G) \leqslant (\frac{4}{7} + \eta)n$, and consequently $\mathrm{sp}(n, \delta + \eta n) < (\frac{2}{7} + 2\eta)n$. We first prove two statements about the structure of G which are forced by (3.14).

(Ψ) If a vertex u has sets of neighbours U, U' on edges in exactly two different triangle components with $|U| \geqslant |U'|$ then $(\frac{1}{7} - 4\eta)n < |U'| < (\frac{1}{7} + 6\eta)n$ and $(\frac{3}{7} - 8\eta)n < |U| < (\frac{3}{7} + 2\eta)n$.

Proof of (Ψ). For the lower bound on $|U'|$, observe that by (c) of Lemma 3.10 we have $\deg^{\min}(G[U']) \geqslant 2\delta - n \geqslant (\frac{1}{7} - 4\eta)n$. To obtain the upper bound, again by Lemma 3.10(c) we have $\deg^{\min}(G[U]) \geqslant 2\delta - n$, and since the sets U and U' are neighbours of u in different triangle components C and C', there are no edges from U to U'. Furthermore, since any edge in $G[U]$ forms a triangle with u using an edge from u to U, all edges in $G[U]$ are in C. Now by Claim 3.8.3 we have

$$\mathrm{CTF}(G) \geqslant \min(3\lfloor |U|/2 \rfloor, 3(2\delta - n), 2\delta - n + |U'|)\,.$$

Since $|U| \geqslant \delta/2$ we have $3\lfloor |U|/2 \rfloor \geqslant (\frac{3}{7} - 3\eta)n - 2 > \mathrm{sp}(n, \delta + \eta n)$. By (3.7) we have $3(2\delta - n) > \mathrm{sp}(n, \delta + \eta n)$. Because (3.14) holds, we have $2\delta - n + |U'| < \mathrm{sp}(n, \delta + \eta n) < (\frac{2}{7} + 2\eta)n$, and therefore $|U'| < (\frac{1}{7} + 6\eta)n$. Now the claimed lower and upper bounds on $|U|$ follow from $U = \mathrm{N}(u) - U'$, and from the fact that no vertex in U' has a neighbour in U, respectively. □

(Ξ) If a vertex u has sets of neighbours U_1, U_2, U_3 on edges in exactly three different triangle components then $(\frac{4}{21} + 2\eta)n > |U_i| > (\frac{4}{21} - 6\eta)n$ for $i \in [3]$.

Proof of (Ξ). Assume that U_1 is the largest of the three sets. By (c) of Lemma 3.10 we have $\deg^{\min}(G[U_i]) \geqslant 2\delta - n \geqslant (\frac{1}{7} - 4\eta)n$ for each i, so $|U_i| > (\frac{1}{7} - 4\eta)n$ for each i. As in the previous case, there can be no edge from U_1 to $U_2 \dot\cup U_3$, and all edges in U_1 are triangle-connected. Thus by Claim 3.8.3 we have

$$\mathrm{CTF}(G) \geqslant \min\left(3\lfloor |U_1|/2 \rfloor, 3(2\delta - n), 2\delta - n + |U_2 \dot\cup U_3|\right).$$

Now since $\mathrm{sp}(n, \delta + \eta n) < (\frac{3}{7} - 10\eta)n$ and (3.14) holds, we have

$$3\lfloor |U_1|/2 \rfloor < \mathrm{sp}(n, \delta + \eta n) \leqslant (\frac{2}{7} + 2\eta)n$$

which implies $|U_1| < (\frac{4}{21} + 2\eta)n$. Since $|U_2|, |U_3| \leqslant |U_1|$ this completes the desired upper bounds. The lower bounds follow from $|U_1| + |U_2| + |U_3| \geqslant \delta \geqslant (\frac{4}{7} - 2\eta)n$. □

Next we show that

(Θ) $\mathrm{int}(G)$ is an independent set.

Proof of (Θ). Assume for a contradiction that there is an edge $uv \in \mathrm{int}(G)$. By Claim 3.8.4 one of the vertices of this edge, say u, is in only 2 triangle components. Let its neighbours be U_1 and U_2 in these two triangle components, and let the neighbours of v be partitioned into sets $V_1, \ldots, V_k$ according to the triangle component containing the edge to v. Assume further that $\mathrm{N}^\wedge(u,v) \subseteq U_1 \cap V_1$, so that $U_2, V_2, \ldots, V_k$ are pairwise disjoint. Let $x \in \mathrm{N}^\wedge(u,v)$. Since x has neighbours in neither U_2 nor V_2, and since by Lemma 3.10(c) we have $|V_2| > (\frac{1}{7} - 4\eta)n$, we conclude that $\delta \leqslant \deg(x) \leqslant n - 1 - |U_2| - |V_2|$. In particular, $|U_2| < (\frac{3}{7} - 8\eta)n$ because $\delta \geqslant (\frac{4}{7} - 2\eta)n$, and therefore by (Ψ) we have

$$(\frac{1}{7} - 4\eta)n < |U_2| < (\frac{1}{7} + 6\eta)n\,.$$

Next we want to derive analogous bounds for $|V_2|$. For this purpose we first show that $k = 2$.

Indeed, if we had $k = 3$, then by (Ξ)

$$\begin{aligned}\deg(x) &\leqslant n - 1 - |U_2| - |V_2| - |V_3| \\ &\leqslant n - 1 - (\tfrac{1}{7} - 4\eta)n - 2(\tfrac{4}{21} - 6\eta)n < (\tfrac{10}{21} + 16\eta)n < \delta\,,\end{aligned}$$

and this contradicts $\deg^{\min}(G) \geqslant \delta$. Similarly, if $k \geqslant 4$, then by Lemma 3.10(c) we have $|V_i| \geqslant (\frac{1}{7} - 4\eta)n$ for each i, and hence

$$\deg(x) \leqslant n - 1 - |U_2| - |V_2| - |V_3| - |V_4| < (\tfrac{3}{7} + 16\eta)n < \delta\,,$$

which too is a contradiction. It follows that $k = 2$ as claimed.

Hence, we can argue analogously as before (for U_2) that $|V_2| > (\frac{3}{7} - 8\eta)$ would contradict $\deg(x) \geqslant \delta$. Consequently, by (Ψ) we have

$$(\frac{1}{7} - 4\eta)n < |V_2| < (\frac{1}{7} + 6\eta)n\,.$$

We now argue that this yields a contradiction to (3.14) in much the same way as we argued in the $r = r' + 1 = 3$ case. Every vertex of U_2 is adjacent to all but at most $n - |U_1| - \delta \leqslant 10\eta n$ vertices of $V(G) - U_1$. Since $|U_2| > 20\eta n$, by Proposition 3.11(a) there is a matching M_u in U_2 covering all but at most one vertex of U_2. Each edge of M_u has at least $\delta - 10\eta n \geqslant (\frac{4}{7} - 12\eta)n$ common neighbours outside U_1. Similarly, in V_2 there is a matching M_v covering all but at most one vertex of V_2, each edge of which has at least $(\frac{4}{7} - 12\eta)n$ common neighbours outside V_1. Since $\mathrm{N}^\wedge(u,v) = U_1 \cap V_1$, we have $U_1 \cap V_2 = \emptyset$. It follows that every edge of M_v has more than half of the edges of M_u in its common neighbourhood, and thus the edges $M_u \dot\cup M_v$ are triangle connected. Choosing greedily for each edge in $M_u \dot\cup M_v$ in succession a distinct common neighbour outside $M_u \dot\cup M_v$, we obtain a connected triangle factor with as many triangles as there are edges in $M_u \dot\cup M_v$.

Since $|U_2|, |V_2| > (\frac{1}{7} - 4\eta)n$, we have $\mathrm{CTF}(G) > (\frac{3}{7} - 12\eta)n - 3 > \mathrm{sp}(n, \delta + \eta n)$, contradicting (3.14). This completes the proof that $\mathrm{int}(G)$ is an independent set. $\square$

It remains to show that each vertex $u \in \mathrm{int}(G)$ is contained in at least $r' = 3$ triangle components. Assume for a contradiction that this is not the case and that some vertex u is only contained in 2 triangle components, C and C'. Let U and U', respectively, be the neighbours of u on edges in C and C'. Without loss of generality $|U| \geqslant |U'|$. Because $\mathrm{int}(G)$ is an independent set, U and U' are contained in the exteriors of C and C'. By Lemma 3.10(*b*) there are thus no edges between U and $\partial(C')$. By Lemma 3.10(*c*) we have $\deg^{\min}(G[U]) \geqslant 2\delta - n$, and since $U \subseteq \partial(C)$ every edge of $G[U]$ is in C. It follows that we may apply Claim 3.8.3 to obtain

$$\mathrm{CTF}(G) \geqslant \min\left(3\lfloor |U|/2 \rfloor, 3(2\delta - n), 2\delta - n + |\partial(C')|\right).$$

Since $|U| \geqslant \delta/2$ we have $3\lfloor |U|/2 \rfloor \geqslant (\frac{3}{7} - 3\eta)n - 2 > \mathrm{sp}(n, \delta + \eta n)$. By (3.7) we have $3(2\delta - n) > \mathrm{sp}(n, \delta + \eta n)$. Since (3.14) holds, we conclude that $2\delta - n + |\partial(C')| < \mathrm{sp}(n, \delta + \eta n) < (\frac{2}{7} + 2\eta)n$, and therefore $|\partial(C')| < (\frac{1}{7} + 6\eta)n$.

Now any vertex in $\partial(C')$ has neighbours only in $\partial(C') \dot\cup \mathrm{int}(G)$, and therefore $|\mathrm{int}(G)| \geqslant \delta - |\partial(C')| \geqslant (\frac{3}{7} - 8\eta)n$. The vertex u has neighbours only in $U' \subseteq \partial(C')$ and U, and therefore

$$|U| \geqslant \delta - |U'| \geqslant \delta - |\partial(C')| \geqslant (\tfrac{3}{7} - 8\eta)n\,.$$

By Lemma 3.10(*c*) we have $\deg^{\min}(G[U]) \geqslant 2\delta - n \geqslant (\frac{1}{7} - 4\eta)n$, and since $|U| > (\frac{2}{7} - 8\eta)n$ we obtain by Proposition 3.11(*a*) a matching M in U with at least $(\frac{1}{7} - 4\eta)n$ edges. Now each vertex in $\mathrm{int}(G)$ is adjacent to all but at most $n - \delta - |\mathrm{int}(G)| \leqslant 10\eta n$ vertices outside $\mathrm{int}(G)$. In particular, each vertex in $\mathrm{int}(G)$ is adjacent to all but at most $10\eta n$ vertices of M, and is therefore a common neighbour of all but at most $10\eta n$ edges of M. We now match greedily vertices of $\mathrm{int}(G)$ with distinct edges of M forming triangles. Since $|\mathrm{int}(G)| > |M|$, we will be forced to halt only when we come to a vertex $x \in \mathrm{int}(G)$ which is not a common neighbour of any remaining edge of M, i.e., when we have used all but at most $10\eta n$ edges of M. It follows that we obtain a triangle factor T with at least $(\frac{1}{7} - 14\eta)n$ triangles. Since each triangle uses an edge of $M \subseteq G[U] \subseteq G[\partial(C)]$, T is a connected triangle factor, and we have $\mathrm{CTF}(G) \geqslant (\frac{3}{7} - 42\eta)n > \mathrm{sp}(n, \delta + \eta n)$ in contradiction to (3.14). $\square$

3.4 Near-extremal graphs

In this section we provide the proof of Lemma 3.9. To prepare this proof we start with two useful lemmas. The first will be used to construct squared paths and squared cycles from simple paths and cycles.

Lemma 3.12. *Given a graph G, let $T = (t_1, t_2, \dots, t_{2l})$ be a path in G and W a set of vertices disjoint from T. Let $Q_1 = (t_1, t_2)$, $Q_i = (t_{2i-3}, t_{2i-2}, t_{2i-1}, t_{2i})$ for all $1 < i \leqslant l$, and $Q_{l+1} = (t_{2l-1}, t_{2l})$. If there exists an ordering σ of $[l+1]$ such that for each i the vertices in $Q_{\sigma(i)}$ have at least i common neighbours in W, then there is a squared path*

$$(q_1, t_1, t_2, q_2, t_3, t_4, q_3, \dots, t_{2\ell}, q_{\ell+1})$$

in G, with $q_i \in W$ for each i, using every vertex of T.

If T is a cycle on $2l$ vertices we let instead $Q_1 = (t_{2l-1}, t_{2l}, t_1, t_2)$, $Q_i = (t_{2i-3}, t_{2i-2}, t_{2i-1}, t_{2i})$ for all $1 < i \leqslant l$, and σ be an ordering on $[l]$. Then, under the same conditions, we obtain a squared cycle C^2_{3l}.

Proof. We need only ensure that for each i one can choose q_i such that q_i is a common neighbour of Q_i and the q_i are distinct. This is possible by choosing for each i in succession $q_{\sigma(i)}$ to be any so far unused common neighbour of $Q_{\sigma(i)}$. $\square$

The second lemma is a variant on Dirac's theorem and permits us to construct paths and cycles of desired lengths which keep some 'bad' vertices far apart.

Lemma 3.13. *Let H be a graph on h vertices and $B \subseteq V(H)$ be of size at most $h/100$. Suppose that every vertex in B has at least $9|B|$ neighbours in H, and every vertex outside B has at least $h/2 + 9|B| + 10$ neighbours in H. Then for any given $3 \leqslant \ell \leqslant h$ we can find a cycle T_ℓ of length ℓ in H on which no four consecutive vertices contain more than one vertex of B. Furthermore, if x and y are any two vertices not in B and $5 \leqslant \ell \leqslant h$, we can find an ℓ-vertex path T_ℓ whose endvertices are x and y on which no four consecutive vertices contain more than one vertex of $B \cup \{x, y\}$.*

Proof. If we seek a path in H from x to y then we create a 'dummy edge' between x and y. If we seek a cycle, let xy be any edge of $H - B$.

First we construct a path P in H covering B with the desired property. Let $B = \{b_1, b_2, \dots, b_{|B|}\}$. For each $1 \leqslant i \leqslant |B| - 1$, choose a vertex $u_i \in H - B$ adjacent to b_i and a vertex $v_i \in H - B$ adjacent to b_{i+1}. Because both u_i and v_i have $h/2 + 9|B| + 10$ neighbours in H, they have at least $18|B| + 20$ common neighbours. At most $3|B|$ of these are either in B or amongst the chosen u_j, v_j, and so we can find a so far unused vertex w_i adjacent to u_i and v_i. Since we require only $|B|-1$ vertices $w_1, \dots, w_{|B|-1}$ we can pick the vertices greedily.

We let v_0 be yet another vertex adjacent to b_1, and $u_{|B|}$ adjacent to $b_{|B|}$, and choose any further vertices $w_0, v_0, w_{|B|}, u_{|B|}$ such that

$$P = (x, y, u_0, w_0, v_0, b_1, u_1, w_1, v_1, b_2, \dots, v_{|B|-1}, b_{|B|}, u_{|B|}, w_{|B|}, v_{|B|})$$

is a path on $4|B|+5$ vertices.

Now we let P' be a path extending P in H of maximum length. We claim that P' is in fact spanning. Suppose not: let u be an end-vertex of P' and v a vertex not on P'. Since P' is maximal every neighbour of u is on P', so $v(P') > h/2+9|B|+10$. If there existed an edge $u'v'$ of $P'-P$ with $u'u$ and $v'v$ edges of H, with v' closer to u on P' than u', then we would have a longer path extending P in H. Counting the edges leaving u and v yields a contradiction.

Finally we let u and v be the end-vertices of the spanning path P'. If uv is an edge of H, or if $u'v'$ is an edge of $P'-P$, with u' nearer to u on P' than v', such that uv' and $u'v$ are edges of H, then we obtain a cycle T spanning H and containing P as a subpath. Again edge counting reveals that such an edge must exist.

To obtain a cycle T_ℓ with $h-|B|-2 \leqslant \ell < h$ we take u to be an end-vertex of the path $T-P$ and v its successor on $T-P$. If we can find two further vertices u' and v' on $T-P$ (in that order from u along $T-P$) with $h-\ell$ vertices between them and with uu' and vv' edges of H then we would obtain a cycle T_ℓ of length ℓ. Again simple edge counting reveals that such a pair of vertices exists. To obtain a cycle T_ℓ with $3 \leqslant \ell < h-|B|-2$ we note that $H-B$ has minimum degree $h/2+8|B|+10 > (h-|B|)/2+1$ and thus contains a cycle of every possible length using the edge xy.

The cycle T_ℓ satisfies the condition that no four consecutive vertices contain more than one vertex of B, since either it preserves P as a subpath or it contains no vertices of B at all. Similarly the path from x to y within T_ℓ satisfies the required conditions. □

Before embarking upon the proof of Lemma 3.9 we give an outline of the method. We recall that the Szemerédi partition supplied to the Lemma is essentially the extremal structure. We shall show that the underlying graph either also has an extremal structure, or possesses features which actually lead to longer squared paths and cycles than required for the conclusion of the Lemma. This is complicated by the fact that the Szemerédi partition is insensitive both to mis-assignment of a sublinear number of vertices and to editing of a subquadratic number of edges: we must assume, for example, that although the vertex set I in the reduced graph R is independent, the vertex set $\bigcup I$ may fail to contain some vertices of G with no neighbours in $\bigcup I$, and may contain a small number of edges meeting every vertex. However, observe that by the definition of an (ε, d)-regular partition, there are no vertices of $\bigcup I$ with more than $(\varepsilon+d)n$ neighbours in $\bigcup I$. Fortunately, it is possible to apply the following strategy in this case.

We start by separating those vertices with 'few' neighbours in $\bigcup I$, which we shall collect in a set W, and those with 'many'. We are then able to show (as Claim 3.9.2 below) that, if there are two vertex disjoint edges in W, then the sets $\bigcup B_1$ and $\bigcup B_2$ are in the same triangle component of G ('unexpectedly', since B_1 and B_2 are in different triangle components in R). We shall show that in this case it is possible to construct very long squared paths and cycles by making use of Lemma 3.6.

Hence we can assume that there are not two disjoint edges in W, which in turn implies that W is almost independent and will give us rather precise control about the size of W. In addition, the minimum degree condition will guarantee that almost every edge from W to the remainder of G is present. We would like to then say that in $V(G) - W$ we can find a long path, which together with vertices from W forms a squared path (and similarly for squared cycles). Unfortunately since $G[W, V(G)-W]$ is not necessarily a complete bipartite graph, this statement is not obviously true: although by definition no vertex outside W has very few neighbours in W, it is certainly possible that two vertices outside W could fail to have a common neighbour in W. But the statement is true for a path possessing sufficiently nice properties—specifically, satisfying the conditions of Lemma 3.12—and the purpose of Lemma 3.13 is to provide paths and cycles with those nice properties. The remainder of our proof, then, consists of setting up conditions for the application of Lemma 3.13.

Proof of Lemma 3.9. Given $\nu > 0$, suppose the parameters $\eta > 0$ and $d > 0$ satisfy the following inequalities.

$$\eta \leqslant \frac{\nu^4}{10^8} \qquad \text{and} \qquad d \leqslant \frac{\nu^4}{10^8} \tag{3.16}$$

Given $d > 0$, Lemma 3.6 returns a constant $\varepsilon_{\mathrm{EL}} > 0$. We set

$$\varepsilon_0 := \min\left(\frac{\nu^4}{10^8}, \varepsilon_{\mathrm{EL}}\right). \tag{3.17}$$

Given $m_{\rm EL}$ and $0 < \varepsilon < \varepsilon_0$, Lemma 3.6 returns a constant $n_{\rm EL}$. We set

$$N := \max\left(1000m_{\rm EL}^4, 100\eta^{-1}\nu^{-1}, n_{\rm EL}\right). \tag{3.18}$$

Now let G, R, and the partition $V(R) = I\dot\cup B_1\dot\cup\dots\dot\cup B_k$ satisfy conditions (i)–(vi) of the lemma.

If $\deg^{\min}(G) = \delta \geqslant \frac{2n-1}{3}$ then we can appeal to Theorem 3.1 to find a spanning squared path in G; if $\delta \geqslant \frac{2n}{3}$ then we can appeal to Theorem 3.2 to find C_ℓ^2 for each $\ell \in [3,n] - \{5\}$. Therefore, the definition of $\mathrm{sp}(n,\delta)$ and $\mathrm{sc}(n,\delta)$ imply that we may assume $\delta < 2n/3$ in the following, and that we only need to find

$$\text{squared paths and squared cycles of length at most } 11n/20. \tag{3.19}$$

We now start by investigating the sizes of I and of the B_i. Define $\delta' := (\delta/n - d - \varepsilon)m$. Since R is an (ε, d)-reduced graph we have

$$\deg^{\min}(R) \geqslant \delta' = (\delta/n - d - \varepsilon)m\,. \tag{3.20}$$

Observe that moreover

$$|I| \leqslant m - \delta' \leqslant \left(1 - \frac{\delta}{n} + d + \varepsilon\right)m\,, \tag{3.21}$$

by (v) because clusters in I have δ' neighbours outside I in R. For $i \in [k]$, fix a cluster $C \in B_i$. By assumption (vi) C has neighbours only in $B_i \cup I$ in R. Since

$$\delta' \leqslant \deg(C) = \deg(C, B_i \cup I) \leqslant \deg(C, B_i) + |I| \leqslant \deg(C, B_i) + m - \delta'\,,$$

we have

$$\begin{aligned} |B_i| > \deg(C, B_i) \geqslant 2\delta' - m &\geqslant \frac{m}{n}\big(2(\delta - dn - \varepsilon n) - n\big) \\ &= \frac{m}{n}\big(2\delta - n - (d+\varepsilon)n\big)\,. \end{aligned}$$

Now since $2\delta - n \geqslant 2\nu n$ by (i), we conclude from (3.16) and (3.17) that

$$|B_i| \geqslant \frac{2m(2\delta - n)}{3n} \geqslant \frac{4}{3}\nu m\,. \tag{3.22}$$

We next show that each B_i is part of exactly one triangle component of R.

Claim 3.9.1. *For each $1 \leqslant i \leqslant k$ the following holds. All edges in $R[B_i]$ are triangle connected in R.*

Proof of Claim 3.9.1. By assumption (vi) we have

$$|B_i| \leqslant 19m(2\delta - n)/(10n) \leqslant 39(2\delta' - m)/20\,, \tag{3.23}$$

where the second inequality comes from (3.16) and (3.17). Since we have $\deg^{\min}{}_R(B_i) \geqslant 2\delta' - m > |B_i|/2$, the graph $R[B_i]$ is connected. It follows that if there are two edges in $R[B_i]$ which are not triangle-connected, then there are two adjacent edges in $R[B_i]$ with this property. That is, there are vertices P, Q and Q' of B_i such that PQ is in triangle component C and PQ' is in triangle component C' with $C \neq C'$.

We now show that there are at least $2\delta' - m$ edges leaving P in $R[B_i]$ which are in C. There are two possibilities. First, suppose there are no C-edges from P to I. In this case, the common neighbourhood $\mathrm{N}^\wedge(PQ)$ must lie entirely in B_i. Every vertex of $\mathrm{N}^\wedge(PQ)$ makes a C-edge with P, and we have $|\mathrm{N}^\wedge(PQ)| \geqslant 2\delta' - m$ as required. Second, suppose that there is a C-edge PP' with $P' \in I$. Since I is an independent set in R, the set $\mathrm{N}^\wedge(PP')$ lies entirely within B_i, and has size at least $2\delta' - m$. Again, every edge from P to $\mathrm{N}^\wedge(PP')$ is a C-edge, as desired.

By the identical argument, there are at least $2\delta' - m$ edges leaving P in $R[B_i]$ which are in C'. Since no edge is in both C and C', there are at least $2(2\delta' - m)$ edges leaving P in $R[B_i]$, so $|B_i| \geqslant 2(2\delta' - m)$. This contradicts (3.23). It follows that all edges of B_i are triangle connected, as desired. $\square$

We next define a set W of those vertices in G which have few neighbours in $\bigcup I$. The intuition is that W consists of $\bigcup I$ and only a few more vertices of G. To simplify notation, we set $\xi := \sqrt[4]{\varepsilon + d + 11\eta}$. By (3.16) and (3.17), we have

$$\xi \leqslant \nu/100\,. \tag{3.24}$$

Let W be the vertices of G which do not have more than ξn neighbours in $\bigcup I$. Since $\xi > d + \varepsilon$, by the independence of I and by the definition of an (ε, d)-regular partition, we have $\bigcup I \subseteq W$. By assumption (v) we have $|I| \geqslant (n - \delta - 11\eta n)m/n$. Hence every edge in W has at least

$$2(\delta - \xi n) - \Big(n - |\bigcup I|\Big) > \frac{\delta - (2\delta - n)}{16} \tag{3.25}$$

common neighbours outside $\bigcup I$, where we use assumption (i) that $2\delta - n > 2\nu n$, (3.16) and (3.24).

By this observation, the next fact implies that we are done if there are two vertex disjoint edges in W.

Claim 3.9.2. *If u_1v_1 and u_2v_2 are vertex disjoint edges of G such that for $i = 1, 2$ the edge u_iv_i has at least $\delta - (2\delta - n)/16$ common neighbours outside $\bigcup I$, then G contains $P^2_{\mathrm{sp}(n,\delta)}$ and C^2_ℓ for each $\ell \in [3, \mathrm{sc}(n, \delta)] - \{5\}$.*

Proof of Claim 3.9.2. Let D' be the set of clusters $C \in V(R) - I$ such that u_1v_1 has at most $2dn/m$ common neighbours in C. By the hypothesis, u_1v_1 has at least $\delta - (2\delta - n)/16$ common neighbours outside $\bigcup I$. Of these, at most εn are in the exceptional set V_0 of the regular partition, and at most $2dn|D'|/m$ are in $\bigcup D'$. The remaining common neighbours must all lie in $\bigcup(V(R) - (I \cup D')$, and hence we have the inequality

$$\begin{aligned}\delta - \frac{2\delta - n}{16} - \varepsilon n - \frac{2dn|D'|}{m} &\leqslant (m - |I| - |D'|)\frac{n}{m} \\ &\overset{(v)}{\leqslant} n - (n - \delta - 11\eta n) - |D'|\frac{n}{m}\,.\end{aligned}$$

Simplifying this, we obtain

$$\frac{n - 2dn}{m}|D'| \leqslant 11\eta n + \varepsilon n + \frac{2\delta - n}{16}\,,$$

and by (3.16) and (3.17), we get $|D'| \leqslant (2\delta - n)m/(14n)$.

Now let D be the set of clusters $C \in V(R) - I$ such that either u_1v_1 or u_2v_2 has at most $2dn/m$ common neighbours in C. The same analysis holds for u_2v_2, so we obtain

$$|D| \leqslant \frac{(2\delta - n)m}{7n}\,. \tag{3.26}$$

Therefore, we conclude from (3.22) that $B_1 - D \neq \emptyset$. Take $X \in B_1 - D$ arbitrarily. We have

$$\begin{aligned}\deg(X, B_1) &\overset{(vi)}{\geqslant} \deg(X) - |I| \geqslant \delta' - |I| \overset{(3.21)}{\geqslant} \delta' - \Big(1 - \frac{\delta}{n} + d + \varepsilon\Big)m \\ &\overset{(3.20)}{\geqslant} \Big(\frac{\delta}{n} - d - \varepsilon\Big)m - \Big(1 - \frac{\delta}{n} + d + \varepsilon\Big)m \\ &\overset{(3.16),(3.17)}{\geqslant} \frac{1}{2}(2\delta - n)\frac{m}{n} \overset{(3.26)}{>} |D|\,.\end{aligned}$$

Thus there exists a cluster $Y \in \mathrm{N}(X) \cap (B_1 - D)$. Hence we have clusters $X, Y \in B_1 - D$ such that $XY \in E(R)$. Analogously, we can find clusters $X', Y' \in B_2 - D$ such that $X'Y' \in E(R)$.

Since ${\deg^{\min}}_R(B_1), {\deg^{\min}}_R(B_2) \geqslant \delta' - |I| \geqslant 2\delta' - m$, we can find greedily a matching M in $R[B_1 \cup B_2]$ with $\delta' - |I|$ edges. Since every cluster in I has at most $m - |I| - \delta'$ non-neighbours outside I, every cluster in I forms a triangle with at least $|M| - (m - |I| - \delta') = 2\delta' - m$ edges of M. In addition, by assumption (v), (3.16), and since $\delta < 2n/3$ we have $|I| > (\frac{1}{3} - 11\eta)m \geqslant \frac{1}{4}m$. Therefore we may choose greedily clusters in I to obtain a set T of at least

$$\min\left\{2\delta' - m, |I|\right\} \geqslant \min\left\{2\delta' - m, \frac{1}{4}m\right\}$$

vertex-disjoint triangles formed from edges of M and clusters of I. Let T_1 be the triangles of T contained in $B_1 \cup I$, and T_2 those contained in $B_2 \cup I$.

By Claim 3.9.1, since each triangle in T_1 contains an edge of B_1, all triangles in T_1 are in the same triangle component as the edge XY. Similarly all the triangles in T_2 are in the same triangle component as the edge $X'Y'$.

Noting that ε satisfies (3.17) and $n > N$ satisfies (3.18), we can apply Lemma 3.6 with $X_1 = X_2 = X$, $Y_1 = Y_2 = Y$ to find a squared path starting with u_1v_1 and finishing with u_2v_2 using the triangles T_1. Similarly, using Lemma 3.6 with $X_1 = X_2 = X'$, $Y_1 = Y_2 = Y'$ we find a squared path (intersecting the first only at u_1, v_1, u_2, and v_2) starting with u_2v_2 and finishing with u_1v_1 using the triangles T_2. Choosing appropriate lengths for these squared paths and concatenating them we get a squared cycle C_ℓ^2 in G, for any $36(m_{\mathrm{EL}}+2)^3 \leqslant \ell \leqslant 3(1-d)\min\{2\delta'-m, m/4\}n/m$. Applying Lemma 3.6 to the copy of K_4 in B_1 directly we obtain C_ℓ^2 for each $\ell \in [3, 3n/m] - \{5\}$. By (3.18) we have $3n/m > 36(m_{\mathrm{EL}} + 2)^3$, and by (3.7), (3.16), (3.17), and (3.19) we have $3(1-d)(2\delta' - m)n/m > \mathrm{sp}(n,\delta) \geqslant \mathrm{sc}(n,\delta)$ and $3(1-d)n/4 \geqslant 11n/20 > \mathrm{sp}(n,\delta) \geqslant \mathrm{sc}(n,\delta)$. It follows that G contains both $P^2_{\mathrm{sp}(n,\delta)}$ and C_ℓ^2 for each $\ell \in [3, \mathrm{sc}(n,\delta)] - \{5\}$ as required. $\square$

By (3.25), if there are two vertex disjoint edges in W, then we are done by Claim 3.9.2. Thus we assume in the following that no such two edges exist. This implies that there are two vertices in W which meet every edge in W. Since neither of these two vertices has more than ξn neighbours in $\bigcup I \subseteq W$, while $|I| > (\frac{1}{3} - 11\eta)m$ by (v) and because $\delta < 2n/3$, there is a vertex in W adjacent to no vertex of W. We conclude that

$$n - \delta - 11\eta n \leqslant |\bigcup I| \leqslant |W| \leqslant n - \delta. \tag{3.27}$$

Our next goal is to extract from each set $\bigcup B_i$ a large set A_i of vertices which are adjacent to almost all vertices in W and are such that $G[A_i]$ has minimum degree somewhat above $|A_i|/2$. Because at least $|W|\delta - 2|W|$ edges leave W, the total number of non-edges between W and $V(G) - W$ is at most

$$|W||V(G) - W| - |W|(\delta - 2) \leqslant (n - \delta)(\delta + 11\eta n - \delta + 2) \overset{(3.27)}{\leqslant} 11\eta n^2 + 2n\,.$$

In particular, by the definition of ξ, by (3.16) and (3.18),

$$\Big|\big\{v \in V(G) - W\colon\ \deg(v, W) < |W| - \xi^2 n\big\}\Big| \leqslant \xi^2 n\,. \tag{3.28}$$

In addition, by assumption (*vi*) we have $|B_i| \leqslant 19m(2\delta - n)/(10n)$, which together with $\delta \leqslant 2n/3$, (3.16), (3.17) and (3.24) implies

$$\Big|\bigcup B_i\Big| \leqslant \frac{19}{10}(2\delta - n) \leqslant \frac{19}{20}\delta < \delta - \xi n - (d + \varepsilon)n\,. \tag{3.29}$$

However, by assumption (*vi*) and the definition of an (ε, d)-regular partition, vertices in $\bigcup B_i$ send at most $(d + \varepsilon)n$ edges to $V(G) - \bigcup B_i - \bigcup I$. It follows from the definition of W that

$$\bigcup B_i \cap W = \emptyset \qquad \text{for all } i \in [k]\,.$$

Furthermore, (3.16), (3.17) and (3.24) imply that $v \in \bigcup B_i$ has at least

$$\delta - |W| - (d + \varepsilon)n \overset{(3.27)}{\geqslant} 2\delta - n - (d + \varepsilon)n \overset{(3.29)}{>} |\bigcup B_i|/2 + 32\xi^2 n \tag{3.30}$$

neighbours in $\bigcup B_i$.

Now, for each $i \in [k]$ we let A_i be the set of vertices in $\bigcup B_i$ which are adjacent to at least $|W| - \xi^2 n$ vertices of W. In the rest of this paragraph we determine some important properties of the sets A_i. By (3.28) we have

$$\Big|\bigcup_{i \in [k]} \big(\bigcup B_i\big) - A_i\Big| \leqslant \xi^2 n \qquad \text{for all } i \in [k]\,. \tag{3.31}$$

But the vertices which are neither in W nor any of the sets A_i must be either in the exceptional set V_0 or in $\bigcup B_i - A_i$ for some i. Hence we have

$$\Big|V_0 \cup \bigcup_{i \in [k]} \big(\bigcup B_i\big) - A_i\Big| \leqslant \varepsilon n + \xi^2 n < 2\xi^2 n\,. \tag{3.32}$$

Accordingly (3.30) implies that

$$\deg^{\min}(G[A_i]) \geqslant |A_i|/2 + 30\xi^2 n\,, \tag{3.33}$$

and since $|B_i| > \delta' - |I| \geqslant 2\delta' - m$ we have

$$|A_i| \geqslant |\bigcup B_i| - 2\xi^2 n \geqslant (1 - \varepsilon)\frac{n}{m}|B_i| - 2\xi^2 n \geqslant 2\delta - n - 3\xi^2 n \tag{3.34}$$

for each $i \in [k]$, where we used the definition of ξ, (3.16), 3.17, and (3.20) in the last inequality.

In the remainder of the proof we utilize the sets A_i in order to find the desired squared path and squared cycles. We start by showing that we obtain squared cycles on ℓ vertices for each $\ell \in [3, \frac{3}{2}|A_1|] - \{5\}$. To see this note first that by Lemma 3.13 (with $B = \emptyset$) we find in A_1 a copy of $C_{2\ell'}$ for each $2\ell' \in \left[4, \min\{|A_1|, 2\frac{n}{4}\}\right]$. By the definition of A_1 every quadruple of consecutive vertices on such a cycle has at least $|W| - 4\xi^2 n$ common neighbours in W, and by the definition of ξ, (3.16), (3.17), and (3.27) we have $|W| - 4\xi^2 n \geqslant n/4$. Hence we can apply Lemma 3.12 to G and W to square this cycle. This gives us squared cycles of lengths ℓ with

$$3 \leqslant \ell \leqslant \min\left\{\frac{3}{2}|A_1|, 3\frac{n}{4}\right\} \overset{(3.19)}{=} \frac{3}{2}|A_1|$$

such that ℓ is divisible by three, but not of lengths not divisible by three.

If we seek a squared cycle $C^2_{3\ell'+1}$ or $C^2_{3\ell'+2}$ (with $3\ell' + 2 \neq 5$) then we need to perform a process which we will call *parity correction* and which we explain in the following two paragraphs. We shall use this parity correction process also in all remaining steps of the proof to obtain squared cycles of lengths not divisible by 3.

For obtaining a squared cycle of length $3\ell' + 1$ we proceed as follows. We pick (using Turán's theorem) a triangle abc in A_1 and *clone* the vertex b, i.e., we insert a dummy vertex b' into G with the same adjacencies as b. Then we apply Lemma 3.13 to $A_1 - \{b\}$ to find a path $P = (a, p_2, p_3, \dots, p_{2\ell'-1}, c)$ on $2\ell'$ vertices whose end-vertices are a and c. Finally we apply Lemma 3.12 to the path bPb', taking $Q_1 = (b, a)$, $Q_2 = (b, a, p_2, p_3)$ as the first quadruple and thereafter every other set of four consecutive vertices on P, finishing with $(p_{2\ell'-2}, p_{2\ell'-1}, c, b')$. This yields a squared path $(q_1, b, a, \dots, c, b')$ on $3(\ell' + 1)$ vertices, which gives a squared cycle $(b, a, \dots, c)$ in G (without q_1 and the clone vertex b') on $3\ell' + 1$ vertices as required.

If we seek a squared cycle of length $3\ell'+2$ with $\ell' > 1$ on the other hand, then we perform a similar process, except that we identify not one triangle in A_1 but two triangles abc, xyz connected with an edge cx (which we obtain by the Erdős-Stone theorem). We apply Lemma 3.13 to find a path $P = (a, \dots, z)$ in $A_1 - \{b, c, y, z\}$ on $2\ell'$ vertices. We then apply Lemma 3.12 once to the path bPy and once to (b, c, x, y). Omitting the first vertex on each of the resulting squared paths and concatenating, we get a squared cycle $C^2_{3\ell'+2}$.

Hence we do indeed obtain squared cycles C^2_ℓ for all $\ell \in [3, \frac{3}{2}|A_1|] - \{5\}$. It remains to show that we can also find C^2_ℓ for all ℓ with $\frac{3}{2}|A_1| \leqslant \ell \leqslant \mathrm{sc}(n, \delta)$ and that we can find $P^2_{\mathrm{sp}(n,\delta)}$.

For this purpose, we first re-incorporate the vertices that are neither in W nor in any of the sets A_i by examining in which of the A_i they have many neighbours. More precisely, for each $i \in [k]$, we let X_i be A_i together with all vertices in $V(G)-W$ which are adjacent to at least $30\xi^2 n$ vertices of A_i. Because every vertex in $V(G)-W$ has at least $\delta - |W|$ neighbours outside W, by (3.27) every vertex in $G - W$ is in X_i for at least one i. Moreover, by the definition of an (ε, d)-regular partition, assumption (vi) and since $A_j \subseteq \bigcup B_j$, we have for all $j \in [k]$ with $j \neq i$ that

$$A_j \cap X_i = \emptyset\,. \tag{3.35}$$

Hence it follows from (3.32) that

$$|X_i| < |A_i| + 2\xi^2 n \qquad \text{and} \qquad |X_1 - A_1| \leqslant 2\xi^2 n\,. \tag{3.36}$$

We finish the proof by distinguishing three cases.

Case 1: $|X_i \cap X_j| \geqslant 2$ for some $i \neq j$. Let v_1 and v_2 be distinct vertices of $X_i \cap X_j$. Let u_1 and u_2 be distinct neighbours in A_i of v_1 and v_2 respectively, and similarly y_1 and y_2 in A_j. Applying Lemma 3.13 to A_i we can find a path from u_1 to u_2 of length ℓ' for any $4 \leqslant \ell' \leqslant |A_i| - 2$. We can find a similar path in A_j from y_1 to y_2. Concatenating these paths with v_1 and v_2 we can find a $2\ell'$-vertex cycle $T_{2\ell'}$ in $X_1 \cup X_2$ for any $10 \leqslant 2\ell' \leqslant |A_i| + |A_j| - 2$. There are no quadruples of consecutive vertices on $T_{2\ell'}$ using both v_1 and v_2. The four quadruples that use either v_1 or v_2 each have at least $(\xi - 3\xi^2)n > 100k$ common neighbours in W, where the inequality follows from (3.18), (3.24), from

$$k \leqslant \nu^{-1}\,, \tag{3.37}$$

and from $\xi - 3\xi^2 > 0$. All other quadruples have at least $|W| - 4\xi^2 n$ common neighbours in W. So applying Lemma 3.12 we obtain a squared cycle on $3\ell'$ vertices. Again it is possible to perform parity corrections (prior to applying Lemma 3.13) so that in this case we have $C_\ell^2 \subseteq G$ for every $\ell \in [3, \frac{3}{2}(|A_i| + |A_j| - 10)] - \{5\}$. By (3.34), we have $\mathrm{sc}(n, \delta) \leqslant \mathrm{sp}(n, \delta) < \frac{3}{2}(|A_i| + |A_j| - 10)$.

Case 2: for some i every vertex of A_i is adjacent to at least one vertex outside $X_i \cup W$. Since

$$|A_i| \overset{(3.31)}{\geqslant} \Big|\bigcup B_i\Big| - \xi^2 n \overset{(3.22)}{\geqslant} \frac{4}{3}\nu(1-\varepsilon)n - \xi^2 n \overset{(3.24)}{\geqslant} 13\xi n \overset{(3.24),(3.37)}{>} 31k\xi^2 n$$

we can certainly find $j \neq i$ such that there are $31\xi^2 n$ vertices in A_i all adjacent to vertices of $X_j - X_i$. Since no vertex of $X_j - X_i$ is adjacent to $30\xi^2 n$ vertices of A_i

(by definition of X_i), we find two disjoint edges u_1v_1 and u_2v_2 from $u_1, u_2 \in A_i$ to $v_1, v_2 \in X_j$. Choosing distinct neighbours y_1 of v_1 and y_2 of v_2 in A_j and applying the same reasoning as in the previous case we are done.

Case 3: for each $i \neq j$ we have $|X_i \cap X_j| \leqslant 1$, and for each i some vertex in A_i is adjacent only to vertices in $W \cup X_i$. Thus $|X_i| \geqslant \delta - |W| + 1$ for each i. We now first focus on finding a squared path on $\mathrm{sp}(n, \delta)$ vertices in G, and then turn to the squared cycles which will complete the proof. If for some $i \neq j$ we have $|X_i \cap X_j| = 1$ then we obtain a squared path of the desired length as in Case 1. There we required two vertices in $X_i \cap X_j$ to obtain a squared cycle (which must return to its start), but one vertex suffices for a squared path to cross from X_i to X_j.

So, assume that the sets X_i are all disjoint. It follows that $k \leqslant (n - |W|)/(\delta - |W| + 1)$. Since $|W| \leqslant n - \delta$ by (3.27), this implies

$$k \leqslant \frac{n - (n - \delta)}{\delta - (n - \delta) + 1} = \frac{\delta}{2\delta - n + 1}\,.$$

Now if $k \geqslant r_p(n, \delta) + 1$ then we would have

$$r_p(n, \delta) + 1 \leqslant k \leqslant \frac{\delta}{2\delta - n + 1}\,,$$

and so

$$r_p(n, \delta) \leqslant \frac{n - \delta - 1}{2\delta - n + 1}\,,$$

but by (3.5) we have $r_p(n, \delta) \geqslant \frac{n-\delta}{2\delta-n+1}$, so

$$k \leqslant r_p(n, \delta)\,.$$

Thus the largest of the sets X_i, say X_1, has at least

$$|X_1| \geqslant \frac{n - |W|}{k} \overset{(3.27)}{\geqslant} \frac{\delta}{k} \geqslant \frac{\delta}{r_p(n, \delta)} \tag{3.38}$$

vertices.

We now want to apply Lemma 3.13 with $H = G[X_1]$ and 'bad' vertices $B = X_1 - A_1$. Note that by (3.36) there are at most $2\xi^2 n$ vertices in $B = X_1 - A_1$, and so we have

$$|B| \overset{(3.36)}{\leqslant} 2\xi^2 n \overset{(3.24)}{\leqslant} \frac{\nu\delta}{100} \overset{(3.37)}{\leqslant} \frac{\delta}{100k} \leqslant \frac{|H|}{100}\,.$$

Moreover, $\deg^{\min}(H) = \deg^{\min}(G[X_1]) \geqslant 30\xi^2 n$ by definition of X_1, and therefore every vertex in B has at least $30\xi^2 n \geqslant 9 \cdot 2\xi^2 n \geqslant 9|B|$ neighbours in H. In addition, vertices v in $H - B \subseteq A_1$ satisfy

$$\begin{aligned} \deg(v, X_1) &\overset{(3.33)}{\geqslant} \frac{|A_1|}{2} + 30\xi^2 n \overset{(3.36)}{>} \frac{|X_1|}{2} + 25\xi^2 n \\ &= \frac{|H|}{2} + 25\xi^2 n \overset{(3.18)}{\geqslant} \frac{|H|}{2} + 9|B| + 10\,. \end{aligned}$$

Hence we can indeed apply Lemma 3.13, to obtain a path T covering $\min\{X_1, n/2\}$ vertices on which every quadruple of consecutive vertices contains at most one 'bad' vertex. Finally we want to apply Lemma 3.12 to the graph $G[X_1 \cup W]$ and the cycle T with the following ordering σ of the quadruples of consecutive vertices in T: σ is such that all quadruples containing vertices from B come first, followed (by an arbitrary ordering of) all other quadruples. There are at most $2 \cdot 2\xi^2 n$ quadruples containing vertices from $B = X_1 - A_1$, and by the definition of A_1 and of W, each of them has at least $(\xi - 3\xi^2)n \geqslant \xi^2 n$ common neighbours in W. All remaining quadruples have, by the definition of A_1, by (3.27) and since $\delta \leqslant 2n/3$, at least $|W| - 4\xi^2 n \geqslant \frac{n}{4} \geqslant \frac{1}{2}\min\{|X_1|, \frac{n}{2}\}$ common neighbours in W. Hence, we can indeed apply Lemma 3.12 to obtain a squared path on at least $\frac{3}{2}\min\{|X_1|, n/2\} \geqslant \mathrm{sp}(n, \delta)$ vertices, where the inequality follows from the definition of $\mathrm{sp}(n, \delta)$, from (3.19), and from (3.38).

At last, we show that we can find in G the desired long squared cycles in Case 3. Assume first that there is a cycle of sets (relabelling the indices if necessary) $X_1, X_2, \ldots, X_s$ for some $3 \leqslant s \leqslant k$ such that $X_i \cap X_{i+1 \bmod s} = \{v_i\}$ for each i, and the v_i are all distinct, then for each i we may choose neighbours $u_i \in A_i$ and y_i in $A_{i+1 \bmod s}$ of v_i, and we may insist that all these $3s$ vertices are distinct. Similarly as before we can apply Lemma 3.13 to each $G[A_i]$ in turn and concatenate the resulting paths, in order to find a cycle $T_{2\ell'}$ for every $8s \leqslant 2\ell' \leqslant |A_1| + |A_2|$ on which there are no quadruples using more than one vertex outside $\bigcup_i A_i$. Again (checking the conditions similarly as before) we may apply Lemma 3.12 to $T_{2\ell'}$ to obtain a squared cycle on 3ℓ vertices. Finally by performing parity corrections we obtain C^2_ℓ for every $\ell \in [3, \frac{3}{2}(|A_1| + |A_2|)] - \{5\}$.

If there exists no such cycle of sets, then $\sum_{i=1}^{k} |X_i| \leqslant n - |W| + k - 1$. Since we have also $|X_i| \geqslant \delta - |W| + 1$ for each i and $|W| \leqslant n - \delta$, it follows from the definition of $r_c(n, \delta)$ (by establishing a relation similar to (3.5)) that $k \leqslant r_c(n, \delta)$, and by averaging, that the largest of the sets X_i, say X_1, contains at least $2\,\mathrm{sc}(n, \delta)/3$ vertices. As before, we can apply Lemma 3.13 to X_1 to obtain a cycle $T_{2\ell'}$ for each $4 \leqslant 2\ell' \leqslant |X_1|$ on which the 'bad' vertices from $B = X_1 - A_1$ are separated, and apply Lemma 3.12 to it to obtain a squared cycle $C^2_{3\ell'}$ for each $6 \leqslant 3\ell' \leqslant \mathrm{sc}(n, \delta)$

as required. Again the parity correction procedure is applicable, so we get C_ℓ^2 for every $\ell \in [3, \mathrm{sc}(n,\delta)] - \{5\}$. $\square$

3.5 Concluding remarks

3.5.1 The proof of Theorem 3.4

Our results were most difficult to prove for $\delta \approx 4n/7$. This is somewhat surprising given the experience from the partial and perfect packing results of Komlós [59] and Kühn and Osthus [71]. In the setting of these results it becomes steadily more difficult to prove packing results as the minimum degree of the graph (and hence the required size of a packing) increases, with perfect packings as the most difficult case. Yet in our setting it is relatively easy to prove our results when the minimum degree condition is large. This difference occurs because we have to embed triangle-connected graphs; as the minimum degree increases the possibilities for bad behaviour when forming triangle-connections are reduced.

This is paralleled by the behaviour of K_4-free graphs: For any minimum degree $\deg^{\min}(G) > 2v(G)/3$ the graph G is not K_4-free. However, if $\deg^{\min}(G) > 5v(G)/8$ then by the Andrásfai-Erdős-Sós theorem [11] the K_4-free graph G is forced to be tripartite, while for smaller values of $\deg^{\min}(G)$ there exist more possibilities.

3.5.2 Extremal graphs

It is straightforward to check (from our proofs) that up to some trivial modifications the graphs $G_p(n,\delta)$ and $G_c(n,\delta)$ are the only extremal graphs. We believe that the graph $G_p(n,\delta)$ remains extremal for squared paths even when δ is not bounded away from $n/2$, although as noted in Section 3.1 the same is not true for $G_c(n,\delta)$ and squared cycles.

However it is not the case that the only extremal graph excluding some C_ℓ^2 of chromatic number four is $K_{n-\delta,n-\delta,2\delta-n}$ (cf. (*ii*) of our main theorem, Theorem 3.4). Let us briefly explain this. Suppose ℓ is not divisible by three. Since C_ℓ^2 has no independent set on more than $\lfloor \ell/3 \rfloor$ vertices, whenever we remove an independent set from C_ℓ^2 we must leave some three consecutive vertices, which form a triangle. Now suppose that we can find a graph H on δ vertices with minimum degree $2\delta - n$ which is both triangle-free and contains no even cycle on more than $2(2\delta - n)$ vertices. Then the graph G obtained by adding an independent set of size $n - \delta$ to H, all of whose vertices are adjacent to all of H, contains no squared cycle of length indivisible by three and no squared cycle with more than $3(2\delta - n)$ vertices.

To mention one possible H, take $\delta := \frac{6n}{11}$ and let H be obtained as follows. We take the disjoint union of three copies of $K_{n/11,n/11}$ and fix a bipartition. Now we add three vertex disjoint edges within one of the two partition classes, one between each copy of $K_{n/11,n/11}$. The resulting triangle-free graph has no even cycle leaving any copy of $K_{n/11,n/11}$. Hence all even cycles have at most $\frac{2n}{11}$ vertices. However, it has odd cycles of up to $\frac{6n}{11} - 3$ vertices.

3.5.3 Long squared cycles

Theorem 3.5 (ii) states that if any of various odd cycles are excluded from G we are guaranteed even cycles of every length up to $2\mathrm{deg}^{\min}(G)$, whereas the equivalent statement in our Theorem 3.4 contains an error term. We believe this error term can be removed, but at the cost of significantly more technical work, requiring both a new version of the stability lemma and new extremal results.

3.5.4 Higher powers of paths and cycles

We note that Theorem 3.2 has a natural generalisation to higher powers of cycles, the so called Pósa-Seymour Conjecture. This conjecture was proved for all sufficiently large n by Komlós, Sárközy and Szemerédi [63]. We conjecture a natural generalisation of Theorem 3.4 for higher powers of paths and cycles.

Given k, n and δ, we construct an n-vertex graph $G_p^{(k)}(n,\delta)$ by partitioning the vertices into an 'interior' set of $\ell := (k-1)(n-\delta)$ vertices upon which we place a complete balanced $k-1$-partite graph, and an 'exterior' set of $n-\ell$ vertices upon which we place a disjoint union of $\lfloor (n-\ell)/(\delta-\ell+1) \rfloor$ almost-equal cliques. We then join every 'interior' vertex to every 'exterior' vertex. We construct $G_c^{(k)}(n,\delta)$ similarly, permitting the cliques in the 'exterior' vertices to overlap in cut-vertices of the 'exterior' set if this reduces the size of the largest clique while preserving the minimum degree δ.

Conjecture 3.14. *Given $\nu > 0$ and k there exists n_0 such that whenever $n \geqslant n_0$ and G is an n-vertex graph with $\mathrm{deg}^{\min}(G) = \delta > \frac{k-1}{k}n + \nu n$, the following hold.*

(i) *If $P_\ell^k \subseteq G_p^{(k)}(n,\delta)$ then $P_\ell^k \subseteq G$.*

(ii) *If $C_{(k+1)\ell}^k \subseteq G_c^{(k)}(n,\delta)$ for some integer ℓ, then $C_{(k+1)\ell}^k \subseteq G$.*

(iii) *If $C_\ell^k \subseteq G_c^{(k)}(n,\delta)$ with $\chi(C_\ell^k) = k+2$ and $C_\ell^k \nsubseteq G$ for some integer ℓ, then $C_{(k+1)\ell}^k \subseteq G$ for each integer $\ell < k\delta - (k-1)n - \nu n$.*

It seems likely that again the νn error term in the last statement is not required, but again (at least for powers of cycles) it is required in the minimum degree condition.

Chapter 4

Turánnical hypergraphs

4.1 Introduction

There are many extensions of Turán's Theorem (Theorem 1.9); we have presented some of these in Section 1.5.1. These extensions, however, do not deviate from the original result as far as the following aspect is concerned. The restrictions they impose on the class of objects under study are *global* and *dense*. More concretely, they require for *every* k-tuple of vertices that these vertices do not host a copy of a given graph K on k vertices. In this chapter we are interested in the question of how weakening these restrictions to less global or sparser ones (that is, forbidding K-copies only for certain k-tuples but not all) can influence the conclusion of the original Turán theorem.

A natural way of formalising the relaxation from Turán's theorem is to introduce a hypergraph which contains a hyperedge for every restriction and then ask for the maximal number k of edges in a graph respecting these restrictions. The following definition makes this precise. We shall distinguish between the case when k is still the Turán number and when it is bigger by a certain percentage.

Definition 4.1 (Turánnical)**.** *Let $r \geqslant 3$ be an integer. Let $\mathcal{F} = (V, \mathcal{E})$ be an n-vertex, r-uniform hypergraph with vertex set V, which we also occasionally call* restriction hypergraph*. The hypergraph $\mathcal{F}$* detects *a graph $G = (V, E)$ if some $F \in \mathcal{E}$ induces a copy of K_r in G. We say that $\mathcal{F}$ is* exactly Turánnical *or simply* Turánnical*, if for all graphs $G = (V, E)$ with $e(G) > t_r(n)$ the hypergraph $\mathcal{F}$ detects G. In addition, $\mathcal{F}$ is* ε-approximately Turánnical *or simply* ε-Turánnical *if for all graphs $G = (V, E)$ with $e(G) > (1 + \varepsilon)t_r(n)$ the hypergraph $\mathcal{F}$ detects G.*

In other words, a restriction hypergraph is Turánnical if it detects all graphs whose density is large enough that *one* copy of K_r is forced to exist, and it is approximately Turánnical if it detects all graphs whose density forces a *positive density* of copies of K_r to exist (cf. the so-called super-saturation theorem, Theorem 4.9, by Erdős and Simonovits [38]).

In this language Turán's theorem states that the complete r-uniform hypergraph is Turánnical.

A natural question is whether the dense complete r-uniform restriction hypergraph from Turán's theorem may be replaced by a much sparser one. Here, hypergraphs formed by *random restrictions* might appear promising candidates. And in fact, we will show that $\mathcal{R}^{(r)}(n,p)$ for appropriate values of $p = p_n$ produces the Turánnical hypergraphs and ε-Turánnical hypergraphs with the fewest number of hyperedges, up to constant factors (compare Proposition 4.2 with Theorems 4.3 and 4.4). In addition, building on the aforementioned work of Schacht [89] we obtain a corresponding result for the random graphs version of Turán's theorem (see Theorem 4.8).

Before we state and explain these results in detail in the following section, let us remark that the observed behaviour concerning the *evolution* of $\mathcal{R}^{(r)}(n,p)$ as we decrease the density of the random restrictions is as follows. When p decreases from 1 to 0, then $\mathcal{R}^{(r)}(n,p)$ stays (asymptotically almost surely) Turánnical for a long time, until $p_n \sim n^{3-r}$. Then, between $p_n \sim n^{3-r}$ and $p_n \sim n^{2-r}$ the hypergraph $\mathcal{R}^{(r)}(n,p)$ is ε-Turánnical for *arbitrarily* small (but fixed) $\varepsilon > 0$, and for even smaller p_n the hypergraph $\mathcal{R}^{(r)}(n,p)$ fails to be ε-Turánnical for *any* non-trivial ε. As we shall see later, this sudden change of behaviour is caused by the supersaturation property of graphs (cf. Theorem 4.9). Put differently, there is a qualitative difference between random restriction sets detecting graphs with enough edges to force a single K_r to exist and restriction sets detecting graphs with enough edges to force a positive K_r-density, but the value of this density is not of big influence.

The remainder of this chapter is organised as follows. In Section 4.2 we state our results. In Section 4.3 we then prove some general deterministic lower bounds on the number of hyperedges in Turánnical and approximately Turánnical hypergraphs. The proofs for our results concerning random restrictions for general graphs are contained in Sections 4.4 and 4.5 and those concerning random restrictions for random graphs in Section 4.6. In Section 4.7 we discuss the so-called sharp thresholds in our setting. In Section 4.8, finally, we explain how the concept of random restrictions generalises to other problems besides Turán's theorem. We provide an outlook on which phenomena may be observed with regard to questions of this type and the corresponding evolution of random restrictions, and how they may differ from the Turán case treated in this chapter.

4.2 Results

In this section we give our results concerning Turánnical hypergraphs which are included in this chapter. Note that Chapter 5 deals with Turánnical hypergraphs as well, this time from a deterministic (i.e., constructive) point of view.

4.2.1 Sparse restrictions

Next we consider sparser hypergraphs. An easy counting argument (which we defer to Section 4.3) gives the following lower bounds for the density of Turánnical and approximately Turánnical hypergraphs.

Proposition 4.2. *Let $r \geqslant 3$ and $n \geqslant 5$ be integers, let ε be a real with $0 < \varepsilon \leqslant 1/(2r)$, and let $\mathcal{F} = ([n], \mathcal{E})$ be an r-uniform hypergraph.*

(*a*) *If $|\mathcal{E}| < \frac{n(n-1)(n-2)}{r(r-1)^2(r-2)}$ then $\mathcal{F}$ is not Turánnical.*

(*b*) *If $|\mathcal{E}| \leqslant (1 - r\varepsilon)\frac{1}{4r}n^2$, then $\mathcal{F}$ is not ε-Turánnical.*

These density bounds are sharp up to constant factors. In fact, in random r-uniform hypergraphs their magnitudes provide thresholds for being Turánnical and approximately Turánnical, respectively, as the following two results show. We first state the result concerning the threshold for being approximately Turánnical.

Theorem 4.3. *For every integer $r \geqslant 3$ and every $0 < \varepsilon \leqslant 1/(2r)$ there are $c = c(r, \varepsilon) > 0$ and $C = C(r, \varepsilon) > 0$ such that for any sequence $p = p_n$ of probabilities*

$$\lim_{n\to\infty} \mathbb{P}\big(\mathcal{R}^{(r)}(n,p) \text{ is } \varepsilon\text{-Turánnical}\big) = \begin{cases} 0, & \text{if } p_n \leqslant cn^{2-r} \text{ for all } n \in \mathbb{N}, \\ 1, & \text{if } p_n \geqslant Cn^{2-r} \text{ for all } n \in \mathbb{N}. \end{cases}$$

Clearly, a random r-uniform hypergraph with hyperedge probability $p = cn^{2-r}$ asymptotically almost surely (a.a.s.) has less than $\frac{3c}{r!}\binom{n}{2}$ hyperedges. Thus part (*b*) of Proposition 4.2 does indeed imply the 0-statement in Theorem 4.3. A proof of the 1-statement is provided in Section 4.4.

Using part (*a*) of Proposition 4.2, a similar calculation shows that a random r-uniform hypergraph with hyperedge probability $p = cn^{3-r}$ with $c > 0$ sufficiently small is asymptotically almost surely not Turánnical. The corresponding 1-statement is given in the following theorem. For the case $r = 3$ the threshold probability is a constant, which we determine precisely.

Theorem 4.4. *For $r = 3$ and p constant we have*

$$\lim_{n\to\infty} \mathbb{P}\big(\mathcal{R}^{(3)}(n,p) \text{ is Turánnical}\big) = \begin{cases} 0, & \text{if } p \leqslant 1/2, \\ 1, & \text{if } p > 1/2. \end{cases}$$

For every integer $r > 3$ there are $c = c(r) > 0$ and $C = C(r) > 0$ such that for any sequence $p = p_n$ of probabilities

$$\lim_{n\to\infty} \mathbb{P}\big(\mathcal{R}^{(r)}(n,p) \text{ is Turánnical}\big) = \begin{cases} 0, & \text{if } p_n \leqslant cn^{3-r} \text{ for all } n \in \mathbb{N}, \\ 1, & \text{if } p_n \geqslant Cn^{3-r} \text{ for all } n \in \mathbb{N}. \end{cases}$$

This theorem is proven in Section 4.5. As a side remark we mention that, for its proof we shall need a structural lemma (Lemma 4.14) which classifies graphs with at least $t_r(n)$ edges and has the following direct consequence which might be of independent interest.

Lemma 4.5. *For every integer $r \geqslant 3$ and real $\tilde{\varepsilon} > 0$ there exists $\delta > 0$ such that for all n-vertex graphs G with $e(G) > t_r(n)$ one of the the following is true.*

(*i*) *Some vertex in G is contained in at least δn^{r-1} copies of K_r.*

(*ii*) *Some edge in G is contained in at least $(1-\tilde{\varepsilon})(n/(r-1))^{r-2}$ copies of K_r.*

An edge contained in b triangles is sometimes called a *book of size* b. Lemma 4.5 in the case $r = 3$ thus states that if $e(G) > t_3(n)$ and no vertex of G is contained in many K_3-copies, then G contains a book of size almost $\frac{n}{2}$. We remark that Mubayi [80] recently showed that for every $\alpha \in (\frac{1}{2}, 1)$, if G has $e(G) > t_3(n)$ and less than $\alpha(1-\alpha)n^2/4 - o(n^2)$ triangles, then G contains a book of size at least $\alpha n/2$. This result is harder, but does not imply Lemma 4.5.

Finally, it follows from Friedgut's celebrated result [43] that the property of being Turánnical considered in Theorem 4.4 has a sharp threshold. This is detailed in Section 4.7.

4.2.2 Sparse restrictions for sparse random graphs

In the previous subsection we examined the effect of random restrictions on Turán's theorem. A version of Turán's theorem for the Erdős-Rényi random graph $G(n,q)$ was recently proved by Schacht [89] and independently by Conlon and Gowers [28]. To understand this theorem, one should view Turán's theorem as the statement that the fraction of the edges one must delete from the complete graph K_n to remove all copies of K_r is approximately $\frac{1}{r-1}$. One can replace K_n with any graph G, and ask which graphs G have the property that deletion of a fraction of approximately $\frac{1}{r-1}$ of the edges is necessary to remove all copies of K_r.

Theorem 4.6 (Schacht [89], Conlon & Gowers [28])**.** *Given $\varepsilon > 0$ and r there exists a constant C such that the following is true. For $q \geqslant Cn^{-2/(r+1)}$, a.a.s. $G = G(n,q)$ has the property that every subgraph of G with at least $(1+\varepsilon)\frac{r-2}{r-1}e(G)$ edges contains a copy of K_r.*

Prior to the recent breakthroughs [89] and [28], Theorem 4.6 was known for $r = 3, 4, 5$ (see [42, 57, 47], respectively).

This result is best possible in the sense that it ceases to be true for values of q growing more slowly than $n^{-2/(r+1)}$. Moreover, ε cannot be replaced by 0.

Again, the restriction set in Theorem 4.6 is the complete r-partite hypergraph (sequence). So, extending Theorem 4.3, we would like to analyse what happens when this is replaced by a sparser set of random restrictions and investigate the influence of the two independent probability parameters (coming from the random restrictions and the random graph) on each other. Thus, we will be dealing with two random objects: namely a random r-uniform hypergraph $\mathcal{R}^{(r)}(n,p)$ and a random graph $G(n,q)$, picked at the same time. Furthermore, since we wish to prove asymptotically almost sure results, we need to refer not to single n-vertex hypergraphs but to sequences of hypergraphs and graphs.

Before we can formulate our result, we first need to generalise the concept of being Turánnical or approximately Turánnical from (copies of K_r in) the complete graph K_n to arbitrary graphs G. Observe that, in Theorem 4.6 we are interested in graphs G for which any subgraph with at least $(1+\varepsilon)\frac{r-2}{r-1}\cdot e(G)$ edges contains a copy of K_r. Hence it is natural to say that the r-uniform hypergraph $\mathcal{F}$ is ε-Turánnical for G when $\mathcal{F}$ detects every such subgraph.

For finding a similarly suitable definition of Turánnical hypergraphs for G we need some additional observations. Recall that ε cannot be 0 in Theorem 4.6. In other words an *exact* version of Turán's theorem for random graphs cannot be expressed in terms of the number of its edges. Instead it has to utilise the structure provided by Turán's theorem: the maximal K_r-free subgraph of $G = G(n,q)$ should have exactly as many edges as the biggest $(r-1)$-partite subgraph of G. Accordingly, we will call a hypergraph Turánnical for G if it detects all subgraphs with more edges. The following definition summarises this.

Definition 4.7 (Turánnical for G)**.** *Let $r \geqslant 3$ be an integer, G an n-vertex graph, and $\mathcal{F}$ an r-uniform hypergraph on the same vertex set. Then we call $\mathcal{F}$* exactly

Turánnical for G *when the following holds. Every subgraph of G with more edges than are contained in a maximum $(r-1)$-partition of G has a copy of K_r induced by an edge of $\mathcal{F}$. We say that $\mathcal{F}$ is* ε-approximately Turánnical for G, *or simply* ε-Turánnical for G, *if every subgraph of G with more than $(1+\varepsilon)\frac{r-2}{r-1}e(G)$ edges has a copy of K_r induced by an edge of $\mathcal{F}$.*

In this language, Theorem 4.6 becomes the statement that, given r and $\varepsilon > 0$, there exists C such that the complete r-uniform hypergraph is a.a.s. ε-Turánnical for $G(n,q)$, whenever $q \geqslant Cn^{-2/(r+1)}$. Moreover, according to a result of Brightwell, Panagiotou and Steger [22], for every r there exists $\mu > 0$ such that the complete r-uniform hypergraph is a.a.s. exactly Turánnical for $G(n,q)$ whenever $q > n^{-\mu}$.[1]

In our last theorem we determine the relationship between r, $\varepsilon > 0$, p and q such that the random r-uniform hypergraph $\mathcal{R}^{(r)}(n,p)$ is a.a.s. ε-Turánnical for $G(n,q)$. Not surprisingly, a suitable combination of the two threshold probabilities from Theorem 4.3 and Theorem 4.6 determines the threshold in this case.

Theorem 4.8. *Given $r \in \mathbb{N}$, $r \geqslant 3$, and $\varepsilon \in (0, 1/(r-2))$, there exist $c = c(r,\varepsilon) > 0$ and $C = C(r,\varepsilon) > 0$ such that for any pair of sequences $p = p_n$ and $q = q_n$ of probabilities and for $\vartheta_q(n) := (nq^{(r+1)/2})^{2-r}$ we have*

$$\lim_{n\to\infty} \mathbb{P}\big(\mathcal{R}^{(r)}(n,p) \text{ is } \varepsilon\text{-Turánnical for } G(n,q)\,\big) = \begin{cases} 0, & \text{if } p_n \leqslant c\vartheta_q(n) \text{ for all } n \in \mathbb{N}, \\ 1, & \text{if } p_n \geqslant C\vartheta_q(n) \text{ for all } n \in \mathbb{N}. \end{cases}$$

This theorem states that for a fixed q_n the threshold probability for $\mathcal{R}^{(r)}(n,p)$ to be ε-Turánnical for $G(n,q)$ is $\vartheta_q(n)$. Equivalently, if instead we fix the hyperedge probability p_n then $\vartheta_p(n) := (np^{1/(r-2)})^{-2/(r+1)}$ is the threshold probability for $G(n,q)$ such that $\mathcal{R}^{(r)}(n,p)$ is ε-Turánnical for $G(n,q)$. In particular, $\vartheta_q(n)$ is constant when q_n is the threshold probability from Theorem 4.3 and $\vartheta_p(n)$ is constant when p_n is the threshold probability from Theorem 4.6.

We note that the requirement $\varepsilon < 1/(r-2)$ in Theorem 4.8 is necessary for the 0-statement. Indeed, if $\varepsilon > 1/(r-2)$ then $(1+\varepsilon)\frac{r-2}{r-1}e(G) > e(G)$. Therefore the premise in Definition 4.7 is never met, and consequently every hypergraph is ε-Turánnical.

In order to establish Theorem 4.8 we employ in Section 4.6 Schacht's machinery from [89]. However we need to modify this machinery to allow working with

[1] However, Brightwell, Panagiotou and Steger do not believe that their result is best possible: for example, for $r = 3$ their proof works for $\mu = 1/250$, but they suggest the result might hold for any $\mu < 1/2$.

two sources of randomness: graphs $G(n,q)$ and hypergraphs $\mathcal{R}^{(r)}(n,p)$. We believe that this might prove useful in the future.

We believe that a similar result as Theorem 4.8 should be true if ε-Turánnical is replaced by exactly Turánnical in this theorem. More precisely, we think that for $r \geqslant 3$ the hypergraph $\mathcal{R}^{(r)}(n,p)$ is a.a.s. exactly Turánnical for $G(n,q)$, if p and q are both sufficiently large. For obtaining a result of this type, possibly a modification of the methods used in [22] may be of assistance.

4.3 The proof of Proposition 4.2

In this section we provide the proof of Proposition 4.2.

Let $\mathcal{F} = (V, \mathcal{E})$ be an r-uniform hypergraph and X be a subset of its vertices of size $|X| = s < r$. The *link hypergraph* $\mathrm{Link}_{\mathcal{F}}(X) = (V, \mathcal{E}')$ of X is the $(r-s)$-uniform hypergraph with hyperedges $\mathcal{E}' = \{Y \in \binom{V}{r-s} : Y \cup X \in \mathcal{E}\}$. If $X = \{x_1, \ldots, x_s\}$ we also write $\mathrm{Link}_{\mathcal{F}}(x_1, \ldots, x_s)$ for $\mathrm{Link}_{\mathcal{F}}(X)$. When the underlying hypergraph $\mathcal{F}$ is clear from the context we write $\mathrm{Link}(X)$ instead of $\mathrm{Link}_{\mathcal{F}}(X)$.

Proof of Proposition 4.2. Let the r-uniform hypergraph $\mathcal{F} = ([n], \mathcal{E})$ be given. We start with the proof of (a) and first consider the case $r > 3$. We have

$$\sum_{\{u,v\} \in \binom{[n]}{2}} e\big(\mathrm{Link}(u,v)\big) = \binom{r}{2} |\mathcal{E}| < \binom{r}{2} \frac{n(n-1)(n-2)}{r(r-1)^2(r-2)} \leqslant \frac{\binom{n}{2} n}{(r-2)(r-1)},$$

Accordingly there are two vertices $u, v \in [n]$ such that $(r-2)e\big(\mathrm{Link}(u,v)\big) \leqslant n/(r-1)$. Let

$$L := \Big\{ w \in [n] : w \in Y \text{ for some } Y \in E\big(\mathrm{Link}(u,v)\big) \Big\}$$

be the set of vertices covered by the hyperedges of $\mathrm{Link}(u,v)$. Because $\mathrm{Link}(u,v)$ is an $(r-2)$-uniform hypergraph, it follows from the choice of u and v that $|L| \leqslant n/(r-1)$. Now suppose the graph $G = ([n], E)$ is a copy of the $(r-1)$-partite Turán graph $\mathrm{T}_r(n)$ such that u and v are in the same partition class of $\mathrm{T}_r(n)$ and L is entirely contained in another partition class. The graph G exists because some partition class of $\mathrm{T}_r(n)$ has at least $n/(r-1)$ vertices, and at least two partition classes of $\mathrm{T}_n(r)$ have at least two vertices (unless $n \leqslant r$, in which case $L = \emptyset$). As $r > 3$, we can add the edge uv to G without creating a copy of K_r on any hyperedge of $\mathcal{F}$. Therefore $G + uv$ witnesses that $\mathcal{F}$ is not Turánnical.

For the case $r = 3$ of (a) we proceed similarly and infer from $|\mathcal{E}| < \frac{1}{2}\binom{n}{3}$ that there are distinct vertices $u, v \in [n]$ with $e\big(\mathrm{Link}(u,v)\big) < \frac{n}{2} - 1$ (observe that the hyperedges in $\mathrm{Link}(u,v)$ are singletons). Accordingly we can place the vertices u, v together with $E\big(\mathrm{Link}(u,v)\big)$ into one partition class of the bipartite graph $\mathrm{T}_3(n)$ and subsequently add the edge uv. $\mathcal{F}$ does not detect G, even though $e(G) = t_3(n) + 1$.

For (b) an even simpler construction for $G = ([n], E)$ suffices. We start with the complete graph $K_n =: G$. Then, for each hyperedge Y of $\mathcal{F}$ we pick two arbitrary vertices $u, v \in Y$ and delete the edge uv from G (if it is still present). Using $|\mathcal{E}| \leqslant (1 - r\varepsilon)\frac{1}{4r}n^2$ and $r \geqslant 3$, $n \geqslant 5$, it is easy to check that the resulting graph G has more than $(1+\varepsilon)t_r(n)$ edges, and by construction G contains no copies of K_r on hyperedges of $\mathcal{F}$. Hence $\mathcal{F}$ is not ε-Turánnical. $\square$

4.4 Approximately Turánnical random hypergraphs

In this section we prove Theorem 4.3. As noted in Section 4.1, the simple deterministic part (b) of Proposition 4.2, that no too sparse hypergraph $\mathcal{F}$ can be ε-approximately Turánnical, gives the 0-statement. We therefore focus on the proof of the 1-statement. To this end we use the following theorem of Erdős and Simonovits [38].

Theorem 4.9 (Erdős & Simonovits [38])**.** *Given any $r \in \mathbb{N}$ and $\varepsilon > 0$, there exists $\delta > 0$ such that the following is true. If G is any n-vertex graph with $e(G) \geqslant (1+\varepsilon)t_r(n)$, then there are at least δn^r copies of K_r in G.*

Proof of Theorem 4.3. Given $\varepsilon > 0$, by Theorem 4.9, there exists $\delta > 0$ such that if G is any graph with $e(G) \geqslant (1+\varepsilon)t_r(n)$, then G contains at least δn^r copies of K_r.

Let $p \geqslant \binom{n}{2} n^{-r}/\delta$. Given one graph G with at least δn^r copies of K_r, the probability that G is not detected by $\mathcal{R}^{(r)}(n,p)$ is at most

$$(1-p)^{\delta n^r} \ .$$

Summing over the at most $2^{\binom{n}{2}}$ such graphs G, we see that the probability that there exists an n-vertex graph G, with at least δn^r copies of K_r, which is undetected by $\mathcal{R}^{(r)}(n,p)$, is at most

$$2^{\binom{n}{2}}(1-p)^{\delta n^r} < 2^{\binom{n}{2}} e^{-p\delta n^r} \leqslant 2^{\binom{n}{2}} e^{-\binom{n}{2}} \ ,$$

which tends to zero as n tends to infinity. In particular, with probability tending to 1, any graph G with $e(G) \geqslant (1+\varepsilon)t_r(n)$ is detected by $\mathcal{R}^{(r)}(n,p)$. $\square$

4.4.1 Sparse ε-Turánnical hypergraphs explicitly

In Chapter 5 we provide an explicit construction of an r-uniform Turánnical hypergraph on n vertices with $\binom{n}{r} - \binom{\lfloor \frac{(r-2)n}{r-1} \rfloor}{r}$ edges. The number of edges of this hypergraph differs from the complete r-uniform hypergraph (which is a trivial example of a Turánnical hypegraph) by a constant factor. Even when we relax to ε-Turánnicalicity instead, we are not aware of a sparser construction. On the other hand, Theorems 4.3 and Theorems 4.4 assert that even a very sparse random hypergraph is Turánnical. Thus the following question comes to mind.

Question 4.10. *Give an explicit construction of an infinite family of Turánnical (or ε-Turánnical) hypergraphs with edge density $o(1)$.*

Let us briefly discuss a possible approach towards Question 4.10 in its easier version about ε-Turánnical hypergraphs. (The problem of constructing exactly Turánnical hypergraphs seems much harder and we omit any discussion entirely.) In view of Theorem 4.3 we believe that the theory of quasirandom hypergraphs is the key for any success. First, let us recall that a *graph* G is *quasirandom* if its overall density is approximately[2] inherited to any subgraph induced by a large[3] set U. Suppose that $U \subseteq \binom{V(\mathcal{H})}{r-1}$. Then an edge e of a hypergraph $\mathcal{H}$ is *induced* by U if $e - \{v\} \in U$ for each $v \in e$. An r-tuple $f \in \binom{V(\mathcal{H})}{r}$ (which is not necessarily an edge of $\mathcal{H}$) is called an *r-clique induced by U* if $f - \{v\} \in U$ for each $v \in f$. Following Gowers[4] [49, 50], an r-uniform hypergraph $\mathcal{H}$ is *quasirandom* if for any "substantial"[5] set U of $(r-1)$-tuples the ratio of the number of edges induced by U in $\mathcal{H}$ to the number of r-cliques induced by U is approximately the edge density of $\mathcal{H}$. (A somewhat different view on hypergraph quasirandomness was independently at around the same time given Rödl and his collaborators, see for example [85], and reference therein.)

The above definition of quasirandom hypergraphs gives actually a seemingly stronger property for free. Let $\mathcal{H} = (V, \mathcal{E})$ be an r-uniform hypergraph. Let $U \subseteq \binom{V}{p}$, where $p < r$. We can then construct inductively a sequence of sets $U_p := U, U_{p+1}, \ldots, U_{r-1}$ such that U_{i+1} is the set of $(i+1)$-cliques induced by U_i. We say that an edge $e \in \mathcal{E}$ is *induced* by U, if it is induced by U_{r-1}. It can then be checked that in a quasirandom hypergraph $\mathcal{H}$ the density of edges induced by any

[2] We omit numerical parameters which should accompany the notion of quasirandomness and which would quantify what is "approximately" in this brief exposition.

[3] meaning containing at least a fixed fraction of all the vertices

[4] Gowers develops his theory only for dense hypergraphs. However these definition are likely to have sensible counterparts for sparse hypergraphs.

[5] meaning that U should contain a positive fraction of all possible $(r-1)$-tuples

set $U \subseteq \binom{V}{p}$ normalized by the number of r-cliques induced by U_{r-1} should be approximately the edge density of $\mathcal{H}$.

Suppose that we can construct a sparse r-uniform quasirandom hypergraph $\mathcal{H} = (V, \mathcal{E})$ (or, more precisely, a family of them). We now claim that $\mathcal{H}$ is approximately Turánnical. Indeed, as in the proof of Theorem 4.3, suppose that $G = (V, E)$ is a graph which has more than $(1+\varepsilon)t_r(n)$ edges. Then by Theorem 4.9 the edge set E induces many r-cliques. Therefore, by the property of quasirandom hypergraphs discussed above many of these r-cliques are edges of $\mathcal{H}$. In particular, $\mathcal{H}$ detects G.

To the best of our knowledge, no construction of sparse quasirandom hypergraphs is known. There is one tempting approach which we were unable to pursue to a successful end. Gowers [51] constructed sparse quasirandom graphs as Cayley graphs over groups Γ which admit no non-trivial low-dimensional representation. Actually, he showed that then *any* Cayley graph over such a group Γ is quasirandom. Gowers' proof is spectral, i.e., it relies on a well-known connection between quasirandomness and eigenvalues of the adjacency matrix of the graph (see [26] for a thorough treatment on the topic). The spectral theory for hypergraphs is much more limited than for graphs. Still, we wonder, whether it is possible to infer quasirandomness of some families of Cayley hypergraphs[6], thus providing examples of sparse approximately Turánnical hypergraphs.

4.5 Exactly Turánnical random hypergraphs

In this section we prove Theorem 4.4. The 0-statement of Theorem 4.4 follows from Proposition 4.2 (a) for $r > 3$, and from Lemma 4.11 below for $r = 3$.

Lemma 4.11. *For $p \leqslant \frac{1}{2}$, we have $\mathbb{P}(\mathcal{R}^{(3)}(n,p)$ is Turánnical $) = o(1)$.*

Proof. By monotonicity, we may assume that $p = \frac{1}{2}$. As in the proof of Proposition 4.2 it suffices to show that there is a.a.s. a pair of vertices $u, v \in V(\mathcal{R}^{(3)}(n,p))$ with $e(\mathrm{Link}(u,v)) \leqslant \frac{n}{2} - 2$ (we remark that the hypergraph $\mathrm{Link}(u,v)$ is 1-uniform in this case). So choose two arbitrary vertices u and v. Observe that from the properties binomial distribution $\mathbb{P}\left(e(\mathrm{Link}(u,v)) > \frac{n}{2} - 2\right) \leqslant 0.6$, for large enough n. Let $\{u_1, v_1\}, \ldots, \{u_{\lfloor \frac{n}{2} \rfloor}, v_{\lfloor \frac{n}{2} \rfloor}\}$ be disjoint pairs of vertices. Using the independence of the variables $e(\mathrm{Link}(u_i, v_i))$, we obtain that $\mathbb{P}\left(\forall i\colon e(\mathrm{Link}(u_i, v_i) > \frac{n}{2} - 2\right) \leqslant 0.6^{\lfloor \frac{n}{2} \rfloor} = o(1)$. ☐

For the 1-statement of Theorem 4.4 we shall, in Lemma 4.14, investigate the structural properties of graphs with more edges than a Turán graph has, and

[6] Cayley hypergraph is a natural counterpart to Cayley graph; we omit the definition.

classify them into three possible categories. We then treat these three types of graphs separately, and show for each of them that with high probability a random restriction hypergraph $\mathcal{R}^{(r)}(n,p)$ detects *each* of the graphs of this type. Let us first take a small detour.

The Erdős-Simonovits theorem, Theorem 4.9, states that graphs G with many more edges than a Turán graph $\mathrm{T}_r(n)$ contain a positive fraction of the possible r-cliques. This is not true anymore when G has just one edge more than $\mathrm{T}_r(n)$. However, as the well-known stability theorem of Simonovits [92] shows, we can still draw the same conclusion when we know in addition that G looks very different from $\mathrm{T}_r(n)$. To state the result of Simonovits we need the following definition. Let ε be a positive constant and G and H be graphs on n vertices. If G cannot be obtained from H by adding and deleting together at most εn^2 edges, then we say that G is *ε-far* from H.

Theorem 4.12 (Simonovits [92])**.** *For every $r \geqslant 3$ and $\varepsilon > 0$ there exists $\delta > 0$ such that any n-vertex graph G with $e(G) \geqslant t_r(n)$ which is ε-far from $\mathrm{T}_r(n)$ contains at least δn^r copies of K_r.*

If a graph G is not far from a Turán graph, on the other hand, we have a lot of structural information about G: we know that its vertex set can be partitioned into $r-1$ sets which are almost of the same size and almost independent, such that most of the edges between these sets are present. If in addition almost all vertices of G have many neighbours in all partition classes other than their own, then we say that G has an ε-close $(r-1)$-partition. The following definition makes this precise.

Definition 4.13 (ε-close $(r-1)$-partition)**.** *Let $G=(V,E)$ be a graph. An* ε-close $(r-1)$-partition *of G is a partition $V = V_0 \dot\cup V_1 \dot\cup \dots \dot\cup V_{r-1}$ of its vertex set such that*

(i) $|V_0| \leqslant \varepsilon^2 n$ *and* $|V_i| \geqslant (1-\varepsilon)\frac{n}{r-1}$ *for all* $i \in [r-1]$,

(ii) *for all* $v \in V_0$ *we have* $\deg(v) \leqslant (1-\varepsilon^2)\frac{r-2}{r-1}n$, *and for all* $i, j \in [r-1]$ *with* $i \neq j$ *and for all* $v \in V_i$ *we have* $\deg(v, V_j) \geqslant (1-\varepsilon)|V_j|$.

The edges (non-edges) in such a partition that run between two different parts V_i and V_j with $1 \leqslant i, j \leqslant r-1$, are called crossing, *and those that lie within a partition class V_i with $1 \leqslant i \leqslant r-1$, are* non-crossing.

The following lemma states that a graph which has at least as many edges as $\mathrm{T}_r(n)$ either contains a vertex whose neighbourhood has a positive K_{r-1}-density, or has an ε-close $(r-1)$-partition. See [19] for a somewhat related result.

Lemma 4.14. *For every integer $r \geqslant 3$ and real $0 < \varepsilon \leqslant 1/(16r^2)$ there exists a positive constant δ such that for every n-vertex graph G with $e(G) \geqslant t_r(n)$ one of the the following is true.*

(*i*) *Some vertex in G is contained in at least δn^{r-1} copies of K_r.*

(*ii*) *G has an ε-close $(r-1)$-partition.*

We postpone the proof of Lemma 4.14 and first sketch that it implies Lemma 4.5.

Proof of Lemma 4.5. Suppose we are given r and $\tilde{\varepsilon}$. By monotonicity we may assume that $\tilde{\varepsilon} < 1/16$. Let δ be given by Lemma 4.14 with input parameters r and $\varepsilon := \tilde{\varepsilon}/r^2$. By Lemma 4.14 it suffices to show that in each n-vertex graph G with

$$e(G) > t_r(n) \tag{4.1}$$

which possesses an ε-close $(r-1)$-partition $V(G) = V_0 \dot{\cup} V_1 \dot{\cup} \ldots \dot{\cup} V_{r-1}$ there is an edge contained in at least $(1-\tilde{\varepsilon})(n/(r-1))^{r-2}$ copies of K_r. First observe that by (4.1) and (*ii*) of Definition 4.13 we have $e(G - V_0) > t_r(n - |V_0|)$. Thus, by Turán's Theorem, there is an edge $uv \subseteq V_i$ for some $i \in [r-1]$. The edge uv has at least $(1-2\varepsilon)|V_j|$ common neighbours in each V_j, $j \neq i$, creating at least

$$\Big(\big(1-(r-1)\varepsilon\big)(1-\varepsilon)\frac{n}{r-1}\Big)^{r-2} \geqslant (1-r\varepsilon)^{r-2}\Big(\frac{n}{r-1}\Big)^{r-2} \geqslant (1-\tilde{\varepsilon})\Big(\frac{n}{r-1}\Big)^{r-2}$$

copies of K_r. □

Proof of Lemma 4.14. Given r and ε, let G be an n-vertex graph with $e(G) \geqslant t_r(n)$. By Theorem 4.12, there exists $\gamma = \gamma(\varepsilon, r) > 0$ such that if G is $\varepsilon^3/(16r^3)$-far from $\mathrm{T}_r(n)$, then G contains γn^r copies of K_r. We set

$$\delta := \min\Big\{\gamma,\ \frac{1}{r!2^r r^r},\ \frac{\varepsilon}{4^r r^r},\ \Big(\frac{\varepsilon}{2r}\Big)^{r-1}\Big\}.$$

Since $e(G) \geqslant t_r(n)$, either $G = \mathrm{T}_r(n)$, which clearly has an ε-close $(r-1)$-partition, or G contains a copy of K_r. Observe that the last term in this minimum ensures that if $n < \frac{2r}{\varepsilon}$, then $\delta n^{r-1} < 1$, and thus that one copy of K_r in G is enough to satisfy the Lemma. It follows that we may henceforth assume $n \geqslant \frac{2r}{\varepsilon}$.

If G contains γn^r copies of K_r then there is a vertex lying in $\gamma n^{r-1} \geqslant \delta n^{r-1}$ copies of K_r. Thus we may assume that G is not $\varepsilon^3/(16r^3)$-far from $\mathrm{T}_r(n)$. So there exists a balanced partition $V(G) = U_1 \dot{\cup} \ldots \dot{\cup} U_{r-1}$ such that the total number of non-edges between the parts is at most $\varepsilon^3 n^2/(16r^3)$.

Now for each $1 \leqslant i \leqslant r-1$, we define

$$V_i = \left\{ v \in V(G)\colon\ \deg(v, V(G) - U_i) \geqslant \left(\frac{r-2}{r-1} - \frac{\varepsilon}{4r}\right) n \right\} . \tag{4.2}$$

We let $V_0 := V(G) - (V_1 \cup \ldots \cup V_{r-1})$. We aim to show that either there is some vertex of G which lies in at least δn^{r-1} copies of K_r, or that $V_0 \dot\cup V_1 \dot\cup \ldots \dot\cup V_{r-1}$ is an ε-close $(r-1)$-partition.

For each $1 \leqslant i \leqslant r-1$, every vertex in $U_i - V_i$ lies in at least $\varepsilon n/(4r)$ non-edges crossing the partition $(U_1, \ldots, U_{r-1})$. It follows that

$$|U_i - V_i| \leqslant \frac{\varepsilon^2 n}{4r^2} , \tag{4.3}$$

since there are at most $\varepsilon^3 n^2/(16r^3)$ such non-edges. Summing over $i = 1, \ldots, r-1$ we get

$$|V_0| \leqslant \frac{(r-1)\varepsilon^2 n}{4r^2} < \frac{\varepsilon^2 n}{4r} < \varepsilon^2 n . \tag{4.4}$$

Since $n \geqslant 2r/\varepsilon$ we also have, for each $1 \leqslant i, j \leqslant r-1$ with $i \neq j$, and each $v \in V_i$, that

$$\begin{aligned} |V_i| &\geqslant |U_i| - \frac{\varepsilon^2 n}{4r^2} > (1-\varepsilon)\frac{n}{r-1} , \qquad \text{and} \\ \deg(v, V_j) &\overset{(4.2),(4.3)}{\geqslant} |U_j| - 1 - \frac{\varepsilon n}{4r} - \frac{\varepsilon^2 n}{4r^2} \\ &\geqslant |V_j| - 1 - (r-2)\frac{\varepsilon^2 n}{4r^2} - \frac{\varepsilon n}{4r} - \frac{\varepsilon^2 n}{4r^2} \geqslant (1-\varepsilon)|V_j| , \end{aligned} \tag{4.5}$$

where we use $\varepsilon \leqslant \frac{1}{10}$ to obtain the last inequality.

We claim that a vertex u lying in more than one of the sets $V_1, \ldots, V_{r-1}$ must lie in at least δn^{r-1} copies of K_r. To see this, observe that u must have at least $(1-\varepsilon)|V_i|$ neighbours in V_i for each $1 \leqslant i \leqslant r-1$. Now consider the following method of constructing a copy of K_r in G using u. We choose a neighbour v_1 of u in V_1, a common neighbour v_2 of u and v_1 in V_2, and so on. Since $\varepsilon \leqslant 1/(16r)$, the common neighbourhood of $u, v_1, \ldots, v_{i-1}$ in V_i contains at least $(1 - i\varepsilon)|V_i| > \frac{n}{2(r-1)}$ vertices for each i, there are at least $\frac{n}{2(r-1)}$ choices at each of the $r-1$ steps (and in particular this construction is possible). This procedure may construct the same copy of K_r more than once (since at this point we do not yet know that the sets $V_1, \ldots, V_{r-1}$ are disjoint), but not more than $(r-1)!$ times. It follows that u lies in at least

$$\frac{1}{(r-1)!}\left(\frac{n}{2(r-1)}\right)^{r-1} \geqslant \delta n^{r-1}$$

copies of K_r.

Hence, we can assume from now on that the sets $V_1, \ldots, V_{r-1}$ are disjoint. Next we claim that a vertex u in V_0 whose degree exceeds $(1-\varepsilon^2)\frac{r-2}{r-1}n$ must lie in at least δn^{r-1} copies of K_r. Without loss of generality, we may assume that we have $\deg(u, V_1) \leqslant \deg(u, V_2) \leqslant \ldots \leqslant \deg(u, V_{r-1})$. Since $u \notin V_1$, we have

$$\begin{aligned}
\deg(u, V_1) &= \deg(u) - \deg(u, V(G) - V_1) \\
&\geqslant \deg(u) - \deg(u, U_2 \dot\cup \ldots \dot\cup U_{r-1}) - |U_1 - V_1| \\
&\overset{(4.2),(4.3)}{>} (1-\varepsilon^2)\frac{r-2}{r-1}n - \left(\frac{r-2}{r-1} - \frac{\varepsilon}{4r}\right)n - \frac{\varepsilon^2 n}{4r^2} \qquad (4.6) \\
&\geqslant -\varepsilon^2 n + \frac{\varepsilon n}{4r} - \frac{\varepsilon^2 n}{4r^2} \geqslant \frac{\varepsilon n}{16r},
\end{aligned}$$

where the last inequality follows from $\varepsilon \leqslant 1/(16r)$. Since $\deg(u, V_2) \geqslant \deg(u, V_1)$ and u has at most $\frac{n}{r-1} + \varepsilon^2 n$ non-neighbours by assumption, we infer that $\deg(u, V_2) \geqslant \frac{n}{3(r-1)}$, using again $\varepsilon \leqslant 1/(16r)$. Hence

$$\deg(u, V_i) \geqslant \frac{n}{3(r-1)} \qquad \text{for each } 2 \leqslant i \leqslant r-1. \tag{4.7}$$

Now consider the same inductive construction of copies of K_r containing u as before. This time we know that there are at least $\frac{\varepsilon n}{16r}$ choices for v_1, and at least

$$\frac{n}{3(r-1)} - (i-1)\varepsilon|V_i| > \frac{n}{4(r-1)}$$

choices for v_i, for each $2 \leqslant i \leqslant r-1$. Since the sets $V_1, \ldots, V_{r-1}$ are disjoint, each copy of K_r can be constructed in only one way. Thus u does indeed lie in at least

$$\frac{\varepsilon n}{16r}\left(\frac{n}{4(r-1)}\right)^{r-2} \geqslant \delta n^{r-1}$$

copies of K_r.

Accordingly, we can assume that $\deg(u) \leqslant (1-\varepsilon^2)\frac{r-2}{r-1}n$, for all u in V_0. Together with (4.4) and (4.5) this implies that the partition $V_0 \dot\cup \ldots \dot\cup V_{r-1}$ satisfies (*i*) and (*ii*) of Definition 4.13 and hence is an ε-close $(r-1)$-partition of G. $\square$

We need a more precise structural result to handle the case $r = 3$ of Theorem 4.4. As we shall see, this is a simple consequence of the above proof.

Corollary 4.15. *For every $0 < \varepsilon \leqslant 1/144$ there exists a positive constant δ such that for all n-vertex graphs G with $e(G) \geqslant t_3(n)$ one of the the following is true.*

(*i*) *G contains at least δn^3 triangles.*

(*ii*) *There is a vertex u of G such that* $\mathrm{N}(u) \supset X \dot\cup Y$, *where* $|X||Y| \geqslant \varepsilon n^2/288$ *and* $e(X,Y) \geqslant (1-4\varepsilon)|X||Y|$.

(*iii*) *G has an ε-close 2-partition.*

Proof. We follow the previous proof with $r = 3$, using the same value for δ. If G contains less than δn^3 triangles we obtain the three sets V_0, V_1, V_2 (as defined in (4.2)). If these sets do not form a partition of $V(G)$, then there is a vertex v in both V_1 and V_2. Then we let $X := \mathrm{N}(v) \cap V_1$ and $Y := \mathrm{N}(v) \cap V_2$. By (4.5) we have $|X||Y| \geqslant (1-\varepsilon)^2|V_1||V_2| \geqslant (1-\varepsilon)^4 n^2/4 > \varepsilon n^2/32$ because $\varepsilon \leqslant 1/2$. Since each vertex of X is adjacent to all but at most $\varepsilon|V_2|$ vertices of Y by (4.5), we also have $e(X,Y) \geqslant (1-4\varepsilon)|X||Y|$ as required.

Hence we may assume that V_0, V_1, V_2 form a partition of $V(G)$. The only remaining barrier to V_0, V_1, V_2 being an ε-close 2-partition of G is the existence of a vertex v in V_0 with degree more than $(1-\varepsilon^2)\frac{n}{2}$. As in the previous proof, if this vertex exists we may without loss of generality presume by (4.6) that it has at least $\varepsilon n/48$ neighbours in V_1, and by (4.7) that it has at least $n/6$ neighbours in V_2. Again we let $X := \mathrm{N}(v) \cap V_1$, and $Y := \mathrm{N}(v) \cap V_2$, and get $|X||Y| \geqslant \varepsilon n^2/288$ as required. Now since $|Y| > |V_2|/4$, and since every vertex in X is adjacent to all but at most $\varepsilon|V_2|$ vertices of Y, we have $e(X,Y) \geqslant (1-4\varepsilon)|X||Y|$ as required. $\square$

Our next lemma counts the number of graphs with an ε-close $(r-1)$-partition and a given number of non-crossing edges. In addition it estimates the number of r-cliques in such a graph.

Lemma 4.16. *Let $\ell \geqslant 0$ and $r \geqslant 3$ be integers, $0 < \varepsilon < 1/(2r)$ be a real and $n \geqslant 2r^3/\varepsilon^2$ be an integer. Let $\mathcal{G}$ be the family of all graphs on a fixed vertex set of size n with $e(G) > t_r(n)$ which have an ε-close $(r-1)$-partition with exactly ℓ non-crossing edges. Then*

(*a*) *if $\ell = 0$ then $|\mathcal{G}| = 0$,*

(*b*) $|\mathcal{G}| \leqslant r^{5\ell n}$, *and*

(*c*) *every $G \in \mathcal{G}$ contains at least $\ell\big(\frac{n}{2r-2}\big)^{r-2}$ copies of K_r.*

Proof. In the following, let $G \in \mathcal{G}$. We fix an ε-close $(r-1)$-partition $V_0, \ldots, V_{r-1}$ of G with ℓ non-crossing edges. Let the number of crossing non-edges be k.

First we show (*c*). Let e be a non-crossing edge of G. Without loss of generality, we may presume e lies in V_1. We can construct an r-clique using e as follows: we choose any common neighbour v_2 of e in V_2, then a common neighbour

v_3 of e and v_2 in V_3, and so on. By definition of an ε-close $(r-1)$-partition, for each $2 \leqslant i \leqslant r-1$, the common neighbourhood of $e, v_2, \ldots, v_{i-1}$ in V_i has size at least $(1-i\varepsilon)|V_i| > \frac{1}{2}n/(r-1)$ because $\varepsilon < 1/(2r)$. It follows that e lies in at least $(n/(2r-2))^{r-2}$ copies of K_r in G. Further, if e' is a second non-crossing edge of G, then no r-clique of G using e' can be one of the r-cliques through e given by the above construction. It follows that G contains $\ell(n/(2r-2))^{r-2}$ copies of K_r.

Now we prove (a) and (b). We first show that

$$\ell \geqslant |V_0| + k + 1\,. \tag{4.8}$$

If $V_0 = \emptyset$, then we have $t_r(n)+1 \leqslant e(G) \leqslant t_r(n)+\ell-k$, and therefore $\ell \geqslant |V_0|+k+1$. If $V_0 \neq \emptyset$ on the other hand, then, since every vertex in V_0 has degree at most $(1-\varepsilon^2)\frac{r-2}{r-1}n$, we have

$$t_r(n)+1 \leqslant e(G) \leqslant (1-\varepsilon^2)\frac{r-2}{r-1}n|V_0| + \left(\frac{n-|V_0|}{r-1}\right)^2 \binom{r-1}{2} + \ell - k\,.$$

Using the facts $|V_0| \leqslant \varepsilon^2 n$ and $\left(\frac{n}{r-1}\right)^2\binom{r-1}{2} \leqslant t_r(n) + r^2$, we infer

$$\begin{aligned} &t_r(n)+1 \\ &\quad\leqslant (1-\varepsilon^2)\tfrac{r-2}{r-1}n|V_0| + \left(\tfrac{n}{r-1}\right)^2\tbinom{r-1}{2} - \tfrac{r-2}{r-1}n|V_0| + \tfrac{(r-2)}{2(r-1)}|V_0|^2 + \ell - k \\ &\quad\leqslant t_r(n) + r^2 - \varepsilon^2\tfrac{r-2}{r-1}n|V_0| + \varepsilon^2\tfrac{r-2}{2(r-1)}n|V_0| + \ell - k \\ &\quad= t_r(n) + r^2 - \varepsilon^2\tfrac{r-2}{2(r-1)}n|V_0| + \ell - k\,. \end{aligned}$$

It follows from $n \geqslant 2r^3/\varepsilon^2$ that $\varepsilon^2\frac{r-2}{2(r-1)}n|V_0| \geqslant r^2 + |V_0|$, and so we again obtain $\ell \geqslant |V_0| + k + 1$.

Now, if $G \in \mathcal{G}$ exists, then (4.8) clearly implies $\ell > 0$, proving (a). It remains to show (b). We can construct any graph G in $\mathcal{G}$ as follows. We choose $k \in \{0, \ldots, \ell-1\}$. We partition $[n]$ into r sets $V_0, \ldots, V_{r-1}$ such that V_0 satisfies (4.8). For each pair of vertices intersecting V_0, we choose whether or not to make it an edge of G; there are at most $2^{|V_0|n} \leqslant 2^{\ell n}$ such choices. Then we choose k pairs of vertices crossing the partition to be non-edges of G, and make all other crossing pairs edges of G. Finally, we choose ℓ pairs of vertices within partition classes to be the ℓ non-crossing edges of G. The total number of choices in this process is at most

$$\sum_{0 \leqslant k \leqslant \ell-1} r^n 2^{\ell n} \binom{\binom{n}{2}}{k}\binom{\binom{n}{2}}{\ell} \overset{(4.8)}{\leqslant} \ell r^n 2^{\ell n} n^{2\ell+2\ell} \leqslant r^{5\ell n}\,,$$

as required. $\square$

With these tools at hand we can proceed to the proof of Theorem 4.4. For a binomially distributed random variable X we will use the following Chernoff bound which can be found, e.g., in [55, Theorem 2.1]. For each $\gamma \in (0, \frac{1}{3})$ we have

$$\mathbb{P}\big(X \leqslant (1-\gamma)\mathbb{E}X\big) \leqslant \exp(-\gamma^2 \mathbb{E}X/2)\ . \tag{4.9}$$

Proof of the 1*-statements of Theorem 4.4.* We shall first prove the case $r = 3$ and then turn to the case $r > 3$. In both cases we will consider the class $\mathcal{G}_r$ of all n-vertex graphs G with $e(G) > t_r(n)$. In the case $r = 3$, $\mathcal{G}_3$ can be written as the union of three sub-classes $\mathcal{G}_{\mathrm{A}}$, $\mathcal{G}_{\mathrm{B}}$, and $\mathcal{G}_{\mathrm{C}}$ defined by the properties in (i), (ii), and (iii) of Corollary 4.15, respectively. Similarly, for $r > 3$ Lemma 4.14 allows us to write $\mathcal{G}_r = \mathcal{G}_{\mathrm{D}} \cup \mathcal{G}_{\mathrm{E}}$, where the graphs $\mathcal{G}_{\mathrm{D}}$ and $\mathcal{G}_{\mathrm{E}}$ enjoy properties given by Lemma 4.14(i) and Lemma 4.14(ii), respectively. We will prove that for each of these sub-classes a.a.s. the random hypergraph $\mathcal{R}^{(r)}(n,p)$ with p as required detects all graphs in this sub-class. The result then follows from the union bound.

<u>*Case $r = 3$:*</u> Let $p > 1/2$ be fixed and set

$$\varepsilon := \min\left\{\frac{1}{144}, \frac{p}{8}, \frac{2p-1}{4p+3}\right\}.$$

Let $\delta > 0$ be guaranteed by Corollary 4.15 for this ε. Observe that this choice of ε and n allows the application of Corollary 4.15. Further, let $\mathcal{G}_3 = \mathcal{G}_{\mathrm{A}} \cup \mathcal{G}_{\mathrm{B}} \cup \mathcal{G}_{\mathrm{C}}$ be as defined above. We will now show for each of the graph classes $\mathcal{G}_{\mathrm{A}}$, $\mathcal{G}_{\mathrm{B}}$, and $\mathcal{G}_{\mathrm{C}}$ that a.a.s. $\mathcal{R}^{(3)}(n,p)$ detects all their members.

Suppose a graph $G \in \mathcal{G}_{\mathrm{A}}$ is given. Then Corollary 4.15(i) the graph G contains at least δn^3 triangles. The probability that $\mathcal{R}^{(3)}(n,p)$ does not detect G is at most

$$(1-p)^{\delta n^3} \leqslant e^{-p\delta n^3} \leqslant e^{-\delta n^3/2}\ ,$$

and since $|\mathcal{G}_{\mathrm{A}}| < 2^{\binom{n}{2}}$, applying the union bound, the probability that there is a graph in $\mathcal{G}_{\mathrm{A}}$ which $\mathcal{R}^{(3)}(n,p)$ does not detect is at most

$$2^{\binom{n}{2}} e^{-\delta n^3/2}\ ,$$

which tends to zero as n tends to infinity.

Recall that $\mathcal{G}_{\mathrm{B}}$ is the sub-class of $\mathcal{G}_3$ with graphs in which there is a vertex u and disjoint set $X, Y \subseteq \mathrm{N}(u)$ with both $|X||Y| \geqslant \varepsilon n^2/288$ and $e(X,Y) \geqslant (1 - 4\varepsilon)|X||Y|$. Suppose that a 3-uniform n-vertex hypergraph $\mathcal{H}$ has the property that for every vertex v and disjoint sets W and Z with $|W||Z| \geqslant \varepsilon n^2/288$, there are more than $4\varepsilon|W||Z|$ hyperedges of $\mathcal{H}$, each consisting of v, a vertex of W, and a vertex of Z. Then, clearly for any $G \in \mathcal{G}_{\mathrm{B}}$ the hypergraph $\mathcal{H}$ detects G. Hence it remains to show that a.a.s. $\mathcal{R}^{(3)}(n,p)$ has this property.

Given one vertex v and pair of disjoint vertex sets X and Y of $\mathcal{R}^{(3)}(n,p)$ with $|X||Y| \geqslant \varepsilon n^2/288$ the expected size of $E\big(\operatorname{Link}_{\mathcal{R}^{(3)}(n,p)}(v)\big) \cap (X \times Y)$ in $\mathcal{R}^{(3)}(n,p)$ is $p|X||Y|$. Using the Chernoff bound (4.9), the probability that we have

$$e\big(\operatorname{Link}_{\mathcal{R}^{(3)}(n,p)}(v) \cap (X \times Y)\big) < 4\varepsilon|X||Y| \leqslant p|X||Y|/2$$

is at most $e^{-p|X||Y|/8} \leqslant e^{-\varepsilon n^2/5000}$. By the union bound, the probability that there exists any such vertex and pair of disjoint subsets in $\mathcal{R}^{(3)}(n,p)$ is at most

$$n 2^n 2^n e^{-\varepsilon n^2/5000}$$

which tends to zero as n tends to infinity.

Finally, $\mathcal{G}_{\mathrm{C}}$ is the class of n-vertex graphs $G \in \mathcal{G}_3$ which possess an ε-close 2-partition $V_0 \dot\cup V_1 \dot\cup V_2$. Since $e(G) \geqslant t_r(n)+1$ there is at least one non-crossing edge e in this partition by Lemma 4.16(a). Without loss of generality, we may presume e lies in V_1. Then the common neighbourhood of e contains more than $(1-2\varepsilon)|V_2| \geqslant (1-3\varepsilon)\frac{n}{2}$ vertices. In particular, if $\mathcal{R}^{(3)}(n,p)$ has the property that every pair of vertices is in at least $(1+3\varepsilon)\frac{n}{2}$ hyperedges, then $\mathcal{R}^{(3)}(n,p)$ detects every graph in $\mathcal{G}_{\mathrm{C}}$. We will show that a.a.s. $\mathcal{R}^{(3)}(n,p)$ has this property.

Given one pair of vertices u, v, we have

$$\mathbb{E}\big(e(\operatorname{Link}_{\mathcal{R}^{(3)}(n,p)}(u,v))\big) = p(n-2)\ .$$

Using the fact that $\varepsilon \leqslant \frac{2p-1}{4p+3}$ we note that

$$(1+3\varepsilon)\frac{n}{2} \leqslant \Big(1+3\frac{2p-1}{4p+3}\Big)\frac{n}{2} = \Big(1-2\frac{2p-1}{4p+3}\Big)pn < (1-\varepsilon)p(n-2)\ ,$$

for large enough n. The Chernoff bound (4.9) then gives

$$\begin{aligned}&\mathbb{P}\left(e(\operatorname{Link}_{\mathcal{R}^{(3)}(n,p)}(u,v)) \leqslant (1+3\varepsilon)\frac{n}{2}\right) \leqslant \\ &\quad \mathbb{P}\left(e(\operatorname{Link}_{\mathcal{R}^{(3)}(n,p)}(u,v)) \leqslant (1-\varepsilon)p(n-2)\right) \leqslant e^{-\varepsilon^2 p(n-2)/2}\ .\end{aligned}$$

By the union bound, the probability that there exists any such pair of vertices in $\mathcal{R}^{(3)}(n,p)$ is at most $\binom{n}{2}e^{-\varepsilon^2 p(n-2)/2}$, which tends to zero as n tends to infinity.

<u>*Case $r > 3$:*</u> Let $\varepsilon := 1/(16r^2)$, and let $\delta > 0$ be the positive constant guaranteed by

Lemma 4.14 for this ε. Let $\mathcal{G}_r = \mathcal{G}_\mathrm{D} \cup \mathcal{G}_\mathrm{E}$ be classes of n-vertex graphs satisfying (i) and (ii) of Lemma 4.14, respectively. Set

$$C := \max\left\{\frac{1}{\delta}, 6r(2r-2)^{r-2}\right\}, \qquad \text{and let} \quad p \geqslant Cn^{3-r}\,.$$

Again, we will prove that a.a.s. $\mathcal{R}^{(r)}(n,p)$ detects all graphs in $\mathcal{G}_\mathrm{D}$ and $\mathcal{G}_\mathrm{E}$.

The class $\mathcal{G}_\mathrm{D}$ contains the graphs from $\mathcal{G}_r$ in which there is a vertex contained in at least δn^{r-1} copies of K_r. Given one such graph G, the probability that G is not detected by $\mathcal{R}^{(r)}(n,p)$ is at most

$$(1-p)^{\delta n^{r-1}} < e^{-Cn^{3-r}\delta n^{r-1}} = e^{-C\delta n^2} \leqslant e^{-n^2}\,,$$

and since there are at most $2^{\binom{n}{2}}$ graphs in $\mathcal{G}_\mathrm{D}$, the probability that there is a graph in $\mathcal{G}_\mathrm{D}$ undetected by $\mathcal{R}^{(r)}(n,p)$ is at most

$$2^{\binom{n}{2}}e^{-n^2}\,,$$

which tends to zero as n tends to infinity.

It remains to consider the class $\mathcal{G}_\mathrm{E}$ of graphs $G \in \mathcal{G}_r$ with ε-close $(r-1)$-partition. For $1 \leqslant \ell \leqslant \binom{n}{2}$ let $\mathcal{G}_\mathrm{E}(\ell) \subseteq \mathcal{G}_\mathrm{E}$ be the class of graphs that have an ε-close $(r-1)$-partition with exactly ℓ non-crossing edges. By Lemma 4.16(a) we have

$$\bigcup_{1\leqslant \ell \leqslant \binom{n}{2}} \mathcal{G}_\mathrm{E}(\ell) = \mathcal{G}_\mathrm{E}\,. \tag{4.10}$$

Now fix $\ell \in \{1,\dots,\binom{n}{2}\}$. Lemma 4.16$(b)$ asserts that $|\mathcal{G}_\mathrm{E}(\ell)| \leqslant r^{5\ell n}$. Moreover, each graph in $\mathcal{G}_\mathrm{E}(\ell)$ contains at least $\ell(n/(2r-2))^{r-2}$ copies of K_r by Lemma 4.16(c). Hence, by the union bound, the probability that $\mathcal{R}^{(r)}(n,p)$ fails to detect at least one graph in $\mathcal{G}_\mathrm{E}(\ell)$ is at most

$$\begin{aligned} r^{5\ell n}(1-p)^{\left(\frac{n}{2r-2}\right)^{r-2}\ell} &< r^{5\ell n}\exp\left(-Cn^{3-r}\ell\Big(\frac{n}{2r-2}\Big)^{r-2}\right) \\ &\leqslant r^{5\ell n}e^{-6r\ell n} < e^{-\ell n}\,. \end{aligned}$$

Finally, applying the union bound in conjunction with (4.10), we conclude that $\mathcal{R}^{(r)}(n,p)$ detects all graphs in $\mathcal{G}_\mathrm{E}$ with probability at least $1-\binom{n}{2}e^{-n}$, which tends to one as n tends to infinity. $\square$

4.6 Turánnical hypergraphs for random graphs

In this section we prove Theorem 4.8. For this purpose we shall use the machinery developed by Schacht [89] for proving Theorem 4.6. Conlon and Gowers [28] obtained independently (using different methods) a result very similar to Schacht's. While either result is equally suited for proving Theorem 4.8 we follow notation introduced in [89]. Schacht formulates a powerful abstract result, a so-called *transference theorem* (Theorem 3.3 in [89]; see also Theorem 4.5 in [28]), which is phrased in the language of hypergraphs and gives very general conditions under which a result from extremal combinatorics may be transferred to an analogue for sparse random structures. Actually, Theorem 4.6 mentioned above is only one of several results where the transference theorem applies. Schacht, and Conlon and Gowers give further applications to transfer the multidimensional Szemerédi theorem, a result on Schur's equation, and others. Here we are interested in a transference of Theorem 4.3.

Below we will state a special version of Schacht's transference theorem, tailored to our situation. For formulating this theorem we need some definitions. We remark that in these definitions we slightly deviate from Schacht's setting. More precisely, the transference theorem uses a certain sequence of hypergraphs which encode the classical extremal problem under consideration. In the case of Turán's problem for K_r, the n-th hypergraph in this sequence has vertex set $E(K_n)$ and a hyperedge for every $\binom{r}{2}$-tuple of elements from $E(K_n)$ which form a copy of K_r in K_n in Schacht's setting. Instead, we shall work with r-uniform hypergraphs $\mathcal{H}_n$ on vertex set $V(K_n)$, making use of the fact that a copy of K_r is uniquely identified by its vertices. The corresponding modifications of the definitions and of the transference theorem are straightforward.

The transference theorem requires the sequence of hypergraphs to satisfy two conditions. The first one is a requirement upon the extremal problem to be transferred, namely, that it has a certain 'super-saturation' property (similar to the one given in Theorem 4.9). The following definition makes this precise.

Definition 4.17 ($(\alpha, \varepsilon, \zeta)$-dense)**.** *Let $\mathbf{H} = (\mathcal{H}_n)_{n\in\mathbb{N}}$ be a sequence of n-vertex r-uniform hypergraphs, $\alpha \geqslant 0$ and $\varepsilon, \zeta > 0$ be constants. We say $\mathbf{H}$ is* $(\alpha, \varepsilon, \zeta)$-dense *if the following is true. There exists n_0 such that for every $n \geqslant n_0$ and every graph G on the vertex set $V(\mathcal{H}_n)$ with at least $(\alpha + \varepsilon)\binom{n}{2}$ edges, the number of copies of K_r in G induced by hyperedges of $\mathcal{H}_n$ is at least $\zeta e(\mathcal{H}_n)$.*

The second condition determines the sparseness of a random graph to which one may transfer the extremal result. Given an r-uniform hypergraph $\mathcal{H}$, a graph G on the same vertex set, and a pair of distinct vertices u and v of $V(G)$, we let

$\deg_i(u, v, G)$ be the number of hyperedges of $\mathcal{H}$ containing u, v and at least i edges of G, not counting the possible edge uv. If $u = v$ we let $\deg_i(u, v, G) := 0$. The hypergraph $\mathcal{H}$ itself is suppressed from the notation as it will be clear from the context. We set

$$\mu_i(\mathcal{H}, q) := \mathbb{E}\Big[\sum\nolimits_{u,v} \deg_i^2\left(u, v, G(n, q)\right)\Big],$$

where the expectation is taken over the space of random graphs $G(n, q)$.

Definition 4.18 ($(K, \mathbf{q})$-bounded)**.** *Let* $\mathbf{H} = (\mathcal{H}_n)_{n\in\mathbb{N}}$ *be a sequence of* n*-vertex* r*-uniform hypergraphs,* $\mathbf{q} = (q_n)_{n\in\mathbb{N}}$ *be a sequence of probabilities, and* $K \geqslant 1$ *be a constant. We say that* $\mathbf{H}$ *is* $(K, \mathbf{q})$-bounded *if the following holds. For each* $i \in [\binom{r}{2} - 1]$ *there exists* n_0 *such that for each* $n \geqslant n_0$ *and* $q \geqslant q_n$ *we have*

$$\mu_i(\mathcal{H}_n, q) \leqslant Kq^{2i} \cdot \frac{e(\mathcal{H}_n)^2}{n^2}.$$

We can now state (a special case of) Schacht's transference theorem.

Theorem 4.19 (transference theorem, Schacht [89])**.** *For all* $r \geqslant 3$*,* $K \geqslant 1$*,* $\delta > 0$*,* $\zeta > 0$ *and* $(\omega_n)_{n\in\mathbb{N}}$ *with* $\omega_n \to \infty$ *as* $n \to \infty$*, there exists* $C > 1$ *such that the following holds. Let* $\varepsilon := 8^{-r(r-1)/2}\delta$*, and let* $\mathbf{H} = (\mathcal{H}_n)_{n\in\mathbb{N}}$ *be a sequence of* n*-vertex* r*-uniform hypergraphs which is* $(\frac{r-2}{r-1}, \varepsilon, \zeta)$*-dense. Let* $\mathbf{q} = (q_n)_{n\in\mathbb{N}}$ *be a sequence of probabilities with* $q_n^{r(r-1)/2} \cdot e(\mathcal{H}_n) \to \infty$ *such that* $\mathbf{H}$ *is* $(K, \mathbf{q})$*-bounded.*

Then the following holds a.a.s. for $G = G(n, Cq_n)$*. Every subgraph of* G *with at least* $(\frac{r-2}{r-1} + \delta) \cdot e(G)$ *edges contains an* r*-clique induced by a hyperedge of* $\mathcal{H}_n$*.*

We remark that the quantification in this theorem and the $(\alpha, \varepsilon, \zeta)$-denseness condition given here is not the same as in [89] (in fact, in [89] the two parameters ε and ζ are not made explicit in the concept of α-denseness used in [89]). The statement in [89] is certainly cleaner, but for our purposes it is necessary that we check the denseness condition only for a special ε (as opposed to all $\varepsilon > 0$, which is necessary for the original definition of α-denseness), and that the constant C does not depend on the sequences $\mathbf{H}$ or $\mathbf{q}$. That Theorem 4.19 is valid, however, follows easily from the proof of [89, Theorem 3.3]. This can be checked as follows. It is clearly stated in the proof of [89, Theorem 3.3] that the requirement of $(\alpha, \varepsilon, \zeta)$-denseness is necessary only once, namely for the base case of the induction performed there, with the value $\varepsilon = 8^{-r(r-1)/2}\delta$ given above. The values of the various constants are also explicitly stated in the proof. In particular, the value of C does indeed depend only upon r, K, δ and ζ as claimed.

To prove the 1-statement of Theorem 4.8, we need to further modify the setting from [89]: we do not have a sequence of fixed hypergraphs, but instead a sequence of random objects $\mathcal{R}^{(r)}(n, p_n)$. We describe how to modify the above definitions appropriately, and explain why the transfer result we require, Corollary 4.22, follows from Theorem 4.19.

Definition 4.20 ($(\alpha, \varepsilon, \zeta)$-dense for random hypergraphs)**.** *Let* $\mathbf{p} = (p_n)_{n\in\mathbb{N}}$ *be a sequence of probabilities, and let* $\alpha, \varepsilon, \zeta \geqslant 0$ *be constants. We say the random hypergraph* $\mathcal{R}^{(r)}(n, p_n)$ *is* a.a.s. $(\alpha, \varepsilon, \zeta)$-dense *if a.a.s. for* $\mathcal{R}_n = \mathcal{R}^{(r)}(n, p_n)$*, the following is true. For every* n*-vertex graph* G *on* $[n]$ *with at least* $(\alpha + \varepsilon)\binom{n}{2}$ *edges, the number of copies of* K_r *in* G *induced by hyperedges of* $\mathcal{R}_n$ *is at least* $\zeta e(\mathcal{R}_n)$*.*

Next, we modify the definition of boundedness.

Definition 4.21 ($(K, \mathbf{q})$-bounded for random hypergraphs)**.** *Let* $\mathbf{p} = (p_n)_{n\in\mathbb{N}}$ *and* $\mathbf{q} = (q_n)_{n\in\mathbb{N}}$ *be sequences of probabilities and* $K \geqslant 1$ *be a constant. We say that the random hypergraph* $\mathcal{R}^{(r)}(n, p_n)$ *is* a.a.s. $(K, \mathbf{q})$-bounded *if the following holds a.a.s. for* $\mathcal{R}_n = \mathcal{R}^{(r)}(n, p_n)$*. For each* $i \in [\binom{r}{2} - 1]$ *and* $\tilde{q} \geqslant q_n$*, we have*

$$\mu_i(\mathcal{R}_n, \tilde{q}) \leqslant K\tilde{q}^{2i} \cdot \frac{e(\mathcal{R}_n)^2}{n^2}\,.$$

Using these definitions we obtain the following transference result using random hypergraphs as a corollary to Theorem 4.19.

Corollary 4.22. *Given* $r \geqslant 3$*,* $K \geqslant 1$*,* $\delta > 0$*,* $\zeta > 0$ *and* $(\omega_n)_{n\in\mathbb{N}}$ *with* $\omega_n \to \infty$ *as* $n \to \infty$*, let* $\varepsilon := \delta/8\binom{r}{2}$*. There exists* $C > 1$ *such that the following is true. Let* $\mathbf{p} = (p_n)_{n\in\mathbb{N}}$ *be a sequence of probabilities such that* $\mathcal{R}^{(r)}(n, p_n)$ *is a.a.s.* $\left(\frac{r-2}{r-1}, \varepsilon, \zeta\right)$*-dense. Let* $\mathbf{q} = (q_n)_{n\in\mathbb{N}}$ *be a sequence of probabilities such that* $Cq_n < 1/\omega_n$*, such that for every integer* L*, a.a.s.* $q_n^{r(r-1)/2} \cdot e\big(\mathcal{R}^{(r)}(n, p_n)\big) > L$*, and such that* $\mathcal{R}^{(r)}(n, p_n)$ *is a.a.s.* $(K, \mathbf{q})$*-bounded. Then for* $G = G(n, Cq_n)$ *and* $\mathcal{R}_n = \mathcal{R}^{(r)}(n, p_n)$ *a.a.s.* $\mathcal{R}_n$ *is* δ*-Turánnical for* G*.*

Proof. Given $r \geqslant 3$, $K \geqslant 1$, $\delta > 0$, $\zeta > 0$ and $(\omega_n)_{n\in\mathbb{N}}$ with $\omega_n \to \infty$ as $n \to \infty$, let C be the constant returned by Theorem 4.19. Let $\mathbf{p}$ and $\mathbf{q}$ be sequences of probabilities satisfying the conditions of the corollary.

We define a property $\mathcal{A}_n$ of r-uniform hypergraphs as follows. An n-vertex hypergraph $\mathcal{H}_n$ has property $\mathcal{A}_n$ if for all n-vertex graphs H with $V(H) = V(\mathcal{H}_n)$ and $e(H) \geqslant \left(\frac{r-2}{r-1} + \varepsilon\right)\binom{n}{2}$ the number of copies of K_r in H induced by hyperedges of $\mathcal{H}_n$ is at least $\zeta e(\mathcal{H}_n)$.

We claim that there is a monotone function $\nu(n)$ tending to zero as n tends to infinity with the following properties.

(*a*) Let $P_1(n)$ be the probability that $\mathcal{R}_n = \mathcal{R}^{(r)}(n, p_n)$ has the property $\mathcal{A}_n$. Then $P_1(n) \geqslant 1 - \nu(n)$.

(*b*) There is a function $L(n)$ tending to infinity such that the probability $P_2(n)$ that for $\mathcal{R}_n = \mathcal{R}^{(r)}(n, p_n)$

$$q_n^{r(r-1)/2} \cdot e\big(\mathcal{R}_n\big) > L(n) \tag{4.11}$$

is at least $1 - \nu(n)$.

(*c*) The probability $P_3(n)$ that, for $\mathcal{R}_n = \mathcal{R}^{(r)}(n, p_n)$, we have for each $i \in [\binom{r}{2} - 1]$ and $\tilde{q} \geqslant q_n$

$$\mu_i(\mathcal{R}_n, \tilde{q}) \leqslant K\tilde{q}^{2i} \cdot \frac{e(\mathcal{R}_n)^2}{n^2}\,, \tag{4.12}$$

is at least $1 - \nu(n)$.

Items (*a*) and (*c*) are immediate from the definitions of $(\frac{r-2}{r-1}, \varepsilon, \zeta)$-denseness and $(K, \mathbf{q})$-boundedness, respectively. Item (*b*) is immediate from the fact that for each L, a.a.s. $q_n^{r(r-1)/2} \cdot e\big(\mathcal{R}_n\big) > L$ holds.

Let n_0 be such that $\nu(n_0) < \frac{1}{3}$. We fix a sequence $\mathbf{R} = (\mathcal{R}_n)_{n\in\mathbb{N}}$ of hypergraphs in the following way. For each $n \geqslant n_0$, consider the set of all n-vertex hypergraphs satisfying Property $\mathcal{A}_n$, (4.11), and (4.12). This set is non-empty by choice of n_0. Now let $\mathcal{R}_n$ be the element of this set which maximises the probability $P_4(n)$ that the random graph $G = G(n, Cq_n)$ possesses a subgraph with at least $(\frac{r-2}{r-1} + \delta) \cdot e(G)$ edges which is undetected by $\mathcal{R}_n$. For $n < n_0$ let $\mathcal{R}_n$ be an arbitrary n-vertex hypergraph.

We deduce from Property $\mathcal{A}_n$ that $\mathbf{R}$ is $\big(\frac{r-2}{r-1}, \varepsilon, \zeta\big)$-dense (in the sense of Definition 4.17), from (4.12) that $\mathbf{R}$ is $(K, \mathbf{q})$-bounded (in the sense of Definition 4.18), and from (4.11) that $\mathbf{R}$ satisfies $q_n^{r(r-1)/2} \cdot e(\mathcal{R}_n) \to \infty$. It follows that we can apply Theorem 4.19 to $\mathbf{R}$, which implies that the probability $P_4(n)$ tends to zero as n tends to infinity. Consequently, with probability at least $1 - \big((1 - P_1(n)) + (1 - P_2(n)) + (1 - P_3(n))\big) - P_4(n) \geqslant 1 - 3\nu(n) - P_4(n) = 1 - o(1)$, the random hypergraph $\mathcal{R}^{(r)}(n, p_n)$ detects every subgraph of $G = G(n, Cq_n)$ with at least $(\frac{r-2}{r-1} + \delta) \cdot e(G)$ edges. Hence $\mathcal{R}^{(r)}(n, p_n)$ is a.a.s. δ-Turánnical for $G(n, Cq_n)$. $\square$

To prove the 1-statement of Theorem 4.8 it now suffices to check that the conditions of Theorem 4.8 guarantee that $\mathcal{R}^{(r)}(n, p)$ satisfies the conditions of Corollary 4.22. We will make use of the Chernoff bound for a binomial random variable

X (see, e.g., [55, Theorem 2.1])

$$\mathbb{P}\big(X \geqslant (1+\gamma)\mathbb{E}X\big) \leqslant \exp(-\gamma^2\mathbb{E}X/3)\,, \qquad \text{for } \gamma \leqslant 1/2\,. \tag{4.13}$$

The last tool we shall need for our proof of Theorem 4.8 is a counterpart of Theorem 4.9 for random graphs due to Kohayakawa, Rödl and Schacht.

Theorem 4.23 (Kohayakawa, Rödl & Schacht [58])**.** *Given any $r \in \mathbb{N}$ and $\varepsilon > 0$, there exists $\delta > 0$ such that for any sequence of probabilities $(q_n)_{n\in\mathbb{N}}$ with $\liminf_n q_n > 0$ the following is a.a.s. true for the random graph $G = G(n, q_n)$. If $G' \subseteq G$ is a graph with at least $(1+\varepsilon)\left(\frac{r-2}{r-1}\right) e(G)$ edges, then there are at least $\delta q_n^{\binom{r}{2}} n^r$ copies of K_r in G'.*

Kohayakawa, Rödl and Schacht prove their result for a wider range of probabilities allowing q_n's decrease roughly at the speed $n^{-\frac{1}{r-1}}$; however we do not need this stronger result. Actually, in our setting when $\liminf_n q_n > 0$, Theorem 4.23 has a relatively simple proof using Szemerédi's Regularity Lemma. Let us remark that Theorem 4.23 was one of the early contributions to the Kohayakawa-Łuczak-Rödl conjecture, and thus it is in a sense obsolete. However, it turns out that the recent approaches due to Conlon and Gowers, and Schacht do not handle this easiest case of dense random graphs.

Proof of Theorem 4.8. Given r and $\varepsilon \in (0, 1/(r-2))$, set $\delta' := \varepsilon$ and $\varepsilon' := \delta'/8^{\binom{r}{2}}$. Let $\zeta > 0$ be the constant provided by Theorem 4.9 for r and ε'. Now set

$$K' := r^{2r+5} 2^{r^2+3} \tag{4.14}$$

and let C' be the constant returned by Corollary 4.22 for input r, K', δ' and ζ. Let δ^* be given by Theorem 4.23 for input parameters ε and r. Set

$$c := \tfrac{1}{16}\left(\tfrac{1}{r-1} - \varepsilon\tfrac{r-2}{r-1}\right) \qquad \text{and} \qquad C := \max\left\{\tfrac{8}{\zeta}, C'^{(r+1)(r-2)/2}, \tfrac{2}{\delta^*}\right\}. \tag{4.15}$$

The constants c and C from (4.15) define the thresholds for the 0-statement and 1-statement of Theorem 4.8. Let $\mathbf{p} = (p_n)_{n\in\mathbb{N}}$ and $\mathbf{q} = (q_n)_{n\in\mathbb{N}}$ be given. We let $\mathcal{T}_n$ denote the event that $\mathcal{R}^{(r)}(n, p_n)$ is ε-Turánnical for $G(n, q_n)$.

First we prove the 0-*statement.* Since adding hyperedges to a sequence of hypergraphs does not destroy their property of being a.a.s. ε-Turánnical for $G(n, q_n)$, we can assume that

$$p_n = c\big(nq_n^{(r+1)/2}\big)^{2-r} \quad \text{and hence} \quad q_n = c'\big(np_n^{1/(r-2)}\big)^{-2/(r+1)}\,, \tag{4.16}$$

where $c' := c^{2/((r+1)(r-2))}$. In particular, since $1 \geqslant p_n$, we have that

$$q_n \gg \frac{1}{n} \,. \tag{4.17}$$

Let Y be the random variable counting the hyperedges of $\mathcal{R}^{(r)}(n, p_n)$ which induce copies of K_r in $G = G(n, q_n)$. Since $\mathcal{R}^{(r)}(n, p_n)$ chooses each of the copies of K_r in G independently with probability p_n we have

$$\mathbb{E}Y = \binom{n}{r} q_n^{\binom{r}{2}} p_n \,. \tag{4.18}$$

Plugging in (4.16) we have

$$n \overset{(4.17)}{\leqslant} \frac{c}{2r!} n^2 q_n \leqslant \mathbb{E}Y \leqslant n^2 q_n \,.$$

Recall that we are dealing with two random objects, $G(n, q_n)$ and $\mathcal{R}^{(r)}(n, p_n)$. In the following argumentation we shall first perform the random experiment for $G(n, q_n)$ and then the one for $\mathcal{R}^{(r)}(n, p_n)$.

Let us first expose the graph $G(n, q_n)$. The Chernoff bound (4.9) implies that the probability that $G(n, q_n)$ has less than $q_n n^2/4$ edges tends to zero. Moreover, the random variable X counting copies of K_r in $G(n, q_n)$ has expectation $\binom{n}{r} q_n^{r(r-1)/2}$ and variance $\mathcal{O}\big(n^r q_n^{r(r-1)/2}\big)$ (see for example Lemma 3.5 of [55]). Hence, applying Chebyshev's inequality, we obtain that

$$\mathbb{P}\big[X \geqslant 2\tbinom{n}{r} q_n^{r(r-1)/2}\big] = o(1) \,. \tag{4.19}$$

We now expose the hypergraph $\mathcal{R}^{(r)}(n, p_n)$. Observe that the random variable Y has distribution $\mathrm{Bin}\,\big(X, p_n\big)$. From the Chernoff bound (4.13) and from (4.19) we infer that a.a.s. Y does not exceed $4\binom{n}{r} q_n^{r(r-1)/2} p_n$. Consequently, we a.a.s. have

$$\begin{aligned}
Y &\leqslant 4\tbinom{n}{r} q_n^{\binom{r}{2}} p_n \\
\Big[\text{using } \tbinom{n}{r} < n^r \text{ and } \tbinom{r}{2} = 1 + \tfrac{(r-2)(r+1)}{2}\Big] &< 4 q_n n^2 n^{r-2} q_n^{\frac{(r-2)(r+1)}{2}} p_n \\
&\overset{(4.16)}{=} 4 q_n n^2 n^{r-2} c \big(n p_n^{\frac{1}{r-2}}\big)^{-(r-2)} p_n \\
&\overset{(4.15)}{=} \left(\frac{1}{r-1} - \varepsilon \frac{r-2}{r-1}\right) \frac{q_n n^2}{4} \\
&\leqslant \left(\frac{1}{r-1} - \varepsilon \frac{r-2}{r-1}\right) e(G) \,.
\end{aligned}$$

Hence, a.a.s. $\mathcal{R}_n = \mathcal{R}^{(r)}(n, p_n)$ does not detect some subgraph G' of G which is obtained by deleting at most $(\frac{1}{r-1} - \varepsilon\frac{r-2}{r-1})e(G)$ edges from G. In particular, $e(G') \geqslant (1+\varepsilon)\frac{r-2}{r-1}e(G)$, which finishes the proof of the 0- statement.

We now turn to the 1-*statement*. Again, by monotonicity, we can assume that

$$p_n = C\big(nq_n^{(r+1)/2}\big)^{2-r} \quad \text{and hence} \quad q_n = C_q\big(np_n^{1/(r-2)}\big)^{-2/(r+1)}, \tag{4.20}$$

where $C_q := C^{2/((r+1)(r-2))} \geqslant C'$. Since $p_n \leqslant 1$ and $q_n \leqslant 1$ we have that

$$q_n \geqslant C_q n^{-2/(r+1)} \quad \text{and} \quad p_n \geqslant Cn^{2-r}. \tag{4.21}$$

We can assume that either $\liminf_n q_n > 0$, or $q_n = o(1)$. In the former case we mimic our proof of Theorem 4.3 while in the latter case we apply Corollary 4.22.

Let us first prove the 1-statement when $\liminf_n q_n > 0$. We repeat the proof strategy of the 1-statement of Theorem 4.3. Suppose that G' is an arbitrary graph on the vertex set $[n]$ with at least $\delta q_n^{\binom{r}{2}} n^r$ copies of K_r. The probability that G' is not detected by $\mathcal{R}^{(r)}(n, p_n)$ is at most

$$(1-p_n)^{\delta q_n^{\binom{r}{2}} n^r}.$$

Suppose now that a random graph $G = G(n, q_n)$ is given. We can assume that G has at most $q_n n^2$ edges as this property is satisfied a.a.s. Consequently, G contains at most $2^{q_n n^2}$ subgraphs G' on the same vertex set. By Theorem 4.23 we a.a.s. have that each such a subgraph with at least $(1+\varepsilon)\left(\frac{r-2}{r-1}\right)e(G)$ edges contains at least $\delta q_n^{\binom{r}{2}} n^r$ copies of K_r. Therefore, the union bound over all such graphs G' gives that

$$\begin{aligned}
\mathbb{P}\Big[\mathcal{R}^{(r)}(n, p_n) \text{ is not } \varepsilon\text{-Turánnical for } G(n, q_n)\Big] &\leqslant 2^{q_n n^2} \times (1-p_n)^{\delta q_n^{\binom{r}{2}} n^r} \\
&\leqslant \exp\left(q_n n^2 - p_n \delta q_n^{\binom{r}{2}} n^r\right) \\
&\overset{(4.20)}{=} \exp\left(q_n n^2 - Cn^2 q_n \delta\right) \\
&\overset{(4.15)}{\to} 0,
\end{aligned}$$

and the statement follows in this case.

Let us now focus on the 1-statement in the case $q_n = o(1)$. The claim will follow from Corollary 4.22 (with parameters r, K', δ', ζ, C') applied to the sequences of probabilities $\mathbf{p}$ and $\mathbf{q}' = (q'_n)_{n\in\mathbb{N}} := \mathbf{q}/C'$, together with the following claim.

Claim 4.8.1. *We have that*

(*a*) *for every L a.a.s.* $(q'_n)^{r(r-1)/2} \cdot e\big(\mathcal{R}^{(r)}(n,p_n)\big) > L$,

(*b*) $\mathcal{R}^{(r)}(n,p_n)$ *is a.a.s.* $\big(\frac{r-1}{r-2},\varepsilon',\zeta\big)$*-dense, and*

(*c*) $\mathcal{R}^{(r)}(n,p_n)$ *is a.a.s.* $(K',\mathbf{q}')$*-bounded.*

Proof of Claim 4.8.1. We first verify (*a*). We have

$$\mathbb{E}\Big(e\big(\mathcal{R}^{(r)}(n,p_n)\big)\Big) = p_n\binom{n}{r},$$

which tends to infinity by (4.21). Consequently, the Chernoff bound (4.9) guarantees that a.a.s. $\mathcal{R}^{(r)}(n,p_n)$ has at least $p_n\binom{n}{r}/2$ hyperedges. Now we have

$$\frac{(q'_n)^{\binom{r}{2}}p_n\binom{n}{r}}{2} \overset{(4.20)}{=} \frac{(q'_n)^{\frac{r^2-r}{2}}Cn^{2-r}q_n^{(r+1)(2-r)/2}\binom{n}{r}}{2} = \Omega\left(q_n n^{2-r}\binom{n}{r}\right),$$

and by (4.21) this tends to infinity.

Now we verify (*b*). Given an n-vertex graph H with $e(H) \geqslant \big(\frac{r-2}{r-1}+\varepsilon'\big)\binom{n}{2}$, by Theorem 4.9, H contains at least ζn^r copies of K_r. It follows that the expected number of hyperedges of $\mathcal{R}_n = \mathcal{R}^{(r)}(n,p_n)$ which induce copies of K_r in H is at least $\zeta n^r p_n$. By the Chernoff bound (4.9), the probability that less than $\zeta n^r p_n/2$ copies of K_r in H are induced by hyperedges of $\mathcal{R}_n$ is at most

$$\exp\left(-\frac{\zeta n^r p_n}{8}\right) \overset{(4.21)}{\leqslant} \exp\left(-\frac{C\zeta n^2}{8}\right) \overset{(4.15)}{=} o(2^{-n^2}).$$

Applying the union bound (on at most $2^{\binom{n}{2}}$ graphs H) we conclude that the probability that there exists any n-vertex graph H with at least $\big(\frac{r-1}{r-2}+\varepsilon'\big)\binom{n}{2}$ edges and less than $3\zeta\binom{n}{r}p_n/2 \leqslant \zeta n^r p_n/2$ copies of K_r on hyperedges of $\mathcal{R}_n$ tends to zero as n tends to infinity. Furthermore, applying the Chernoff bound (4.13) in conjunction with (4.21), the probability that $\mathcal{R}^{(r)}(n,p)$ has more than $3p_n\binom{n}{r}/2$ hyperedges tends to zero as n tends to infinity. It follows that for $\mathcal{R}_n$ a.a.s. every n-vertex graph H with more than $\big(\frac{r-2}{r-1}+\varepsilon'\big)\binom{n}{2}$ edges has at least $\zeta e(\mathcal{R}_n)$ copies of K_r on hyperedges of $\mathcal{R}_n$. Therefore, $\mathcal{R}^{(r)}(n,p_n)$ is a.a.s. $\big(\frac{r-2}{r-1},\varepsilon',\zeta\big)$-dense.

Now we prove (*c*). We need to show that $\mathcal{R}_n = \mathcal{R}^{(r)}(n,p_n)$ a.a.s. has the property that for each $1 \leqslant i \leqslant \binom{r}{2} - 1$ and each $\tilde{q} \geqslant q'_n$, we have

$$\mu_i(\mathcal{R}_n, \tilde{q}) \leqslant K' \tilde{q}^{2i} \frac{e(\mathcal{R}_n)^2}{n^2} . \tag{4.22}$$

We will show that (4.22) holds for all $1 \leqslant i \leqslant \binom{r}{2} - 1$ and $\tilde{q} \geqslant q'_n$ provided that $\mathcal{R}_n$ obeys a simple bound (inequality (4.24) below); this bound will turns out to hold a.a.s. for our random hypergraph.

Given a hypergraph $\mathcal{R}_n$ and two distinct vertices u and v, let F_1 and F_2 be two hyperedges containing u and v and intersecting in a set A of j vertices. Then the probability $P_{i,j}$ that both F_1 and F_2 contain at least i edges of the random graph $G = G(n, \tilde{q})$, not counting uv, can be bounded as follows. We use the random variables $X_A := |E(G[A]) - uv|$, $X_{F_1} := e(G[F_1 - A]) + e(G[F_1 - A, A])$, and $X_{F_2} := e(G[F_2 - A]) + e(G[F_2 - A, A])$. Then

$$\begin{aligned} P_{i,j} &\leqslant \sum_{k=0}^{\binom{j}{2}-1} \mathbb{P}(X_A = k)\mathbb{P}(X_{F_1} \geqslant i-k)\mathbb{P}(X_{F_2} \geqslant i-k) \\ &\leqslant \sum_{k=0}^{\binom{j}{2}-1} \binom{\binom{j}{2}-1}{k} \tilde{q}^k \left(\binom{\binom{r}{2}-\binom{j}{2}}{i-k} \tilde{q}^{i-k} \right)^2 \\ &\leqslant 2^{\binom{j}{2}-1+2(\binom{r}{2}-\binom{j}{2})} \sum_{k=0}^{\binom{j}{2}-1} \tilde{q}^{2i-k} \leqslant j^2 2^{r^2} \cdot \tilde{q}^{2i+1-\binom{j}{2}} . \end{aligned} \tag{4.23}$$

Let $N(j)$ count the number of pairs of hyperedges in $\mathcal{R}_n$ intersecting in exactly j vertices. Then we have

$$\begin{aligned} \mu_i(\mathcal{R}_n, \tilde{q}) = \mathbb{E}\left[\sum_{\substack{u,v \\ u \neq v}} \deg_i^2(u, v, G(n, \tilde{q})) \right] &= \sum_{\substack{u,v \\ u \neq v}} \sum_{\substack{F_1 \in \mathcal{E}(\mathcal{R}_n) \\ F_1 \ni u,v}} \sum_{\substack{F_2 \in \mathcal{E}(\mathcal{R}_n) \\ F_2 \ni u,v}} P_{i,|F_1 \cap F_2|} \\ = \sum_{j=2}^{r} N(j) j(j-1) P_{i,j} &\overset{(4.23)}{\leqslant} r^4 2^{r^2} \sum_{j=2}^{r} N(j) \tilde{q}^{2i+1-\binom{j}{2}} . \end{aligned}$$

It follows that $\mathcal{R}_n$ satisfies (4.22) if we have, for each $2 \leqslant j \leqslant r$ and $\tilde{q} \geqslant q'_n$,

$$r^5 2^{r^2} \cdot N(j) \cdot \tilde{q}^{1-\binom{j}{2}} \leqslant K' \frac{e(\mathcal{R}_n)^2}{n^2} . \tag{4.24}$$

Since $j \geqslant 2$ we have $1 - \binom{j}{2} \leqslant 0$. Therefore, the left-hand side of (4.24) is non-increasing in $\tilde{q}$. The right-hand side of (4.24) does not depend upon $\tilde{q}$. It follows that we need only verify that a.a.s. $\mathcal{R}_n = \mathcal{R}^{(r)}(n, p_n)$ satisfies (4.24) for each $2 \leqslant j \leqslant r$, with $\tilde{q} = q'_n$. We have that a.a.s. $e(\mathcal{R}^{(r)}(n, p_n)) \geqslant p_n\binom{n}{r}/2 \geqslant p_n n^r/(2r^r)$, by the Chernoff bound (4.9). So it is enough to show that a.a.s. for each $2 \leqslant j \leqslant r$ we have

$$N(j) \leqslant \frac{K'}{r^5 2^{r^2}} (q'_n)^{\frac{(j-2)(j+1)}{2}} \frac{p_n^2 n^{2r-2}}{4r^{2r}} \overset{(4.14)}{=} 2(q'_n)^{\frac{(j-2)(j+1)}{2}} p_n^2 n^{2r-2} \,. \tag{4.25}$$

To show that (4.25) holds, we first consider the case $j = r$. Observe that $N(r)$ is simply the number of hyperedges in $\mathcal{R}^{(r)}(n, p_n)$, and is therefore (by the Chernoff bound (4.13)) a.a.s. at most $2p_n\binom{n}{r} \leqslant 2p_n n^r$. Substituting $q'_n \geqslant \left(np_n^{1/(r-2)}\right)^{-2/(r+1)}$ into the right-hand side of (4.25) (for $j = r$), we have

$$2(q'_n)^{\frac{(r-2)(r+1)}{2}} p_n^2 n^{2r-2} \geqslant 2\Big(np_n^{\frac{1}{r-2}}\Big)^{2-r} p_n^2 n^{2r-2} = 2p_n n^r \ .$$

Therefore (4.25) holds for $j = r$.

Suppose now that $2 \leqslant j \leqslant r - 1$. Then we have

$$\mathbb{E}(N(j)) = \binom{n}{r}\binom{r}{j}\binom{n-r}{r-j} p_n^2 = \mathcal{O}(n^{2r-j} p_n^2) \,.$$

We have by (4.21) that $q'_n = \Omega\big(n^{-\frac{2}{r+1}}\big) = \omega\big(n^{-\frac{2}{j+1}}\big)$ for each $2 \leqslant j \leqslant r - 1$. Consequently,

$$\mathbb{E}(N(j)) = \mathcal{O}(n^{2r-j} p_n^2) = \mathcal{O}(n^{2-j} p_n^2 n^{2r-2}) = o\Big((q'_n)^{\frac{(j-2)(j+1)}{2}} p_n^2 n^{2r-2}\Big) \,.$$

By Markov's inequality, (4.25) holds a.a.s. for every $2 \leqslant j \leqslant r - 1$. This completes the proof that $\mathcal{R}^{(r)}(n, p_n)$ is a.a.s. $(K', \mathbf{q}')$-bounded. $\square$

It follows that a.a.s. $\mathcal{R}^{(r)}(n, p_n)$ satisfies the conditions to apply Corollary 4.22, that is, a.a.s. $\mathcal{R}^{(r)}(n, p_n)$ is ε-Turánnical for $G(n, q_n)$. $\square$

4.7 Sharp thresholds

In this section we use Friedgut's [44] condition for sharp thresholds to prove that the threshold we obtained in Theorem 4.4 is sharp. For a background on threshold phenomena we refer the reader to [44]. We show the following result.

Theorem 4.24. *For every integer $r \geqslant 3$ there are $c, C > 0$ and a sequence of* *numbers $(c_n \in (c, C))_{n\in\mathbb{N}}$ such that for every $\gamma > 0$ we have*

$$\lim_{n\to\infty} \mathbb{P}\big(\mathcal{R}^{(r)}\big(n, (c_n-\gamma)n^{3-r}\big) \textit{ is Turánnical}\big) = 0 \qquad \textit{and,}$$
$$\lim_{n\to\infty} \mathbb{P}\big(\mathcal{R}^{(r)}\big(n, (c_n+\gamma)n^{3-r}\big) \textit{ is Turánnical}\big) = 1 \,.$$

As usual it is reasonable to conjecture that the sequence (c_n) in this theorem converges, and as usual in the field we are not able to prove this.

Before we can state Friedgut's result we need to introduce some notation. Given two hypergraphs $\mathcal{G}$ and $\mathcal{M}$ with $v(\mathcal{G}) \geqslant v(\mathcal{M})$ we write $\mathcal{G} \cup \mathcal{M}^*$ for the random hypergraph obtained from the following random experiment. Let ϕ be a (uniformly chosen) random injection from $V(\mathcal{M})$ to $V(\mathcal{G})$ and for each hyperedge F of $\mathcal{M}$ add the hyperedge $\phi(F)$ to $\mathcal{G}$ (without creating multiple hyperedges). A family of r-uniform hypergraphs is called a *hypergraph property* if it is closed under isomorphism and under adding hyperedges.

Friedgut formulates his result for graphs. Here, we use the corresponding hypergraph result, specialised to our situation; see also [43] for a discussion of this result and for extensions to other combinatorial structures.

Theorem 4.25 (Friedgut [44, Theorem 2.4])**.** *Suppose that Theorem 4.24 does not hold for some $r \geqslant 3$. Then there exists a sequence $p = p_n$, $\tau > 0$, a fixed r-uniform hypergraph $\mathcal{M}$ with*

$$\mathbb{P}\big(\mathcal{M} \subseteq \mathcal{R}^{(r)}(n,p)\big) > \tau \,, \tag{4.26}$$

and $\alpha > 0$ with

$$\alpha < \mathbb{P}\big(\mathcal{R}^{(r)}(n,p) \textit{ is Turánnical}\big) < 1 - 3\alpha \,, \tag{4.27}$$

and a constant $\varepsilon > 0$ such that, for every hypergraph property $\mathcal{P}$ which satisfies that $\mathcal{R}^{(r)}(n,p)$ is a.a.s. in $\mathcal{P}$, the following holds. There exists an infinite set $Z \subseteq \mathbb{N}$ and for each $n \in Z$ a hypergraph $\mathcal{G}_n \in \mathcal{P}$ such that

$$\mathbb{P}\big(\mathcal{G}_n \cup \mathcal{M}^* \textit{ is Turánnical}\big) > 1 - \alpha \,, \tag{4.28}$$
$$\mathbb{P}\big(\mathcal{G}_n \cup \mathcal{R}^{(r)}(n,\varepsilon p) \textit{ is Turánnical}\big) < 1 - 2\alpha \,. \tag{4.29}$$

The statement of Theorem 4.25 may seem technical and untransparent, so let us try to clarify it (following [43]). That is we try to motivate the conditions above for a property *not* having a sharp threshold. A primal example of such a property is the *subhypergraph containment property*, i.e., a property $\mathcal{P}_\mathcal{M}$ consisting of all hypergraphs containing a fixed hypergraph $\mathcal{M}$. Observe that the property $\mathcal{P}_\mathcal{M}$ is not very sensitive to adding a sparse random hypergraph, while it is very sensitive to adding a randomly placed copy of $\mathcal{M}$; these two features are captured in (4.28), and (4.29), respectively. Friedgut's result (an instance of which we give as Theorem 4.25) says that each property without a sharp can be approximated by subhypergraph containment properties.

With Theorem 4.25 at hand, we can now give a proof of Theorem 4.24. It turns out that we do not need to utilize Theorem 4.25 in its full strength; in particular we shall not use assertion (4.26).

Proof of Theorem 4.24. Suppose that Theorem 4.24 does not hold for some $r \geqslant 3$. Let p_n, the r-uniform hypergraph $\mathcal{M}$, and $\alpha > 0$ be given by Theorem 4.25. In particular, by (4.27) we have that $\alpha < 1/4$. It follows from (4.27) and from Theorem 4.4 that

$$cn^{3-r} \leqslant p \leqslant Cn^{3-r} ,$$

for some absolute constants $c, C > 0$. Let $\beta := \frac{1}{2e(\mathcal{M})}$ and let $\mathcal{P}$ be the family of n-vertex hypergraphs which detect every n-vertex graph F with at least $\beta\binom{n}{r}$ r-cliques. It follows from the proof of Theorem 4.3 that a.a.s. $\mathcal{R}^{(r)}(n,p) \in \mathcal{P}$.

Let now $Z \subseteq \mathbb{N}$ and $(\mathcal{G}_n)_{n\in Z}$ be given by Theorem 4.25. We will derive a contradiction using just a single hypergraph $\mathcal{G}_n$, $n \in Z$. Indeed, from (4.29) we see that $\mathcal{G}_n$ itself cannot be Turánnical. Let W be a graph which witnesses this, i.e., W is an n-vertex graph with more than $t_r(n)$ edges which is not detected by $\mathcal{G}_n$. By the definition of $\mathcal{P}$ and since $\mathcal{G}_n \in \mathcal{P}$, the graph W contains less than $\beta\binom{n}{r}$ r-cliques. If $\mathcal{G}_n \cup \mathcal{M}^*$ is Turánnical then at least one hyperedge of $\mathcal{M}$ must be placed on an r-clique of W. Therefore we have

$$\mathbb{P}\big(\mathcal{G}_n \cup \mathcal{M}^* \text{ is Turánnical}\big) \leqslant e(\mathcal{M})\beta < \frac{1}{2} ,$$

which contradicts (4.28). □

4.8 Random restrictions

Traditional extremal combinatorics deals with questions in the following framework. Given a combinatorial structure $\mathcal{S}$ (such as the edge set of the complete graph K_n, or the set $2^{[n]}$ of subsets of $[n]$) and a monotone increasing parameter $f\colon 2^{\mathcal{S}} \to \mathbb{N}$ (such as the minimum degree of $H \subseteq K_n$, or the number of sets in the set family $H \subseteq 2^{[n]}$), we ask:

> What is the maximum possible value $f(H)$ for $H \subseteq \mathcal{S}$ satisfying a set of restrictions $\mathcal{R}$?

Often the restrictions $\mathcal{R}$ are simply *all* substructures of $\mathcal{S}$ of a certain type. For example, in the setting of Turán's theorem every r-tuple of vertices forbids a clique; in that of Sperner's theorem [94], every pair of sets $A \subseteq B \subseteq [n]$ is forbidden to be in the set family $H \subseteq 2^{[n]}$.

In this framework there are two places where randomness may come into play. Firstly, one could choose $\mathcal{S}$ to be a random structure (and thus H be a substructure of a random structure). A famous example of this type of randomness is the Kohayakawa-Łuczak-Rödl conjecture concerning a version of Turán's theorem for random graphs (see [57]) mentioned already in Section 1.5. Versions of the famous Erdős-Ko-Rado theorem for random hypergraphs as studied by Balogh, Bohman, and Mubayi [14] form another example.

Secondly, the restriction set can be relaxed to a random subset of all possible restrictions $\mathcal{R}$. This is exemplified in Theorems 4.3 and 4.4 in the context of Turán's theorem. Moreover, the two types of randomness can be combined, as shown in Theorem 4.8.

Obviously, similar randomised versions can be formulated for many other problems. Probably the closest one to the results obtained in this chapter would be a variant of the Erdős-Stone theorem about the extremal number of H-free graphs with random restrictions. While the statement and the proof of Theorem 4.3 translates *mutatis mutandis* to that setting when $\chi(H) \geqslant 3$, obtaining either a proof for $\chi(H) = 2$ or an analogue of Theorem 4.4 seem to be significantly harder. We conclude by mentioning two additional problems which seem interesting for further research.

Ramsey theory. Graph Ramsey theory deals with estimating the parameter $R(H)$, which is the smallest number n such that any two-colouring of edges of the complete graph K_n contains a monochromatic copy of H.

In a randomised version of this problem of the first type mentioned above, we colour the edges of the random graph $G(n,q)$ instead of K_n and search for a monochromatic copy of H in such a colouring. The threshold for this problem was determined by Rödl and Ruciński [86] (see also Friedgut, Rödl, Schacht [45], Conlon and Gowers [28] for some recent progress).

Concerning the second approach for randomisation mentioned above, we suggest considering the following problem. Given n and a probability p, let $\mathcal{R}(n,p)$ be a set of copies of H in K_n obtained by picking H-copies independently at random with probability p from the set of all copies of H in K_n. What is the threshold $p = p_n$ such that a.a.s. $\mathcal{R} = \mathcal{R}(n,p)$ has the property that for every two-edge-colouring of K_n, there is a monochromatic copy of H contained in $\mathcal{R}$?

Very recently, a solution to this problem was announced by Gugelmann, Person, Steger and Thomas [53] (at least in the case $H = K_r$). Their threshold probability is $p_n = \Theta(n^{2-r})$.

VC-dimension. The celebrated Sauer-Shelah Lemma [88, 91] states that if $\mathcal{A}$ is a family of subsets of $[n]$ with $|\mathcal{A}| > \binom{n}{0} + \ldots + \binom{n}{k-1}$ then there is a set $X \subseteq [n]$ of size k which is *shattered* by $\mathcal{A}$, i.e., for every $Y \subseteq X$, there is $A \in \mathcal{A}$ such that $Y = X \cap A$.

A randomised variant of this Lemma of the first type mentioned above would generate a random family $\mathcal{X} = \binom{[n]}{k}_p$ of k-sets in $[n]$, each k-set being present in this family independently with probability $p = p_n$. The question is then: How large must $|\mathcal{A}|$ be in order to guarantee a shattered k-set $X \in \mathcal{X}$?

A randomised version of the second type, instead, would randomise the concept of a shattering in the Sauer-Shelah Lemma. More precisely, a *p-shattering* does not require every subset $Y \subseteq X$ to be represented as $X \cap A$ for some $A \in \mathcal{A}$, but only for each $X \subseteq [n]$ of size k a family of subsets Y which are selected randomly and independently from 2^X with probability p. The question then is: Given $0 < c \leqslant 1$, what is the threshold $p = p_n$ such that a.a.s. there exists a set family with $c\big(\binom{n}{0} + \ldots + \binom{n}{k-1}\big)$ members which does not even p-shatter any k-set in $[n]$?

Chapter 5

A deterministic construction of Turánnical hypergraphs

5.1 Introduction

Turán's Theorem is a primal example of a result which is *stable*: If G is an n-vertex K_r-free graph with almost $t_r(n)$ edges then G looks very similar to $\mathrm{T}_r(n)$. To state the result we need to define "to look very similar". We say that an n-vertex graph G is *ε-close* to a graph H on the same vertex set if H can be obtained from G by editing (deleting/inserting) at most εn^2 edges. In this case we also say that G is *(εn^2)-near* to H. That is, the notion of being close and being near are the same except for a scaling factor which is the square of the order of the graph.

Theorem 5.1 (Erdős & Simonovits [92]). *Suppose that $r \geqslant 3$ and $\varepsilon^* > 0$ are given. Then there exists $\gamma^* > 0$ such that each ℓ-vertex graph G with no K_r and $e(G) > t_r(\ell) - \gamma^* \ell^2$ is ε^*-close to $\mathrm{T}_r(\ell)$.*

The statement "G is ε^*-close to $\mathrm{T}_r(\ell)$" in the theorem above is of course meant "...after a labelling of the vertices of $\mathrm{T}_r(\ell)$ in a suitable way". This is always the case in the sequel.

In this chapter we extend Turán's Theorem by determining the maximum number of edges a graph G on a vertex set V, $|V| = n$, may have subject to not containing any copy of K_r that touches a fixed set $M \subseteq V$. It turns out that when $(r-1)|M| \geqslant n$ the unique extremal graph is $\mathrm{T}_r(n)$; see Theorem 5.2. Therefore, an r-uniform hypergraph on the vertex set V with all the edges touching M present is Turánnical, a notion we introduced and investigated thoroughly in Chapter 4.

The case $(r-1)|M| < n$ is more complicated. In particular, there is not a unique extremal graph for the problem. We identify all the extremal graphs in Theorem 5.2. Furthermore, in Theorem 5.3 we provide a stability version of Theorem 5.2.

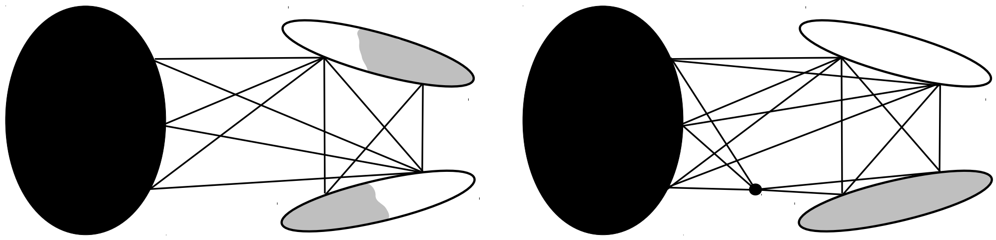
To state our results we need the following definitions. We define a function $t_r(n,m)$ by

$$t_r(n,m) := \begin{cases} t_r(n)\,, & \text{if } n \leqslant (r-1)m\,, \\ \binom{n}{2} - nm + (r-1)\binom{m+1}{2}\,, & \text{otherwise}\,. \end{cases} \tag{5.1}$$

Next, we introduce the graph classes $\mathcal{T}_r(n,m)$. We start off with the more complicated case $n > (r-1)m$. Then the class consists of *common* graphs and *sporadic* graphs. Each common graph $G \in \mathcal{T}_r(n,m)$ is constructed as follows. Initially, we take $G = \mathrm{T}_r((r-1)m)$. We then fix an arbitrary set $M \subseteq V(G)$ of size m and add $n-(r-1)m$ new vertices to G. Finally, for each of the new vertices we add edges to all other vertices except those in M. Clearly, no K_r touches M. We write $\mathcal{T}_r^*(n,m) \subseteq \mathcal{T}_r(n,m)$ for the common graphs.

Now, we describe a construction of all sporadic graphs $G \in \mathcal{T}_r(n,m)$. Again, we start with $G = \mathrm{T}_r((r-1)m)$; let $V_1,\ldots,V_{r-1}$ be its colour classes. We place arbitrarily the set M of size m inside $V_1 \cup \ldots \cup V_{r-1}$. First assume that M coincides with one of the sets $V_1,\ldots,V_{r-1}$, say $M = V_1$. Then we take an arbitrary integer p, $0 \leqslant p \leqslant \min\{r-2, n-(r-1)m\}$. We add to G vertices $v_1,\ldots,v_p$ such that the vertex v_i is adjacent to all v_j's ($j \neq i$), and all the vertices of $V_1 = M$ and all the vertices of a chosen $(r-3)$-tuple of classes $V_2, V_3, \ldots, V_{r-1}$. We require that no two vertices v_i, v_j are adjacent to the same collection of classes $V_1,\ldots,V_{r-1}$ (this is what determined the condition $p \leqslant r-2$). After adding the vertices $v_1,\ldots,v_p$ we add remaining $n-((r-1)m+p)$ vertices, each of them adjacent to all the vertices except M and the vertex itself.

It remains to provide the construction when $M \neq V_1,\ldots,V_{r-1}$. In particular, M intersects at least two sets $V_1,\ldots,V_{r-1}$. Let $0 \leqslant p \leqslant \min\{r-1, n-(r-1)m\}$ be an arbitrary integer. We add to G vertices $v_1,\ldots,v_p$ such that the vertex v_i is adjacent to all v_j's ($j \neq i$), and all the vertices of a chosen $(r-2)$-tuple of classes $V_1,\ldots,V_{r-1}$. We require that no two vertices v_i, v_j are adjacent to the same collection of classes $V_1,\ldots,V_{r-1}$. After adding the vertices $v_1,\ldots,v_p$ we add remaining $n-((r-1)m+p)$ vertices, each of them adjacent to all the vertices except M and the vertex itself. Observe that there is no r-clique touching M in either of the two constructions above. For a reason which will become apparent in our proof (cf. end of Section 5.2.4), we call the vertices $v_1,\ldots,v_p$ in both constructions *pure*.



(a) A common graph. (b) A sporadic graph.

Figure 5.1: Examples of graphs from the class $\mathcal{T}_3(n,m)$ for $n > 2m$. The set M is depicted in grey, the clique part in black.

Examples of graphs from $\mathcal{T}_r(m,n)$ are given on Figure 5.1.

By construction, it is clear that each graph in $\mathcal{T}_r(n,m)$ has n vertices and no copy of K_r intersects M. Moreover, observe that the number of edges of any graph from $\mathcal{T}_r(n,m)$ is given by the function $t_r(n,m)$ defined in (5.1) since

$$m^2\binom{r-1}{2} + m(r-2)\big(n-(r-1)m\big) + \binom{n-(r-1)m}{2} = \binom{n}{2} - nm + (r-1)\binom{m+1}{2}\,.$$

When $n \leqslant (r-1)m$ we define $\mathcal{T}_r(n,m) := \{\mathrm{T}_r(n)\}$ to contain only the $(r-1)$-partite Turán graph of order n.

We are now ready to state our main result, an extension of Turán's Theorem.

Theorem 5.2. *Given $r \geqslant 3$ and $m \leqslant n$, let G be any n-vertex graph and $M \subseteq V(G)$ contain m vertices. If no copy of K_r in G intersects M, then $e(G) \leqslant t_r(n,m)$. Moreover, if $e(G) = t_r(n,m)$ then $G \in \mathcal{T}_r(n,m)$.*

Even though an extension of Turán's Theorem, it should be noted that we do need Turán's Theorem for the proof. Further, we prove a stability version of Theorem 5.2.

Theorem 5.3. *Suppose that $r \geqslant 3$ and $\varepsilon > 0$ are given. Then there exists $\gamma > 0$ with the following property. For $m \leqslant n$, let G be any n-vertex graph and $M \subseteq V(G)$ contain m vertices. If no copy of K_r in G intersects M and $e(G) > t_r(n,m) - \gamma n^2$, then G is ε-close to a graph from $\mathcal{T}_r(n,m)$.*

Theorem 5.3 is used as an auxiliary result for proving Theorem 2.9.

5.2 Proof of Theorem 5.2 and Theorem 5.3

We provide a simultanuous proof of Theorem 5.2 and of Theorem 5.3.

Throughout the proof we use the following notation. Given a graph H and a set $X \subseteq V(H)$ we write $\Gamma_H(X)$ for the *external neighborhood of* X, $\Gamma_H(X) := \{v \in V(H) - X \,:\, \deg(v,X) \geqslant 1\}$.

Let r, n, m be fixed and let G and M satisfy the conditions of Theorem 5.2 or Theorem 5.3.

The definition of $t_r(n,m)$ suggests the following case distinction. We call the situation when $n \leqslant (r-1)m$ as *Case I* which we deal with first (both for Theorem 5.2 and Theorem 5.3). Otherwise, the situation is called *Case II* understanding of which relies on the results of Case I.

In the setting of Theorem 5.3 the parameter γ is set in Case I and Case II separately as a function of r and ε. This causes no problem as the resulting value of γ can be taken as the minimum of the corresponding parameters from each of the two cases.

5.2.1 Case I; bound and uniqueness

Our approach is inductive in essence. Our formalism is adapted so that we can then recycle our proof to get stability of Case I in Section 5.2.2.

We start by iteratively finding vertex disjoint cliques $Q_1, \dots, Q_k$ with at least r vertices in G as follows. Assume, that $Q_1, \dots, Q_{i-1}$ have already been defined for some i. Then let Q_i be an arbitrary maximum clique on at least r vertices in $G - \bigcup_{j<i} Q_j$. If no such clique exists, then set $k := i-1$ and terminate.

Now, let us establish some simple bounds on the number of edges between these cliques and the rest of G. For this purpose, set $q_i := v(Q_i) \geqslant r$ to be the order of the clique Q_i for all $i \in [k]$ and $q := \sum_{i=1}^k q_i$. Clearly, the graph $G - \bigcup_{i=1}^k V(Q_i)$ is K_r-free, and therefore

$$e\Big(G - \bigcup_{i=1}^{k} V(Q_i)\Big) \leqslant t_r(n-q)\,, \tag{5.2}$$

by Turán's Theorem. Moreover, $M \subseteq V(G) - \bigcup_{i=1}^k V(Q_i)$ and we have $\deg(v, Q_i) \leqslant r-2$ for each $v \in M$, as v is not contained in a copy of K_r by assumption. In addition, the maximality of $Q_1, \dots, Q_k$ implies that $\deg(v, Q_i) \leqslant q_i - 1$ for any $v \in V(G) - (M \cup \bigcup_{j=1}^i V(Q_i))$. Putting these three estimates together we obtain

$$\begin{aligned} e(G) \leqslant \sum_{i=1}^{k} \binom{q_i}{2} + \sum_{1 \leqslant i<j \leqslant k} (q_i - 1)q_j + t_r(n-q) + mk(r-2) \\ + (q-k)(n-m-q) =: g_n(q_1, \dots, q_k)\,. \end{aligned} \tag{5.3}$$

Observe that (5.3) defines a functions $g_n(q_1, \dots, q_\ell)$ for each number of arguments ℓ. In particular, we also allow $\ell = 0$, in which case (5.3) asserts that $g_n() = t_r(n)$. In the remainder of this case of the proof we shall investigate the family of functions $g_n(q_1, \dots, q_\ell)$. We shall show, that for all $\ell > 0$ we have $g_n() > g_n(q_1, \dots, q_\ell)$, which is a consequence of the following claim.

Lemma 5.4. *Assuming that $q = \sum_{i=1}^k q_i \leqslant n - m$ and $q_i \geqslant r$ for all $i \in [k]$ we have*

$$g_n(q_1, \dots, q_{k-1}, q_k) < g_n(q_1, \dots, q_{k-1}, q_k - 1) \qquad \text{if } q_k > r\,, \quad \text{and} \tag{5.4}$$

$$g_n(q_1, \dots, q_{k-1}, q_k) < g_n(q_1, \dots, q_{k-1}) \qquad \text{if } q_k = r\,. \tag{5.5}$$

Proof. Adding one or r vertices to a Turán graph $\mathrm{T}_r(n')$ to create a bigger Turán graph and counting the additionally created edges gives

$$t_r(n'+1) - t_r(n') = n' - \left\lfloor \frac{n'}{r-1} \right\rfloor, \quad \text{and} \tag{5.6}$$

$$t_r(n'+r) - t_r(n') = (r-1)n' + \binom{r}{2} - \left\lfloor \frac{n'+r-1}{r-1} \right\rfloor. \tag{5.7}$$

Observe that $m > 1$, or otherwise $r \leqslant q \leqslant n - 1 \leqslant (r-1)m - 1$ would lead to a contradiction. If $q_k > r$ then plugging (5.6) (with $n' = n - q$) into the definition of g in (5.3) we obtain

$$g_n(q_1, \dots, q_{k-1}, q_k - 1) - g_n(q_1, \dots, q_{k-1}, q_k) = m - \left\lfloor \frac{n-q}{r-1} \right\rfloor - 1 > 0\,, \tag{5.8}$$

proving (5.4). Similarly, if $q_k = r$ then (5.7) implies

$$g_n(q_1, \dots, q_{k-1}) - g_n(q_1, \dots, q_{k-1}, q_k) = m - \left\lfloor \frac{n-q}{r-1} \right\rfloor - 1 > 0\,, \tag{5.9}$$

proving (5.5). □

Clearly, applying Lemma 5.4 for sequentially decreasing or discarding the last argument of $g_n(q_1, \dots, q_\ell)$ gives that

$$g_n\big(v(Q_1), v(Q_2), \dots, v(Q_k)\big) = g_n(q_1, \dots, q_k) \leqslant g_n() = t_r(n)\,.$$

Moreover, equality holds only when $k = 0$, that is, when G does not contain any K_r. This proves Theorem 5.2 in the case $n \leqslant (r-1)m$.

5.2.2 Case I; stability

We revise the proof above to get a proof of Theorem 5.3 in the case $n \leqslant (r-1)m$. Let γ^* be given by Theorem 5.1 for the input parameter $\varepsilon^* := \varepsilon/2$. We can assume that $\gamma^* < \min\{1/(2r), \varepsilon/4\}$. Set $\gamma := (\gamma^*/5r)^2$. Further we assume that $n > (4r^2)/\gamma^2$ as small graphs can be dealt with separately by a standard argument[1].

First, we note that (5.8) and (5.9) can be used to strengthen Lemma 5.4 as follows.

Lemma 5.5. *Using the same notation as in Lemma 5.4, and assuming that $q > \gamma^* n/4$ and $n > \frac{4r^2}{\gamma^*}$, we have*

$$g_n(q_1,\dots,q_{k-1},q_k) < g_n(q_1,\dots,q_{k-1},q_k-1) - \frac{\gamma^* n}{4r}\,, \qquad \text{if } q_k > r \text{ and} \tag{5.10}$$

$$g_n(q_1,\dots,q_{k-1},q_k) < g_n(q_1,\dots,q_{k-1}) - \frac{\gamma^* n}{4r}\,, \qquad \text{if } q_k = r\,. \tag{5.11}$$

Let $q^* = \sum_{i=1}^k q_i$ be the total order of the cliques Q_i defined in Section 5.2.1. We claim that

$$q^* < \gamma^* n/2\,. \tag{5.12}$$

Suppose the contrary. Then discarding sequentially the cliques Q_i, Lemma 5.5 is executed with the current $q \geqslant \gamma^* n/4$ at least $\frac{q^*-\gamma^* n/4}{r} \geqslant \frac{\gamma^* n}{4r}$ times. Therefore, we have that

$$e(G) \leqslant t_r(n) - \frac{\gamma^* n}{4r} \times \frac{\gamma^* n}{4r} < t_r(n) - \gamma n^2\,, \tag{5.13}$$

a contradiction.

We use a crude estimate $\deg(v) \leqslant n$ for each $v \in \bigcup_{i=1}^k V(Q_i)$ to get

$$e(G) \leqslant \frac{\gamma^* n^2}{2} + e\Big(G - \bigcup_{i=1}^k V(Q_i)\Big)\,. \tag{5.14}$$

Let $\ell := n - q^*$. Using the fact that the function $f(x) := t_r(x) - \gamma^* x^2$ is increasing

[1] Indeed, setting $\overline{\gamma}$ small forces $t_r(n) - \overline{\gamma} n^2 > t_r(n) - 1$ for each $n < (4r^2)/\gamma^2$. Therefore, by the already established uniqueness result, for each such n, and each n-vertex K_r-free graph G with at least $t_r(n) - \overline{\gamma} n^2$ edges, we have that G must be the Turán graph. In particular, G is ε-close to the Turán graph. Thus the statement of Theorem 5.3 is valid even for $n < (4r^2)/\gamma^2$ in Case I when γ is replaced by $\overline{\gamma}$.

in x we have

$$t_r(\ell) - \gamma^*\ell^2 \leqslant t_r(n) - \gamma^* n^2 < t_r(n) - \gamma n^2 - \frac{\gamma^* n^2}{2}$$
$$\leqslant e(G) - \frac{\gamma^* n^2}{2} \overset{(5.14)}{\leqslant} e\Big(G - \bigcup_{i=1}^{k} V(Q_i)\Big) .$$

By Theorem 5.1, we have that the K_r-free graph $G - \bigcup_{i=1}^{k} V(Q_i)$ is ε^*-close to $\mathrm{T}_r(\ell)$. A straightforward calculation then gives that G must be ε-close to $\mathrm{T}_r(n)$.

5.2.3 Case II; bound

Set $X := \Gamma_G(M)$ and $Y := V(G) - (M \cup X)$.

In two steps we transform G into a graph G_2 which has a clearer structure. We start by transforming G sequentially into an intermediate graph G_1 as follows. First we set $G_1 := G$. If there is a vertex $v \in V(G) - M$ with $\deg_{G_1}(v) < n - m$ then we replace v by a vertex which is adjacent to all the vertices except M and v itself. Observe that these replacements are void when $v \in Y$ and $\deg_G(v) = n - m - 1$; indeed, in that case v was already adjacent to all the vertices except M and v itself. We repeat this process until no more vertices v can be found, and call the resulting graph G_1. Set $X' := \Gamma_{G_1}(M)$, and $Y' := V(G) - (M \cup X')$. Clearly, the following properties hold.

(P1) We have $e(G_1) \geqslant e(G)$. The inequality is strict if there is a vertex $v \in V(G) - M$ with $\deg_G(v) < n - m - 1$. Further, $\deg_G(v) \leqslant \deg_{G_1}(v) = n - m - 1$ for each $v \in Y'$.

(P2) We have $\deg_G(v) < n - m$ for each $v \in X - X'$.

(P3) No copy of K_r touches M in G_1.

(P4) We have $\deg_{G_1}(v) \geqslant n - m$ for each $v \in X'$.

(P5) We have $G[M \cup X'] = G_1[M \cup X]$.

Fix an arbitrary linear order $\prec$ on X'. We start modifying G_1 into a graph G_2; initially we let $G_2 := G_1$. We sequentially repeat the following process. If there are two nonadjacent vertices $v_1, v_2 \in X'$ with either $\deg(v_1) < \deg(v_2)$, or $\deg(v_1) = \deg(v_2)$ and $v_1 \prec v_2$ then we replace v_1 by a clone of v_2. Observe that such a step does not decrease the number of edges in G, does not change the set $\Gamma_{G_2}(M)$, and does not create K_r's touching M. Further, the cloning process cannot produce a vertex inside X' of degree less than $n-m$ by **(P4)**. We stop when each pair v_1, v_2 of nonadjacent vertices in X' have the same neighborhood in G_2; the auxiliary order $\prec$ guarantees that we indeed terminate.

(P6) We have $e(G_2) \geqslant e(G_1)$.

(P7) The graph $G_2[X']$ is a complete t-partite graph for some t. Let $X_1 \dot\cup X_2 \dot\cup \ldots \dot\cup X_t = X'$ be its color classes. Each two vertices $u, v \in X_i$ have the same neighborhoods in G_2.

(P8) We have $\deg_{G_2}(v) \geqslant n-m$ for each $v \in X'$.

For each $v \in M$, let $I_v \subseteq [t]$ be the set of indices i such that v is adjacent to at least one (and hence all) vertices of X_i in the graph G_2. Observe that

$$|I_v| \leqslant r-2\ , \tag{5.15}$$

as otherwise v is contained in an r-clique. For each $i \in [t]$, let $d_i := \deg_{G_2}(x, M)$ be the number of neighbors of a vertex $x \in X_i$ in G_2 in the set M; this number does not depend on the choice of x.

We claim, that for each $i \in [t]$ we have

$$|X_i| \leqslant d_i\ . \tag{5.16}$$

Indeed, consider an arbitrary vertex $x \in X_i$. By **(P8)** the vertex x is adjacent to at least $n-m$ vertices outside X_i. The bound (5.16) follows. Therefore, we have

$$|X'| = \sum_{i=1}^{t} |X_i| \leqslant \sum_{i=1}^{t} d_i \leqslant (r-2)m\,, \tag{5.17}$$

where the last inequality follows from (5.15).

Observe that the graph $G_2[M \cup X']$ together with the set M satisfies by (5.17) the conditions of Case I of Theorem 5.2. Therefore, we have $e(G_2[M \cup X']) \leqslant t_r(|M| + |X'|)$. It now suffices to use trivial bounds $e(G[Y']) \leqslant \binom{|Y'|}{2}$, $e(G_2[M, Y']) = 0$, and $e(G_2[X', Y']) \leqslant |X'||Y'|$ to conclude that

$$e(G) \leqslant e(G_1) \leqslant e(G_2) \leqslant t_r(|X'| + m) + \binom{|Y'|}{2} + |X'||Y'| \leqslant t_r(n, m)\ , \tag{5.18}$$

as desired. Note that the last inequality required optimization over values of $|X'|$ and $|Y'|$, subject to $|X'| + |Y'| = n - m$, $|X'|, |Y'| \geqslant 0$, $|X'| \leqslant (r-2)m$. The unique maximum is attained when $|X'| = (r-2)m$, and $|Y'| = n - (r-1)m$.

5.2.4 Case II; uniqueness

We use the notation from Section 5.2.3. Suppose that G satisfies the assumptions of Theorem 5.2, and $e(G) = t_r(n, m)$. In particular, all the inequalities in (5.18) are equalities. The job may seem easy now: One could think that it is enough to infer from the equalities in (5.18) that Y' must form a complete graph, parts X' and Y' form a complete bipartite graph and $M \cup X'$ be the Turán graph, forcing our graph to be in $\mathcal{T}_r(n, m)$. This is true, but only for the graph G_2. We therefore need to trace back the modifications to be able to conclude something about the structure of G. However, the fact that G_2 is almost determined[2] will be crucial for us.

By the remark after (5.18), we must have $|X'| = (r-2)m$. From **(P1)** and from the fact that $e(G) = e(G_1)$ we have that $\deg_G(v) = n - m - 1$ for each $v \in Y'$.

Lemma 5.6. *All the vertices in X' have degree exactly $n - m$ in the graph G_1.*

Proof. Let us revisit the procedure which modified G_1 to G_2. To get the graph G_2 with properties **(P6)**-**(P8)** several procedures P might have been used subject to the following. First, in each step P chooses a pair of non-adjacent vertices v_1, v_2 in X'. The lower-degree vertex is modified to a clone of the higher-degree vertex. In the case both vertices are of the same degree the choice whether to clone v_1 by v_2 or vice versa is left to P. Second, P terminates whenever each pair of non-adjacent vertices are mutual clones, which must be guaranteed to happen in finite time. We write $\mathcal{P}$ for the class of all the procedures with the properties above, and for a given $P \in \mathcal{P}$ we write $G_{2,P}$ for the graph which was obtained from G_1 by performing P.

Above, we concluded that G_2 must be a common graph in $\mathcal{T}_r(n, m)$ if $e(G) = t_r(n, m)$. This conclusion must be valid for any procedure from $\mathcal{P}$, as otherwise, we would have

$$e(G) \leqslant e(G_1) \leqslant e(G_{2,P}) < t_r(n, m) .$$

Suppose that $P \in \mathcal{P}$ is arbitrary. We write $G_{P,t}$ for the graph obtained from G_1 by running the first t steps of P. We next give a sequence of strengthening claims about the structure of the graphs $G_{P,t}$.

First, we claim that during running any $P \in \mathcal{P}$ and at any step t we have $\deg_{G_{P,t}}(v_1) = \deg_{G_{P,t}}(v_2)$ for each two non-adjacent vertices $v_1, v_2 \in X'$. Indeed, suppose $\deg_{G_{P,t}}(v_1) > \deg_{G_{P,t}}(v_2)$. We can then modify P to clone at time $t + 1$ the vertex v_2 to v_1 and then somehow clone the remaining vertices. For the modified procedure P' we have $e(G_{P',t}) < e(G_{P',t+1})$, and this gives $e(G) \leqslant e(G_1) < e(G_{2,P'}) \leqslant t_r(n, m)$, a contradiction.

[2]By "almost determined" we mean is that we have that G_2 is a common graph in $\mathcal{T}_r(n, m)$.

Second, we claim that all the degrees in X' in the graph $G_{P,t}$ (for any P and t) are equal. Indeed, let $d_1 > d_2 > \ldots > d_k$ be such that each of the sets $V_i := \{v \in X' : \deg_{G_{P,t}}(v) = d_i\}$ is non-empty, and the sets $V_1, \ldots, V_k$ form a partition of X'. By the claim above, $G_{P,t}[V_i, V_j]$ must be a complete bipartite graph for each $i \neq j$. Therefore no cloning between V_i and V_j can ever happen. We now modify P (from time t onwards) as follows. Take for each i a vertex $w_i \in V_i$, and let $W_i \subseteq V_i$ be the vertices in $G_{P,t}$ which are not adjacent to w_i (we have $w_i \in W_i$). We clone sequentially all the vertices of W_i by w_i (for $i = 1, \ldots, k$). Note that after cloning all the vertices W_i (say at time t_i) it must be that the modified cloning procedure P' produces a graph G_{P',t_i} such that $G_{P',t_i}[X' - W_i, W_i]$ is complete bipartite. As all the vertices of W_i are mutual clones at that point, we have a guarantee that the adjacencies of the vertices of W_i will never be modified. We finish off by cloning the remaining vertices. If $k > 1$ then $\deg_{G_{2,P'}}(w_1) = d_1 \neq d_2 = \deg_{G_{2,P'}}(w_2)$. This contradicts the fact that the all the degrees inside X' in the graph $G_{2,P'} \in \mathcal{T}_r^*(m, n)$ are the same. Therefore all the degree in X' in $G_{P,t}$ are the same; let this degree be d.

Obviously, $d \leqslant n - m$. Indeed, the cloning procedure never decreases the degree of the cloned vertex, and we know that the resulting graph $G_{2,P} \in \mathcal{T}_r^*(n, m)$ has degrees $n - m$ inside X'. On the other hand if $d < n - m$ then (again, as $G_{2,P}$ has degree $n - m$ in X') there must a moment t when there are some vertices of degree d and some vertices of degree more than d inside X' in the graph $G_{P,t}$. This contradicts the claim above. $\square$

Lemma 5.6 gives us sufficient information to actually conclude that $G_1[M \cup X']$ must be the Turán graph.

Lemma 5.7. *$G[M \cup X'] = G_1[M \cup X'] = \mathrm{T}_r((r-1)m)$.*

Proof. The equality $G[M \cup X'] = G_1[M \cup X']$ comes from **(P5)**. Thus it suffices to show that $G_1[M \cup X'] = \mathrm{T}_r((r-1)m)$.

For two vertices $v_1, v_2 \in X'$ let $v_1 \sim v_2$ denote the fact that v_1 is not adjacent to v_2 in G_1. We claim that $\sim$ is an equivalence. The symmetry and reflexivity are obvious. Suppose that the transitivity fails, i.e., there are three distinct vertices $x, y, z \in X'$ such that $xy \in E(G_1)$, and $xz, yz \notin E(G_1)$. Let us focus on an arbitrary procedure $P \in \mathcal{P}$ which first makes z a clone of x. (As by Lemma 5.6 we have $\deg_{G_1}(z) = \deg_{G_1}(x)$, such a P exists.) This step increases the degree of y by one, a contradiction to Lemma 5.6.

Let $V_1, \ldots, V_{r-1}$ be the colour classes of the Turán graph $G_2[M \cup X']$. Let $W_1 \dot\cup \ldots \dot\cup W_k = X'$ be the equivalence classes of $\sim$. As the structure of G_1 cannot be modified inside X' by any procedure P we have that under a suitable labelling of $W_1, \dots, W_k$ we have either $k = r-2$ and $W_i = V_i$ $(i = 1, \dots, r-2)$, or $k = r-1$ and $W_i = V_i \cap X'$ $(i = 1, \dots, r-1)$. There is no r-clique in $G_1[M \cup X']$. Indeed, by the above, if there was one, then it would have to touch M, a contradiction. Therefore $G_1[M \cup X']$ is a graph on $(r-1)m$ vertices with no K_r and $t_r((r-1)m)$ edges. The claim follows by the uniqueness part of Turán's Theorem. □

By the above, we have $G_1 \in \mathcal{T}_r^*(n, m)$. It only remains to trace back (relatively simple) changes between the graph G and G_1 to reveal the structure of G. To this end we need to understand properties of the vertices of $X - X'$.

If $X - X' = \emptyset$ then by **(P1)** we have $G = G_1$, and therefore $G \in \mathcal{T}_r(n, m)$ is a common graph.

Next, we suppose that $X - X' \neq \emptyset$. Let $\{v_1, v_2, \dots, v_p\} = X - X'$. Let $V_1, \dots, V_{r-1}$ be the colour classes of the Turán graph $G_1[M \cup X'] = G[M \cup X']$. We have $|V_j| = m$. We call the vertex v_i *j-pure* if v_i is adjacent in G to all the vertices of $M \cup X'$ except V_j.

For each $i \in [p]$ there exists j_i such that v_i is j_i-pure. Indeed, suppose v_i is not j-pure for any j. First, suppose that v_i is adjacent to suitable vertices $u_1 \in V_1, \dots, u_{r-1} \in V_{r-1}$. As $v_i \in X$, we can assume that $\{u_1, \dots, u_{r-1}\} \cap M \neq \emptyset$. Consequently, the vertices $v_i, u_1, u_2, \dots, u_{r-1}$ form an r-clique touching M, a contradiction. Therefore, there exists $j_i \in [r-1]$ such that v_i is not adjacent to any vertex of V_{j_i}. Since $|V_{j_i}| = m$, and $\deg_G(v_i) = n - m - 1$ by **(P1)** we have that v_i is j_i-pure.

Next, we claim that $j_i \neq j_{i'}$ for $i \neq i'$. Indeed, with no loss of generality, assume for contradiction that v_1 and v_2 are 1-pure. Let $u_2 \in V_2, u_3 \in V_3, \dots, u_{r-1} \in V_{r-1}$ be arbitrary vertices such that at least one of $u_2, \dots, u_{r-1}$ touches M. Then the vertices $v_1, v_2, u_2, u_3, \dots, u_{r-1}$ form a K_r touching M, a contradiction.

Observe that the pure vertices $v_1, v_2, \dots, v_p$ have exactly properties as those in our construction of sporadic graphs. This proves that $G \in \mathcal{T}_r(n, m)$.

5.2.5 Case II; stability

We could have revised the proof from Section 5.2.3 as we had done with the proof of Section 5.2.1 in Section 5.2.2. However we decided rather to reduce the stability of Case II to already established stability of Case I.

Let $0 < \gamma \ll \delta \ll \beta \ll \alpha \ll \varepsilon$. In particular, below we make use of the following relations: $2\sqrt{\gamma} < \delta$, and $2\delta < \beta^2$. Further, we assume that β is set so small compared to ε, that Case I of Theorem 5.3 asserts the following.

Fact 5.8. *If H is an s-vertex graph where no r-clique touches a fixed set M, $|M| \geqslant s/(r-1)$, and $e(H) \geqslant t_r(s) - \frac{20\beta}{\varepsilon}s^2$ then H is $\varepsilon/10$-close to $\mathrm{T}_r(s)$.*

Suppose that $e(G) \geqslant t_r(n,m) - \gamma n^2$.

While in Section 5.2.3 we defined X to be the neighbors of M it turns out that for the stability result we need to consider "robust neighbors" of M. To this end we define

$$\begin{aligned} X &:= \{v \in V(G) - M \ : \ \deg(v, M) \geqslant \beta n\} \text{ , and} \\ Y &:= V(G) - (M \cup X) \text{ .} \end{aligned}$$

Lemma 5.9. *All but at most δn vertices of X have degrees at least $n - m - \delta n$.*

Proof. Indeed, suppose for the contrary that there is a set $W \subseteq X$ of more than δn vertices v with $\deg(v) < n - m - \delta n$. Make each such vertex v adjacent to all the vertices except M and v itself; thus we have $\deg(v) = n - m - 1$. Clearly, the modified graph G' gained at least $(\delta n - 1) \times \delta n$ vertices. On the other hand no K_r can touch M in G', and thus $e(G') \leqslant t_r(n,m)$. This contradicts the fact that $e(G) \geqslant t_r(n,m) - \gamma n^2$. $\square$

Let $X_0 \subseteq X$ be the exceptional set from Lemma 5.9. We then modify edges incident to remaining vertices $X - X_0$ as follows. If there are two vertices $v_1, v_2 \in X - X_0$ with $\deg(v_1) \leqslant \deg(v_2)$ then we replace v_1 by a clone of v_2; we do this in such a way that we avoid running into loops. Clearly, after the process terminates we are left with a complete t-partite with color classes $X_1 \dot\cup \ldots \dot\cup X_t = X - X_0$. We claim that

$$t \leqslant \frac{1}{\beta} \text{ .} \tag{5.19}$$

Indeed, suppose for contradiction that (5.19) does not hold. Let $x_1 \in X_1, x_2 \in X_2, \ldots, x_t \in X_t$ be arbitrary. By the definition of X, there are at least $t \times \beta n > n$ edges between $T := \{x_1, \ldots, x_t\}$ and M. In particular, at least one vertex u_0 of M is adjacent to more than $\frac{n}{m} > r - 1$ vertices of T. These vertices together with u_0 form a clique of order at least r touching M, a contradiction.

As in Section 5.2.3 we consider the number $d_i := \deg(x_i, M)$, where $i \in [t]$. We employ the fact that $x_i \notin X_0$ and get a weaker version of the bound (5.16):

$$|X_i| \leqslant d_i + \delta n \text{ .} \tag{5.20}$$

We are now in a position to obtain a stability version of (5.17).

$$
\begin{aligned}
|X| = |X_0| + \sum_{i=r}^{t} |X_i| &\overset{\text{Lemma 5.9, (5.20)}}{\leqslant} \delta n + \sum_{i=1}^{t}(d_i + \delta n) \\
&\overset{(5.15)}{\leqslant} \delta n + (r-2)m + t \times \delta n \overset{(5.19)}{\leqslant} (r-2)m + \beta n \,.
\end{aligned}
\tag{5.21}
$$

Let $X' \subseteq X$ be an arbitrary set of size at most βn such that $|X - X'| \leqslant (r-2)m$. We write $\ell := |X|$.

Recall that our goal is to show that G is ε-close to a graph from $\mathcal{T}_r(n,m)$. To this end we write $e_1 := e(G[Y])$, $e_2 := e(G[X,Y])$, $e_3 := e(G[M \cup (X - X')])$, $e_4 := \sum_{v \in X'} \deg(v)$, and $e_5 := e(G[M,Y])$. Observe that

$$
e(G) \leqslant \sum_{i=1}^{5} e_i \,.
$$

The preparatory steps above allow us to give a good bound on each of the quantities e_i, and further give a stability to each of these bounds. This is summarized in Lemmas 5.10-5.13 below.

We begin with Lemmas 5.10 and 5.11 which are based on trivial bounds. Their "furthermore" stability parts follow directly from the definition of nearness.

Lemma 5.10. *We have $e_1 \leqslant \binom{n-m-\ell}{2}$. Furthermore, $e_1 \leqslant \binom{n-m-\ell}{2} - 4\beta n^2$, unless $G[Y]$ is $4\beta n^2$-near to a complete graph.*

Lemma 5.11. *We have $e_2 \leqslant \ell(n-m-\ell)$. Furthermore, $e_2 \leqslant \ell(n-m-\ell) - 4\beta n^2$, unless $G[X,Y]$ is $4\beta n^2$-near to a complete bipartite graph with colour classes X and Y.*

Lemma 5.12 below employs the bound and the corresponding stability result from Case I. We note that it is essential for these results to apply that $|X - X'| \leqslant (r-2)|M|$.

Lemma 5.12. *We have $e_3 \leqslant t_r(|M \cup (X - X')|) \leqslant t_r(\ell + m, m)$. Furthermore, $e_3 \leqslant t_r(\ell + m, m) - 4\beta n^2$, unless $G[M \cup (X - X')]$ is $(\varepsilon n^2/10)$-near to the graph $\mathrm{T}_r(|M \cup (X - X')|)$.*

Proof. The bound $e_3 \leqslant t_r(|M \cup (X - X')|) \leqslant t_r(\ell + m, m)$ follows from Case I.

For the stability part, we distinguish two cases. If $n \geqslant (5/\varepsilon)^{1/2}(\ell + m)$ then $G[M \cup (X - X')]$ is $(\varepsilon n^2/10)$-near to any graph on $|M \cup (X - X')|$ vertices and the statement is void. We can therefore assume that $n < (5/\varepsilon)^{1/2}(\ell + m)$. Assume that $e_3 > t_r(\ell+m,m) - 4\beta n^2 \geqslant t_r(\ell+m,m) - \frac{20\beta}{\varepsilon}(\ell+m)^2$. Then Fact 5.8 asserts that $G[M \cup (X - X')]$ is $\varepsilon(\ell+m)^2/10$-near to $\mathrm{T}_r(|M \cup (X - X')|)$ and the claim follows. $\square$

The bounds on e_4 and e_5 follow directly.

Lemma 5.13. *We have $e_4 \leqslant \beta n^2$, and $e_5 \leqslant \beta n^2$.*

Therefore, we have

$$e(G) \leqslant \sum_{i=1}^{5} e_i \leqslant \binom{n-m-\ell}{2} + \ell(n-m-\ell) + t_r(\ell+m,m) + 2\beta n^2 \ .$$

Optimizing over values of ℓ, one obtains that the maximum of the right-hand side is achieved whenever $\ell \geqslant (r-2)m$, and is equal to $t_r(n,m) + 2\beta n^2$. Furthermore, when $\ell < (r-2)m - \alpha n$, then the right-hand side is less than $t_r(n,m) - \gamma n^2$. By the assumption of the theorem this cannot occur. On the other hand, recall that by (5.21), we have $\ell \leqslant (r-2)m + \beta n$. Using the "furthermore" parts of Lemmas 5.10-5.13, we therefore have that

$$e(G) \leqslant t_r(n,m) - 2\beta n^2 \ ,$$

unless $G[Y]$ is $4\beta n^2$-near to a complete graph, $G[X,Y]$ is $4\beta n^2$-near to a complete bipartite graph with colour classes X and Y, and $G[M \cup (X - X')]$ is $(\varepsilon n^2/10)$-near to the graph $\mathrm{T}_r(|M \cup (X - X')|)$. Note that the graph $G[Y,M]$ is $(\varepsilon n^2/10)$-near to the empty graph by the definition of Y. It follows that G is εn^2-close to a graph from $\mathcal{T}_r(n,m)$. We note that we actually proved that G is εn^2-close to a graph from $\mathcal{T}_r^*(n,m)$. This is not surprising (or contradictory) as any sporadic graph is close to a common graph.

Chapter 6

Hamilton cycles in dense vertex-transitive graphs

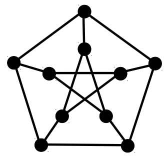
6.1 Introduction

The decision problems of whether a graph contains a Hamilton cycle or a Hamilton path are two of the most famous NP-complete problems, and so it is unlikely that there exist good characterizations of such graphs. For this reason, it is natural to ask for sufficient conditions which ensure the existence of a Hamilton cycle or a Hamilton path. To this direction, the following well-known conjecture of Lovász is still wide open.

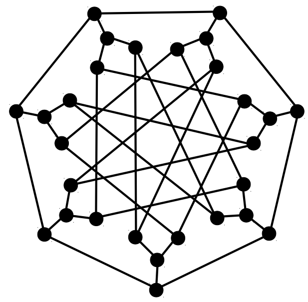
Conjecture 6.1. *Every connected vertex-transitive graph has a Hamilton path.*

In contrast to common belief, Lovász in 1969 [75] asked for the construction of a connected vertex-transitive graph containing no Hamilton path. Traditionally however, the Lovász conjecture is always stated in the positive.

At the moment no counterexample is known. Moreover, there are only five known examples of connected vertex-transitive graphs having no Hamilton cycle. These are K_2, the Petersen graph (see Figure 6.1(a)), the Coxeter graph (see Figure 6.1(b)) and the graphs obtained from the Petersen and Coxeter graphs by replacing every vertex with a triangle. Apart from K_2, the other four examples are not Cayley graphs and this leads to the conjecture that every connected Cayley graph on at least three vertices is Hamiltonian. Similarly as with Conjecture 6.1 this is now folklore, and its origin may be difficult to trace back, but probably the first conjecture in this direction is due to Thomassen (see [16]), and asserts that there are only finitely many connected vertex-transitive graphs that are non hamiltonian. At



(a) The Petersen graph.



(b) The Coxeter graph.

Figure 6.1: The Petersen and the Coxeter graphs.

the moment however, the best known general result which is due to Babai [12] states that every connected vertex-transitive graph on n vertices has a cycle of length at least $\sqrt{3n}$.

The conjecture has attracted a lot of interest from researchers and there is no common agreement as to its validity. For example, in the negative direction, Babai [13] conjectured that there is an absolute constant $c > 0$ and infinitely many connected Cayley graphs G without cycles of length greater than $(1 - c)|G|$.

We give a very limited overview of the vast research these questions have motivated, referring the reader to the following surveys [99, 30, 73, 82] and their references. Some of the most important results in the field are listed below.

- It is known that a connected Cayley graph is hamiltonian whenever the underlying group is one of the following types: abelian [76, §12, Problem 17] or p-group [98]. For dihedral groups, the positive result is known only when the order of the group is divisible by four [10].

- There has also been a substantial work done regarding hamiltonicity of a Cayley graph of a given group, and a *suitable* ("small") generating set, see: [100] and [82], and a *randomly* chosen generating set [68].

We discuss two other results which are more related to our current contribution further below, in Section 6.1.1.

We prove that every sufficiently large dense connected vertex-transitive graph is Hamiltonian.

Theorem 6.2. *For every $\alpha > 0$ there exists an n_0 such that every connected vertex-transitive graph on $n \geqslant n_0$ vertices of valency at least αn contains a Hamilton cycle.*

6.1.1 Relation to previous results

As said above, we do not aim to survey results related to Conjecture 6.1. However, it turns out that Theorem 6.2 is implied in several settings by other results. We want to describe these and pinpoint some situations when the Hamiltonicity given by Theorem 6.2 was not known before. We will restrict the discussion to the family of Cayley graphs.

Recall that Fleischner's Theorem [41] asserts that the (distance-)square of a 2-connected graph is Hamiltonian. Suppose that G is a connected Cayley graph over a group Γ with a generating set X. It is easy to check that connectivity of a vertex-transitive graph already implies its 2-connectivity. If we find a set $Y \subseteq X$ which generates Γ, and such that $Y^2 \subseteq X$, then Fleischner's Theorem applies and the Hamiltonicity of G follows. This is a 'typical'[1] situation when X is dense in Γ. However, there are examples, when the set Y does not exist.

There are two important classes of groups where Hamiltonicity of the corresponding Cayley graph follows by other methods. One class is abelian groups. In the abelian setting, the Hamiltonicity of the Cayley graph is known for all generating sets. The argument has been pushed further by Pak and Radoičić [82] to groups which are close to abelian. Another important class is groups with no non-trivial irreducible representations of low dimension. This family for example, contains all non-abelian simple groups. For these groups, Gowers [51], proved that the corresponding Cayley graph is quasirandom (in the sense of Chung-Graham-Wilson [27]), no matter what the set X of generators is taken to be (provided that X is dense). In this case, the Hamiltonicity follows from the well-known fact (see e.g. [67, Proposition 4.19]) that dense pseudorandom graphs are Hamiltonian. However, there are groups which are very far from abelian and yet have non-trivial low-dimensional representations. Soluble groups are one such example.

6.1.2 Overview

Here is an overview of the rest of the chapter. Section 6.2 contains some notation that we are going to use. Our proof will use Szemerédi's Regularity Lemma. In using the Regularity Lemma, we would like some properties of the original graph G to be inherited by the reduced graph obtained from the application of the lemma. In Section 6.3 we discuss some results from matching theory in this direction. These results will enable us to show that the reduced graph (after a minor modification) contains an almost perfect matching. Short Section 6.4 contains a version of the

[1] In the sense that most examples that come to mind are of this sort

Blow-up Lemma tailored to our needs. In Section 6.5 we discuss two non-standard notions of connectivity: robustness and iron connectivity. The main result of Section 6.5 is Theorem 6.9 which says that G can be partitioned into a bounded number of isomorphic vertex-transitive pieces each of which is iron connected. This is a much stronger notion than the standard notion of vertex connectivity. In particular, iron connectivity is inherited by the reduced graph as well. It will turn out that if G 'looks very much like a bipartite graph' then there are some additional difficulties that need to be overcome. In Section 6.6 we quantify what we mean by the phrase 'looks very much like a bipartite graph' and prove that in this case the vertex set of G can be partitioned into two equal parts such that every automorphism of G respects this partition. In Section 6.7 we apply the Regularity Lemma to show that every sufficiently large iron connected vertex-transitive graph contains a Hamilton cycle. In fact, we will need and prove a somewhat stronger property. Finally, in Section 6.8 we put all the pieces together. We first partition G into the bounded number of vertex-transitive, iron connected pieces, then find a Hamilton cycle in each of these pieces, and then show how to glue these pieces together. It turns out that what we need for the glueing is not Hamilton cycles but rather what we call ℓ-pathitions which their existence is also guaranteed from our work in Section 6.7.

It turns out that all the steps of our proof of Theorem 6.2 can be performed algorithmically. In Section 6.9 we discuss how to turn the proof into a polynomial time algorithm for finding a Hamilton cycle in dense vertex-transitive graphs.

6.2 Notation and preliminaries

We denote the automorphism group of G by $\mathrm{Aut}(G)$. We will usually denote the elements of $\mathrm{Aut}(G)$ by f or g.

Recall that a graph G is *Hamilton-connected* if for any pair of distinct vertices x, y there is a Hamilton path with x and y as terminal vertices. Another important connectivity notion is that of linkedness: G is *ℓ-linked* if for any set of distinct vertices $x_1, \dots, x_\ell, y_1, \dots, y_\ell \in V(G)$ there exist vertex-disjoint paths $P_1, \dots, P_\ell$ such that x_i and y_i are terminal vertices of P_i. For our proof of Theorem 6.2, we will need a combination of the two notions above. Given a graph G and a subset U of the vertex set of G, we say that G is *ℓ-pathitionable with exceptional set U* if for any $\ell' \in [\ell]$, and for any set of distinct vertices $x_1, \dots, x_{\ell'}, y_1, \dots, y_{\ell'} \in V(G) - U$ there exist vertex-disjoint paths $P_1, \dots, P_{\ell'}$ such that x_i and y_i are terminal vertices of P_i. Furthermore, we require that the paths $P_1, \dots, P_{\ell'}$ cover all the vertices of G. So a graph is 1-pathitionable with exceptional set $\emptyset$ if and only if it is Hamilton-connected.

Observe that for example the complete bipartite graph $K_{n,n}$ is not 1-pathitionable. Indeed, we cannot connect two vertices of the same colour class of $K_{n,n}$ by a Hamilton path. Yet, we will need to deal with graphs which are bipartite or even almost bipartite. To this end we introduce a modification of pathitionability to bipartite setting. Suppose that a graph G together with a partition $V(G) = A \cup B$ is given. We say that G is *ℓ-bipathitionable with exceptional set U with respect to the partition A, B* if for any $\ell' \in [\ell]$, and for any set of distinct vertices $x_1, \dots, x_{\ell'}, y_1, \dots, y_{\ell'} \in V(G) - U$ such that

$$|\{x_1, \dots, x_{\ell'}, y_1, \dots, y_{\ell'}\} \cap A| = |\{x_1, \dots, x_{\ell'}, y_1, \dots, y_{\ell'}\} \cap B| \tag{6.1}$$

there exist vertex-disjoint paths $P_1, \dots, P_{\ell'}$ such that x_i and y_i are terminal vertices of P_i. Furthermore, we require that the paths $P_1, \dots, P_{\ell'}$ cover all the vertices of G.

Suppose that $\mathcal{S} = \{P_1, \dots, P_\ell\}$ is a system of vertex-disjoint paths in a graph G. We then say that a system of paths $\mathcal{S}' = \{P'_1, \dots, P'_\ell\}$ is an *extension of $\mathcal{S}$* if the paths P'_i are vertex-disjoint, and for each $i \in [\ell]$ we have $V(P'_i) \supset V(P_i)$, and P_i and P'_i have the same endvertices. If $\mathcal{S}'$ covers all the vertices of G then we say that $\mathcal{S}'$ is a *complete extension*.

Given a graph G and a natural number ℓ, the *ℓ-blow-up of G*, denoted $\ell \times G$ is the graph in which every vertex of G is replaced by an independent set of size ℓ, and each edge of G is replaced by a complete bipartite graph between the two corresponding independent sets.

As an auxiliary tool we will need to work with *digraphs* as well. For basic terminology about digraphs we refer the reader to [15]. In particular we do not allow loops or multiple edges. (We do however allow edges between the same two vertices which have different direction.) Recall that a digraph G is strongly connected if for any pair of distinct vertices $a, b \in V(G)$ there is a directed walk from a to b. We will also need the following extension of the notion of strong connectedness: we say that a digraph D is *ℓ-strongly connected* if for every set $U \subseteq V(D)$, $|U| \leqslant \ell$ and for any pair of distinct vertices $a, b \in V(G) - U$ there exists a directed walk from a to b avoiding U.

6.3 Some matching theory

Let us recall that a function $f : V \to [0, 1]$ is a *fractional vertex cover* of a graph $G = (V, E)$ if $f(x) + f(y) \geqslant 1$ for every $xy \in E$. We write $\tau^*(G)$ for the weight of the minimum fractional vertex cover, i.e.

$$\tau^*(G) := \min\{\|f\|_1 : f \text{ is a fractional vertex cover of } G\} .$$

A function $M : E \to [0,1]$ is a *fractional matching* of a graph $G = (V, E)$ if for every $v \in V$ we have $\sum_{e \ni v} M(e) \leqslant 1$, where the summation is taken over all edges $e \in E$ containing the vertex v. We write $\nu^*(G)$ for the weight of the maximum fractional matching, i.e.

$$\nu^*(G) := \max\{\|M\|_1 : M \text{ is a fractional matching of } G\}.$$

The fractional matching M is said to be *half-integral* if $M(e) \in \{0, \frac{1}{2}, 1\}$ for every $e \in E$.

It is easy to see that for every graph G we have $\tau^*(G) \geqslant \nu^*(G)$. The duality of linear programming guarantees that in fact we have equality. Moreover, the half-integrality property of fractional matchings (cf. [90, Theorem 30.2]) says that there is a half-integral matching with weight $\nu^*(G)$.

Theorem 6.3.

(a) For every graph G we have $\tau^*(G) = \nu^*(G)$.

(b) For every graph G there is a half-integral matching M of G with $\|M\|_1 = \nu^*(G)$.

The next lemma asserts that removal of a small fraction of edges from a vertex-transitive graph G does not decrease $\tau^*(G)$ much.

Lemma 6.4. *Let G be a vertex-transitive graph on n vertices. Suppose G' is a spanning subgraph of G such that $e(G') \geqslant (1-\delta)e(G)$. Then $\tau^*(G') \geqslant (1-\delta)\tau^*(G)$.*

Proof. Let $f : V(G) \to [0,1]$ be an arbitrary fractional vertex cover of G'. To prove the lemma, it suffices to show that the there is a function $f' : V(G) \to [0,1]$ such that

(a) $\|f\|_1 = \|f'\|_1$;

(b) $f'(x) + f'(y) \geqslant 1 - \delta$ for every edge $xy \in E(G)$.

Indeed, if the above hold then the function $g : V(G) \to [0,1]$ defined by $g(x) := f(x)/(1-\delta)$ is a fractional vertex cover of G with $(1-\delta)\|g\|_1 = \|f\|_1$ and the claim of the lemma follows.

To show that such an f' exists, we define

$$f'(v) := \frac{1}{|\mathrm{Aut}(G)|} \sum_{g \in \mathrm{Aut}(G)} f(g(v)) \ .$$

Observe that f' is constant, and that (a) is satisfied. Suppose for contradiction that (b) fails for some edge xy of G. Since f' is constant, we get that (b) fails for every edge of G. Thus,

$$\sum_{uv \in E(G)} (f'(u) + f'(v)) < (1-\delta)e(G) \leqslant e(G') \leqslant \sum_{uv \in E(G')} (f(u) + f(v)) \ , \qquad (6.2)$$

where the last inequality follows from the fact that f is a fractional vertex cover of G'. Plugging the defining formula for f' in (6.2) we get

$$\sum_{g \in \mathrm{Aut}(G)} \sum_{uv \in E(G)} (f(g(u)) + f(g(v))) < \sum_{g \in \mathrm{Aut}(G)} \sum_{uv \in E(G')} (f(u) + f(v)) \ ,$$

However, observe that due to the vertex-transitivity of G, the sum $\sum_{uv \in E(G)} (f(g(u)) + f(g(v)))$ does not depend on g. Therefore, $\sum_{uv \in E(G)} (f(u)+f(v)) < \sum_{uv \in E(G')} (f(u)+f(v))$, a contradiction. $\square$

The following lemma asserts that $\tau^*(G) = \frac{n}{2}$ for every non-empty vertex-transitive graph of order n. This is easy and well-known; nevertheless we include a proof for completeness.

Lemma 6.5. *Suppose that G is a vertex-transitive graph of order n and at least one edge. Then $\tau^*(G) = \frac{n}{2}$.*

Proof. The constant one-half function is a fractional vertex cover of G, thus establishing $\tau^*(G) \leqslant \frac{n}{2}$.

Suppose for contradiction that there exist a fractional vertex cover $f : V(G) \to [0,1]$ such that $\|f\|_1 < \frac{n}{2}$. The function $f' : V(G) \to [0,1]$ defined by $f'(v) := \frac{1}{|\mathrm{Aut}(G)|} \sum_{g \in \mathrm{Aut}(G)} f(g(v))$ is a constant function, which is a fractional vertex cover. Since $\|f'\|_1 = \|f\|_1 < \frac{n}{2}$, we have $f'(v) < \frac{1}{2}$ for each $v \in V(G)$. In particular, $f'(x) + f'(y) < 1$ for an edge $xy \in E(G)$, a contradiction. $\square$

The next lemma asserts that 2-blow-up graphs contain an integral matching which is twice the weight of the maximum fractional matching of the original graph.

Lemma 6.6. *There exists a matching of weight $2\nu^*(H)$ in the graph $2 \times H$.*

Proof. Suppose that each vertex v in H was replaced by two vertices v^1 and v^2 in the graph $2 \times H$.

Consider a half-integral matching M in the graph H of weight $\nu^*(H)$. Such a matching exists by Theorem 6.3(b). We now construct an integral matching (i.e. a matching) M' in $2 \times H$ of weight $2\nu^*(H)$ as follows: For any edge uv with weight 1 in M, we add the edges u^1v^1 and u^2v^2 in M'. The set of edges with weight $\frac{1}{2}$ in M form a subgraph of R which is a union of paths and cycles. For every such path $v_1 \cdots v_r$ we add in M' all edges of the form $v_j^s v_{j+1}^s$ with $1 \leqslant s \leqslant 2, 1 \leqslant j \leqslant r-1$ and $j+s$ even. Finally, for every such cycle $v_1 \cdots v_r v_1$ we add in M' all edges of the form $v_j^s v_{j+1}^s$ with $1 \leqslant s \leqslant 2, 1 \leqslant j \leqslant r-1$ and $j+s$ even, together with either the edge $v_r^1 v_1^2$ if r is odd or the edge $v_r^2 v_1^2$ if r is even. It is immediate by the construction that M' is indeed a matching of $2 \times H$ of weight $\|M'\|_1 = 2\|M\|_1 = 2\nu^*(H)$. □

The last lemma says that the property of containing a large matching is inherited on the reduced graph as well. Here we formulate it without referring to the Regularity lemma.

Lemma 6.7. *Suppose that a graph $\tilde{R}$ is given and let $\tilde{G}$ be a subgraph of its m-blow-up. Then $\nu^*(\tilde{R}) \geqslant \frac{\nu^*(\tilde{G})}{m}$.*

Proof. Suppose that a fractional matching M in $\tilde{G}$ is given. We can then define a fractional matching $M_{\tilde{R}}$ in $\tilde{R}$ by defining its weight on an edge $AB \in E(\tilde{R})$ as

$$M_{\tilde{R}}(AB) := \frac{1}{m} \sum_{a \in A, b \in B, ab \in E(\tilde{G})} M(ab) \ .$$

This is indeed a fractional matching as for each $A \in V(\tilde{R})$ we have

$$\sum_{B : AB \in E(\tilde{R})} M_{\tilde{R}}(AB) = \frac{1}{m} \sum_{a \in A} \sum_{b \in V(\tilde{G})} M(ab) \leqslant \frac{1}{m} \sum_{a \in A} \sum_{b \in V(\tilde{G})} M(ab) \frac{1}{m} \sum_{a \in A} 1 \leqslant 1 \ .$$

Moreover,

$$\|M_{\tilde{R}}\|_1 = \frac{1}{m} \sum_{e \in E(\tilde{G})} M(e) \ ,$$

and the lemma follows. □

6.4 A blow-up type lemma for Hamilton paths

In this short section we give a version of the Blow-up Lemma tailored to our proof of Theorem 6.2.

Lemma 6.8. *Suppose $0 < \varepsilon \ll d$ and let (A, B) be an (ε, d)-super-regular pair with $|A| = |B|$. Let $a \in A$ and $b \in B$. Then $A \cup B$ contains a Hamilton path with endvertices a and b.*

Proof. The lemma follows from Lemma 1.8. We need to deal with one minor difficulty which does not allow a direct application of Lemma 1.8, namely that we are prescribing exactly the images a and b of the endvertices of the Hamilton path.

Recall that by the "furthermore" part of Lemma 1.8 we can impose additional restriction on a small number of target sets of vertices of the graph we are trying to embed in the super-regular pair. We thus proceed as follows.

We can assume that $|A|$ is sufficiently large. Otherwise, setting ε small, we can force (A, B) to form a complete bipartite graph, and then the statement is trivial.

Let $A' \subseteq A$ and $B' \subseteq B$ be the neighborhood of b and a, respectively. We have $|A' - \{a\}| \geqslant \frac{d|A|}{2}$, and $|B' - \{b\}| \geqslant \frac{d|B|}{2}$. Observe also, that the pair $(A - \{a\}, B - \{b\})$ is $(2\varepsilon, \frac{d}{2})$-super-regular. By Lemma 1.8 we can find a Hamilton path P in the pair $(A - \{a\}, B - \{b\})$. Furthermore, we can require the endvertices of the path to lie in the sets A' and B'. The path aPb is a Hamilton path in (A, B) satisfying the assertions of the lemma. $\square$

6.5 Robustness and iron connectivity

We introduce two non-standard notions of connectivity: robustness and iron connectivity. These notions turn out to be suitable in combination with the Regularity Lemma — roughly speaking, when a graph has high iron connectivity, then the reduced graph corresponding to it also has high iron connectivity.

We say that a graph G is *ℓ-robust* if G remains connected even after removal of an arbitrary set $E' \subseteq E(G)$ with $\deg^{\max}(E') \leqslant \ell$. We say that G is *ℓ-iron* if G stays connected after simultaneous removal of an arbitrary edge-set $E' \subseteq E(G)$ with $\deg^{\max}(E') \leqslant \ell$ and an arbitrary vertex-set $U \subseteq V(G)$ with $|U| \leqslant \ell$.

Our main aim in this chapter is to show that every dense vertex-transitive graph can be partitioned into not too many isomorphic vertex-transitive subgraphs which have high iron connectivity. This is stated in the following theorem.

Theorem 6.9. *For every $\alpha > 0$ there exist $\beta, R, N_0 > 0$ such that the following holds: Suppose G is a vertex-transitive graph of order $n > N_0$ and valency at least αn. Then there exists a partition $V(G) = V_1 \cup \cdots \cup V_r$ into $r < R$ parts such that all the graphs $G[V_i]$ are isomorphic to a graph G' which is vertex-transitive* *and (βn)-iron. Furthermore, for each $g \in \mathrm{Aut}(G)$ and each $1 \leqslant j \leqslant r$ we have $g(V_j) \in \{V_1, \ldots, V_r\}$.*

A typical example of a connected vertex-transitive graph G with very low iron connectivity (and even robustness) is a graph formed by two disjoint cliques of order $n/2$ — say on vertex sets V_1 and V_2 — with a perfect matching between V_1 and V_2. The sets V_1 and V_2 are likely to be the decomposition of G given by Theorem 6.9 — and indeed this is the decomposition our proof would give.

The first step towards the proof of the above theorem would be to gather together vertices of G which cannot be separated from the removal of an edge set of small maximum degree. To this end, given two vertices u and v of G we say that u and v are *ℓ-robustly adjacent* if whenever we remove from G an arbitrary set $E' \subseteq E(G)$ with $\deg^{\max}(E') \leqslant \ell$ then u and v are still in the same connected component. We write $u \sim_{(\ell)} v$ in this case.

We shall also associate to a graph G an auxiliary graph H, called *k-codeg graph of G*. H is on the same vertex set as G. Two distinct vertices $v_1, v_2 \in V(H)$ are adjacent in H if and only if $|\mathrm{N}_G(v_1) \cap \mathrm{N}_G(v_2)| \geqslant k$.

The following lemma summarizes properties of the relation $\sim_{(\ell)}$, and of k-codeg graphs.

Lemma 6.10. *We have the following properties.*

(a) The relation $\sim_{(\ell)}$ is an equivalence relation on $V(G)$. The equivalence classes of $\sim_{(\ell)}$ are called ℓ-islands.

(b) Suppose that a vertex v of G has more than ℓ neighbors in some ℓ-island L. Then $v \in L$.

(c) If G is vertex-transitive then all ℓ-islands induce mutually isomorphic, vertex-transitive graphs.

(d) If G is vertex-transitive then the k-codeg graph H of G is vertex-transitive as well. We have $\deg(H) \geqslant \frac{\deg(G)^2}{n} - k$.

(e) Suppose that $n \geqslant 10\alpha^{-2}$. If G is a vertex-transitive graph on n vertices with valency at least αn then each $(\alpha^2 n/5)$-island contains at least $\alpha^2 n/2$ vertices.

Proof. Parts (a)-(b) are trivial. As for part (c), it is obvious that all ℓ-islands induce mutually isomorphic graphs. To see the vertex-transitivity of the islands, we note that each automorphism of G maps an ℓ-island again onto an ℓ-island. Therefore, taking the set $A \subseteq \mathrm{Aut}(G)$ of automorphisms which map a given ℓ-island L onto itself and considering a restriction $A_{|L} := \{g_{|L} \ : \ g \in A\}$ on L, we get a subgroup $A_L \leqslant \mathrm{Aut}(G[L])$ which witnesses vertex-transitivity of $G[L]$.

The first part of (d) is obvious. For the second part we count the number of triples (x, y, z) with z adjacent to both x and y in two different ways to get

$$\begin{aligned} n \deg(G)^2 &= \sum_{x,y \in V(G)} |\mathrm{N}_G(x) \cap \mathrm{N}_G(y)| \\ &\leqslant \sum_{x,y \in V(G), xy \in E(H)} (n-2) + \sum_{x,y \in V(G), xy \notin E(H)} (k-1) \\ &= n(n-2)\deg(H) + n(n-1-\deg(H))(k-1) \\ &\leqslant n^2 \deg(H) + n^2(k-1) \ , \end{aligned}$$

and the claim follows.

To prove Part (e), consider the $(\alpha^2 n/2)$-codeg graph H of G. By Part (d), H is vertex-transitive of valency $\deg(H) \geqslant \alpha^2 n/2$. Observe now that if $|\mathrm{N}_G(u) \cap \mathrm{N}_G(v)| \geqslant 2\frac{\alpha^2 n}{5} + 1$ then u and v lie in the same $(\alpha^2 n/5)$-island; in particular the conclusion applies when uv is an edge of H. Since $\deg(H) \geqslant \alpha^2 n/2$ we deduce that each $(\alpha^2 n/5)$-island of G contains at least $\alpha^2 n/2$ vertices. □

As a corollary of Lemma 6.10 we get the following.

Lemma 6.11. *Suppose G is a vertex-transitive graph on n vertices with valency at least αn. If G is not $(\alpha^4 n/40)$-robust, then there exists a partition $V(G) = V_1 \cup \ldots \cup V_r$ with $2 \leqslant r \leqslant \frac{2}{\alpha^2}$ such that all the graphs $G[V_i]$ are isomorphic to the same vertex-transitive graph G' of order n' and valency at least $4\alpha n'/3$.*

Proof. Let $V_1, \ldots, V_r$ be the $(\alpha^4 n/40)$-islands of G. If $r = 1$ then G is $(\alpha^4 n/40)$-robust and there is nothing to prove. Thus we assume that $r > 1$.

Observe that since $\alpha^4/40 < \alpha^2/5$, each $(\alpha^4 n/40)$-island consists of several $(\alpha^2 n/5)$-islands. In conjunction with Part (e) of Lemma 6.10, we get that $r \leqslant 2a^{-2}$. By Part (b) of Lemma 6.10 each vertex $v \in V_1$ sends at most $\alpha^4 n/40$ edges to V_i for $i \neq 1$. It follows that

$$\deg(v, V_1) \geqslant \alpha n - (r-1)\frac{\alpha^4 n}{40} \geqslant \alpha n - \frac{\alpha^2 n}{20} \geqslant \frac{2\alpha n}{3} \ .$$

On the other hand, for $n' := |V_1|$ we have $n' = \frac{n}{r} \leqslant \frac{n}{2}$. Therefore the valency of the graph $G' := G[V_1]$ is at least $4\alpha n'/3$. This proves the lemma. □

Lemma 6.11 says that if G is not robust then we can partition it into a few island each having higher (by a constant factor) density than G. Repeating this process, it will follow that every dense vertex-transitive graph can be partitioned in a symmetric way into a bounded number of robust graphs.

Lemma 6.12. *For every $\alpha > 0$ there exist μ, R, N_0 such that the following holds: Suppose G is a vertex-transitive graph of order $n > N_0$ and valency at least αn. Then there exists a partition $V(G) = V_1 \cup \dots \cup V_r$, into $r < R$ parts such that all the graphs $G[V_i]$ are isomorphic to a graph G' which is vertex-transitive and (μn)-robust. Furthermore, for each $g \in \mathrm{Aut}(G)$ and each $1 \leqslant j \leqslant r$ we have $g(V_j) \in \{V_1, \dots, V_r\}$.*

Proof. We first set up necessary constants. Let $Q := \lceil \log_{4/3}(\frac{1}{\alpha}) \rceil$, and $\alpha_i := (4/3)^i \alpha$ for $i = 0, 1, \dots$. Let $R := \prod_{i=0}^{Q}(2\alpha_i^{-2})$, and $\mu := \alpha^4/40R$. Last, let N_0 be sufficiently large.

Set $G_0 := G$, and $n_0 := n$. Inductively, in steps $i = 0, 1, \dots$ we either get that G_i is $(\alpha_i^4 n_i/40)$-robust, or by Lemma 6.11 that there is a partition $V(G_i) = V_{i,1} \cup \dots \cup V_{i,r_i}$ (with $r_i \leqslant 2/\alpha_i$) such that each graph $G_i[V_{i,j}]$ ($j = 1, \dots, r_i$) is isomorphic to a vertex-transitive graph G_{i+1} of order n_{i+1}, thus allowing a next step of the iteration. By induction, and the properties of the partition output by Lemma 6.11 the vertex set of the original graph G can be partitioned into vertex-sets inducing graphs isomorphic to G_{i+1}. Observe that it is guaranteed by Lemma 6.11 and induction that G_{i+1} has valency at least $\alpha_{i+1} n_{i+1}$.

Since $\alpha_Q \geqslant 1$, the above procedure must terminate in step $i_{\mathrm{stop}} < Q$. It is easily checked that the partition of $V(G)$ into copies of $G_{i_{\mathrm{stop}}}$ satisfies the assertions of the lemma. ∎

Observe that ℓ-iron connectivity implies ℓ-robustness. If the converse was true then we could immediately deduce Theorem 6.9 from Lemma 6.12. However, the converse is very far from being true. For example, the union of two cliques of size $2m$ having exactly one common vertex is $(m-1)$-robust but it is not even 1-iron as the common vertex of the two cliques is a cut-vertex. The following lemma gives a partial converse for the class of vertex-transitive graphs.

Lemma 6.13. *Let G be a (μn)-robust vertex-transitive graph of order n and valency at least αn. Let $\lambda := \min\left\{\frac{\alpha}{2^{3+2/\alpha}}, \frac{\mu}{2^{2+2/\alpha}}\right\}$. Then G is (λn)-iron.*

Before diving into the proof of Lemma 6.13 let us give a heuristic why the lemma ought to hold. The graph G is robust by the assumptions of the lemma. On the other hand it is known ([48, Theorem 3.4.2]) that connected vertex-transitive graphs of high valency have high vertex connectivity. Therefore one can hope for a combination of the two properties, that is for iron connectivity.

Proof of Lemma 6.13. Let $d \geqslant \alpha n$ be the valency of G. Suppose for contradiction that G is not (λn)-iron. That is, we have a partition $V(G) = A_0 \cup U_0 \cup B_0$, $|U_0| \leqslant \lambda n$, $\deg^{\max}{}_G(A_0, B_0) \leqslant \lambda n$. We proceed with an iterative procedure described below. For $i \geqslant 0$ we are given a partition $V(G) = A_i \cup U_i \cup B_i$. We further have the following properties:

$(\mathrm{I1})_i$ $|U_i| \leqslant 2^i \lambda n$,

$(\mathrm{I2})_i$ $\deg^{\max}{}_G(A_i, B_i) \leqslant 2^i \lambda n$, and

$(\mathrm{I3})_i$ $0 < |A_i| \leqslant n - \frac{i\alpha n}{2}$.

We terminate this iterative procedure when for each $g \in \mathrm{Aut}(G)$, if there is an $a \in A_i$ such that $g(a) \in A_i$ then for each $b \in B_i$ we have that $g(b) \notin A_i$. Otherwise, as we shall show below, we can produce a partition $V(G) = A_{i+1} \cup U_{i+1} \cup B_{i+1}$ satisfying $(\mathrm{I1})_{i+1}$, $(\mathrm{I2})_{i+1}$, and $(\mathrm{I3})_{i+1}$. Note that from (I3) it follows that we must terminate in $i_{\mathrm{stop}} < \frac{2}{\alpha}$ steps.

Suppose we did not terminate in step i. Then there exists $g \in \mathrm{Aut}(G)$, $a \in A_i$, $b \in B_i$ such that $g(a), g(b) \in A_i$. Observe that $(\mathrm{I2})_i$ gives $|\mathrm{N}(b) - (B_i \cup U_i)| \leqslant 2^i \lambda n$, and consequently with the help of $(\mathrm{I1})_i$ we have $|\mathrm{N}(b) - B_i| \leqslant 2^{i+1} \lambda n$. Similarly, $|\mathrm{N}(g(b)) - A_i| \leqslant 2^{i+1} \lambda n$. We conclude that

$$\begin{aligned} |A_i \cap g(B_i)| &\geqslant |\mathrm{N}(g(b))| - |\mathrm{N}(g(b)) - A_i| - |\mathrm{N}(g(b)) - g(B_i)| \\ &= d - |\mathrm{N}(g(b)) - A_i| - |\mathrm{N}(b) - B_i| \\ &\geqslant \alpha n - 2^{i+2} \lambda n \geqslant \frac{\alpha n}{2}, \end{aligned} \tag{6.3}$$

where the last inequality follows since $\alpha \geqslant 2^{3+2/\alpha} \lambda \geqslant 2^{3+i} \lambda$.

Define $A_{i+1} := A_i \cap g(A_i)$, $U_{i+1} := U_i \cup g(U_i)$, and $B_{i+1} := (B_i \cup g(B_i)) - U_{i+1}$. This is a partition of $V(G)$; see Figure 6.2. $(\mathrm{I1})_{i+1}$ and $(\mathrm{I2})_{i+1}$ are obviously satisfied. The lower bound in $(\mathrm{I3})_{i+1}$ follows from the fact that $g(a) \in A_i \cap g(A_i)$. The upper bound is then established through the following chain of inequalities:

$$|A_i \cap g(A_i)| \leqslant |A_i| - |A_i \cap g(B_i)| \overset{(6.3)}{\leqslant} |A_i| - \frac{\alpha n}{2} .$$

This finishes the iterative step.

We now deal with the situation of termination in the step $i_{\mathrm{stop}} < \frac{2}{\alpha}$. For simplicity, we write $A := A_{i_{\mathrm{stop}}}$, $B := B_{i_{\mathrm{stop}}}$, and $U := U_{i_{\mathrm{stop}}}$. We have

$$|U| \leqslant 2^i \lambda n < 2^{2/\alpha} \lambda n \leqslant \frac{1}{4} \mu n \quad \text{and similarly} \quad \deg^{\max}{}_G(A, B) \leqslant \frac{1}{4} \mu n . \tag{6.4}$$

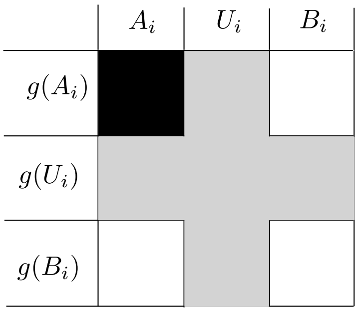


Figure 6.2: The sets A_{i+1}, U_{i+1} and B_{i+1} as intersections of the sets A_i, U_i, B_i, $g(A_i)$, $g(U_i)$, and $g(B_i)$. The set A_i is represented by black, U_i by grey, and B_i by white.

Furthermore, we have

$$\text{For every } g \in \mathrm{Aut}(G)\text{, if } g(a') \in A \text{ for some } a' \in A\text{, then } g(b') \notin A \text{ for each } b' \in B. \tag{6.5}$$

We first prove that each vertex $u \in U$ has either almost all its neighbors in A, or in B.

Claim 6.13.1. *For each $u \in U$, either $|\mathrm{N}(u) \cap A| \geqslant d - \frac{3}{4}\mu n$, or $|\mathrm{N}(u) \cap B| \geqslant d - \frac{3}{4}\mu n$.*

Proof of Claim 6.13.1. Suppose the statement fails for some $u \in U$. Then we have

$$|\mathrm{N}(u) \cap A| = |\mathrm{N}(u)| - |\mathrm{N}(u) \cap B| - |\mathrm{N}(u) \cap U| > \frac{\mu}{2} n \text{ , and} \tag{6.6}$$
$$|\mathrm{N}(u) \cap B| = |\mathrm{N}(u)| - |\mathrm{N}(u) \cap A| - |\mathrm{N}(u) \cap U| > \frac{\mu}{2} n \text{ .} \tag{6.7}$$

Let $a \in A$ be arbitrary and take a $g \in \mathrm{Aut}(G)$ such that $g(u) = a$. We then have $\mathrm{N}(a) = \mathrm{N}(g(u)) = g(\mathrm{N}(u))$, and in particular $g(\mathrm{N}(u) \cap A) \subseteq \mathrm{N}(a)$.

We claim that there exists an $a' \in \mathrm{N}(u) \cap A$ such that $g(a') \in A$. Indeed, if this was not the case, then $g(x) \in B \cup U$ for each $x \in \mathrm{N}(u) \cap A$. Therefore, we would then have

$$\begin{aligned} |\mathrm{N}(a) \cap (B \cup U)| &= |g(\mathrm{N}(u)) \cap (B \cup U)| \geqslant |(g(\mathrm{N}(u) \cap A) \cap (B \cup U)| \\ &= |g(\mathrm{N}(u) \cap A)| = |\mathrm{N}(u) \cap A| \overset{(6.6)}{>} \mu n/2 \text{ ,} \end{aligned}$$

contradicting (6.4).

Similarly, using (6.7) and the fact that $g(\mathrm{N}(u) \cap B) \subseteq g(\mathrm{N}(u)) = \mathrm{N}(a)$, we get that there exists a $b' \in \mathrm{N}(u) \cap B$ such that $g(b') \in A$. The properties of g, a' and b' contradict (6.5). $\square$

By Claim 6.13.1 we have a partition $U = U_A \cup U_B$, where $U_A := \{u \in U : \deg(u, A) \geqslant d - \frac{3}{4}\mu n\}$ and $U_B := \{u \in U : \deg(u, B) \geqslant d - \frac{3}{4}\mu n\}$. Define $V_1 := A \cup U_A$ and $V_2 := B \cup U_B$. We have $V_1, V_2 \neq \emptyset$. It is straightforward to verify that ${\deg^{\max}}_G(V_1, V_2) \leqslant \mu n$. This contradicts the fact that G is (μn)-robust. □

Observe now that Lemma 6.13 together with Lemma 6.12 immediately imply Theorem 6.9.

We conclude this section with three easy lemmas which are tailored for applications later in the proof of Theorem 6.20.

Lemma 6.14. *Suppose that a graph H is ℓ-iron. Then the 2-blow-up $2 \times H$ is also ℓ-iron.*[2]

Proof. Observe first, that the minimum degree of H is at least $2\ell + 1$. Indeed, if there existed a vertex v with $\deg(v) \leqslant 2\ell$ then this vertex could be isolated from the rest of the graph by deletion of at most ℓ edges incident with v, and at most ℓ vertices in the neighbourhood of v.

Observe that there are two natural vertex disjoint copies of H in $2\times H$, say H_1 and H_2. Consider any sets $E' \subseteq E(2 \times H)$, with $\deg^{\max}(E') \leqslant \ell$ and $V' \subseteq V(2 \times H)$ with $|V'| \leqslant \ell$. Since H is ℓ-iron, both H_1 and H_2 remain connected after the removal of V' and E'. Since the minimum degree of H is at least $2\ell + 1$, then every vertex of H_1 has at least $2\ell + 1$ neighbours in H_2. In particular after the removal of V' and E' there is still an edge between H_1 and H_2 and therefore $(2 \times H) - (V' \cup E')$ is still connected. Therefore $2 \times H$ is ℓ-iron. □

Lemma 6.15. *Let R' be a graph on k' vertices. Suppose that there exist sets $L_1, L_2 \subseteq V(R')$ such that $|L_1| \leqslant \sqrt{\varrho}k'$, and $e(L_2, V(R') - (L_1 \cup L_2)) \leqslant \varrho k'^2$. If there exists disjoint sets $W_1, W_2 \subseteq V(R') - (L_1 \cup L_2)$, such that $\mathrm{N}(W_2) \subseteq L_1 \cup L_2$, and $\min\{|W_1|, |W_2|\} > 2\sqrt{\varrho}k'$, then R' is not $(2\sqrt{\varrho}k')$-iron.*

Proof. Let $L := \{v \in L_2 : \deg(v, V(R') - (L_1 \cup L_2)) \geqslant 2\sqrt{\varrho}k'\}$, and $P := \{v \in V(R') - (L_1 \cup L_2) : \deg(v, L_2) \geqslant 2\sqrt{\varrho}k'\}$. We have $\max\{|L|, |P|\} \leqslant \sqrt{\varrho}k'/2$. In particular,

$$W_1 - (L_1 \cup L \cup P) \neq \emptyset \quad \text{and} \quad W_2 - (L_1 \cup L \cup P) \neq \emptyset\,. \tag{6.8}$$

Define $E' \subseteq E(R')$ to be edges running between $L_2 - L$ and $V(R') - (L_1 \cup L_2 \cup P)$. We have ${\deg^{\max}}_{R'}(E') \leqslant 2\sqrt{\varrho}k'$. By (6.8), R' is not connected after removal of the

[2]In fact it is not much more difficult to show that the 2-blow-up is 2ℓ-iron but ℓ-iron connectivity is enough for our purposes and has a clearer proof.

vertex set $L_1 \cup L \cup P$ and the edge set E'. Indeed, after the removal of E' we have that there are no more edges between $W_2-(L_1\cup L\cup P)$ and $V(R')-(W_2\cup L_1\cup L\cup P)$. Therefore, R' is not $(2\sqrt{\varrho}k')$-iron. $\square$

Lemma 6.16. *Let H be an n-vertex h-strongly connected digraph and let x, y be two distinct vertices of H. Then there exists a (directed) path from x to y of length at most $\frac{n}{h}+1$.*

Proof. By directed version of Menger's Theorem (cf. [15, Theorem 7.3.1(b)]), there exist h internally vertex-disjoint directed paths from x to y. Therefore one of these paths must contain at most $\frac{n-2}{h}$ internal vertices and so must have length at most $\frac{n-2}{h}+1 \leqslant \frac{n}{h}+1$. $\square$

6.6 Bipartite case

In this section we give a fine description of dense vertex-transitive graphs which are almost bipartite. Their properties are stated in Lemma 6.17.

The *edit distance* $dist(G_1, G_2)$ between two n-vertex graph is the number of edges one needs to edit (i.e. to either remove or add) to get G_2 from G_1, minimized over all identification of $V(G_1)$ with $V(G_2)$. Given an n-vertex graph G, we say that it is ε-*close* to a graph property $\mathcal{P}$ if there exists an n-vertex graph $H \in \mathcal{P}$ such that $dist(G,H) < \varepsilon n^2$. Otherwise we say that it is ε-*far* from $\mathcal{P}$.

Lemma 6.17. *Let $c \in (0, \frac{1}{17})$ be arbitrary. Suppose that G is a cn-iron vertex-transitive graph G on n vertices which is c^4-close to bipartiteness. Then there exist a bipartition $V(G) = A\dot{\cup}B$ such that $|A| = |B|$, for each $u \in A$ and each $v \in B$ we have $\deg(u,A) \leqslant 6c^2n$, and $\deg(v,B) \leqslant 6c^2n$. Furthermore, we have $g(A) = A$ or $g(A) = B$ for each $g \in \mathrm{Aut}(G)$.*

Proof. We write Δ for the valency of G. Observe that since G is cn-robust, then $\Delta \geqslant cn$. Let $A \cup B = V(G)$ be the bipartition which maximizes $e(A,B)$. We have

$$e(A) + e(B) < c^4n^2 \; . \tag{6.9}$$

We claim that

$$\min\{|A|,|B|\} \geqslant \frac{n}{3} \; . \tag{6.10}$$

Indeed, suppose for contradiction that, for example, $|A| > \frac{2n}{3}$ and $|B| < \frac{n}{3}$. Counting $e(A,B)$ in two ways we arrive to $\sum_{v\in A}\deg(v) - 2e(A) = \sum_{v\in B}\deg(v) - 2e(B)$, and therefore

$$\frac{2\Delta n}{3} < \Delta|A| \leqslant \Delta|B| + 2c^4n^2 < \frac{\Delta n}{3} + 2c^4n^2,$$

a contradiction as $\Delta \geqslant cn$ and c is sufficiently small. This proves (6.10).

Define $A' := \{v \in A : \deg(v, A) \geqslant c^2 n\}$, and $B' := \{v \in B : \deg(v, B) \geqslant c^2 n\}$. By (6.9) we have $|A'|, |B'| < 2c^2 n$. Together with (6.10) this gives that

$$A - A' \neq \emptyset \quad \text{and} \quad B - B' \neq \emptyset \, . \tag{6.11}$$

Claim 6.17.1. *For each $g \in \mathrm{Aut}(G)$ we either have $|A \cap g(A)| \geqslant |A| - 5c^2 n$ or $|A \cap g(B)| \geqslant |A| - 5c^2 n$. Also, for each $g \in \mathrm{Aut}(G)$ we either have $|B \cap g(A)| \geqslant |B| - 5c^2 n$ or $|B \cap g(B)| \geqslant |B| - 5c^2 n$.*

Proof of Claim 6.17.1. It is enough to prove the first statement.

We start with some general calculations. We shall later use them to show that if $g \in \mathrm{Aut}(G)$ failed to fulfil the assertions we would get a contradiction to cn-iron connectivity.

Let $\tilde{A} := A - A'$ and $\tilde{B} := B - B'$. Consider the partition $V(G) = X \cup Y \cup U$, where $X := (\tilde{A} \cap g(\tilde{A})) \cup (\tilde{B} \cap g(\tilde{B}))$, $Y := (\tilde{A} \cap g(\tilde{B})) \cup (\tilde{B} \cap g(\tilde{A}))$, and $U := V(G) - (X \cup Y)$. We have

$$|U| \leqslant |A'| + |B'| + |g(A')| + |g(B')| \leqslant 4c^2 n \leqslant cn \, . \tag{6.12}$$

We claim that

$$\deg^{\max}{}_G(X, Y) \leqslant cn \, . \tag{6.13}$$

To prove this it suffices to prove that

$$\max\left\{\Delta_{\tilde{A}\tilde{A},\tilde{A}\tilde{B}}, \Delta_{\tilde{A}\tilde{A},\tilde{B}\tilde{A}}, \Delta_{\tilde{B}\tilde{B},\tilde{A}\tilde{B}}, \Delta_{\tilde{B}\tilde{B},\tilde{B}\tilde{A}}, \Delta_{\tilde{A}\tilde{B},\tilde{A}\tilde{A}}, \Delta_{\tilde{B}\tilde{A},\tilde{A}\tilde{A}}, \Delta_{\tilde{A}\tilde{B},\tilde{B}\tilde{B}}, \Delta_{\tilde{B}\tilde{A},\tilde{B}\tilde{B}}\right\} \leqslant \frac{cn}{2} \, , \tag{6.14}$$

where $\Delta_{CD,EF} := \max\{\deg(v, E \cap g(F)) : v \in C \cap g(D)\}$ defines the eight new symbols above. Here we only prove $\Delta_{\tilde{A}\tilde{A},\tilde{A}\tilde{B}} \leqslant \frac{cn}{2}$, the other seven inequalities are similar. Consider an arbitrary $v \in \tilde{A} \cap g(\tilde{A})$. In particular, we have $v \notin A'$. We then have $\deg(v, \tilde{A} \cap g(\tilde{B})) \leqslant \deg(v, \tilde{A}) \leqslant \deg(v, A) < c^2 n$, where the last inequality follows from the definition of the set A'. This establishes (6.14).

Suppose now that the statement of the Claim fails for $g \in \mathrm{Aut}(G)$. We then have $X \neq \emptyset$ and $Y \neq \emptyset$. Indeed, to show for example that $X \neq \emptyset$ we proceed as follows. First, we note that

$$|A \cap g(A)| = |A| - |A \cap g(B)| > 5c^2 n \, .$$

Therefore, we have

$$|X| \geqslant |A \cap g(A)| - |A'| - |g(A')| > 5c^2n - 2c^2n - 2c^2n > 0 \,.$$

Let E' be the edges of G running between X and Y. Now if we remove U and E' from G we get a disconnected graph. Together with the bounds (6.12) and (6.13) this proves that G is not cn-iron, a contradiction. $\square$

Claim 6.17.2. *For every $v \in A$ we have $\deg(v, A) \leqslant 6c^2n$. Also, for every $v \in B$ we have $\deg(v, B) \leqslant 6c^2n$.*

Proof of Claim 6.17.2. By symmetry, it suffices to prove the first part of the statement. Let $w \in B - B'$ be arbitrary; such a choice possible by (6.11). Let $v \in A$ and take $g \in \mathrm{Aut}(G)$ be such that $g(v) = w$. Let $P := \mathrm{N}(v) \cap A$, and $Q := \mathrm{N}(v) \cap B$. Suppose for contradiction that $|P| > 6c^2n$. Since the bipartition $A \cup B$ was chosen to maximize $e(A, B)$, we must have $|Q| \geqslant \frac{cn}{2}$. Since $\mathrm{N}(w) = g(P) \cup g(Q)$ and since also $w \notin B'$ we have that $|g(A) \cap A| \geqslant |g(P) \cap A| > 5c^2n$ and so $|g(A) \cap B| < |B| - 5c^2n$. Similarly, we also have $|g(B) \cap A| \geqslant |g(Q) \cap A| > 5c^2n$ and so $|g(B) \cap B| < |A| - 5c^2n$. But these contradict Claim 6.17.1. $\square$

Claim 6.17.3. *For every $g \in \mathrm{Aut}(G)$ we either have $g(A) = A$, or $g(A) = B$.*

Proof of Claim 6.17.3. Let $C, D \in \{A, B\}$. Let $C' := V(G) - C$, and $D' := V(G) - D$. (Thus $C', D' \in \{A, B\}$.)

Suppose that $C \cap g(D) \neq \emptyset$. We can take a $v \in C$ with $g^{-1}(v) \in D$. Using Claim 6.17.2 for $g^{-1}(v)$, and then for v we get.

$$\begin{aligned}
6c^2n \geqslant &\deg(g^{-1}(v), D) = |\mathrm{N}(g^{-1}(v)) \cap D| = |\mathrm{N}(v) \cap g(D)| \geqslant |\mathrm{N}(v) \cap C' \cap g(D)| \\
&= |\mathrm{N}(v) \cap C'| - |\mathrm{N}(v) \cap C' \cap g(D')| \geqslant |\mathrm{N}(v)| - |\mathrm{N}(v) \cap C| - |C' \cap g(D')| \\
&\geqslant \Delta - 6c^2n - |C' \cap g(D')| \geqslant cn - 6c^2n - |C' \cap g(D')|.
\end{aligned}$$

Thus $|C' \cap g(D')| \geqslant cn - 12c^2n > 5c^2n$. Hence $C' \cap g(D') \neq \emptyset$. Repeating the previous argument for C' and D' yields $|C \cap g(D)| > 5c^2n$.

Therefore for every $C, D \in \{A, B\}$ we have $|C \cap g(D)| = 0$ or $|C \cap g(D)| > 5c^2n$. Combining this with Claim 6.17.1 finishes the proof. $\square$

Claims 6.17.2 and 6.17.3 show that the bipartition A, B satisfies the conclusion of Lemma 6.17. $\square$

Remark 6.18. In the above proof we showed that the partition maximizing $e(A, B)$ satisfies the conclusion of Lemma 6.17. In fact we only used the following two properties of the partition:

1. The partition satisfies (6.9).
2. For every $v \in A$ we have $\deg(v, A) \leqslant \deg(v, B)$ and for every $v \in B$ we have that $\deg(v, B) \leqslant \deg(v, A)$.

In particular any partition satisfying the above two properties also satisfies the conclusion of Lemma 6.17. This fact will be important in the proof of Theorem 6.22 which provides an algorithmic version of Theorem 6.2.

Remark 6.19. Note that the bipartite subgraph $G[A, B]$ obtained from the partition A, B given by Lemma 6.17 by removing all edges within the parts A and B is itself vertex-transitive. Indeed observe that for any automorphism $g \in \mathrm{Aut}(G)$ and any edge e between the parts A and B we have that $g(e)$ also lies between these parts. Therefore every automorphism of G restricted to $G[A, B]$ is also an automorphism and so $G[A, B]$ is vertex-transitive.

6.7 Hamilton cycles in iron connected vertex-transitive graphs

In this section, we prove a stronger version of Theorem 6.2 under the additional assumption of high iron connectivity of the host graph. This is stated in Theorem 6.20 in the non-bipartite setting, and in Theorem 6.21 in the bipartite setting

The basic idea is to follow Łuczak's 'connected matching argument' [77]. The novel ingredient in our work is an innocent looking modification of this technique: we observe that we can extend the argument to work with fractional matchings as well. This allows one to use the LP-duality. We believe that this observation will find further important applications in the future. (After the first version of [23] was posted on the arXiv we learned that Rödl and Ruciński announced a solution of a certain Dirac-type problem for hypergraphs using Farkas' Lemma, an approach similar to our linear programming approach. The corresponding paper later appeared as [8]) The use of the LP-duality in conjunction with the Regularity Lemma originated in discussion with Dan Král' and Diana Piguet. In our current setting of vertex-transitive graphs it turns out that the full strength of our LP-duality machinery is not needed, and that – as was pointed to me by Deryk Osthus – there is a simpler proof of a stronger statement than our Theorems 6.20 and 6.21, first used in [69].

Theorem 6.20. *For every $\beta, \gamma > 0$ and every $C \in \mathbb{N}$, there exists an N_1 such that every βn-iron vertex-transitive graph of order $n \geqslant N_1$ which is β-far from bipartiteness is C-pathitionable with an exceptional set $U \subseteq V(G)$ with $|U| < \gamma n$.*

Theorem 6.21. *For every $c \in (0, \frac{1}{17})$, $\gamma > 0$ and $C \in \mathbb{N}$ there an exists N_2 such that for every vertex-transitive graph G of order $n \geqslant N_2$ the following holds. Suppose G is cn-iron and c^4-close to bipartiteness. Let A, B be the bipartition of G given by Lemma 6.17. Then there exists a set $U \subseteq V(G)$ with $|U| < \gamma n$ such that G is C-bipathitionable with exceptional set U with respect to partition A, B.*

After proving Theorem 6.20 in detail below, we indicate necessary changes to make an analogous proof of Theorem 6.21 work as well.

Proof of Theorem 6.20. We begin by fixing additional constants $\varepsilon, d_1, d_2, \gamma_1, \gamma_2$ satisfying

$$0 < \varepsilon \ll d_1 \ll \gamma_1 \ll \gamma_2 \ll d_2 \ll \gamma, \beta.$$

Let $N' := 1/\varepsilon$. Let $N(\varepsilon, N', 1)$ and $n_0(\varepsilon, N', 1)$ be the numbers given by Lemma 1.1. Set

$$n_0 := \max\left\{\frac{N(\varepsilon, N', 1)}{\gamma_1}, n_0(\varepsilon, N', 1)\right\} . \tag{6.15}$$

Let G be any βn-iron connected vertex-transitive graph on $n \geqslant n_0$ vertices of valency Δ. Apply Lemma 1.1 (see also Remark 1.3) with parameters $\varepsilon, N', \ell := 1$ and d_1, d_2 to G to obtain a partition $V_0, V_1, \ldots, V_k$ of $V(G)$. Let $G_1, G_2 \subseteq G$ be the spanning subgraphs of G given by Lemma 1.1 corresponding to the densities d_1 and d_2 respectively. Let also R_1 and R_2 be the reduced graphs of G with respect to the above partition, the parameters ε and d_1, d_2 and the subgraphs G_1 and G_2 respectively. We write $m := |V_1|$.

We first claim that R_1 has a large fractional matching.

Claim 6.20.1. $\nu^*(R_1) \geqslant (1 - \frac{\gamma_1}{2})\frac{k}{2}$.

Proof of Claim 6.20.1. Observe that by Lemma 6.5 we have that $\tau^*(G) = n/2$. We also have that

$$e(G_1) \geqslant e(G) - \frac{(d_1 + \varepsilon)n^2}{2} \geqslant \left(1 - \frac{\gamma_1}{2}\right) e(G) ,$$

where in the first inequality we used property (1.1) and in the second one we used the fact that $e(G) \geqslant \beta n^2/2$. By Lemma 6.4 we obtain that $\tau^*(G_1) \geqslant (1 - \frac{\gamma_1}{2})\frac{n}{2}$. Observe that $\nu^*(G_1) = \nu^*(G_1 - V_0)$ by the fact that $G_1[V_i]$ is an empty graph.

Therefore, combining Lemma 6.7 with Theorem 6.3(a) we have

$$\nu^*(R_1) \geqslant \frac{\nu^*(G_1)}{m} = \frac{\tau^*(G_1)}{m} \geqslant \left(1 - \frac{\gamma_1}{2}\right)\frac{n}{2m} \geqslant \left(1 - \frac{\gamma_1}{2}\right)\frac{k}{2}\,. \qquad \square$$

The density d_1 was used to find a large matching in R_1 (cf. Claim 6.20.1). On the other hand, it is more convenient to work with the higher threshold density d_2 to infer some connectivity properties of certain graphs that will be derived from R_2 (most importantly, to deduce Claim 6.20.5).

Since G is βn-iron and $\varepsilon, d_2 \ll \beta$, properties of the (ε, d_2)-regular partition imply that $G_2[V - V_0]$ is $(\beta n/2)$-iron. We claim that the iron connectivity is inherited by the reduced graph R_2 as well.

Claim 6.20.2. *R_2 is $(\beta k/2)$-iron.*

Proof of Claim 6.20.2. Indeed, suppose we could disconnect R_2 by removing a set of clusters S of size at most $\beta k/2$ together with an edge set $F \subseteq E(R)$ with $\deg^{\max}(F) \leqslant \beta k/2$. Let $E' \subseteq E(G_2[V - V_0])$ be the set of edges contained in the regular pairs corresponding to F. Then we could also disconnect $G_2[V - V_0]$ by removing all vertices belonging to the clusters of S together with the edge set E'. However, the clusters of S contain at most $\beta km/2 \leqslant \beta n/2$ vertices and also $\deg^{\max}(E') \leqslant \beta km/2 \leqslant \beta n/2$. This would contradict the $(\beta n/2)$-iron connectivity of $G_2[V - V_0]$. $\square$

For each $1 \leqslant i \leqslant k$, we arbitrarily partition V_i into two parts V_i^1 and V_i^2 of equal sizes. In the case that the V_i's have odd sizes, then before the partitioning we move an arbitrary vertex from each cluster into V_0. We denote the new exceptional set obtained by V_0'. We also define a new graph R_1' on vertex set $\{V_1^1, V_1^2, \ldots, V_k^1, V_k^2\}$ where V_i^s is adjacent to V_j^t if and only of V_i was adjacent to V_j in R_1. Similarly, we define a graph R_2' on the same vertex set as R_1' to be the 2-blow-up of R_2. By Lemma 1.5 every edge of R_1' corresponds to a $(3\varepsilon, d_1/2)$-regular pair, and every edge of R_2' corresponds to a $(3\varepsilon, d_2/2)$-regular pair. We have $R_1' = 2 \times R_1$, $R_2' = 2 \times R_2$, and $R_2' \subseteq R_1'$. Consider a matching M in R_1' of weight at least $(1 - \frac{\gamma_1}{2})k$. Such a matching exists by Claim 6.20.1 and by Lemma 6.6.

Observe that R_1' is itself a reduced graph of the partition $V_0', V_1^1, V_1^2, \ldots, V_k^1, V_k^2$ with respect to the parameters 3ε and $d_1/2$ and some subgraph G_1' of G. In particular, we can apply Lemma 1.6 to R_1' and the matching M to remove exactly $3\varepsilon m$ vertices from each cluster of R_1' so that every pair of clusters corresponding to an edge of M is $(6\varepsilon, d_1/4)$-super-regular while every pair of clusters corresponding to an edge of R'_1 is $(6\varepsilon, d_1/4)$-regular. It also follows that every pair of these modified clusters corresponding to an edge of R'_2 is $(6\varepsilon, d_2/4)$-regular.

We now move all clusters of R'_1 which are not incident to the matching M into the exceptional set. This modification is also performed in the graph R'_2. Let k' be the number of clusters of this modified graph R'_1, and m' be the size of each of its clusters, which are denoted by $V'_1, \dots, V'_{k'}$ (and we write V'_0 for the exceptional set).

The modified graph R'_2 is obtained from $2 \times R_2$ by removing a small number of vertices. From Claim 6.20.2 and Lemma 6.14 we get that $2 \times R_2$ is $(\beta k/2)$-iron. Therefore

$$R'_2 \text{ is } (\tfrac{\beta k'}{10})\text{-iron.} \tag{6.16}$$

By the above, there is a partition of the vertices of G into $k' + 1$ classes $V'_0, V'_1, \dots, V'_{k'}$, and a spanning subgraph G' of G with the following properties:

(a) $1/\varepsilon \leqslant k' \leqslant 2N(\varepsilon, N', 1) \leqslant 2\gamma_1 n$ (using the bound (6.15)).

(b) $|V'_0| \leqslant 2\gamma_1 n, |V'_1| = \dots = |V'_{k'}| = m'$.

(c) $\deg_{G'}(v) \geqslant \deg_G(v) - 3\gamma_1 n$ for every $v \in V(G) - V'_0$.

(d) $G'[V'_i]$ is empty for every $0 \leqslant i \leqslant k'$.

(e) All pairs (V'_i, V'_j) with $1 \leqslant i < j \leqslant k'$ are 6ε-regular with density either 0 or at least $d_2/4$.

(f) There is a $\beta k'/10$-iron graph R'_2 on vertex set $V'_1, \dots, V'_{k'}$ such that every edge of R'_2 corresponds to a $(6\varepsilon, d_2/4)$-regular pair in G.

(g) There is a perfect matching M on the complete graph formed by the clusters $V'_1, \dots, V'_{k'}$. Further, every edge of M corresponds to a $(6\varepsilon, d_1/4)$-super-regular pair in G.

Let us denote the edges of M by A_iB_i for $1 \leqslant i \leqslant k'/2$. Using Lemma 1.7 with $\theta := d_2$ for each $1 \leqslant i \leqslant k'/2$, we find A^*_i and B^*_i with $|A^*_i| = |B^*_i| = d_2 m'$, such that (A^*_i, B^*_i) is an $(6\varepsilon/d_2, d_1 d_2/16)$-ideal for (A_i, B_i). The set $U := V'_0 \cup \bigcup_{i=1}^{k'/2} (A^*_i \cup B^*_i)$ is the exceptional set in the statement of the theorem. Observe that $|U| \leqslant 2\gamma_1 n + d_2 n \leqslant \gamma n$. Suppose now we are in the setting of the theorem, that is, we are given distinct vertices $x_1, y_1, \dots, x_\ell, y_\ell \in V(G) - U$ (where $1 \leqslant \ell \leqslant C$), and our task is to find a system $\mathcal{S}$ of ℓ vertex-disjoint of paths which partition $V(G)$. Furthermore it is required that x_j and y_j are the endvertices of the j-th path.

Our first aim is to find a system $\mathcal{S}'$ of ℓ vertex-disjoint paths, the j-th path having endvertices x_j and y_j with the following properties.

(A1) $V(\mathcal{S}')$ contains all vertices of V_0';

(A2) for each $i \in [k'/2]$, we have that $|V(\mathcal{S}') \cap A_i| = |V(\mathcal{S}') \cap B_i|$;

(A3) for each $i \in [k'/2]$, there is an edge of $\mathcal{S}'$ whose respective endvertices lie in A_i and B_i;

(A4) for each $i \in [k'/2]$, we have that $|V(\mathcal{S}') \cap A_i^*| = |V(\mathcal{S}') \cap B_i^*| = 0$.

Having obtained this system $\mathcal{S}'$ then we can find a complete extension $\mathcal{S}$ of $\mathcal{S}'$ as follows: For each $i \in [k'/2]$ let $e_i := a_ib_i$ be an edge of $\mathcal{S}'$ with $a_i \in A_i$ and $b_i \in B_i$ as guaranteed by (A3). Since (A_i^*, B_i^*) is an $(\frac{6\varepsilon}{d_2}, \frac{d_1d_2}{16})$-ideal for (A_i, B_i) and since by property (A4) the system $\mathcal{S}'$ does not meet $A_i^* \cup B_i^*$, we have that the pair $(A_i - V(\mathcal{S}'), B_i - V(\mathcal{S}'))$ is $(\frac{6\varepsilon}{d_2}, \frac{d_1d_2}{16})$-super-regular. By property (A2) we also have that $|A_i - V(\mathcal{S}')| = |B_i - V(\mathcal{S}')|$ so we can apply Lemma 6.8 to deduce that $G[(A_i \cup B_i) - V(\mathcal{S}')]$ contains a Hamilton path P_i with endvertices a_i and b_i. We now replace the edges e_i by the paths P_i for each $1 \leqslant i \leqslant k'/2$ to obtain a new system $\mathcal{S}$ containing all vertices of $V_1' \cup \cdots \cup V_{k'}'$. Since by property (A1) it also contains all vertices of V_0', then $\mathcal{S}$ is a complete extension of $\mathcal{S}'$ as asserted by the theorem.

It therefore remains to prove that we can find a system $\mathcal{S}'$ satisfying the properties (A1)-(A4). In order to prove that, it will be actually more convenient to demand $\mathcal{S}'$ to satisfy the following strengthening of property (A2) as well:

(A2′) for each $i \in [k'/2]$, we have that $|V(\mathcal{S}') \cap A_i| = |V(\mathcal{S}') \cap B_i| \leqslant 2C\sqrt{\gamma_1}m'$.

Let $z_1, \ldots, z_r$ denote the vertices of the exceptional set V_0'.

Claim 6.20.3. *There exist distinct vertices $u_1, v_1, \ldots, u_r, v_r \in V(G) - V_0'$ such that $u_i, v_i \in \mathrm{N}_G(z_i)$ for each $i \in [r]$. Furthermore, we have $|V_j' \cap \{u_1, v_1, \ldots, u_r, v_r\}| \leqslant \sqrt{\gamma_1}m'$ for each $j \in [k']$.*

Proof of Claim 6.20.3. The vertices $u_1, v_1, \ldots, u_r, v_r$ can be chosen greedily. We proceed sequentially for $i = 1, \ldots, r$. When choosing the neighbors u_i and v_i of z_i, there are at most d_2n vertices which are not allowed to be chosen because they belong to some A_i^* or some B_i^*, at most $3r \leqslant 6\gamma_1 n$ vertices which are not allowed to be chosen because they either belong to V_0' or they have been already chosen as neighbors of another z_j $(j < i)$, and finally there are at most $4\sqrt{\gamma_1}n$ vertices which are not allowed to be chosen because they belong to clusters from which we have already chosen $\sqrt{\gamma_1}m'$ vertices. To see the last claim observe that since we will choose a total of $2r \leqslant 4\gamma_1 n$ vertices u_i, v_i, there are at most $4\gamma_1 n/\sqrt{\gamma_1}m'$ clusters from which we have already chosen $\sqrt{\gamma_1}m'$ vertices, and these clusters contain at most $4\sqrt{\gamma_1}n$ vertices.

So in total there are at most $(d_2+6\gamma_1+4\sqrt{\gamma_1})n$ vertices which are not allowed to be chosen. But since the valency of G is $\Delta \geqslant \beta n$ and $d_2, \gamma_1 \ll \beta$ we can indeed choose the vertices u_i and v_i greedily. $\square$

The system $\mathcal{S}' = \{P_1, \ldots, P_\ell\}$ will be such that the path P_1 will contain all the 2-paths $u_i z_i v_i$ (for $i = 1, \ldots, r$) and edges $u_i v_i$ (for $i = r+1, \ldots, r'$). Therefore, the path P_1 alone will guarantee (A3), i.e., for every $i \in [k'/2]$ there is an edge of P_1 between A_i and B_i. Further, the path P_1 alone will absorb all the vertices of V_0'. The paths P_j (for $j \in \{2, \ldots, \ell\}$) will be the shortest connections between x_j and y_j (subject to further requirements, to be specified later). We describe in detail the construction of the path P_1; the construction of the paths $P_2, \ldots, P_\ell$ is easier as they do not have to absorb any special vertices.

For each $i \in [k'/2]$, we take an edge $u_{r+i}v_{r+i} \in E(G)$ such that $u_{r+i} \in A_i - A_i^*$ and $v_{r+i} \in B_i - B_i^*$. Furthermore, we require that u_{r+i} and v_{r+i} are distinct from $\{u_1, v_1, \ldots, u_r, v_r\}$. Such a choice is possible as (A_i, B_i) forms a dense regular pair. Set $r' := r + k'/2$. The bounds $|V_0'| \leqslant 2\gamma_1$, and $k' \leqslant \gamma_1 n$ (which is implied by (a)) give that

$$r' \leqslant 3\gamma_1 n \ . \tag{6.17}$$

Our next task on our way to constructing the path P_1 is for each $0 \leqslant i \leqslant r'$ to find a path Q_i in G with endvertices v_i and u_{i+1}; here $v_0 := x_1$ and $u_{r'+1} := y_1$. The path P_1 will be the union of these paths together with the 2-paths $u_i z_i v_i$ (for $i \in [r]$) and the edges $u_i v_i$ (for $i = r+1, \ldots, r'+1$). To guarantee that $\mathcal{S}'$ satisfies properties (A1)-(A4) and (A2$'$) we will require that the paths Q_i satisfy the following properties:

(B1) the paths Q_i are disjoint and do not contain any vertex from V_0';

(B2) for each $0 \leqslant i \leqslant r'$ and each $1 \leqslant j \leqslant k'/2$ we have that $|V(Q_i) \cap A_j| = |V(Q_i) \cap B_j|$;

(B3) for each $1 \leqslant j \leqslant k'/2$, we have that $|V(\cup_i Q_i) \cap A_j|, |V(\cup_i Q_i) \cap B_j| \leqslant 2\sqrt{\gamma_1}m'$;

(B4) for each $0 \leqslant i \leqslant r'$ and each $1 \leqslant j \leqslant k'/2$ we have that $|V(Q_i) \cap A_j^*| = |V(Q_i) \cap B_j^*| = 0$;

To achieve this aim we will further demand that the following property is also satisfied:

(B5) for each $0 \leqslant i \leqslant r$, the path Q_i has length at most $\gamma_1^{-1/3}$.

Let us now show how can this be done. Suppose we have already found the paths $Q_0, Q_1, \ldots, Q_{i-1}$ and we are now at the stage where we require a path from v_i to u_{i+1}.

We use (B5) and (6.17) and infer that the paths $Q_0, Q_1, \ldots, Q_{i-1}$ contain at most $i'\gamma_1^{-1/3} \leqslant r'\gamma_1^{-1/3} \leqslant 3\gamma_1^{2/3}n$ vertices. In particular, we have the following.

Claim 6.20.4. *There are at most $3\gamma_1^{2/3}n/(\gamma_1^{1/2}m') \leqslant 4\gamma_1^{1/6}k'$ indices $j \in [k'/2]$ for which $|V(Q_1 \cup \cdots \cup Q_{i-1}) \cap A_j| \geqslant \sqrt{\gamma_1}m$ or $|V(Q_1 \cup \cdots \cup Q_{i-1}) \cap B_j| \geqslant \sqrt{\gamma_1}m$.*

When choosing Q_i, we will make sure that no vertex of Q_i is contained in such clusters except possibly the first four and the last four vertices of Q_i. (It might happen that v_i or u_{i+1} belong to such clusters so in this case Q_i definitely cannot avoid these clusters completely. By using at most four vertices, and the high min-degree of R_1' we will be able to get out of these forbidden clusters and then we will make sure that we never visit them again.) If we can achieve this then we can guarantee that for each $1 \leqslant j \leqslant k'/2$, we have that

$$|V(\cup_i Q_i) \cap A_j|, |V(\cup_i Q_i) \cap B_j| \leqslant \sqrt{\gamma_1}m' + \gamma_1^{-1/3} + 8(r'+1) \leqslant 2\sqrt{\gamma_1}m',$$

as required by property (B3).

For finding the paths Q_i we will need to use an auxilary digraph R^*, which should be viewed as a "shifted version" of R_2'. The vertex set of R^* is the same as the vertex set of R_2' while its edge set is defined as

$$\begin{aligned} E(R^*) := &\left\{\overrightarrow{B_jA_i}, \overrightarrow{B_iA_j} : A_iA_j \in E(R_2')\right\} \\ &\cup \left\{\overrightarrow{A_jB_i}, \overrightarrow{A_iB_j} : B_iB_j \in E(R_2')\right\} \cup \left\{\overrightarrow{A_jA_i}, \overrightarrow{B_iB_j} : A_iB_j \in E(R_2'), i \neq j\right\} . \end{aligned}$$

Claim 6.20.5. *The digraph R^* is $\left(\frac{d_2\beta^2k'}{1000}\right)$-strongly connected.*

Proof of Claim 6.20.5. Suppose that R^* is not $\left(\frac{d_2\beta^2k'}{1000}\right)$-strongly connected. That means that we can write $V(R^*) = S_0 \cup S_1 \cup S_2$, where $|S_0| < d_2\beta^2k'/1000$, $S_1, S_2 \neq \emptyset$, and there are no directed edges from S_1 to S_2. If $XY \in M$, then we call Y the *partner* of X. We partition further each S_i $(i = 1, 2)$ into three sets:

$$S_i^0 := \{X \in S_i : \text{partner of } X \text{ is in } S_0\},$$
$$S_i^1 := \{X \in S_i - S_i^0 : \text{partner of } X \text{ is in } S_{3-i}\},$$
$$S_i^2 := \{X \in S_i - S_i^0 : \text{partner of } X \text{ is in } S_i\}\ .$$

(See Figure 6.3(a).) For the set $L_1 := S_0 \cup S_1^0 \cup S_2^0$ we have

$$|L_1| \leqslant \frac{d_2\beta^2 k'}{500}\ . \tag{6.18}$$

The graph R_1' can be viewed as an edge-weighted graph, where the weight of each its edge is the density of the corresponding regular pair. Thus the weights used on the edges of R_1' are in the interval $[d_1, 1]$. In particular, we have the notion of weighted degree which is defined for a cluster $X \in V(R_1')$ as the sum of weights of edges incident with X, and is denoted $\overline{\deg}(X)$. Observe that the property that all vertices of G have the same degree gets inherited on the weighted graph R_1', i.e., for each cluster V_i', $(i \in [k'])$ we have that its weighted degree satisfies

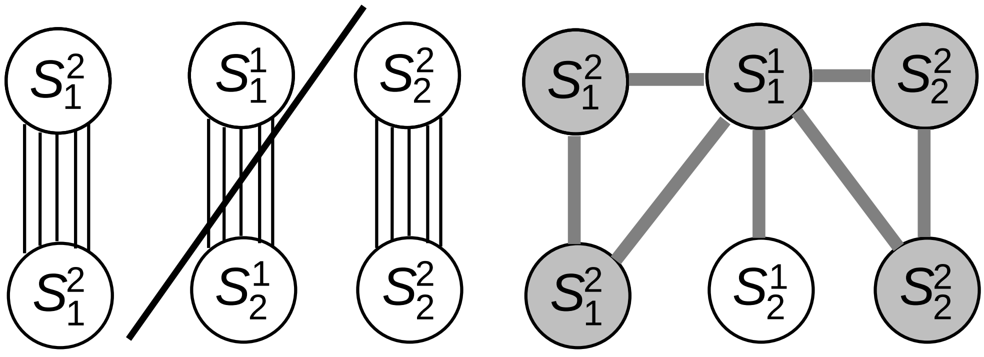
$$(1-\gamma_2)\frac{\Delta k'}{n} \leqslant \overline{\deg}(V_i') \leqslant (1+\gamma_2)\frac{\Delta k'}{n}\ . \tag{6.19}$$

The set S_2^1 is independent in R_2' by the definition of the graph R^*. Indeed, suppose that there is an edge $XY \in E(R_2')$ inside S_2^1. Then, by the definition of R^*, there is a directed edge from the partner of X (which is in S_1^1) to Y, a contradiction to the assumption that there are no directed edges frmo S_1^1 to S_2^1. Further it can be similarly checked, that there are no edges between S_2^1 and $S_1^2 \cup S_2^2$, or between S_1^2 and S_2^2. This is depicted on Figure 6.3(b).

At this point, we distinguish three cases. Suppose first that $S_1^1 = \emptyset$. Then the set L_1 witnesses (using the bound (6.18)) that R_2' is not $\left(\frac{d_2\beta^2 k'}{500}\right)$-vertex connected, and therefore not $\left(\frac{d_2\beta^2 k'}{500}\right)$-iron. This contradicts (6.16). It remains to consider

- Case A: $S_1^1 \neq \emptyset$ and $\max\{|S_1^2|, |S_2^2|\} > \frac{\beta k'}{2}$, and
- Case B: $S_1^1 \neq \emptyset$ and $\max\{|S_1^2|, |S_2^2|\} \leqslant \frac{\beta k'}{2}$.

Before diving into Case A and Case B separately, we make some calculations which will turn out to be useful in both cases.



(a) Separation of the digraph R^*. There are no directed edges crossing from left to right. Vertices of $S_0 \cup S_1^0 \cup S_2^0$ are omitted from the picture.

(b) The situation in the graph R_2'. Allowed edges are depicted in grey.

Figure 6.3: The digraph R^* and the graph R_1'. The sets S_i^2 are split into two according to an arbitrary orientation given by the matching M.

We have

$$\begin{aligned}\sum_{W \in S_2^1} \mathrm{d}\overline{\mathrm{eg}}(W, S_1^1) &\geqslant \sum_{W \in S_2^1} (\mathrm{d}\overline{\mathrm{eg}}(W) - |L_1|) \\ &\overset{(6.19),(6.18)}{\geqslant} |S_2^1| \left((1-\gamma_2)\frac{\Delta k'}{n} - \frac{d_2 \beta^2 k'}{500} \right) . \end{aligned} \tag{6.20}$$

Using the facts that $R_2' \subseteq R_1'$ and that edges of R_2' correspond to pairs of density at least d_2 we have

$$\begin{aligned} e_{R_2'}\left(S_1^2 \cup S_2^2, S_1^1\right) + e_{R_2'}\left(S_1^1\right) &\leqslant \frac{1}{d_2}\left(\sum_{W \in S_1^1} \mathrm{d}\overline{\mathrm{eg}}(W) - \sum_{W \in S_1^1} \mathrm{d}\overline{\mathrm{eg}}(W, S_2^1) \right) \\ &\overset{(6.19)}{\leqslant} \frac{1}{d_2}\left(|S_1^1|(1+\gamma_2)\frac{\Delta k'}{n} - \sum_{W \in S_1^1} \mathrm{d}\overline{\mathrm{eg}}(W, S_2^1) \right) \\ \text{[hand-shaking lemma]} \quad &\overset{(6.20)}{\leqslant} \frac{|S_1^1|}{d_2}\left(2\gamma_2 \frac{\Delta k'}{n} + \frac{d_2 \beta^2 k'}{500} \right) \leqslant \frac{\beta^2 k'^2}{400} . \end{aligned} \tag{6.21}$$

We now turn to dealing with Case A. In this case it is our aim to show that R_2' is not $(\frac{\beta k'}{10})$-iron.

First, we show that $|S_2^1| > \frac{\beta k'}{2}$. Indeed, consider $A \in S_2^1$ arbitrary. As $|S_2^1| = |S_1^1| > 0$, such an A exists. As $\deg_{R_2'}(A) \geqslant (1-2\gamma_2)\frac{\Delta k'}{n}$, and as A can send edges (in the graph R_2') only to L_1 and S_1^1, we get

$$|S_2^1| = |S_1^1| \geqslant (1-2\gamma_2)\frac{\Delta k'}{n} - |L_1| \overset{(6.18)}{>} \frac{\beta k'}{2} .$$

We now utilize the assumptions of Case A. Without loss of generality, assume that $|S_1^2| > \frac{\beta k'}{2}$. Set $\varrho := \frac{\beta^2}{400}$. The set $L_2 := S_1^1$ satisfies by (6.21) that $e_{R_2'}(L_2, V(R_2') - (L_1 \cup L_2)) \leqslant \varrho(k')^2$. Further, we have two disjoint sets $W_1 := S_2^1$ and $W_2 := S_1^2$ with $\mathrm{N}(W_2) \subseteq L_1 \cup L_2$, and $\min\{|W_1|, |W_2|\} > 2\sqrt{\varrho}k'$. Therefore, Lemma 6.15 applies, and we get that R_2' is not $(2\sqrt{\varrho}k')$-iron. This contradicts (6.16).

It remains to consider Case B. In this case we get a contradiction by showing that G is close to a bipartite graph.

Indeed, consider first a partition $W \dot\cup S_2^1 = V(R')$, where $W := S_1^2 \cup S_1^1 \cup S_2^2 \cup L_1$. The graph R_2' is almost bipartite with respect to the partition $W \dot\cup S_2^1$ since S_2^1 is independent and W is very sparse as the following calculation shows:

$$\begin{aligned} e_{R_2'}(W) &\leqslant e_{R_2'}(S_1^1) + k'\left(|S_1^2 \cup S_2^2 \cup L_1|\right) \\ \text{[by Case B, (6.18)]} \quad &\leqslant e_{R_2'}(S_1^1) + \frac{d_2\beta^2 k'^2}{500} \\ &\overset{(6.21)}{\leqslant} \frac{\beta k'^2}{3} . \end{aligned}$$

The partition $W \dot\cup S_2^1 = V(R_2')$ induces a partition $A \dot\cup B$ of G (placing the vertices of V_0' to the sets A and B arbitrarily), such that $e_G(A) + e_G(B) \leqslant \frac{\beta n^2}{3} + 4d_2 n^2 < \beta n^2$. This is a contradiction to the fact that G is β-far from bipartitness. $\square$

Recall that we were looking for a path Q_i from v_i to u_{i+1}. Let us write X for the cluster containing v_i, Y for the cluster containing u_{i+1} and Z for the partner of Y. By Claim 6.20.4 there were at most $4\gamma_1^{1/6}k'$ clusters which we wanted to make sure that their vertices are avoided by Q_i (except perhaps the first four and last four vertices of Q_i). Let us write S for the set of these clusters. Since by Claim 6.20.5 R^* is $\left(\frac{d_2\beta^2 k'}{1000}\right)$-strongly connected and since also $4\gamma_1^{1/6}k' \ll \frac{d_2\beta^2 k'}{1000}$, we have that the digraph $R^* - S$ is $\left(\frac{d_2\beta^2 k'}{2000}\right)$-strongly connected. By Lemma 6.16 there is a directed path Q_i' in R^* from X to Z avoiding S of length at most $\frac{2000}{d_2\beta^2} + 1 \ll \gamma_1^{-1/3}$. Suppose $Q_i' = X_1X_2\cdots X_t$ where $X_1 := X$ and $X_t := Z$. For $i \in [t]$, let Y_i be the partner of X_i. It follows from the definition of $E(R^*)$ that $Q_i'' := X_1Y_1X_2Y_2\cdots X_tY_t$ is a path in R. Observe that by our construction, if a cluster belongs to S then so does its partner. Therefore, since Q_i' avoids S, so does Q_i''. Since for each $j \in [t]$ the pair X_jY_j is $(6\varepsilon, d/4)$-super-regular and for each $j \in [t-1]$ the pair Y_jX_{j+1} is $(6\varepsilon, d_1/4)$-

regular, it follows that we can find a path $Q_i = p_1q_1r_1s_1p_2q_2r_2s_2 \cdots p_tq_tr_ts_t$ in G, where $p_1 := v_i, s_t := u_{i+1}$, and for each $j \in [t]$, $p_j, r_j \in X_j$ and $q_i, s_j \in Y_j$. Furthermore, we can assume that Q_i avoids all vertices of $Q_1, \dots, Q_{i-1}$ and all vertices of $A_1^*, B_1^*, \dots, A_{k'/2}^*, B_{k'/2}^*$. To see that this is indeed the case, for every $j \in [t-1]$ we first fix the edges $s_jp_{j+1} \in X_jY_{j+1}$ and then for each $j \in [t]$, within each super-regular pair (A_j, B_j) we find the paths $p_jq_jr_js_j$. These paths indeed exist by the super-regularity of the pair. Observe that by construction, the paths $Q_0, Q_1 \dots, Q_r$ satisfy all properties (B1)-(B5).

Construction of other paths P_i $(i > 1)$ again uses the auxiliary graph R^* in the same manner. □

Sketch of the proof of Theorem 6.21. Let A, B be the partition given by Lemma 6.17. By passing to the subgraph $G[A, B]$ we can assume that the input graph G is bipartite. Remark 6.19 guarantees that this modified graph is still vertex-transitive and Lemma 6.17 guarantees that it has high iron connectivity.

The proof works very similar to the proof of Theorem 6.20. We just draw attention to three small differences:

First, the Regularity Lemma must be applied with prepartition A, B. Let $\mathcal{A}$ and $\mathcal{B}$ be the clusters inside A, and B, respectively.

Second, when finding good partners u_i and v_i for exceptional vertex z_i, we require that

$$u_i, v_i \in B \text{ if } z_i \in A \text{ and } u_i, v_i \in A \text{ if } z_i \in B. \tag{6.22}$$

Last, Claim 6.20.5 need not hold in the bipartite setting. Indeed, typically clusters in $\mathcal{A}$ form one component and clusters inside $\mathcal{B}$ form another component of the auxiliary digraph R^*. It can be proven (using the same methods) that both graphs $R^*[\mathcal{A}]$ and $R^*[\mathcal{B}]$ have high strong connectivity. This is sufficient in the bipartite case. The key for the entire embedding working is that (6.1), (6.22) and the fact that all edges of M cross between $\mathcal{A}$ and $\mathcal{B}$ guarantee that all the paths will automatically occupy the same number of vertices in A as in B. □

6.8 The proof of Theorem 6.2

We first set up constants. Let $\beta_{\mathrm{T6.9}}$, $R_{\mathrm{T6.9}}$, and N_0 be given by Theorem 6.9 for input parameter α. Let N_1 be given by Theorem 6.20 for input parameters $\beta_{\mathrm{T6.20}} := \beta_{\mathrm{T6.9}}^4$, $C_{\mathrm{T6.20}} := R_{\mathrm{T6.9}}$, and $\gamma_{\mathrm{T6.20}} := \frac{1}{10R_{\mathrm{T6.9}}}$. Let N_2 be given by Theorem 6.21 for input parameters $c_{\mathrm{T}6.21} := \min\{\beta_{\mathrm{T}6.20}, \frac{1}{18}\}$ and $C_{\mathrm{T}6.21} := 4R_{\mathrm{T}6.9}$. Let

$$n_0 := \max\{N_0, 100R^3_{\mathrm{T}6.9}, 10R_{\mathrm{T}6.9}N_1, 10R_{\mathrm{T}6.9}N_2\} .$$

Suppose now we are in the setting of the theorem.

Consider a partition $V_1, \ldots, V_r$ of $V(G)$ given by Theorem 6.9. We have $r < R_{\mathrm{T}6.9}$. We call the sets $V_1, \ldots, V_r$ *continents*. If $r = 1$ then the existence of a Hamilton cycle follows. Indeed, consider first the case when G is $c^4_{\mathrm{T}6.21}$-far from bipartiteness. Let $U_1 \subseteq V(G)$ be the exceptional set given by Theorem 6.20. There exist an edge $xy \in E(G - U_1)$. Using 1-pathitionability of G there exists a Hamilton path from x to y. This path together with the edge xy forms a Hamilton cycle. If on the other hand G is $c^4_{\mathrm{T}6.21}$-close to bipartiteness, then an analogous construction using Theorem 6.21 instead of Theorem 6.20 works.

It remains to consider the case $r > 1$. Let $m := |V_1|$. The proof now splits into two cases. The first case deals with the situation when the graphs $G[V_i]$ are far from bipartiteness. The second case deals with the setting when the graphs $G[V_i]$ are close to bipartiteness. In both cases one needs to glue paths of the graphs $G[V_i]$ (these paths are guaranteed by pathitionability and bipathitionability, respectively) into one Hamilton cycle.

Case I: *All the graphs $G[V_i]$ are $c^4_{\mathrm{T}6.21}$-far from bipartiteness.*

We write $k := \frac{2}{n}\sum_{1\leqslant i<j\leqslant r} e(V_i, V_j)$. By the symmetry of our partition, each vertex sends exactly k edges outside its own continent. A pair V_iV_j is *fat* is there exists a matching of size at least $\frac{m}{r}$ in $G[V_i, V_j]$. If $e(V_i, V_j) > 0$ but V_iV_j is not fat then we say that V_iV_j is *thin*. Let k' be the number of edges any vertex v sends into thin pairs. By vertex-transitivity, k' does not depend on the choice of v.

Claim 6.2.1. *We have $e(V_i, V_j) < \frac{k'm}{r}$ for each thin pair V_iV_j.*

Proof of Claim 6.2.1. Suppose that

$$e(V_i, V_j) \geqslant \frac{k'm}{r} . \tag{6.23}$$

We claim that V_iV_j is fat. To this end it suffices by König's Matching Theorem to show that there is no vertex cover of $G[V_i, V_j]$ of size less than $\frac{m}{r}$. This is in turn implied by (6.23) and by the fact that $\deg^{\max}{}_G(V_i, V_j) \leqslant k'$. □

Claim 6.2.2. *There does not exist any thin pair.*

Proof of Claim 6.2.2. Let K be the number of edges in thin pairs incident to V_1. We have $K = mk'$. On the other hand, using Claim 6.2.1, we have $K \leqslant (r-1)\frac{k'm}{r}$.

Therefore, $mk' \leqslant \frac{r-1}{r}mk'$, and consequently $k' = 0$. $\square$

We construct an auxiliary graph H on the vertex set $\mathcal{V} := \{V_1, \ldots, V_r\}$. The edges of H are formed by fat pairs. From the fact that G is connected, and from Claim 6.2.2 we get that H is connected. Let T be a spanning tree of H. Rooting T at its vertex V_1 we get the notion of *children* of a continent V_i, and of a *parent* $\mathrm{Par}(V_i)$ of V_i (the parent $\mathrm{Par}(V_i)$ is defined only when $i \neq 1$).

Let $U_1 \subseteq V_1, \ldots, U_r \subseteq V_r$ be the exceptional sets given by Theorem 6.20. We have $|U_i| < \gamma_{\mathrm{T6.20}} m$. Each graph $G[V_i]$ is $C_{\mathrm{T6.20}}$-pathitionable with exceptional set U_i. For each fat pair V_iV_j let $M_{i,j} \subseteq G[V_i, V_j]$ be a matching of size at least $\frac{m}{r}$.

Claim 6.2.3. *There exists a family M consisting of two matching edges $x^-_{i,j}y^-_{i,j}, x^+_{i,j}y^+_{i,j}$ from each $M_{i,j}$ with $V_iV_j \in E(T)$ and $V_j := \mathrm{Par}(V_i)$ having the following properties:*

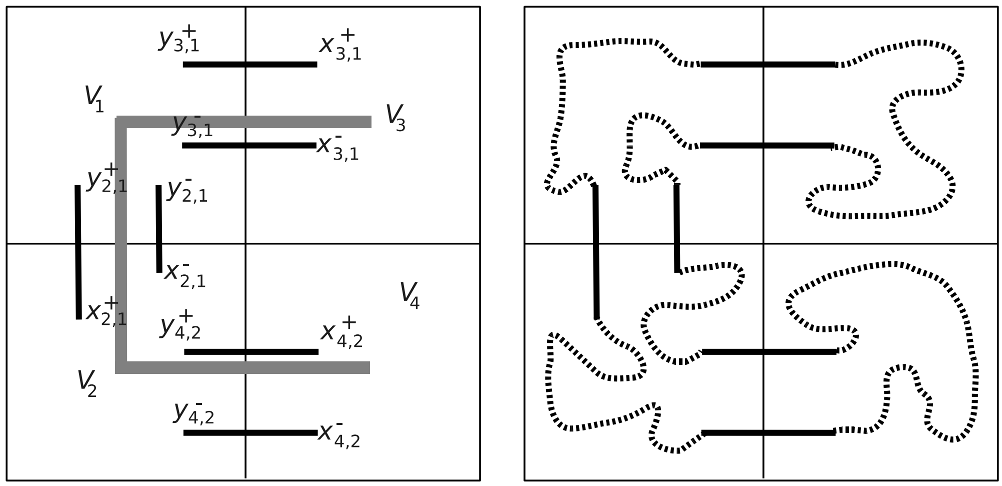
- *$x^-_{i,j}, x^+_{i,j} \in V_i$ and $y^-_{i,j}, y^+_{i,j} \in V_j$ for any $V_iV_j \in E(T), V_j = \mathrm{Par}(V_i)$,*
- *M is a matching in G, and*
- *$V(M) \cap \bigcup_{i=1}^r U_i = \emptyset$.*

Proof of Claim 6.2.3. The statement follows by greedily choosing two edges from each matching $M_{i,j}$ subject to restrictions above. Since the sets U_i and U_j each forbids at most $\gamma_{\mathrm{T6.20}} m$ edges of $M_{i,j}$, and the already chosen edges $x^-_{i',j'}y^-_{i',j'}, x^+_{i',j'}y^+_{i',j'}$ (where $(i', j') \neq (i, j)$) forbid at most $4(r-1)$ edges, and since we have $2\gamma_{\mathrm{T6.20}} m + 4(r-1) + 2 \leqslant |M_{i,j}|$, the choice of $x^-_{i,j}y^-_{i,j}$ and $x^+_{i,j}y^+_{i,j}$ is possible. $\square$

Given the family $M = \{x^-_{i,j}y^-_{i,j}, x^+_{i,j}y^+_{i,j} \subseteq M_{i,j}\}_{V_iV_j \in E(T), V_j = \mathrm{Par}(V_i)}$ from Claim 6.2.3 we are now ready to construct the desired Hamilton cycle. The first step is to decompose each continent V_i into a system of paths $\mathcal{S}_i$. To describe $\mathcal{S}_i$ we need to distinguish three cases based on the position of V_i in T.

- <u>*V_i is the root of T (i.e., $i = 1$).*</u>
  Let $V_{i_1}, \ldots, V_{i_p}$ be the children of V_1. As $p \leqslant r \leqslant C_{\mathrm{T6.20}}$, we have that $G[V_i]$ is p-pathitionable with exceptional set U_i. Define $V_{i_{p+1}} = V_{i_1}$. Let $\mathcal{S}_1$ be a decomposition of V_1 into p paths such that the j-th path begins in $y^+_{i_j,1}$ and ends in $y^-_{i_{j+1},1}$. Such a system of paths exists thanks to the p-pathitionability of $G[V_1]$.

- <u>*V_i is a leaf of T, and $i \neq 1$.*</u>
  Let $V_{i'}$ be the parent of V_i. Let $\mathcal{S}_i$ consist of a (single) Hamilton path starting in $x^-_{i,i'}$ and ending in $x^+_{i,i'}$. Such a path exists thanks to the 1-pathitionability of $G[V_i]$.



(a) An example of a partition of G into continents $V_1, \ldots, V_4$ together with tree T (depicted in grey), and edges $x^-_{i,j}y^-_{i,j}$, $x^+_{i,j}y^+_{i,j}$.

(b) The final Hamilton cycle. The systems $\mathcal{S}_i$ are depicted by dotted lines.

Figure 6.4: Gluing together the paths $\mathcal{S}_i$ and M.

- *V_i is an internal vertex of T, and $i \neq 1$.*
  Let $V_{i'}$ be the parent of V_i. Let $V_{i_1}, \ldots, V_{i_q}$ be the children of V_1. As $q < r \leqslant C_{\mathrm{T6.20}}$, we have that $G[V_i]$ is $(q+1)$-pathitionable with exceptional set U_i. Then let $\mathcal{S}_i$ consist of $q+1$ paths $P_0, P_1, \ldots, P_q$ which decompose V_i. We require that P_0 has endvertices $x^+_{i,i'}$ and $y^+_{i_1,i}$. The endvertices of the path P_j ($j \in [q-1]$) are required to be $y^-_{i_j,i}$ and $y^+_{i_{j+1},i}$. Last, the endvertices of the path P_q are required to be $y^-_{i_q,i}$ and $x^-_{i,i'}$. Such a system of path exists thanks to the $(q+1)$-pathitionability of $G[V_i]$.

It can be easily checked that M together with the system $\{\mathcal{S}_i\}_{i=1}^r$ forms a Hamilton cycle in G. See Figure 6.4 for an example.

Case II: *All the graphs $G[V_i]$ are $c^4_{\mathrm{T6.21}}$-close to bipartiteness.*
Let A_i and B_i be the partition of each graph $G[V_i]$ given by Lemma 6.17 with input constant $c_{\mathrm{T6.21}}$. Let $\mathcal{W} := \{A_1, B_1, A_2, B_2, \ldots, A_r, B_r\}$. Elements of $\mathcal{W}$ are called *bicontinents*. A pair XY of elements of $\mathcal{W}$ is said to be *bifat* if $G[X,Y]$ contains a matching of size at least $\frac{m}{2r}$. If $e(X,Y) > 0$ but XY is not bifat then we call the pair XY *bithin*.

Claim 6.2.4. *There does not exist a bithin pair.*

Proof of Claim 6.2.4. The proof translates of the claim mutatis mutandis from the proof of Claim 6.2.2. □

Let H be a graph on the vertex set $\mathcal{W}$, where a pair XY forms an edge of H if XY is bifat. Observe that $A_iB_i \in E(H)$ for every $i \in [r]$. In particular, since G is connected, H is connected as well.

As in Case I we can find matching $M_{XY} = M_{YX}$ for each $XY \in E(H)$ with the following properties:

- $M_{XY} \subseteq G[X, Y]$, $|M_{XY}| = 2$,
- $M = \bigcup_{XY \in E(H)} M_{XY}$ is a matching in G, and
- $V(M) \cap \bigcup_{i=1}^{r} U_i = \emptyset$.

As it will turn out the role of the edges in matchings $M_{A_iB_i}$ is somewhat inferior: they are just used to guarantee connectivity of H, and – unlike other matchings M_{XY} – they are not guaranteed to lie on the resulting Hamilton cycle. Therefore, we write $M' := M - \bigcup_{i=1}^{r} M_{A_iB_i}$.

Let H' be a clone of H with each original edge of H replaced by two parallel edges. Since H' is connected and all its degrees are even we can find an Eulerian circuit $\mathcal{E}$ in H'. Also, observe that H' is vertex-transitive, and in particular, we have

$$\deg_{H'}(A_i) = \deg_{H'}(B_i) \;, \tag{6.24}$$

for any $i \in [r]$.

The aim is to use $\mathcal{E}$ to find a Hamilton cycle in G. To this end we find requirements for systems of paths $\mathcal{S}_i$ within each graph $G[V_i]$.

We identify (in a natural way) edges of H' with edges in M. Therefore, $\mathcal{E}$ may be viewed as moving between bicontinents. During each (say, j-th) visit of $X \in \mathcal{W}$ we remember vertex $a_{X,j} \in V(M) \cap X$ which was used to enter X, and vertex $b_{X,j} \in V(M) \cap X$ which was used to leave X. We view $\mathcal{E}$ cyclically. In other words, for the starting bicontinent Y of the circuit $\mathcal{E}$ the vertex $b_{Y,1}$ is the vertex coming from the first matching edge along $\mathcal{E}$ while $a_{Y,1}$ coming from the very last step in $\mathcal{E}$.

Let C_X be the number of times bicontinent X was visited. We have $C_X < 2r$. Observe also that by (6.24) we have $C_{A_i} = C_{B_i}$ for each $i \in [r]$. Therefore, by $4r$-bipathitionability of $G[V_i]$ there exist for each $i \in [r]$ a system of $\mathcal{S}_i$ of $C_{A_i} + C_{B_i}$ paths decomposing V_i such that:

- The j-th path (for $j \in [C_{A_i}]$) starts in vertex $a_{A_i,j}$ and ends in $b_{A_i,j}$.
- The $(j + C_{A_i})$-th path (for $j \in [C_{B_i}]$) starts in vertex $a_{B_i,j}$ and ends in $b_{B_i,j}$.

It can be easily verified that the system $\{\mathcal{S}_i\}$ together with the matching M' forms a Hamilton cycle in G.

6.9 Algorithmic aspects

As said in the Introduction, the problem of deciding whether a graph is Hamiltonian is NP-hard. Even when the hamiltonicity of a graph G is guaranteed, finding a Hamilton cycle in G cannot be done in polynomial time unless P=NP. Yet in many situation there is an efficient algorithm for finding a Hamilton cycle in graphs satisfying certain conditions. See for example [20, 87, 24].

In this short section we note that the tools we use to prove Theorem 6.2 can be turned into an efficient algorithm for finding a Hamilton cycle in dense vertex-transitive graphs.

Theorem 6.22. *For every $\alpha > 0$ there is an n_0 such that every connected vertex-transitive graph on $n \geqslant n_0$ vertices and valency at least αn contains a Hamilton cycle. Moreover there is a polynomial time algorithm for finding a Hamilton cycle in such a graph.*

Recall the main steps of the proof of Theorem 6.2:

(A) By Theorem 6.9, the input graph G is partitioned into the continents $V_1, \ldots, V_r$.

(B) It is checked whether the graphs $G[V_i]$ is close to bipartiteness or not. In the first case, partitions satisfying the conclusion of Lemma 6.17 are found.

(C) For each $G[V_i]$, an exceptional set U_i is found so that the consequence of Theorem 6.20 or Theorem 6.21 is satisfied. (Depending on whether $G[V_i]$ is far from bipartite or not.)

(D) A way to connect certain systems of paths into one Hamilton cycle in G is devised. (In *Case I and Case II in the proof of Theorem 6.2 in the non-bipartite and the bipartite case, respectively.*)

(E) A system of paths (with prescribed end-vertices) is found in the graphs $G[V_i]$. (In *Theorem 6.20 and Theorem 6.21 in the non-bipartite and the bipartite case, respectively.*)

(F) A Hamilton cycle is found in G. (In *the final part of the proof of Theorem 6.2.*)

We now discuss the algorithmic versions of the steps above, thus providing a proof of Theorem 6.22.

For step (A) observe that in the proof of Theorem 6.9 it was crucial to be able to tell whether a graph is robustly connected. However, the obvious algorithm for testing robust connectivity requires exponentially many steps. We can overcome this obstacle with the help of codeg-graphs. We claim that there is a partition $V_1, \dots, V_r$ satisfying the conclusion of Thoerem 6.9 and moreover each V_i is a union of components of the $(19\alpha^2 n/20)$-codeg graph F of G. To see this consider the construction of the $V_1, \dots, V_r$ as given by Lemma 6.12. At step i, if G_i is not $(\alpha_i^4 n_i/40)$-robust, then we partition G_i into its $(\alpha_i^4 n_i/40)$-islands. By Lemma 6.10(b), every vertex has at most $r_i \alpha_i^4 n_i/40 \leqslant \alpha_i^2 n_i/20$ neighbours outside its island. In particular, every vertex will have at most

$$\sum_{i=0}^{\infty} \frac{\alpha_i^2 n_i}{20} = \frac{\alpha^2}{20} \sum_{i=0}^{\infty} \left(\frac{16}{9}\right)^i n_i \leqslant \frac{\alpha^2 n}{20} \sum_{i=0}^{\infty} \left(\frac{8}{9}\right)^i = \frac{9\alpha^2 n}{20}$$

neighbours outside its continent. In particular, any two vertices which are neighbours in the $(19\alpha^2 n/20)$-codeg graph F must belong to the same continent. There is an efficient way to construct F and moreover by Lemma 6.10(d) every component of F has minimum degree at least $\alpha^2 n/20$ and so F has at most $20/\alpha^2$ components. In particular, we can construct a bounded number of partitions (depending only on α and not on n) of the vertex set of G by grouping the components of F in all possible ways. At least one of these partitions satisfies the conclusion of Theorem 6.9. From now on the algorithm will work on all these possible partitions concurrently. We will show that for the partition that satisfies the conclusion of Theorem 6.9 it will only take polynomially many steps to construct a Hamilton cycle. Note that it might happen that some of the partitions do not satisfy the conclusion of Theorem 6.9; the algorithm is not required to produce a Hamilton cycle for these partitions as we only have to produce one Hamilton cycle.

For step (B), given a cn-iron vertex-transitive graph G we would like to decide in polynomial time whether it is c^4-close to bipartiteness or not and in the first case exhibit a partition satisfying the conclusion of Lemma 6.17. Unfortunately we cannot do this in polynomial time but not all is lost. Instead, we will show that there is a $0 < c' < c^4$ and a polynomial time algorithm that either proves that $G[V_i]$ is c'-far from bipartiteness or proves that $G[V_i]$ is c-close to bipartiteness and exhibits a partition which satisfies the conclusion of Lemma 6.17. If it so happens that G is both c'-far from and c^4-close to bipartiteness then there is no control as to which of the two possible outcomes will appear. To see how this can be done we apply the Regularity Lemma (Lemma 1.1) to $G[V_i]$ for some appropriate parameters. It is well known that the partition guaranteed by the Regularity Lemma can be found in polynomial time [7]. If the reduced graph is not bipartite (this can be checked in constant time) then the counting lemma shows that $G[V_i]$ is far from bipartite. If on the other hand the reduced graph is bipartite then it is immediate that $G[V_i]$ must be close to bipartite. It remains to show how to exhibit a bipartition satisfying the conclusions of Lemma 6.17. From the reduced graph we can exhibit a partition A', B' of $G[V_i]$ which satisfies (6.9). If every vertex has at least as many neighbours in the opposite part rather than its own part then by Remark 6.18 the partition has the required properties. If this was not the case then we move one such vertex to the opposite part and repeat the process. This process has to end (in polynomially many steps) as after each move the number of edges between the two parts strictly increases.

For step (C) we have already noted that there is an algorithmic version of the Regularity Lemma [7]. There are however two issues that need to be addressed. The first one is that for our proof of Theorem 6.21 it was important that the partition given by the Regularity Lemma was a refinement of the partition A, B of the vertex set. The statement of the algorithmic version of the Regularity Lemma in [7] does not deal with this issue. From the proof of the statement however it is immediate that we can start with any such prepartition. The second issue is that the algorithmic version of the Regularity Lemma in [7] is not stated in the degree form. The usual argument used to deduce the degree form from the standard form is algorithmic provided one knows which pairs are ε-regular. In principle, it is not easy to check algorithmically whether a pair is ε-regular or not and in fact the algorithmic proof of the Regularity Lemma does not say which pair are ε-regular and which are not. It does however produce a big enough (but possibly) incomplete list of ε-regular pairs and this is enough for our purpose of constructing a graph of regular pairs G'. The graphs R_1, R_2, R_1', R_2' in the proof of Theorem 6.20 can now be easily constructed algorithmically. It is also well-known that there is a polynomial-time algorithm for finding a maximum matching and so the matching M of R_1' can be constructed. The next step in our proof of Theorem 6.20 is an application of Lemma 1.6 in order to make the pairs corresponding to the matching M super-regular. We only stated Lemma 1.6 as an existence result but in the proof of the result one removes from each cluster the εm vertices which have the smallest degree inside its neighbouring cluster in M. Thus this can also be done algorithmically. Finally, we have already given an algorithmic proof of Lemma 1.7 and so the exceptional sets U_i can be constructed in polynomial time.

For step (D) we observe that the fat or bifat pairs can be easily recognized and so the auxiliary graph H can be constructed efficiently. The global connections in this step are based either on a spanning tree (in the non-bipartite case), or on an Eulerian circuit (in the bipartite case) in H. Since H is bounded these can be found in a bounded number of steps. The large matchings between the fat or bifat pairs can also be found in polynomial time and the matching M of Claim 6.2.3 (or the corresponding matching in the bipartite case) is constructed from these matchings greedily.

For step (E), the system of paths is constructed from the paths $P_1, \dots, P_\ell$ using the Blow-up Lemma. An algorithmic version of the Blow-up Lemma appears in [62]. For the construction of P_1 first note that the neighbours of the vertices of the exceptional set V_0' were selected greedily according to some restrictions. At each step it is easy to verify which vertices are not allowed to be chosen as neighbours. Similarly, the edges $u_{r+1}v_{r+1}, \dots, u_{r+k'/2}v_{r+k'/2}$ are also chosen greedily. To complete the construction of P_1 we need to construct some auxiliary paths Q_i. Each such path was arising from a path Q_i' which was the shortest path in a subdigraph of R^*. The digraph R^* and also the set of vertices of R^* which Q_i' is not allowed to pass can be constructed efficiently and hence so can the path Q_i'. It is now immediate how to construct the path Q_i'' in R. Finally, another greedy argument constructs the path Q_i from the path Q_i''. The other paths $P_2, \dots, P_\ell$ are constructed in a similar way.

Finally, step (F) is just putting steps (D) and (E) together.

6.10 Concluding remarks

Here we present three remarks on possible extensions of our proof of Theorem 6.2. To this end we summarize again the proof we presented.

(a) Firstly, we decompose the input graph G into pieces with linear (in their size) iron-connectivity – called islands. Recall that this step is based on a density pumping-up argument. In each stage we either do have the desired iron-connectivity of the working partition, or we can refine it making each its piece almost twice as dense.

(b) Secondly, using the Regularity Lemma and the Blow-up technique we find a system of paths in each of these islands.

(c) The last step is gluing these systems of paths together.

Our remarks concern the density requirement in Theorem 6.2, a question of finding Hamilton decompositions in vertex-transitive graphs, and a directed version of Lovász' conjecture.

- We wonder whether the approach we used to prove Theorem 6.2 cannot be pushed further to weaken the requirement $\deg(G) \geqslant \alpha n$ to $\deg(G) \geqslant f(n)$ for some sublinear function f. It particular, there seems to exist an analogue of step (a) whenever $\deg(G) \geqslant n^{2/3+\varepsilon}$.

  Step (b) is an obvious bottleneck for any success as the strength of Regularity method is very limited for sparse graphs. However there are alternative approaches which replace the Regularity Method; in particular, the paper [74] seems relevant.

- Besides the existential question of finding one Hamilton cycle, there is a substantial interest in *Hamilton decompositions* of graphs. A Hamilton decomposition is a collection of edge-disjoint Hamilton cycles which cover all the edges. In the context of this chapter, the following conjecture due to Alspach [9] is open: *Each connected Cayley graph over an abelian group of even valency has a Hamilton decomposition.* The requirement on even valency is obviously necessary. There are examples of Cayley graphs of even valency over non-abelian groups with no Hamilton decomposition. However it seems possible that graphs of even valency appearing in Theorem 6.2 posses a Hamilton decomposition.

- A directed version of Conjecture 6.1 is known not to be true: there are examples of connected vertex-transitive digraphs not containing a directed Hamilton path (that is, a path where all the edges are directed from its first end-vertex to its other end-vertex). However, we do not see a counterexample to the following.

  Conjecture 6.23. *For every $\alpha > 0$ there exists an n_0 such that every connected vertex-transitive digraph on $n \geqslant n_0$ vertices of valency at least αn contains every possible orientation of a Hamilton cycle.*

  This was suggested to us by Nešetřil.

  While some methods we used to prove Theorem 6.2 can probably be adapted to attack Conjecture 6.23 some important steps are definitely missing.

Index

Bibliography

[1] M. Aigner and G. M. Ziegler. *Proofs from The Book*. Springer-Verlag, Berlin, fourth edition, 2010.

[2] P. Allen. Minimum degree conditions for cycles. Preprint (2009).

[3] P. Allen, J. Böttcher, and J. Hladký. Filling the gap between Turán's theorem and Pósa's conjecture. Preprint (arXiv:0906.3299), to appear in *J. London Math. Soc.*

[4] P. Allen, J. Böttcher, J. Hladký, and D. Piguet. A density Corrádi-Hajnal theorem. Preprint.

[5] P. Allen, J. Böttcher, J. Hladký, and D. Piguet. An extension of Turán's Theorem, uniqueness, and stability. Preprint.

[6] P. Allen, J. Böttcher, J. Hladký, and D. Piguet. Turánnical hypergraphs. Preprint (arXiv:1011.1483), to appear in *Random Structures Algorithms*.

[7] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster. The algorithmic aspects of the regularity lemma. *J. Algorithms*, 16(1):80–109, 1994.

[8] N. Alon, P. Frankl, H. Huang, V. Rödl, A. Ruciński, and B. Sudakov. Large matchings in uniform hypergraphs and the conjectures of Erdős and Samuels. Preprint (arXiv:1107.1219).

[9] B. Alspach, G. Sárközy, and E. Szemerédi. Research problem 59. *Discrete Math.*, 50:115, 1984.

[10] B. Alspach and C. Q. Zhang. Hamilton cycles in cubic Cayley graphs on dihedral groups. *Ars Combin.*, 28:101–108, 1989.

[11] B. Andrásfai, P. Erdős, and V. T. Sós. On the connection between chromatic number, maximal clique and minimal degree of a graph. *Discrete Math.*, 8:205–218, 1974.

[12] L. Babai. Long cycles in vertex-transitive graphs. *J. Graph Theory*, 3(3):301–304, 1979.

[13] L. Babai. Automorphism groups, isomorphism, reconstruction. In *Handbook of combinatorics, Vol. 1, 2*, pages 1447–1540. Elsevier, Amsterdam, 1995.

[14] J. Balogh, T. Bohman, and D. Mubayi. Erdős-Ko-Rado in random hypergraphs. *Combin. Probab. Comput.*, 18(5):629–646, 2009.

[15] J. Bang-Jensen and G. Gutin. *Digraphs*. Springer Monographs in Mathematics. Springer-Verlag London Ltd., London, second edition, 2009. Theory, algorithms and applications.

[16] J.-C. Bermond. *Hamiltonian Graphs*, chapter 6, pages 127–167. Selected Topics in Graph Theory. Academic Press, 1979.

[17] T. Bohman and P. Keevash. The early evolution of the H-free process. *Invent. Math.*, 181(2):291–336, 2010.

[18] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[19] B. Bollobás and A. Thomason. Dense neighbourhoods and Turán's theorem. *J. Combin. Theory Ser. B*, 31(1):111–114, 1981.

[20] J. A. Bondy and V. Chvátal. A method in graph theory. *Discrete Math.*, 15(2):111–135, 1976.

[21] J. Böttcher. *Embedding large graphs - The Bollobás-Komlós conjecture and beyond*. PhD thesis, Technical University Munich, 2009.

[22] G. Brightwell, K. Panagiotou, and A. Steger. On extremal subgraphs of random graphs. In *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 477–485, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.

[23] D. Christofides, J. Hladký, and A. Máthé. Hamilton cycles in dense vertex-transitive graphs. Preprint (arXiv:1008.2193).

[24] D. Christofides, P. Keevash, D. Kühn, and D. Osthus. Finding Hamilton cycles in robustly expanding digraphs. submitted.

[25] D. Christofides, P. Keevash, D. Kühn, and D. Osthus. A semiexact degree condition for Hamilton cycles in digraphs. *SIAM J. Discrete Math.*, 24:709–756, 2010.

[26] F. R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1997.

[27] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[28] D. Conlon and T. Gowers. Combinatorial theorems in sparse random sets. Preprint (arXiv:1011.4310).

[29] K. Corrádi and A. Hajnal. On the maximal number of independent circuits in a graph. *Acta Math. Acad. Sci. Hungar.*, 14:423–439, 1963.

[30] S. J. Curran and J. A. Gallian. Hamiltonian cycles and paths in Cayley graphs and digraphs—a survey. *Discrete Math.*, 156(1-3):1–18, 1996.

[31] R. Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, third edition, 2005.

[32] G. A. Dirac. Some theorems on abstract graphs. *Proc. London Math. Soc. (3)*, 2:69–81, 1952.

[33] P. Erdős. Über ein Extremalproblem in der Graphentheorie. *Arch. Math.*, 13:222–227, 1962.

[34] P. Erdős and A. H. Stone. On the structure of linear graphs. *Bulletin of the American Mathematical Society*, 52:1087–1091, 1946.

[35] P. Erdős. Problem 9. In M. Fiedler, editor, *Theory of Graphs and its Applications (Proc. Sympos. Smolenice, 1963)*, page 159. Publ. House Czechoslovak Acad. Sci., Prague, 1964.

[36] P. Erdős and T. Gallai. On maximal paths and circuits of graphs. *Acta Math. Acad. Sci. Hungar*, 10:337–356 (unbound insert), 1959.

[37] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[38] P. Erdős and M. Simonovits. Supersaturated graphs and hypergraphs. *Combinatorica*, 3(2):181–192, 1983.

[39] G. Fan and H. A. Kierstead. The square of paths and cycles. *J. Combin. Theory Ser. B*, 63(1):55–64, 1995.

[40] G. Fan and H. A. Kierstead. Hamiltonian square-paths. *J. Combin. Theory Ser. B*, 67(2):167–182, 1996.

[41] H. Fleischner. The square of every two-connected graph is hamiltonian. *J. Combin. Theory Ser. B*, 16(1):29–34, 1974.

[42] P. Frankl and V. Rödl. Large triangle-free subgraphs in graphs without K_4. *Graphs Combin.*, 2(2):135–144, 1986.

[43] E. Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999. With an appendix by Jean Bourgain.

[44] E. Friedgut. Hunting for sharp thresholds. *Random Structures Algorithms*, 26(1-2):37–51, 2005.

[45] E. Friedgut, V. Rödl, and M. Schacht. Ramsey properties of random discrete structures. *Random Structures Algorithms*, 37(4):407–436, 2010.

[46] Z. Füredi. Turán type problems. In *Surveys in combinatorics, 1991 (Guildford, 1991)*, volume 166 of *London Math. Soc. Lecture Note Ser.*, pages 253–300. Cambridge Univ. Press, Cambridge, 1991.

[47] S. Gerke, T. Schickinger, and A. Steger. K_5-free subgraphs of random graphs. *Random Structures Algorithms*, 24(2):194–232, 2004.

[48] C. Godsil and G. Royle. *Algebraic graph theory*, volume 207 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2001.

[49] W. T. Gowers. Quasirandomness, counting and regularity for 3-uniform hypergraphs. *Combin. Probab. Comput.*, 15(1-2):143–184, 2006.

[50] W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Ann. of Math. (2)*, 166(3):897–946, 2007.

[51] W. T. Gowers. Quasirandom groups. *Combin. Probab. Comput.*, 17(3):363–387, 2008.

[52] C. Grosu and J. Hladký. The extremal function for partial bipartite tilings. Preprint (arXiv:0910.1064), to appear in *European J. Combin.*

[53] L. Gugelmann, Y. Person, A. Steger, and H. Thomas. A randomized version of Ramsey's Theorem. to appear in proceedings of EuroComb 2011.

[54] A. Hajnal and E. Szemerédi. Proof of a conjecture of P. Erdős. In *Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969)*, pages 601–623. North-Holland, Amsterdam, 1970.

[55] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs.* Wiley-Interscience, New York, 2000.

[56] P. Keevash. Hypergraph Turán problems. to appear in Surveys in Combinatorics 2011.

[57] Y. Kohayakawa, T. Łuczak, and V. Rödl. On K^4-free subgraphs of random graphs. *Combinatorica*, 17(2):173–213, 1997.

[58] Y. Kohayakawa, V. Rödl, and M. Schacht. The Turán theorem for random graphs. *Combin. Probab. Comput.*, 13(1):61–91, 2004.

[59] J. Komlós. Tiling Turán theorems. *Combinatorica*, 20(2):203–218, 2000.

[60] J. Komlós, G. N. Sárközy, and E. Szemerédi. On the square of a Hamiltonian cycle in dense graphs. *Random Structures Algorithms*, 9(1-2):193–211, 1996.

[61] J. Komlós, G. N. Sárközy, and E. Szemerédi. Blow-up lemma. *Combinatorica*, 17(1):109–123, 1997.

[62] J. Komlós, G. N. Sárközy, and E. Szemerédi. An algorithmic version of the blow-up lemma. *Random Structures Algorithms*, 12(3):297–312, 1998.

[63] J. Komlós, G. N. Sárközy, and E. Szemerédi. Proof of the Seymour conjecture for large graphs. *Annals of Combinatorics*, 2:43–60, 1998.

[64] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi. The regularity lemma and its applications in graph theory. In *Theoretical aspects of computer science (Tehran, 2000)*, volume 2292 of *Lecture Notes in Comput. Sci.*, pages 84–112. Springer, Berlin, 2002.

[65] J. Komlós and M. Simonovits. Szemerédi's regularity lemma and its applications in graph theory. In *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, volume 2 of *Bolyai Soc. Math. Stud.*, pages 295–352. János Bolyai Math. Soc., Budapest, 1996.

[66] T. Kövári, V. Sós, and P. Turán. On a problem of K. Zarankiewicz. *Colloq. Math.*, 3:50–57, 1954.

[67] M. Krivelevich and B. Sudakov. Pseudo-random graphs. In *More sets, graphs and numbers*, volume 15 of *Bolyai Soc. Math. Stud.*, pages 199–262.

[68] M. Krivelevich and B. Sudakov. Sparse pseudo-random graphs are Hamiltonian. *J. Graph Theory*, 42(1):17–33, 2003.

[69] D. Kühn and D. Osthus. Packings in dense regular graphs. *Combin. Probab. Comput.*, 14(3):325–337, 2005.

[70] D. Kühn and D. Osthus. Embedding large subgraphs into dense graphs. In *Surveys in combinatorics 2009*, volume 365 of *London Math. Soc. Lecture Note Ser.*, pages 137–167. Cambridge Univ. Press, Cambridge, 2009.

[71] D. Kühn and D. Osthus. The minimum degree threshold for perfect graph packings. *Combinatorica*, 29(1):65–107, 2009.

[72] D. Kühn, D. Osthus, and A. Taraz. Large planar subgraphs in dense graphs. *J. Comb. Theory, Ser. B*, 95(2):263–282, 2005.

[73] K. Kutnar and D. Marušič. Hamilton cycles and paths in vertex-transitive graphs—current directions. *Discrete Math.*, 309(17):5491–5500, 2009.

[74] I. Levitt, G. Sárközy, and E. Szemerédi. How to avoid using the Regularity Lemma; Pósa's conjecture revisited. *Discrete Math.*, 310:630–641, 2010.

[75] L. Lovász. Problem 11. In R. Guy, H. Hanani, N. Sauer, and J. Schönheim, editors, *Combinatorial Structures and their Applications*, volume 1969 of *Proceedings of the Calgary International Conference on Combinatorial Structures and their Applications held at the University of Calgary, Calgary, Alberta, Canada, June*, pages xvi+508. Gordon and Breach Science Publishers, New York, 1970.

[76] L. Lovász. *Combinatorial problems and exercises.* AMS Chelsea Publishing, Providence, RI, second edition, 2007.

[77] T. Łuczak. $R(C_n, C_n, C_n) \leqslant (4+o(1))n$. *J. Combin. Theory Ser. B*, 75(2):174–187, 1999.

[78] W. Mantel. Problem 28, solution by H. Gouwentak, W. Mantel, J. Teixeira de Mattes, F. Schuh and W.A. Wythoff. *Wiskundige Opgaven*, 10:60–61, 1907.

[79] J. W. Moon. On independent complete subgraphs in a graph. *Canad. J. Math.*, 20:95–102, 1968.

[80] D. Mubayi. Books versus triangles. Preprint (arXiv:1002.1492).

[81] V. Nikiforov and R. H. Schelp. Cycle lengths in graphs with large minimum degree. *Journal of Graph Theory*, 52(2):157–170, 2006.

[82] I. Pak and R. Radoičić. Hamiltonian paths in Cayley graphs. *Discrete Math.*, 309(17):5501–5508, 2009.

[83] A. A. Razborov. On the minimal density of triangles in graphs. *Combin. Probab. Comput.*, 17(4):603–618, 2008.

[84] I. Reiman. Über ein Problem von K. Zarankiewicz. *Acta Math. Acad. Sci. Hungar.*, 9:269–273, 1958.

[85] V. Rödl, B. Nagle, J. Skokan, M. Schacht, and Y. Kohayakawa. The hypergraph regularity method and its applications. *Proc. Natl. Acad. Sci. USA*, 102(23):8109–8113 (electronic), 2005.

[86] V. Rödl and A. Ruciński. Threshold functions for Ramsey properties. *J. Amer. Math. Soc.*, 8(4):917–942, 1995.

[87] G. N. Sárközy. A fast parallel algorithm for finding Hamiltonian cycles in dense graphs. *Discrete Math.*, 309(6):1611–1622, 2009.

[88] N. Sauer. On the density of families of sets. *J. Combinatorial Theory Ser. A*, 13:145–147, 1972.

[89] M. Schacht. Extremal results for random discrete structures. Submitted.

[90] A. Schrijver. *Combinatorial optimization. Polyhedra and efficiency. Vol. A*, volume 24 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2003. Paths, flows, matchings, Chapters 1–38.

[91] S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41:247–261, 1972.

[92] M. Simonovits. A method for solving extremal problems in graph theory, stability problems. In *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pages 279–319. Academic Press, New York, 1968.

[93] M. Simonovits. How to solve a Turán type extremal graph problem? (linear decomposition). In *Contemporary trends in discrete mathematics (Štiřín Castle, 1997)*, volume 49 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 283–305. Amer. Math. Soc., Providence, RI, 1999.

[94] E. Sperner. Ein Satz über Untermengen einer endlichen Menge. *Math. Z.*, 27:544–548, 1928.

[95] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.

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

[97] P. Turán. Eine Extremalaufgabe aus der Graphentheorie. *Mat. Fiz. Lapok*, 48:436–452, 1941.

[98] D. Witte. Cayley digraphs of prime-power order are Hamiltonian. *J. Combin. Theory Ser. B*, 40(1):107–112, 1986.

[99] D. Witte and J. A. Gallian. A survey: Hamiltonian cycles in Cayley graphs. *Discrete Math.*, 51(3):293–304, 1984.

[100] D. S. Witte. On Hamiltonian circuits in Cayley diagrams. *Discrete Math.*, 38(1):99–108, 1982.