Results in Extremal Graph Theory, Ramsey Theory and Additive Combinatorics

Olivér Noel Janzer

Trinity College
and
Department of Pure Mathematics and Mathematical Statistics
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

June 2020

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared here and specified in the text. It is not substantially the same as any work that has already been submitted before for any degree or other qualification.

Sections 2.4 and 2.5 are based on joint work with David Conlon and Joonkyung Lee.

Section 3.2 is based on joint work with Andrzej Grzesik and Zoltán Lóránt Nagy.

Section 3.3 is based on joint work with Abhishek Methuku and Zoltán Lóránt Nagy.

Chapters 5, 7 and 8 are based on joint work with Timothy Gowers.

Chapter 9 is based on joint work with Debarun Ghosh, Ervin Győri, Addisu Paulos, Nika Salia and Oscar Zamora.

Olivér Noel Janzer

Results in Extremal Graph Theory, Ramsey Theory and Additive Combinatorics

Olivér Noel Janzer

Abstract

This dissertation contains results from various areas of Combinatorics.

In Chapter 2, we consider a central problem in Extremal Graph Theory. The extremal number (or Turán number) $\mathrm{ex}(n, H)$ of a graph H is the maximum number of edges in an H-free graph on n vertices. It is a major area of research to better understand the extremal number of bipartite graphs. In this chapter we develop a new method which allows us to obtain strong (and often best possible) upper bounds in a wide range of cases. Our results answer several conjectures of Conlon and Lee, and Kang, Kim and Liu. Furthermore, they relate to and improve work of (among others) Füredi, Alon, Krivelevich and Sudakov, Kostochka and Pyber, and Jiang and Seiver.

While in Chapter 2 the focus is on subdivided graphs, in Chapter 3 we study the extremal number of blow-ups. In particular, we obtain tight upper bounds for the extremal number of blow-ups of trees. As an extension of this, we pose a general conjecture relating the extremal number of F and that of its blow-up. We prove the conjecture for the 2-blowup of C_6.

In Chapter 4 we study a coloured variant of the Turán problem. The rainbow Turán number of H, denoted by $\mathrm{ex}^*(n, H)$, is the maximum possible number of edges in an n-vertex properly edge-coloured graph without a rainbow subgraph isomorphic to H. We prove that $\mathrm{ex}^*(n, C_{2k}) = O(n^{1+1/k})$, which is tight and establishes a conjecture of Keevash, Mubayi, Sudakov and Verstraëte. We use the same method to answer several further questions in various topics: among others, a question of Conlon and Tyomkyn on colour-isomorphic cycles and a conjecture of Jiang and Newman of blow-ups of cycles. We also disprove an old conjecture of Erdős and Simonovits on (ordinary) extremal numbers.

In Chapter 5, we consider the following problem. Let $2 \le s < t$ be fixed integers. If G is an arbitrary K_t-free graph on n vertices, how large a K_s-free induced subgraph must there exist in G? This number, which is a generalisation of the usual off-diagonal Ramsey numbers, is viewed as a function in n, and is called the Erdős-Rogers function. We obtain new upper bounds in the case $s + 2 \le t \le 2s - 1$, improving results of (among others) Bollobás, Erdős and Krivelevich, and answering a question of Dudek, Retter and Rödl.

In Chapter 6, we investigate the relationship between two well-studied notions of tensor rank. We show that the partition rank of a tensor is bounded above by a polynomial in the analytic rank of the same tensor. This improves Ackermann-type bounds obtained by various authors including Green and Tao, and Bhowmick and Lovett.

In Chapter 7, we use the main technical lemma of Chapter 6 to prove a result about the expansion of subsets of the Cayley graph on the tensor product $\mathbb{F}_2^{n_1} \otimes \cdots \otimes \mathbb{F}_2^{n_d}$ where the generators are the rank 1 tensors. This is motivated by the famous Unique Games Conjecture from Theoretical Computer Science, and is a partial generalisation of a recent breakthrough result of Khot, Minzer and Safra.

In Chapter 8, we ask the following question. Given constants α, β, γ, what is the minimal possible edge density of a graph G on n vertices with the property that every subset $A \subset V(G)$ with $|A| \ge \alpha n$ contains a subset $B \subset A$ with $|B| \ge \beta n$ such that $G[B]$

has edge density at least γ? We also study a bipartite version of this question, obtaining sharp results in both cases.

In Chapter 9, we determine asymptotically the maximum possible number of induced C_5's in a planar graph on n vertices.

Acknowledgements

I am very grateful to my supervisor, Timothy Gowers, for our fruitful collaborations. I have benefitted greatly from our joint projects – working with him allowed me to become familiar with important techniques very early on in my PhD. I am also grateful to him for his helpful advice on various matters that arose over the last three years.

I would also like to thank the others who I have had the pleasure working with: David Conlon, Debarun Ghosh, Andrzej Grzesik, Ervin Győri, Joonkyung Lee, Abhishek Methuku, Zoltán Lóránt Nagy, Addisu Paulos, Nika Salia and Oscar Zamora.

I thank Imre Leader, Žarko Ranđelović and Eero Räty for allowing me to join their mini-seminars which I greatly enjoyed.

Finally, I am immensely grateful to my family for their support. Without their help and encouragement, I would never have been able to study at Cambridge. Special thanks goes to my brother, Barnabás, for many stimulating mathematical discussions.

Contents

Chapter 1

Introduction

Apart from this introduction, the dissertation is organized into eight main chapters.

Chapter 2 is based on four of my papers [23, 65, 66, 69], and contains several results about the extremal number of bipartite graphs, with an emphasis on subdivided graphs. Given a graph H, the extremal number $\mathrm{ex}(n, H)$ denotes the maximum number of edges in a graph on n vertices which does not contain H as a subgraph. The k-subdivision of a multigraph F is obtained by replacing each edge of F with a path of length $k + 1$, and is denoted by F^k.

Section 2.3 is based on [65]. In this section we prove that the 1-subdivision of K_t has extremal number $O(n^{\frac{3}{2}-\frac{1}{4t-6}})$. This proves in a strong form a conjecture proposed by Conlon and Lee [24], and improves the bound $O(n^{\frac{3}{2}-\frac{1}{6t}})$ obtained by them. We shall also briefly discuss the connection of this result with off-diagonal Ramsey numbers.

Sections 2.4 and 2.5 are based on joint work with Conlon and Lee [23]. In Section 2.4 we prove that a $K_{2,2}$-free bipartite graph with maximum degree r on one side has extremal number $o(n^{2-1/r})$. This improves the celebrated result of Füredi [46], which states that any bipartite graph with maximum degree r on one side has extremal number $O(n^{2-1/r})$. In Section 2.5 we give a very short proof of a recent difficult result of Kang, Kim and Liu [79].

Section 2.6 is based on [69]. The main result in this section is an upper bound for the extremal number of the $(k-1)$-subdivision of an arbitrary multigraph. More precisely, we show that if k is even and F is a multigraph, then $\mathrm{ex}(n, F^{k-1}) = O(n^{1+1/k})$, and when F is a simple graph, then $\mathrm{ex}(n, F^{k-1}) = O(n^{1+1/k-c(F,k)})$ for some $c(F, k) > 0$. The former bound is sharp, while the latter is sharp up to the value of $c(F, k)$. These results answer two conjectures of Conlon and Lee [24], and improve results of Kostochka and Pyber [88], Jiang [71] and Jiang and Seiver [76], the most recent of which was $\mathrm{ex}(n, K_t^{k-1}) = O(n^{1+16/k})$.

Section 2.7 is based on [66]. We prove that $\mathrm{ex}(n, K_{s,t}^{k-1}) = O(n^{1+\frac{s-1}{sk}})$, which is tight for t sufficiently large. This result settles a conjecture of Conlon, Janzer and Lee [23], and improves on a substantial body of work by Conlon and Lee [24], Kang, Kim and Liu [79], Jiang and Qiu [75] and the author [65].

In Chapter 3 we continue the study of bipartite extremal numbers. However, the focus in this chapter is on the extremal number of blow-up-like graphs. The r-blowup of a graph F is obtained by replacing the vertices and edges of F with independent sets of size r and copies of $K_{r,r}$, respectively. We denote this graph by $F[r]$. We make the conjecture that if $\mathrm{ex}(n, F) = O(n^{2-\alpha})$, then $\mathrm{ex}(n, F[r]) = O(n^{2-\frac{\alpha}{r}})$.

Section 3.2 is based on joint work with Grzesik and Nagy [57]. In this section we prove that if H is the r-blow-up of a tree, then $\mathrm{ex}(n, H) = O(n^{2-1/r})$, which is tight and confirms the above conjecture when F is a tree. We also establish some generalisations of this result, which extend the theorem of Füredi about the extremal number of bipartite graphs with maximum degree r on one side.

Section 3.3 is based on joint work with Methuku and Nagy [70]. In this section we prove that $\mathrm{ex}(n, C_6[2]) = O(n^{5/3})$, and more generally that for any t, $\mathrm{ex}(n, \theta_{3,t}[2]) = O(n^{5/3})$. This is tight when t is sufficiently large, and proves the above conjecture for $F = \theta_{3,t}$ and $r = 2$.

Chapter 4 is based on [67]. The rainbow Turán number $\mathrm{ex}^*(n, H)$ of a graph H is the maximum possible number of edges in a properly edge-coloured n-vertex graph with no rainbow subgraph isomorphic to H. We settle a conjecture of Keevash, Mubayi, Sudakov and Verstraëte [82] by proving that for any integer $k \geq 2$, $\mathrm{ex}^*(n, C_{2k}) = O(n^{1+1/k})$. This is tight and improves the bound of Das, Lee and Sudakov [26] stating that $\mathrm{ex}^*(n, C_{2k}) = O\left(n^{1+\frac{(1+\varepsilon_k)\log k}{k}}\right)$ where $\varepsilon_k \to 0$ as $k \to \infty$.

We use the same method to prove several other conjectures in various topics. First, we prove that there exists a constant c such that any properly edge-coloured n-vertex graph with more than $cn(\log n)^4$ edges contains a rainbow cycle. It is known that there exist properly edge-coloured n-vertex graphs with $\Omega(n \log n)$ edges which do not contain any rainbow cycle.

Secondly, we prove that in any proper edge-colouring of K_n with $o(n^{\frac{r}{r-1}\cdot\frac{k-1}{k}})$ colours, there exist r colour-isomorphic, pairwise vertex-disjoint copies of C_{2k}. This proves in a strong form a conjecture of Conlon and Tyomkyn [25], and a strenghtened version proposed by Xu, Zhang, Jing and Ge [110]. As a corollary, our theorem generalises a recent result of Fish, Pohoata and Sheffer [45] on the Erdős–Gyárfás function.

Moreover, we answer a question of Jiang and Newman [73] by showing that there exists a constant $c = c(r)$ such that any n-vertex graph with more than $cn^{2-1/r}(\log n)^{7/r}$ edges contains the r-blowup of an even cycle. Finally, by showing that $\mathrm{ex}(n, C_{2k}[r]) = O(n^{2-\frac{1}{r}+\frac{1}{k+r-1}+o(1)})$, we disprove an old conjecture of Erdős and Simonovits [31] which proposed a lower bound for the extremal number of bipartite graphs with given minimum degree.

Chapter 5 is based on joint work with Gowers [52]. Our results concern the following problem. Let $2 \leq s < t$ be integers. The Erdős-Rogers function $f_{s,t}(n)$ measures how large a K_s-free induced subgraph there must be in a K_t-free graph on n vertices. This function has been studied by several authors including Bollobás, Erdős, Krivelevich, Rödl and

Sudakov [13,27–29,37,90,91,105,106,109]. After a sequence of earlier papers, it was proved that by Dudek, Retter and Rödl that for every $s \geq 3$, $f_{s,s+1}(n) = n^{1/2+o(1)}$ [27]. They asked whether it is true that for every $s \geq 3$, $f_{s,s+2}(n) = o(n^{1/2})$. Via a novel probabilistic construction, we improve the upper bound for $f_{s,t}(n)$ in the range $s+2 \leq t \leq 2s-1$. In particular, we show that $f_{s,s+2}(n) = O(n^{\alpha_s})$ for some $\alpha_s < 1/2$, answering the question of Dudek, Retter and Rödl affirmatively. Our bound is close to the best known lower bound, due to Sudakov [105].

Chapter 6 has the same content as [68], which is an improved version of my earlier manuscript [64]. Tensors are generalisations of matrices to higher dimension. Unlike in the case of matrices, there does not exist a unique definition for the rank of a tensor as the different equivalent characterisations of the usual matrix rank lead to different notions for tensors. The main focus of this paper is the relationship between two well-studied notions of rank. More precisely, we show that the partition rank of a tensor T is bounded above by a polynomial in the analytic rank of T. Before our work, the best known bound was an Ackermann-type function. Our result has an essentially equivalent formulation in terms of polynomials over finite fields. In that language, it roughly states that if the distribution of the values of a degree d polynomial (in n variables) over $\mathbb{F}_q$ is far from uniform, then the polynomial can be written as a function of not too many (the number does not depend on n) polynomials of degree less than d. This improves on results of various authors including Green and Tao [56] and Bhowmick and Lovett [9].

Chapter 7 is based on joint work with Gowers [51]. In this paper we were aiming to generalise a recent breakthrough result of Khot, Minzer and Safra [84] which completed the proof of the so-called 2-to-2 Games Conjecture. The Unique Games Conjecture is a central conjecture in Theoretical Computer Science which, if true, implies that for a certain set of constraints it is NP-hard to distinguish between situations where (say) 1% and where 99% of the constraints can be satisfied. The 2-to-2 Games Conjecture (now proven) asserts that it is NP-hard to distinguish between situations where 50% and where 99% of the constraints can be satisfied. This weakening was reduced to the problem of finding a qualitative description of the so called *closed sets* of the group $\mathrm{M}_{m,n}(\mathbb{F}_2)$ of $m \times n$ matrices over $\mathbb{F}_2$. A set $\mathcal{A} \subset \mathrm{M}_{m,n}(\mathbb{F}_2)$ is η-closed if the probability that $A+B \in \mathcal{A}$ is at least η when $A \in \mathcal{A}$ is uniformly random and B is a uniformly random rank 1 matrix. In this paper we consider the same problem in higher dimensions. We say that $\mathcal{A} \subset \mathbb{F}_2^{n_1} \otimes \cdots \otimes \mathbb{F}_2^{n_d}$ is η-closed if $\mathbb{P}(A + u_1 \otimes \cdots \otimes u_d \in \mathcal{A}) \geq \eta$ where $A \in \mathcal{A}$ and $u_i \in \mathbb{F}_2^{n_i}$ are uniformly randomly chosen. We make a conjecture that would describe the closed sets qualitatively, and prove the conjecture in an important special case. In particular, we show that our conjecture holds whenever $A \subset \mathbb{F}_2^{n_1} \otimes \cdots \otimes \mathbb{F}_2^{n_d}$ is a vector space.

Chapter 8 is based on joint work with Gowers [53]. In this paper we consider the following questions. Suppose that a graph G on n vertices has the property that for any $A \subset V(G)$ of size at least αn there is some $B \subset A$ of size at least βn such that $G[B]$ has edge density at least γ. (α, β and γ may depend on n, we only assume that they are not very small.) What is the minimal density of G? We may ask an analogous question about bipartite graphs as well. Suppose that G is a bipartite graph on $n+n$ vertices with parts X and Y such that for any $A \subset X$, $B \subset Y$ with $|A|, |B| \geq \alpha n$, there exist $C \subset A$, $D \subset B$ with $|C|, |D| \geq \beta n$ such that $G[C, D]$ has density at least γ. What is the minimal density of G? In the graph case we give a lower bound $\frac{\beta\gamma}{\alpha}(1 - o(1))$ which is tight when α/β is an integer. In the bipartite case, we show that the answer is between $c\frac{\beta\gamma}{\alpha}\log(1/\alpha)$ and $C\frac{\beta\gamma}{\alpha}\log(1/\alpha)$ for some absolute constants $0 < c < C$. We also prove some structural results about graphs with the above property.

Chapter 9 is based on joint work with Ghosh, Győri, Paulos, Salia and Zamora [49]. In this chapter we prove that the maximum possible number of induced 5-cycles in a planar graph on n vertices is $\frac{n^2}{3} + O(n)$.

Chapter 2

The extremal number of subdivisions

2.1 Introduction

For a family $\mathcal{H}$ of graphs, the extremal number (or Turán number) $\mathrm{ex}(n, \mathcal{H})$ is defined to be the maximal number of edges in a graph on n vertices that does not contain any $H \in \mathcal{H}$ as a subgraph. When $\mathcal{H} = \{H\}$, we write $\mathrm{ex}(n, H)$ for the same number. The Erdős–Stone–Simonovits theorem [32, 33] states that

$$\mathrm{ex}(n, H) = \left(1 - \frac{1}{\chi(H) - 1}\right)\binom{n}{2} + o(n^2),$$

which determines the asymptotics of $\mathrm{ex}(n, H)$ when $\chi(H) > 2$. However, for bipartite graphs H, this theorem only gives $\mathrm{ex}(n, H) = o(n^2)$, and determining the order of magnitude of $\mathrm{ex}(n, H)$ is notoriously difficult. Even for simple graphs such as even cycles and complete bipartite graphs, the problem is not settled. An old result of Bondy and Simonovits [15] states that $\mathrm{ex}(n, C_{2k}) = O(n^{1+1/k})$, but matching lower bounds are only known for $k \in \{2, 3, 5\}$ [8,103]. Also, we have an upper bound $\mathrm{ex}(n, K_{s,t}) = O(n^{2-1/s})$ [89], but this is only known to be tight when $t > (s - 1)!$ [5, 86]. For a survey on the classical results in the area, see [47].

The following conjecture has been made about the order of magnitude of the extremal function.

Conjecture 2.1.1 (Erdős–Simonovits [35]). *For every graph H, there exists a rational number $r \in \{0\} \cup [1, 2]$ such that $\mathrm{ex}(n, H) = \Theta(n^r)$.*

The converse of this statement is one of the most central conjectures in Extremal Graph Theory.

Conjecture 2.1.2 (Rational Exponents Conjecture; Erdős–Simonovits [35]). *For every rational number $r \in (1, 2)$, there exists a graph H with $\mathrm{ex}(n, H) = \Theta(n^r)$.*

We say that $r \in (1, 2)$ is *realisable* (by H) if there exists a graph H such that $\mathrm{ex}(n, H) = \Theta(n^r)$. With this terminology, the Rational Exponent Conjecture states that every rational number between 1 and 2 is realisable. In a recent breakthrough, Bukh and Conlon [18] have proved that for any rational number $r \in (1,2)$ there exists a finite family $\mathcal{H}$ of graphs such that $\mathrm{ex}(n,\mathcal{H}) = \Theta(n^r)$. However, Conjecture 2.1.2 remains wide open. In fact, until recently only very few realisable numbers were known, namely $2 - 1/s$ for $s \in \mathbb{N}$, which are realised by $K_{s,t}$ for t sufficiently large. A few years ago, the family $1 + 1/s$ was also shown to be realisable, by theta graphs. The theta graph $\theta_{s,t}$ is the union of t paths of length s which have the same endpoints but are pairwise internally vertex-disjoint. Note that $\theta_{s,2} = C_{2s}$. A classical result of Faudree and Simonovits [44] states that $\mathrm{ex}(n,\theta_{s,t}) = O(n^{1+1/s})$, and it was proved recently by Conlon [21] that this is tight for sufficiently large t.

Two years ago, Jiang, Ma and Yepremyan [72] enlarged the class of realisable exponents by proving that $7/5$ and $2 - \frac{2}{2s-1}$ for $s \geq 2, s \in \mathbb{N}$ are also realisable. Subsequently, Kang, Kim and Liu [79] proved that for each $a, b \in \mathbb{N}$ with $a < b$ and $b \equiv \pm 1 \pmod{a}$, the number $2 - \frac{a}{b}$ is realisable.

In this chapter we present a method that allows us to obtain further large families of realisable exponents and, perhaps more importantly, can be used to prove strong upper bounds for the extremal number of subdivided graphs.

For a multigraph F, a *subdivision* of F is a graph obtained by replacing the edges of F with pairwise internally vertex-disjoint paths of arbitrary lengths. The k-subdivision of F is the graph obtained by replacing the edges of F with pairwise internally vertex-disjoint paths of length $k+1$, and is denoted by F^k. The 1-subdivision of F is also denoted by F'.

The study of the extremal number of K_t' has been initiated by Conlon and Lee [24] in an attempt to generalise the following celebrated result of Füredi [46] and Alon, Krivelevich and Sudakov [4]. In this theorem and everywhere else in this chapter (unless stated otherwise) the asymptotic notation means that $n \to \infty$ and all other parameters are kept constant.

Theorem 2.1.3 (Füredi, Alon–Krivelevich–Sudakov)**.** *Let H be a bipartite graph such that in one of the parts all the degrees are at most r. Then*

$$\mathrm{ex}(n,H) = O(n^{2-1/r}).$$

This result is tight, since as we have mentioned before, for s sufficiently large in terms of r, $\mathrm{ex}(n,K_{r,s}) = \Omega(n^{2-1/r})$. Moreover, it is conjectured [89] that this should already hold when $s = r$. On the other hand, a recent conjecture of Conlon and Lee [24] says that containing $K_{r,r}$ as a subgraph should be the only reason why Theorem 2.1.3 is tight.

Conjecture 2.1.4 (Conlon–Lee [24])**.** *Let H be a bipartite graph such that in one of the parts all the degrees are at most r and H does not contain $K_{r,r}$ as a subgraph. Then there exists some $\delta > 0$ such that* $\mathrm{ex}(n,H) = O(n^{2-1/r-\delta})$.

It is easy to see that any $K_{2,2}$-free bipartite graph in which every vertex in one part has degree at most two is a subgraph of K_t' for some positive integer t. Conlon and Lee have verified their conjecture in the $r = 2$ case by proving the following result.

Theorem 2.1.5 (Conlon–Lee [24])**.** *For any integer $t \geq 3$,*

$$\mathrm{ex}(n, K_t') = O(n^{3/2-1/6^t}).$$

They have asked for an upper bound of the form $\mathrm{ex}(n, K_t') = O(n^{3/2-\delta_t})$, where $1/\delta_t$ is bounded by a polynomial in t. We prove such a bound even for a linear $1/\delta_t$.

Theorem 2.1.6 (Janzer [65])**.** *For any integer $t \geq 3$,*

$$\mathrm{ex}(n, K_t') = O(n^{3/2-\frac{1}{4t-6}}).$$

Note that this bound is tight for $t = 3$ as $\mathrm{ex}(n, C_6) = \Theta(n^{4/3})$. If Theorem 2.1.6 is tight for every t, that may have very important consequences in Ramsey theory. The Ramsey number $R(t, m)$ is the smallest number N such that any graph on N vertices contains a clique of size t or an independent set of size m. When t is fixed and $m \to \infty$, the best known bounds are of the form $m^{\frac{t+1}{2}-o(1)} \leq R(t, m) \leq m^{t-1-o(1)}$. The exponent in the lower bound was first proved by Spencer [104], while the upper bound m^{t-1} is a classical result of Erdős and Szekeres [41]. The current best bounds (which improve the earlier results by polylogarithmic factors) are due to Bohman and Keevash [11], and Ajtai, Komlós and Szemerédi [1].

Recently, Mubayi and Verstraëte [98] showed that the existence of certain pseudorandom graphs would imply that $R(t, m) = m^{t-1-o(1)}$. An (n, d, λ) graph is a d-regular graph on n vertices whose eigenvalues, apart from the largest one, have absolute value at most λ. It is known that if d/n is bounded away from 1, then we must have $\lambda = \Omega(\sqrt{d})$. When $\lambda = \Theta(\sqrt{d})$, we say that our graph is *pseudorandom*. It is known that any K_t-free pseudorandom d-regular n-vertex graph has $d = O(n^{1-\frac{1}{2t-3}})$. The result of Mubayi and Verstraëte states that if there exists a K_t-free pseudorandom d-regular n-vertex graph with $d = \Theta(n^{1-\frac{1}{2t-3}})$, then $R(t, m) = m^{t-1-o(1)}$. Alon [3] constructed such a graph in the case $t = 3$.

A different construction was found by Conlon [20]. He starts with a C_4-free and C_6-free bipartite graph H with $\Theta(n^{4/3})$ edges (with parts X and Y) and defines a graph G with vertex set Y as follows. For each $x \in X$, we randomly partition $N_H(x)$ into two parts and take a complete bipartite graph between the parts. Then we define G to be the union of these bipartite graphs. Because of the C_6-freeness of H, G is triangle-free. Moreover, since H has $\Theta(n^{4/3})$ edges and it is C_4-free, almost surely G has $\Theta(n^{5/3})$ edges. Conlon showed that G is pseudorandom (up to a logarithmic factor which makes no difference in the Ramsey theory applications). Here comes the connection to our Theorem 2.1.6. If, instead of starting with a C_6-free graph, we start with a K_t'-free bipartite graph H with $\Theta(n^{3/2-\frac{1}{4t-6}})$ edges (which may or may not exist) and define G as above, then

G is K_t-free and (provided that H is C_4-free) G has $\Theta(n^{2-\frac{1}{2t-3}})$ edges. If we can also show that, similarly to the $t=3$ case, G is sufficiently pseudorandom, that would imply $R(t,m) = m^{t-1-o(1)}$.

We now continue our discussion of bipartite extremal numbers. Our next result gives some small progress towards Conjecture 2.1.4 for general r.

Theorem 2.1.7 (Conlon–Janzer–Lee [23]). *Let H be a bipartite graph such that in one of the parts all the degrees are at most r and H does not contain $K_{2,2}$ as a subgraph. Then $\mathrm{ex}(n,H) = o(n^{2-1/r})$.*

Recently, Sudakov and Tomon proved the following stronger result.

Theorem 2.1.8 (Sudakov–Tomon [107]). *Let H be a bipartite graph such that in one of the parts all the degrees are at most r and H does not contain $K_{r,r}$ as a subgraph. Then $\mathrm{ex}(n,H) = o(n^{2-1/r})$.*

In the next two subsections we present the rich history of the study of longer subdivisions of graphs and multigraphs.

2.1.1 Longer subdivisions of (multi)graphs

Many researchers have studied the problem of estimating the number of edges needed in a graph G on n vertices to guarantee that it contains as a subgraph a subdivided copy of a fixed graph. The first result in this direction is due to Mader [94] who proved that for any graph F there exists a constant $C = C(F)$ such that if an n-vertex graph G contains at least Cn edges, then G contains a subdivision of F as a subgraph. In this result the size of the subdivided graph can grow with n, which is necessary since an n-vertex graph with Cn edges need not contain a cycle of bounded length.

Answering a question of Erdős about planar subgraphs [30], Kostochka and Pyber [88] proved that any n-vertex graph with at least $4^{t^2}n^{1+\varepsilon}$ edges contains a subdivided K_t with at most $\frac{7t^2\log t}{\varepsilon}$ vertices. This is the first result that guarantees a subdivided K_t of bounded size.

Let $\mathcal{F}_{t,k}$ be the family of graphs that can be obtained by replacing the edges of K_t with pairwise internally vertex-disjoint paths of length at most k. Jiang [71] proved that for any $t \in \mathbb{N}$ and any $0 < \varepsilon < 1/2$, we have $\mathrm{ex}(n, \mathcal{F}_{t,\lceil 10/\varepsilon \rceil}) = O(n^{1+\varepsilon})$.

Note that Jiang's result improves that of Kostochka and Pyber in two ways. Firstly, any $F \in \mathcal{F}_{t,\lceil 10/\varepsilon \rceil}$ has at most $\frac{Ct^2}{\varepsilon}$ vertices, so a log factor is saved. Secondly, the edges in Jiang's theorem are replaced by uniformly short paths not depending on t. However, they can still have different lengths. The next result of Jiang and Seiver guarantees a subdivided K_t with prescribed path lengths.

Theorem 2.1.9 (Jiang–Seiver [76]). *For any $t \in \mathbb{N}$ and any even $k \in \mathbb{N}$,*

$$\mathrm{ex}(n, K_t^{k-1}) = O(n^{1+\frac{16}{k}}).$$

Note that if k is odd, then K_t^{k-1} is not a bipartite graph, so $\mathrm{ex}(n, K_t^{k-1}) = \Theta(n^2)$.

Conlon and Lee conjectured that the following two strengthenings hold.

Conjecture 2.1.10 (Conlon–Lee [24]). *Let F be a multigraph and let $k \geq 2$ be even. Then*

$$\mathrm{ex}(n, F^{k-1}) = O(n^{1+\frac{1}{k}}).$$

Conjecture 2.1.11 (Conlon–Lee [24]). *Let F be a simple graph and let $k \geq 2$ be even. Then there exists some $\varepsilon > 0$ such that*

$$\mathrm{ex}(n, F^{k-1}) = O(n^{1+\frac{1}{k}-\varepsilon}).$$

In the case $k = 2$, Conjecture 2.1.10 follows from the $r = 2$ case of Theorem 2.1.3, while Conjecture 2.1.11 follows from Theorem 2.1.5. Conlon, Janzer and Lee [23] proved Conjecture 2.1.11 for every bipartite graph F (see Theorem 2.1.14 in the next subsection).

We prove both Conjecture 2.1.10 and Conjecture 2.1.11.

Theorem 2.1.12 (Janzer [69]). *Let F be a multigraph and let $k \geq 2$ be even. Then*

$$\mathrm{ex}(n, F^{k-1}) = O(n^{1+\frac{1}{k}}).$$

Theorem 2.1.13 (Janzer [69]). *Let F be a simple graph and let $k \geq 2$ be even. Then there exists some $\varepsilon > 0$ such that*

$$\mathrm{ex}(n, F^{k-1}) = O(n^{1+\frac{1}{k}-\varepsilon}).$$

Note that Theorem 2.1.12 is tight. Indeed, as we have mentioned above, the theta graph $\theta_{k,\ell}$ (which is the $(k-1)$-subdivision of the multigraph consisting of a multiplicity ℓ edge) has extremal number $\Theta(n^{1+1/k})$ for all $\ell \geq \ell_0(k)$. Moreover, Erdős–Rényi random graphs show that $\mathrm{ex}(n, K_t^{k-1}) = \Omega(n^{1+1/k-c_{k,t}})$ where $c_{k,t} \to 0$ as $t \to \infty$. So the term $1 + 1/k$ in the exponent in Theorem 2.1.13 is also best possible, though our ε is not optimal.

2.1.2 Longer subdivisions of the complete bipartite graph

In this subsection we focus on the extremal number of the subdivisions of the complete bipartite graph. The first few results on this topic concerned the 1-subdivision of the complete bipartite graph. Conlon and Lee [24] proved that if $s \leq t$, then $\mathrm{ex}(n, K'_{s,t}) = O(n^{\frac{3}{2}-\frac{1}{12t}})$. This was improved by the author [65] to $\mathrm{ex}(n, K'_{s,t}) = O(n^{\frac{3}{2}-\frac{1}{4s-2}})$ and the same result was reproved using different methods by Kang, Kim and Liu [79]. Moreover, they conjectured that $\mathrm{ex}(n, K'_{s,t}) = O(n^{\frac{3}{2}-\frac{1}{2s}})$ holds, which is then tight for sufficiently large t by a general result of Bukh and Conlon (see Theorem 2.8.1 below). The conjecture was proved by the author, Conlon and Lee [23]. About longer subdivisions, we proved the following result.

Theorem 2.1.14 (Conlon–Janzer–Lee [23]). *For any integers $s, t, k \geq 2$,*

$$\mathrm{ex}(n, K_{s,t}^{k-1}) = O(n^{1+\frac{s}{sk+1}}).$$

This is nearly sharp for t sufficiently large, since Theorem 2.8.1 implies that there exists $t_0 = t_0(s, k)$ such that for all $t \geq t_0$, $\mathrm{ex}(n, K_{s,t}^{k-1}) = \Omega(n^{1+\frac{s-1}{sk}})$.

Together with Conlon and Lee, we conjectured that this lower bound is tight.

Conjecture 2.1.15 (Conlon–Janzer–Lee [23]). *For any integers $s, t, k \geq 2$,*

$$\mathrm{ex}(n, K_{s,t}^{k-1}) = O(n^{1+\frac{s-1}{sk}}).$$

Jiang and Qiu proved that the conjecture holds for $k = 3$ and $k = 4$ (as mentioned above, the $k = 2$ case had been proved by the author, Conlon and Lee).

Theorem 2.1.16 (Jiang–Qiu [75]). *For any integers $s, t \geq 2$ and $k \in \{3, 4\}$,*

$$\mathrm{ex}(n, K_{s,t}^{k-1}) = O(n^{1+\frac{s-1}{sk}}).$$

We prove Conjecture 2.1.15 for arbitrary k.

Theorem 2.1.17 (Janzer [66]). *For any integers $s, t, k \geq 2$,*

$$\mathrm{ex}(n, K_{s,t}^{k-1}) = O(n^{1+\frac{s-1}{sk}}).$$

Corollary 2.1.18. *For any integers $s, k \geq 2$, there exists $t_0 = t_0(s, k)$ such that for all integers $t \geq t_0$,*

$$\mathrm{ex}(n, K_{s,t}^{k-1}) = \Theta(n^{1+\frac{s-1}{sk}}).$$

This means that $1 + \frac{s-1}{sk}$ is realisable for every $s, k \geq 2$.

The structure of this chapter

The rest of this chapter is organised as follows. In Section 2.2 we present some preliminary lemmas that will be used in the proofs. In Section 2.3 we prove Theorem 2.1.6. In Section 2.4 we prove Theorem 2.1.7. In Section 2.5 we give a short proof of a result of Kang, Kim and Liu, mentioned in the introduction. In Section 2.6 we prove Theorem 2.1.12 and Theorem 2.1.13. In Section 2.7 we prove Theorem 2.1.17. In Section 2.8 we present some concluding remarks.

2.2 Preliminaries

A common feature of our proofs is that we first assume that our host graph is sufficiently regular. Let us say that a graph G is *K-almost-regular* if $\max_{v \in V(G)} \deg(v) \leq$

$K \min_{v \in V(G)} \deg(v)$. The reason why we may assume that our graph is almost regular is the following result of Jiang and Seiver, which is a slight modification of a much earlier result of Erdős and Simonovits [38].

Lemma 2.2.1 (Jiang–Seiver [76]). *Let ε, c be positive reals, where $\varepsilon < 1$ and $c \geq 1$. Let n be a positive integer that is sufficiently large as a function of ε. Let G be a graph on n vertices with $e(G) \geq cn^{1+\varepsilon}$. Then G contains a K-almost-regular subgraph G_{reg} on $m \geq n^{\frac{\varepsilon}{2}\frac{1-\varepsilon}{1+\varepsilon}}$ vertices such that $e(G_{\text{reg}}) \geq \frac{2c}{5}m^{1+\varepsilon}$ and $K = 20 \cdot 2^{\frac{1}{\varepsilon^2}+1}$.*

In Section 2.4 we will need a version of this where c can be smaller than 1.

Lemma 2.2.2. *Let ε, c be positive reals, where $\varepsilon < 1$. Let n be a positive integer that is sufficiently large as a function of ε. Let G be a graph on n vertices with $e(G) \geq cn^{1+\varepsilon}$. Then G contains a K-almost-regular subgraph G_{reg} on $m \geq n^{\frac{\varepsilon-\varepsilon^2}{4+4\varepsilon}}$ vertices such that $e(G_{\text{reg}}) \geq \frac{2c}{5}m^{1+\varepsilon}$ and $K = 20 \cdot 2^{\frac{1}{\varepsilon^2}+1}$.*

The proof of this is the same as the proof of Lemma 2.2.1 with one straightforward modification. Nevertheless, we include it here for completeness.

Proof. For convenience, we will drop ceilings and floors whenever doing so does not affect the analysis in an essential way. Let ε, c be positive reals, where $\varepsilon < 1$. Let n be a positive integer sufficiently large as a function of ε. Let G be a graph on n vertices with $e(G) \geq cn^{1+\varepsilon}$. Set $p = \lceil 2^{\frac{1}{\varepsilon^2}+1} \rceil$. We partition $V(G)$ into $2p$ almost equal parts $B_1, \ldots, B_{2p}$, where B_1 consists of $\lceil \frac{n}{2p} \rceil$ vertices of the highest degrees in G. Suppose first that at most $\frac{c}{2}n^{1+\varepsilon}$ edges of G are incident to B_1. We say that G is of type 1. Let $H = G - B_1$. Then $e(H) \geq \frac{c}{2}n^{1+\varepsilon}$. Successively remove vertices of degree less than $\frac{c}{10}n^{\varepsilon}$ from H until we get stuck; denote the remaining subgraph by G_{reg}. Let $m = |V(G_{\text{reg}})|$. Since at most $\frac{c}{10}n^{\varepsilon} \cdot n = \frac{c}{10}n^{1+\varepsilon}$ edges are removed in the process, we have $e(G_{\text{reg}}) \geq \frac{4c}{10}n^{1+\varepsilon} \geq \frac{2c}{5}m^{1+\varepsilon}$. Also, $\delta(G_{\text{reg}}) \geq \frac{c}{10}n^{\varepsilon}$ by the way we obtained G_{reg}. By our assumption of B_1, $d_G(x) \geq \Delta(G_{\text{reg}})$ for all $x \in B_1$. Also, $\sum_{x \in B_1} d_G(x) \leq cn^{1+\varepsilon}$ since at most $\frac{c}{2}n^{1+\varepsilon}$ edges of G are incident to B_1. We have $\Delta(G_{\text{reg}})(n/2p) \geq \sum_{x \in B_1} d_G(x) \geq cn^{1+\varepsilon}$, from which we get $\Delta(G_{\text{reg}}) \leq 2pcn^{\varepsilon}$. Thus, $\Delta(G_{\text{reg}})/\delta(G_{\text{reg}}) \leq 2pcn^{\varepsilon}/\frac{c}{10}n^{\varepsilon} = 20p$. So G_{reg} is K-almost-regular. Also, $m \geq 2e(G_{\text{reg}})/\Delta(G_{\text{reg}}) \geq \frac{4c}{5}n^{1+\varepsilon}/2pcn^{\varepsilon} = \frac{2}{5p}n \geq n^{\frac{\varepsilon-\varepsilon^2}{4+4\varepsilon}}$ for large n. So, the claim holds.

Suppose now that more than $\frac{c}{2}n^{1+\varepsilon}$ edges of G are incident to B_1. We say that G is of type 2. By an averaging argument, for some $j \in \{2, \ldots, 2p\}$, the subgraph G_1 of G induced by $B_1 \cup B_j$ has more than $\frac{1}{2p}\frac{c}{2}n^{1+\varepsilon} = \frac{c}{4p}n^{1+\varepsilon}$ edges. Let $n_1 = |V(G_1)|$. Then $n_1 \approx n/p$. Note that $cn_1^{1+\varepsilon} = c(\frac{n}{p})^{1+\varepsilon} = \frac{c}{p}n^{1+\varepsilon}\frac{1}{p^{\varepsilon}} \leq \frac{c}{4p}n^{1+\varepsilon}$, using that $p^{\varepsilon} \geq 2^{(\frac{1}{\varepsilon^2}+1)\varepsilon} \geq 4$. So $e(G_1) \geq cn_1^{1+\varepsilon}$.

We can now replace G with G_1 and repeat the analysis. If G_1 is of type 1, we terminate. If G_1 of type 2, we define G_2 from G_1 the way we defined G_1 from G. We continue like this as long as the new graph G_i is of type 2. We terminate when G_i is of type 1 for the first time. With $G_0 = G$, let k be the smallest i such that G_i is of type 1. Then $|V(G_k)| \approx \frac{n}{p^k}$

and $e(G_k) \geq \frac{c}{(4p)^k} n^{1+\varepsilon}$. Since $e(G_k) \leq |V(G_k)|^2$, we have $\frac{c}{(4p)^k} n^{1+\varepsilon} \leq \frac{n^2}{p^{2k}}$. Thus, $(\frac{p}{4})^k \leq \frac{n^{1-\varepsilon}}{c} \leq n^{1-\varepsilon+\frac{\varepsilon(1-\varepsilon)^2}{2(1+\varepsilon^2)}}$ as n is sufficiently large. Hence, $k \leq (1-\varepsilon+\frac{\varepsilon(1-\varepsilon)^2}{2(1+\varepsilon^2)})\frac{\log n}{\log(p/4)}$. Since $n_k = |V(G_k)| \approx n/p^k$, $\log n_k \approx \log n - k\log p \geq (1-(1-\varepsilon+\frac{\varepsilon(1-\varepsilon)^2}{2(1+\varepsilon^2)})\frac{\log p}{\log(p/4)})\log n$. Plugging in $p = 2^{\frac{1}{\varepsilon^2}+1}$, we get $\log n_k \geq (1-(1-\varepsilon+\frac{\varepsilon(1-\varepsilon)^2}{2(1+\varepsilon^2)})\frac{\frac{1}{\varepsilon^2}+1}{\frac{1}{\varepsilon^2}-1})\log n = \frac{\varepsilon-\varepsilon^2}{2+2\varepsilon}\log n$, therefore $n_k \geq n^{\frac{\varepsilon-\varepsilon^2}{2+2\varepsilon}}$. Since G_k is of type 1, by our earlier arguments it contains a subgraph G_{reg} on m vertices where $m \geq \frac{2}{5p}n_k \geq n^{\frac{\varepsilon-\varepsilon^2}{4+4\varepsilon}}$ for large n. Furthermore, $e(G_{\text{reg}}) \geq \frac{2c}{5}m^{1+\varepsilon}$, and G_{reg} is K-almost-regular. This completes the proof. □

We will in fact need a version of this lemma which gives an almost-regular bipartite subgraph. Following Conlon and Lee, we say that a bipartite graph G with a bipartition $A \cup B$ is *balanced* if $\frac{1}{2}|B| \leq |A| \leq 2|B|$.

Lemma 2.2.3. *Let ε, c be positive reals, where $\varepsilon < 1$. Let n be a positive integer that is sufficiently large as a function of ε. Let G be a graph on n vertices with $e(G) \geq cn^{1+\varepsilon}$. Then G contains a K-almost-regular balanced bipartite subgraph G_{reg} on $m \geq n^{\frac{\varepsilon-\varepsilon^2}{4+4\varepsilon}}$ vertices such that $e(G_{\text{reg}}) \geq \frac{c}{10}m^{1+\varepsilon}$ and $K = 60 \cdot 2^{\frac{1}{\varepsilon^2}+1}$.*

The proof of this lemma is almost identical to the proof of Lemma 2.3 in [24] and is therefore omitted.

The notation we will use in the remaining sections is mostly standard. For a graph G and $v \in V(G)$, we write $N_G(v)$ (or $N(v)$ if G is clear) for the neighbourhood of v in G. Also, we write $d_G(v)$ or $d(v)$ for the degree of v. Finally, if $u_1, \ldots, u_r \in V(G)$, then we write $d_G(u_1, \ldots, u_r) = d(u_1, \ldots, u_r) = |N_G(u_1) \cap \cdots \cap N_G(u_r)|$.

2.3 The 1-subdivision of K_t

In this section, we shall prove Theorem 2.1.6. Note that Lemma 2.2.3 reduces Theorem 2.1.6 to the following.

Theorem 2.3.1. *For every $K \geq 1$ and integer $t \geq 3$, there exists a constant $c = c(t, K)$ with the following property. Let n be sufficiently large and let G be a balanced bipartite graph with bipartition $A \cup B$, $|B| = n$ such that the degree of every vertex of G is between δ and $K\delta$, for some $\delta \geq cn^{\frac{t-2}{2t-3}}$. Then G contains a copy of K_t'.*

Given a bipartite graph G with bipartition $A \cup B$, the *neighbourhood graph* is the weighted graph W_G on vertex set A where the weight of the pair uv is $d_G(u, v)$. For a subset $U \subset A$, we write $W(U)$ for the total weight in U, ie. $W(U) = \sum_{uv \in \binom{U}{2}} d_G(u, v)$.

We shall use the following simple lemma of Conlon and Lee [24, Lemma 2.4].

Lemma 2.3.2. *Let G be a bipartite graph with bipartition $A \cup B$, $|B| = n$, and minimum degree at least δ on the vertices in A. Then for any subset $U \subset A$ with $\delta|U| \geq 2n$,*

$$\sum_{uv \in \binom{U}{2}} d_G(u, v) \geq \frac{\delta^2}{2n}\binom{|U|}{2}$$

In other words, the conclusion of Lemma 2.3.2 is that $W(U) \geq \frac{\delta^2}{2n}\binom{|U|}{2}$.

In the next definition, and in the rest of this section, for a weighted graph W on vertex set A, if $u, v \in A$, then $W(u,v)$ stands for the weight of uv. Moreover, we shall tacitly assume throughout the section that $t \geq 3$ is a fixed integer.

Definition 2.3.3. Let W be a weighted graph on vertex set A and let $u, v \in A$ be distinct. We say that uv is a *light edge* if $1 \leq W(u,v) < \binom{t}{2}$ and that it is a *heavy edge* if $W(u,v) \geq \binom{t}{2}$.

Note that if there is a K_t in W_G formed by heavy edges, then there is an K_t' in G.

The next lemma is one of our key observations.

Lemma 2.3.4. *Let G be an K_t'-free bipartite graph with bipartition $A \cup B$, $|B| = n$ and suppose that $W(A) \geq 2t^2 n$. Then the number of light edges in W_G is at least $\frac{W(A)}{2t^4}$.*

Proof. Let $B = \{b_1, \ldots, b_n\}$. Let $k_i = |N_G(b_i)|$ and suppose that $k_i \geq t$ for some i. As G is K_t'-free, there is no K_t in $W[N_G(b_i)]$ formed by heavy edges. Thus, by simple averaging, the number of light edges in $N_G(b_i)$ is at least $\frac{1}{\binom{t}{2}}\binom{k_i}{2}$. But

$$\sum_{i:k_i<t} \binom{k_i}{2} < t^2 n \leq \frac{W(A)}{2},$$

so

$$\sum_{i:k_i \geq t} \binom{k_i}{2} \geq \frac{W(A)}{2}.$$

Since every light edge is present in at most $\binom{t}{2}$ of the sets $N_G(b_i)$, it follows that the total number of light edges is at least

$$\frac{1}{\binom{t}{2}} \sum_{i:k_i \geq t} \frac{1}{\binom{t}{2}} \binom{k_i}{2} \geq \frac{W(A)}{4t^2}.$$

□

Corollary 2.3.5. *Let G be a K_t'-free bipartite graph with bipartition $A \cup B$, $|B| = n$, and minimum degree at least δ on the vertices in A. Then for any subset $U \subset A$ with $|U| \geq \frac{4tn}{\delta}$ and $|U| \geq 2$, the number of light edges in $W_G[U]$ is at least $\frac{\delta^2}{4t^4 n}\binom{|U|}{2}$.*

Proof. By Lemma 2.3.2, we have $W(U) \geq \frac{\delta^2}{2n}\binom{|U|}{2} \geq \frac{\delta^2}{8n}|U|^2 \geq 2t^2 n$. Now the result follows by applying Lemma 2.3.4 to the graph $G[U \cup B]$. □

We are now in a position to complete the proof of Theorem 2.3.1.

Proof of Theorem 2.3.1. Let c be specified later and suppose that n is sufficiently large. Assume, for contradiction, that G is K_t'-free. We shall find distinct $u_1, \ldots, u_t \in A$ with the following properties.

(i) Each u_iu_j is a light edge in W_G.

(ii) If i, j, k are distinct, then $N_G(u_i) \cap N_G(u_j) \cap N_G(u_k) = \emptyset$.

(iii) For each $1 \leq i \leq t-1$, the number of $v \in A$ with the property that for every $j \leq i$, u_jv is a light edge is at least $(\frac{\delta^2}{16t^4n})^i \cdot |A|$.

As n is sufficiently large, we have $|A| \geq n/2 \geq \frac{4tn}{\delta}$, therefore by Corollary 2.3.5 there are at least $\frac{\delta^2}{4t^4n}\binom{|A|}{2}$ light edges in A, so we may choose $u_1 \in A$ such that the number of light edges u_1v is at least $\frac{\delta^2}{4t^4n}(|A|-1) \geq \frac{\delta^2}{16t^4n}|A|$.

Now suppose that $2 \leq i \leq t-1$, and that $u_1, \ldots, u_{i-1}$ have been constructed satisfying (i), (ii) and (iii). Let U_0 be the set of vertices $v \in A$ with the property that u_jv is a light edge for every $j \leq i-1$. By (iii), we have $|U_0| \geq (\frac{\delta^2}{16t^4n})^{i-1}|A|$. Now let U consist of those $v \in U_0$ for which $N_G(u_j) \cap N_G(u_k) \cap N_G(v) = \emptyset$ holds for all $1 \leq j < k \leq i-1$. Since u_ju_k is a light edge for any $1 \leq j < k \leq i-1$, we have that $d_G(u_j, u_k) < \binom{t}{2}$. But the degree of every $b \in B$ is at most $K\delta$, therefore the number of $v \in A$ for which $N_G(u_j) \cap N_G(u_k) \cap N_G(v) \neq \emptyset$ is at most $\binom{t}{2}K\delta$, so $|U_0 \setminus U| \leq \binom{i-1}{2}\binom{t}{2}K\delta$. But note that for sufficiently large n, we have $(\frac{\delta^2}{16t^4n})^{i-1}|A| \geq 2\binom{i-1}{2}\binom{t}{2}K\delta$ because $\delta = o((\delta^2/n)^{t-2}n)$ and $\delta = o((\delta^2/n)n)$. Thus,

$$|U| \geq \frac{1}{2}|U_0| \geq \frac{1}{2}\left(\frac{\delta^2}{16t^4n}\right)^{i-1}|A|.$$

But for sufficiently large $c = c(t, K)$, we have $\frac{1}{2}(\frac{\delta^2}{16t^4n})^{i-1}|A| \geq \frac{4tn}{\delta}$. Indeed, this is obvious when $\delta^2 \geq 16t^4n$, and otherwise, using $\delta \geq cn^{\frac{t-2}{2t-3}}$, we have

$$\frac{1}{2}\left(\frac{\delta^2}{16t^4n}\right)^{i-1}|A| \geq \frac{1}{2}\left(\frac{\delta^2}{16t^4n}\right)^{t-2}|A| \geq \frac{1}{4(16t^4)^{t-2}} \cdot \frac{\delta^{2t-4}}{n^{t-3}} \geq \frac{4tn}{\delta}$$

Thus, by Corollary 2.3.5, there exists some $u_i \in U$ with at least $\frac{\delta^2}{4t^4n}(|U|-1) \geq (\frac{\delta^2}{16t^4n})^i|A|$ light edges adjacent to it in U. This completes the recursive construction of the vertices $\{u_j\}_{1 \leq j \leq t-1}$.

By (iii) for $i = t-1$, there is a set $V \subset A$ consisting of at least $(\frac{\delta^2}{16t^4n})^{t-1}|A|$ vertices v such that for every $j \leq t-1$, u_jv is a light edge. Since every u_iu_j is a light edge, the number of those $v \in A$ with $N_G(u_i) \cap N_G(u_j) \cap N_G(v) \neq \emptyset$ for some $i \neq j$ is at most $\binom{t-1}{2}\binom{t}{2}K\delta$. But for large enough $c = c(t, K)$, this is less than $(\frac{\delta^2}{16t^4n})^{t-1}|A|$, so there exists $u_t \in V$ such that $u_1, \ldots, u_t$ satisfy (i), (ii) and (iii) above.

Now it is easy to see that there exists a K_t' in G containing $u_1, \ldots, u_t$ as vertices. $\square$

2.4 $K_{2,2}$-free bipartite graphs with max degree r on one side

We now use ideas from the previous section to prove Theorem 2.1.7. In order to prove this theorem, we may clearly assume that all the degrees in one part of H are *exactly* r. Then Lemma 2.2.3 reduces Theorem 2.1.7 to the following statement.

Theorem 2.4.1. *Let $r \geq 2$ be an integer, let $K \geq 1$ be fixed and let H be a bipartite graph such that in one of the parts all the degrees are exactly r and H does not contain $K_{2,2}$ as a subgraph. Then, for any constant $c > 0$, there exists n_0 such that if $n \geq n_0$ and G is a K-almost-regular balanced bipartite graph with bipartition $A \cup B$, $|B| = n$, with minimum degree $\delta \geq cn^{1-1/r}$, then G contains a copy of H.*

We need the following generalisation of Lemma 2.3.2.

Lemma 2.4.2. *Let $r \geq 2$ be an integer and let G be a bipartite graph with bipartition $A \cup B$, $|B| = n$, and minimum degree at least δ on the vertices in A. Then, for any subset $U \subset A$ with $|U| \geq \frac{rn}{\delta}$,*

$$\sum_{u_1 \dots u_r \in \binom{U}{r}} d(u_1, \dots, u_r) \geq \frac{\delta^r}{r^r n^{r-1}} |U|^r \geq \frac{\delta^r}{r^r n^{r-1}} \binom{|U|}{r}.$$

Proof. Writing $d_U(v)$ for $|N_G(v) \cap U|$, we have that

$$\begin{aligned} \sum_{u_1 \dots u_r \in \binom{U}{r}} d(u_1, \dots, u_r) &= \sum_{b \in B} \binom{d_U(b)}{r} \geq n \binom{\sum_{b \in B} d_U(b)/n}{r} \\ &= n \binom{\sum_{u \in U} d(u)/n}{r} \geq n \binom{\delta|U|/n}{r} \\ &\geq n \left(\frac{\delta|U|}{rn}\right)^r = \frac{\delta^r}{r^r n^{r-1}} |U|^r, \end{aligned}$$

where the first inequality follows from the convexity of $\binom{x}{r}$ and in the last inequality we used that $|U| \geq \frac{rn}{\delta}$. $\square$

Given a bipartite graph G with bipartition $A \cup B$, the *neighbourhood r-graph* is the weighted r-uniform hypergraph W_G on vertex set A, where the weight of the hyperedge $u_1 \dots u_r$ (for $u_1, \dots, u_r$ distinct) is $d(u_1, \dots, u_r)$. For a subset $U \subset A$, we write $W(U)$ for the total weight in U, i.e., $W(U) = \sum_{u_1 \dots u_r \in \binom{U}{r}} d(u_1, \dots, u_r)$. In this language, the conclusion of Lemma 2.4.2 is that $W(U) \geq \frac{\delta^r}{r^r n^{r-1}} \binom{|U|}{r}$.

In the next definition, for a weighted r-graph W on vertex set A and $u_1, \dots, u_r \in A$, we write $W(u_1, \dots, u_r)$ for the weight of the hyperedge $u_1 \dots u_r$. Moreover, in what follows we fix $r \geq 2$ and a bipartite graph H with the property that in one part all the degrees are exactly r. Let $h = |V(H)|$.

Definition 2.4.3. Let W be a weighted r-graph on vertex set A and let $u_1, \ldots, u_r \in A$ be distinct. We say that $u_1 \ldots u_r$ is a *light edge* if $1 \leq W(u_1, \ldots, u_r) < \binom{h}{r}$ and that it is a *heavy edge* if $W(u_1, \ldots, u_r) \geq \binom{h}{r}$.

Note that if there is a $K_h^{(r)}$ in W_G formed by heavy edges, then clearly there is a copy of H in G. This observation is an important ingredient in our next lemma, which is the generalisation of Lemma 2.3.4.

Lemma 2.4.4. *Let G be an H-free bipartite graph with bipartition $A \cup B$, $|B| = n$, and suppose that $W(A) \geq 2h^r n$. Then the number of light edges in W_G is at least $\frac{W(A)}{2h^{2r}}$.*

Proof. Suppose $B = \{b_1, \ldots, b_n\}$. Let $k_i = |N_G(b_i)|$ and suppose that $k_i \geq h$ for some i. As G is H-free, there is no $K_h^{(r)}$ in $W[N_G(b_i)]$ formed by heavy edges. Since $\mathrm{ex}(t, K_h^{(r)}) \leq (1 - 1/\binom{h}{r})\binom{t}{r}$ holds for $t \geq h$, the number of light edges in $N_G(b_i)$ is at least $\frac{\binom{k_i}{r}}{\binom{h}{r}}$. But

$$\sum_{i:k_i<h} \binom{k_i}{r} < h^r n \leq \frac{W(A)}{2},$$

so

$$\sum_{i:k_i\geq h} \binom{k_i}{r} \geq \frac{W(A)}{2}.$$

Since every light edge is present in at most $\binom{h}{r}$ of the sets $N_G(b_i)$, it follows that the total number of light edges is at least

$$\frac{1}{\binom{h}{r}} \sum_{i:k_i\geq h} \frac{\binom{k_i}{r}}{\binom{h}{r}} \geq \frac{W(A)}{2h^{2r}},$$

as required. □

Corollary 2.4.5. *Let G be an H-free bipartite graph with bipartition $A \cup B$, $|B| = n$, and minimum degree at least δ on the vertices in A. Then, for any subset $U \subset A$ with $|U| \geq \frac{2hrn}{\delta}$, the number of light edges in $W_G[U]$ is at least $\frac{\delta^r}{2h^{2r}r^r n^{r-1}}\binom{|U|}{r}$.*

Proof. By Lemma 2.4.2, we have $W(U) \geq \frac{\delta^r}{r^r n^{r-1}}|U|^r \geq 2h^r n$. Hence, the result follows by applying Lemma 2.4.4 to the graph $G[U \cup B]$. □

We now recall Definition 5 from [85].

Definition 2.4.6. An r-uniform hypergraph $\mathcal{G} = (V, E)$ is (ρ, d)-dense if, for any subset $U \subset V$ of size $|U| \geq \rho|V|$, $e_{\mathcal{G}}(U) \geq d\binom{|U|}{r}$.

Recall also that a *linear hypergraph* is a hypergraph where any two edges intersect in at most one vertex. The following result follows from Theorem 7 in [85].

Theorem 2.4.7 (Kohayakawa–Nagle–Rödl–Schacht). *Let $\mathcal{L}$ be a linear r-uniform hypergraph on ℓ vertices. Then, for every $d > 0$, there exist $\rho = \rho(\mathcal{L}, d) > 0$, $\varepsilon = \varepsilon(\mathcal{L}, d) > 0$ and $n_0 = n_0(\mathcal{L}, d)$ such that every (ρ, d)-dense r-uniform hypergraph $\mathcal{G} = (V, E)$ on $n \geq n_0$ vertices contains at least $\varepsilon|V|^\ell$ copies of $\mathcal{L}$.*

We are now in a position to complete the proof of Theorem 2.4.1.

Proof of Theorem 2.4.1. We may assume that $\delta \leq n^{1-1/(2r)}$, as we already know that $\mathrm{ex}(n, H) = O(n^{2-1/r})$. Suppose that G is H-free. Define $\mathcal{G}$ to be the r-uniform (simple) hypergraph whose vertex set is A and whose edges are precisely the light edges of W_G. By Corollary 2.4.5, for any $U \subset A$ with $|U| \geq \frac{2hrn}{\delta}$, we have

$$e_{\mathcal{G}}(U) \geq \frac{\delta^r}{2h^{2r}r^r n^{r-1}}\binom{|U|}{r} \geq \frac{c^r}{2h^{2r}r^r}\binom{|U|}{r}.$$

Suppose H has bipartition $X \cup Y$ with every vertex in Y having degree r. Define $\mathcal{L}$ to be the r-uniform hypergraph whose vertex set is X and whose edges are the neighbourhoods $N_H(y)$ for $y \in Y$. Since H does not contain a $K_{2,2}$, it follows that $\mathcal{L}$ is linear. Let $d = \frac{c^r}{2h^{2r}r^r}$ and choose $\rho > 0$, $\varepsilon > 0$ and n_0 as in the conclusion of Theorem 2.4.7. Note that for n sufficiently large, we have $\frac{2hrn}{\delta} < \rho|A|$, so $\mathcal{G}$ is (ρ, d)-dense and consequently contains at least $\varepsilon|A|^{|X|}$ copies of $\mathcal{L}$. All these copies of $\mathcal{L}$ provide homomorphic copies of H in G (with vertices in X mapped to vertices in A and vertices in Y mapped to vertices in B), but some of these may be degenerate in the sense that distinct vertices in Y may be mapped to the same vertex in B.

We now give an upper bound for the number of degenerate copies of H, counting only those copies that were obtained by the method above. Any such degenerate copy must contain some $u \in B$ and $v_1, \dots, v_{r+1} \in N_G(u)$ with $v_1 \dots v_r$ a light edge in W_G. The number of possible choices for such a configuration is at most $(2n)^r \cdot \binom{h}{r} \cdot K\delta$, since we can choose $v_1, \dots, v_r$ in at most $(2n)^r$ ways (since $|A| \leq 2n$), then we can choose u in at most $\binom{h}{r}$ ways (since $v_1 \dots v_r$ is a light edge) and, finally, we can choose v_{r+1} in at most $K\delta$ ways (since $\Delta(G) \leq K\delta$). But the number of ways to extend this to a copy of H is at most $(2n)^{|X|-r-1} \cdot \binom{h}{r}^{\binom{|X|}{r}}$, because we can map those vertices in X that have not been mapped in at most $(2n)^{|X|-r-1}$ ways and, given any choice for the images of X, there are at most $\binom{h}{r}$ possible choices for the image of each $y \in Y$, since we are only counting those copies of H in which $N_H(y)$ is mapped to a light edge. Thus, of the $\varepsilon|A|^{|X|}$ copies of H that we found, at most $\binom{h}{r}^{\binom{|X|}{r}+1} K\delta(2n)^{|X|-1}$ are degenerate. Since $\delta \leq n^{1-\frac{1}{2r}}$ and $|A| \geq n/2$, for sufficiently large n we obtain a non-degenerate copy of H. □

2.5 A short proof of a result of Kang, Kim and Liu

As mentioned in the introduction, Kang, Kim and Liu proved that for each $a, b \in \mathbb{N}$ with $a < b$ and $b \equiv \pm 1 \pmod{a}$, the number $2 - \frac{a}{b}$ is realisable. Their main result was a

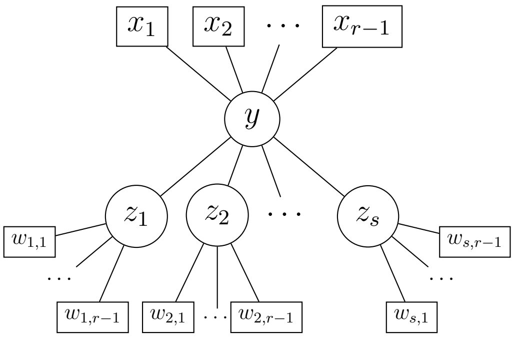

Figure 2.1: $H_{s,1}(r)$

tight upper bound on the extremal number of certain graphs from which the result just mentioned for $b \equiv -1 \pmod a$ follows fairly easily. We now define this family of graphs.

Consider a graph F with a set $R \subsetneq V(F)$ of root vertices. The *rooted t-blowup* of this rooted graph is the graph obtained by taking t vertex-disjoint copies of F and identifying the different copies of v for each $v \in R$. We let $H_{s,1}(r)$ be the graph consisting of vertices x_i $(1 \le i \le r-1)$, y, z_j $(1 \le j \le s)$ and $w_{j,k}$ $(1 \le j \le s, 1 \le k \le r-1)$ and edges $x_i y$ for all i, yz_j for all j and $z_j w_{j,k}$ for all j,k. Then $H_{s,t}(r)$ is the rooted t-blowup of $H_{s,1}(r)$, with the roots being $\{x_i : 1 \le i \le r-1\} \cup \{w_{j,k} : 1 \le j \le s, 1 \le k \le r-1\}$. For a picture, we refer the reader to Figure 2.1, where the root vertices are marked by rectangular boxes. The result of Kang, Kim and Liu [79, Lemma 3.2] is now as follows.

Theorem 2.5.1 (Kang–Kim–Liu). *For any integers $s,t \ge 1$ and $r \ge 2$,*

$$\mathrm{ex}(n, H_{s,t}(r)) = O(n^{2-\frac{s+1}{r(s+1)-1}}).$$

Combined with results of Bukh and Conlon [18] (see Theorem 2.8.1 below), Theorem 2.5.1 easily implies that $2 - \frac{s+1}{r(s+1)-1}$ is realisable for every $s \ge 1, r \ge 2$.

In this section, we illustrate our method by giving a new proof of Theorem 2.5.1 which is significantly shorter than the original one. By Lemma 2.2.1, it suffices to prove the following.

Theorem 2.5.2. *Let $s,t \ge 1$ and $r \ge 2$ be fixed integers and $K \ge 1$ a constant. Suppose that G is a K-almost-regular graph on n vertices with minimum degree $\delta = \omega(n^{1-\frac{s+1}{r(s+1)-1}})$. Then, for n sufficiently large, G contains a copy of $H_{s,t}(r)$.*

In what follows, let $s,t \ge 1$ and $r \ge 2$ be fixed integers and $K \ge 1$ a constant. Let $H = H_{s,t}(r)$ and $\delta = \omega(n^{1-\frac{s+1}{r(s+1)-1}})$. The constant L will be assumed to be sufficiently large in terms of s, t, r and K, while n will always be sufficiently large in terms of s, t, r, K and L. As a shorthand, we will now write $d_G(S)$ for the size of the common neighbourhood $N_G(S)$ of a set S.

Definition 2.5.3. An r-set $S \subset V(G)$ is called an *r-edge* if $d_G(S) > 0$. The weight of S is $d_G(S)$. S is called an *L-light r-edge* if $1 \le d_G(S) \le L$ and an *L-heavy r-edge* if $d_G(S) > L$.

Lemma 2.5.4. *Let G be an H-free K-almost-regular graph on n vertices with minimum degree δ. Then the total weight on L-heavy r-edges is at most an f_L-proportion of the total weight of r-edges, where $f_L \to 0$ as $L \to \infty$.*

Proof. First note that for any $r-1$ distinct vertices $x_1, \dots, x_{r-1}$, we cannot have $m = m_{s,t,r} = t + s(r-1)$ vertices in $N(x_1) \cap \dots \cap N(x_{r-1})$ such that any r of them form an edge of weight at least $c = c_{s,t,r} = |V(H)|$, since then we could find a copy of H. Indeed, if there are vertices y_i for $1 \le i \le t$ and $w_{j,k}$ for $1 \le j \le s$, $1 \le k \le r-1$ such that $N_G(\{y_i, w_{j,1}, \dots, w_{j,r-1}\})$ contains at least c elements for every i, j, then we can choose an element $z_{i,j}$ from each of these sets such that all the $x_i, y_j, z_{k,\ell}$ and $w_{a,b}$ are distinct, yielding a copy of H. Thus, as long as $|N(x_1) \cap \dots \cap N(x_{r-1})| \ge m$, we have that in $N(x_1) \cap \dots \cap N(x_{r-1})$ the proportion of those r-sets with weight at most c is at least $\eta = \eta_{s,t,r} = 1/\binom{m}{r}$. Since each r-set in $N_G(\{x_1, \dots, x_{r-1}\})$ is clearly an r-edge, it follows that the total number of r-edges of weight at most c is at least

$$\frac{1}{\binom{c}{r-1}} \cdot \eta \cdot \sum_{\substack{x_1 \dots x_{r-1} \in \binom{V(G)}{r-1} \\ d_G(x_1, \dots, x_{r-1}) \ge m}} \binom{d_G(x_1, \dots, x_{r-1})}{r}, \tag{2.1}$$

where we used the fact that an r-tuple of weight at most c is in at most $\binom{c}{r-1}$ of the sets $N_G(\{x_1, \dots, x_{r-1}\})$. Note now that

$$\sum_{x_1 \dots x_{r-1} \in \binom{V(G)}{r-1}} d_G(x_1, \dots, x_{r-1}) \ge n \binom{\delta}{r-1} = \Omega(n\delta^{r-1}).$$

Therefore, on average $d_G(x_1, \dots, x_{r-1})$ is $\Omega(n(\delta/n)^{r-1}) = \omega(1)$, so, by Jensen's inequality, we have

$$\begin{aligned}\sum_{x_1 \dots x_{r-1} \in \binom{V(G)}{r-1}} \binom{d_G(x_1, \dots, x_{r-1})}{r} &\ge 2\binom{|V(G)|}{r-1}\binom{m}{r} \\ &\ge 2 \sum_{\substack{x_1 \dots x_{r-1} \in \binom{V(G)}{r-1} \\ d_G(x_1, \dots, x_{r-1}) < m}} \binom{d_G(x_1, \dots, x_{r-1})}{r}.\end{aligned}$$

Thus, together with (2.1), the total number of r-edges of weight at most c (and, therefore, the total weight of r-edges) is at least

$$\frac{1}{2} \cdot \frac{1}{\binom{c}{r-1}} \cdot \eta \cdot \sum_{x_1 \dots x_{r-1} \in \binom{V(G)}{r-1}} \binom{d_G(x_1, \dots, x_{r-1})}{r}. \tag{2.2}$$

On the other hand, the total weight on r-edges of weight at least L is at most

$$\frac{L}{\binom{L}{r-1}} \cdot \sum_{x_1 \dots x_{r-1} \in \binom{V(G)}{r-1}} \binom{d_G(x_1, \dots, x_{r-1})}{r}, \tag{2.3}$$

since an r-edge of weight w is in $\binom{w}{r-1}$ of the sets $N_G(\{x_1, \dots, x_{r-1}\})$ and $w/\binom{w}{r-1}$ is a non-increasing function of w. If $r \geq 3$, then $L/\binom{L}{r-1} \to 0$ as $L \to \infty$ and, hence, the proportion of weight on L-heavy edges tends to 0 as L tends to infinity.

In the $r = 2$ case, (2.3) does not help us, so we take a slightly different approach. For a constant $\varepsilon > 0$, let $\xi = \frac{\varepsilon\eta}{2c}$. If $N(x_1)$ contains more than $\xi\binom{d(x_1)}{2}$ pairs of weight at least c, then, for n sufficiently large, there exists a copy of H. Indeed, the vertex x_1 together with a copy of $K_{s,t}$ in $N(x_1)$ formed by edges of weight at least c easily extend to a nondegenerate copy of H. Thus, for large enough n and $L = c$, the total weight on edges of weight at least L is at most

$$\xi \cdot \sum_{x \in V(G)} \binom{d_G(x)}{2},$$

which is at most ε times (2.2). $\square$

The following definition and lemma contain the key idea in our proof. Note that we continue to abuse notation slightly by referring to the vertices of $H_{s,t}(r)$ and their embedded images in another graph G by the same labels.

Definition 2.5.5. An embedding of $H_{s,1}(r)$ in a graph G is *L-good* if the r-sets $\{x_1, \dots, x_{r-1}, z_i\}$ and $\{y, w_{i,1}, \dots, w_{i,r-1}\}$ are L-light in G for every $1 \leq i \leq s$.

Lemma 2.5.6. *Let G be an H-free K-almost-regular graph on n vertices with minimum degree δ. Then, for L sufficiently large (not depending on n), the number of L-good embeddings of $H_{s,1}(r)$ in G is at least $\frac{1}{2}n\delta^{sr+r-1}$.*

Proof. The total weight on r-edges in G is equal to the number of r-stars, which is at most $n(K\delta)^r$ as $\Delta(G) \leq K\delta$. Thus, Lemma 2.5.4 implies that the number of r-stars whose leaf set is heavy is at most $c_L n\delta^r$, where $c_L \to 0$ as $L \to \infty$.

Since $H_{s,1}(r)$ is a tree on $sr + r$ vertices and every vertex in G has degree at least δ, there are at least $(1 - o(1))n\delta^{sr+r-1}$ copies of $H_{s,1}(r)$ in G. By the first paragraph, $\{x_1, \dots, x_{r-1}, z_1\}$ is not light in at most $rc_L n\delta^r(K\delta)^{sr-1}$ of them. Indeed, there are at most $(K\delta)^{sr-1}$ ways to extend a fixed choice of $x_1, \dots, x_{r-1}, y, z_1$, since $H_{s,1}(r)$ is connected and every vertex in G has degree at most $K\delta$. The factor r accounts for the fact that knowing the vertex set $\{x_1, \dots, x_{r-1}, y, z_1\}$ of the r-star leaves r possibilities for z_1. The same holds for the other r-sets $\{x_1, \dots, x_{r-1}, z_i\}$ and $\{y, w_{i,1}, \dots, w_{i,r-1}\}$, so the number of copies of $H_{s,1}(r)$ which are not suitable is at most $2s \cdot rc_L n\delta^r(K\delta)^{sr-1} = 2rsc_L K^{sr-1}n\delta^{sr+r-1}$. Since $c_L \to 0$ as $L \to \infty$, the result follows. $\square$

We are now in a position to prove Theorem 2.5.2.

Proof of Theorem 2.5.2. Choose L large enough that the conclusion of Lemma 2.5.6 holds. By that lemma and averaging, there exist x_i $(1 \leq i \leq r-1)$ and $w_{j,k}$ $(1 \leq j \leq s, 1 \leq k \leq r-1)$ which extend to at least $\Omega(n^{1-(r-1)-s(r-1)}\delta^{sr+r-1}) = \omega(1)$ L-good embeddings of $H_{s,1}(r)$. Take a maximal set $\mathcal{M}$ of such extensions which are vertex-disjoint apart from the roots. If $\mathcal{M}$ consists of at least t copies of $H_{s,1}(r)$, then their union forms a copy of $H_{s,t}(r)$.

Suppose instead that $\mathcal{M}$ consists of at most $t-1$ extensions. Then any other extension has a non-root vertex which coincides with one of the non-root vertices of some $M \in \mathcal{M}$. Since there are $O(1)$ non-root vertices in the graphs $M \in \mathcal{M}$ and $O(1)$ vertices in $H_{s,1}(r)$, there must exist some non-root vertex of $H_{s,1}(r)$ that is mapped to the same vertex in $\omega(1)$ of the good embeddings of $H_{s,1}(r)$ that extend x_i $(1 \leq i \leq r-1)$ and $w_{j,k}$ $(1 \leq j \leq s, 1 \leq k \leq r-1)$. Suppose first that y is mapped to the same vertex in $\omega(1)$ copies. Since $\{y, w_{j,1}, \dots, w_{j,r-1}\}$ is L-light for every j, this leaves at most $L = O(1)$ possibilities for each z_j, which is a contradiction. Similarly, suppose that some z_j takes the same vertex in $\omega(1)$ copies. Since $\{x_1, \dots, x_{r-1}, z_j\}$ is L-light, this allows only $L = O(1)$ possibilities for y, so y is mapped to some vertex $\omega(1)$ times. As before, this leads to a contradiction. $\square$

2.6 Longer subdivisions of (multi)graphs

In this section we prove Theorems 2.1.12 and 2.1.13.

2.6.1 The high-level structure of the proof

Using Lemma 2.2.1, Theorem 2.1.12 and Theorem 2.1.13 reduce to the following two statements, respectively. For notational convenience, we have dropped the assumption that k is even, and replaced k by $2k$.

Theorem 2.6.1. *Let F be a multigraph and let $k \geq 1$. Suppose that G is a K-almost-regular graph on n vertices with minimum degree $\delta = \omega(n^{\frac{1}{2k}})$. Then, for n sufficiently large, G contains a copy of F^{2k-1}.*

Theorem 2.6.2. *Let F be a simple graph and let $k \geq 1$. Then there exists $\varepsilon > 0$ with the following property. Suppose that G is a K-almost-regular graph on n vertices with minimum degree $\delta = \omega(n^{\frac{1}{2k}-\varepsilon})$. Then, for n sufficiently large, G contains a copy of F^{2k-1}.*

For the rest of this section we let F be an arbitrary fixed multigraph and write $H = F^{2k-1}$. Moreover, throughout the section we tacitly assume that n is sufficiently large.

The next definition was introduced in [23], and was used to prove Theorem 2.1.14.

Definition 2.6.3. Let L be a positive real and let $f(\ell, L) = L^{5^\ell}$ for $1 \le \ell \le 2k$. We recursively define the notions of L*-admissible* and L*-good paths of length* ℓ in a graph. Any path of length 1 is both L-admissible and L-good. For $2 \le \ell \le 2k$, we say a path $P = v_0 v_1 \dots v_\ell$ is L-admissible if every proper subpath of P is L-good, i.e., $v_i v_{i+1} \dots v_j$ is L-good for every $(i,j) \neq (0,\ell)$. The path P is L-good if it is L-admissible and the number of L-admissible paths of length ℓ between v_0 and v_ℓ is at most $f(\ell, L)$.

The next lemma will be used several times later.

Lemma 2.6.4. *Let $\ell \ge 2$ and let $L > \ell$. If a path $P = v_0 \dots v_\ell$ is L-admissible, but not L-good, then there exist at least L pairwise internally vertex-disjoint paths of length ℓ from v_0 to v_ℓ.*

Proof. Take a maximal set of pairwise internally vertex-disjoint paths of length ℓ from v_0 to v_ℓ and assume that it consists of fewer than L paths. These paths contain at most $L(\ell-1)$ internal vertices in total and any path of length ℓ between v_0 and v_ℓ intersects at least one of these vertices. Since there are at least L^{5^ℓ} L-admissible paths of length ℓ between v_0 and v_ℓ, it follows by pigeon hole that there exist some $1 \le i \le \ell-1$ and some $x \in V(G)$ such that there are at least $\frac{L^{5^\ell}}{(\ell-1)L(\ell-1)}$ L-admissible paths of the form $u_0 u_1 \dots u_\ell$ with $u_0 = v_0, u_i = x, u_\ell = v_\ell$. Observe that $\frac{L^{5^\ell}}{(\ell-1)L(\ell-1)} > L^{5^i} L^{5^{\ell-i}}$, so either there are more than L^{5^i} L-good paths of length i between v_0 and x or there are more than $L^{5^{\ell-i}}$ L-good paths of length $\ell - i$ between x and v_ℓ. In either case, we contradict the definition of an L-good path. □

Our strategy will be to prove that, roughly speaking, in any almost regular H-free graph there are many good paths of length $2k$. In Subsection 2.6.2 we prove that almost all paths of length k are good. In Subsection 2.6.3 we extend this to paths of length $2k$ and prove the following lemma.

Lemma 2.6.5. *Let G be an F^{2k-1}-free K-almost-regular graph on n vertices with minimum degree $\delta \ge L^{100^k |V(F)||E(F)|^2 (k+1)}$, and let $S \subset V(G)$. Then, provided that L is sufficiently large compared to $|V(F)|$, $|E(F)|$, k and K, $|S| = \omega(\frac{n}{\delta^{1/2}})$ and $|S| = \omega(\frac{n}{L^{1/2}})$, the number of L-good paths of length $2k$ with both endpoints in S is $\Omega(\frac{|S|^2 \delta^{2k}}{n})$.*

In this result and everywhere else in the section, the asymptotic notation Ω allows the implied constant to depend on $|V(F)|$, $|E(F)|$, k and K, which are thought of as constants, while δ and L are functions of n. Note that this is in contrast with the previous section, where L was independent of n.

With Lemma 2.6.5 in hand, the proof of Theorem 2.6.1 is immediate.

Proof of Theorem 2.6.1. Suppose that G does not contain F^{2k-1} as a subgraph. Since $\delta = \omega(n^{\frac{1}{2k}})$, we may choose L with $L = \omega(1)$, $L^{100^k |V(F)||E(F)|^2 (k+1)} \le \delta$ and $n^2 f(2k, L) = o(n \delta^{2k})$. Then we may apply Lemma 2.6.5 with $S = V(G)$ to get that the number of L-good paths of length $2k$ in G is $\Omega(n \delta^{2k})$, which is $\omega(n^2 f(2k, L))$. However, by the definition of L-goodness, between any two vertices there can be at most $f(2k, L)$ such paths, which is a contradiction. $\square$

The proof of Theorem 2.6.2 is slightly more complicated.

Proof of Theorem 2.6.2. Firstly note that F is a subgraph of K_t for some t, so it suffices to prove the result for $F = K_t$. Let $\varepsilon > 0$ be sufficiently small, to be specified, and let G be a K-almost-regular graph on n vertices with minimum degree $\delta = \omega(n^{\frac{1}{2k}-\varepsilon})$. Assume that G does not contain a copy of $H = F^{2k-1}$.

For vertices $u, v \in V(G)$, let us write $u \sim v$ if there is a path of length $2k$ between u and v. Also, let us say that u and v are *distant* if for every $1 \leq i \leq 4k-2$, the number of walks of length i between u and v is at most $\delta^{i-2k+1/2}$. Observe that for any $u \in V(G)$ the number of walks of length i starting from u is at most $(K\delta)^i$, so the number of vertices $v \in V(G)$ for which there are at least $\delta^{i-2k+1/2}$ walks of length i from u to v is at most $\frac{(K\delta)^i}{\delta^{i-2k+1/2}} = K^i\delta^{2k-1/2}$. Thus, the number of $v \in V(G)$ for which u and v are not distant is $O(\delta^{2k-1/2})$.

Define $c_0 = \varepsilon$ and $c_{\ell+1} = (3 \cdot 5^{2k}+1)c_\ell + 2k\varepsilon$ for $0 \leq \ell \leq t-1$. Assume that ε is small enough so that

$$3 \cdot 100^k |V(F)||E(F)|^2(k+1) \cdot c_\ell \leq \frac{1}{2k} - \varepsilon \tag{2.4}$$

for all $0 \leq \ell \leq t$. Then in particular $c_\ell \leq \frac{1}{4k} - \varepsilon/2$ holds for all $0 \leq \ell \leq t$. For future reference, note that then

$$n^{c_\ell} \leq n^{\frac{1}{4k}-\varepsilon/2} = o(\delta^{1/2}). \tag{2.5}$$

Claim. For any $0 \leq \ell \leq t$, there exist distinct vertices $x_1, \dots, x_\ell \in V(G)$ and a set $S_\ell \subset V(G)$ such that

(i) there is a copy of K_ℓ^{2k-1} in G with the vertices of the subdivided K_ℓ being $x_1, \dots, x_\ell$

(ii) $x_i \sim y$ for every $1 \leq i \leq \ell$ and every $y \in S_\ell$

(iii) $|S_\ell| = \Omega(n^{1-c_\ell})$ and

(iv) x_i and x_j are distant for every $1 \leq i < j \leq \ell$.

Note that in particular for $\ell = t$, condition (i) guarantees the existence of a subgraph K_t^{2k-1}, so it suffices to prove the claim.

Proof of Claim. We proceed by induction on ℓ. For $\ell = 0$, we may take $S_0 = V(G)$. Assume now that we have verified the claim for ℓ.

Suppose that for some $y \in S_\ell$ there exist $1 \leq i < j \leq \ell$ and two paths of length $2k$, one (called P_i) from x_i to y and one (called P_j) from x_j to y, which share a vertex other than y. Let they intersect at some vertex $z \neq y$. Now let the subpath of P_i between x_i and z have length α and let the subpath of P_j between x_j and z have length β. Then there is a walk of length $\alpha + \beta$ from x_i to x_j through z. Moreover, there is a path of length $2k - \alpha$ from z to y. Observe that $2k - \alpha \leq 4k - (\alpha + \beta) - 1$.

Let Y be the set of $y \in S_\ell$ for which there exist some $1 \leq i < j \leq \ell$ and a walk W of length $\gamma \leq 4k - 2$ between x_i and x_j such that for some vertex w on W the distance of y from w is at most $4k - \gamma - 1$. By condition (iv), there are at most $\delta^{\gamma - 2k + 1/2}$ walks of length γ between any x_i and x_j so there are $O(\delta^{\gamma - 2k + 1/2})$ vertices appearing in at least one of these walks. Therefore the number of vertices at distance at most $4k - \gamma - 1$ from at least one of these vertices is $O(\delta^{\gamma - 2k + 1/2} \cdot \delta^{4k - \gamma - 1}) = O(\delta^{2k - 1/2})$. That is, $|Y| = O(\delta^{2k - 1/2})$.

Notice that by the discussion above, for any $y \in S_\ell \setminus Y$ and any $i \neq j$, a path of length $2k$ from x_i to y, and a path of length $2k$ from x_j to y have no common vertex other than y. Thus, by condition (ii) there exist ℓ paths of length $2k$, one from each x_i to y which are pairwise vertex-disjoint apart from at y. Moreover, these paths are also vertex-disjoint from the paths forming the K_ℓ^{2k-1} guaranteed by condition (i), apart from the trivial intersections at $x_1, \dots, x_\ell$ (else, there is a path of length at most $2k - 1$ from y to a point on a path of length $2k$ between some x_i and x_j, which contradicts the fact that $y \notin Y$). Thus, for any $y \in S_\ell \setminus Y$ there is a copy of $K_{\ell+1}^{2k-1}$ in G with the vertices of the subdivided $K_{\ell+1}$ being $x_1, \dots, x_\ell, y$.

Let Z be the set of $z \in S_\ell$ which are not distant to x_i for at least one $1 \leq i \leq \ell$. By the second paragraph in this proof, $|Z| = O(\delta^{2k-1/2})$.

Let $S'_\ell = S_\ell \setminus (Y \cup Z)$. Recall that $|Y| = O(\delta^{2k-1/2})$. Note that if $\delta = \omega(n^{\frac{1}{2k}})$, then, by Theorem 2.6.1, G contains H as a subgraph, so we may assume that $\delta = O(n^{\frac{1}{2k}})$. Then $\delta^{2k-1/2} = O(\frac{n}{\delta^{1/2}})$, which is $o(n^{1-c_\ell})$ by equation (2.5). Thus, $|Y \cup Z| = o(n^{1-c_\ell})$ and so $|S'_\ell| = \Omega(n^{1-c_\ell})$.

Let $L = n^{3c_\ell}$. Then, by equation (2.4), we have $L^{100^k |V(F)||E(F)|^2 (k+1)} \leq n^{\frac{1}{2k} - \varepsilon} = o(\delta)$. Moreover, by equation (2.5), we have $n^{1-c_\ell} = \omega(\frac{n}{\delta^{1/2}})$, and by the definition of L, we have $n^{1-c_\ell} = \omega(\frac{n}{L^{1/2}})$. Hence, by Lemma 2.6.5, the number of L-good paths of length $2k$ with both endpoints in S'_ℓ is $\Omega(\frac{|S'_\ell|^2 \delta^{2k}}{n})$. Between any two vertices in S'_ℓ there are at most $f(2k, L)$ L-good paths of length $2k$, so the number of pairs $(z, y) \in S'_\ell \times S'_\ell$ with $z \sim y$ is $\Omega(\frac{|S'_\ell|^2 \delta^{2k}}{n f(2k, L)})$. Thus, there exists some $x_{\ell+1} \in S'_\ell$ such that the number of $y \in S'_\ell$ with $x_{\ell+1} \sim y$ is $\Omega(\frac{|S'_\ell| \delta^{2k}}{n f(2k, L)}) \geq \Omega(\frac{n^{1-c_\ell-2k\varepsilon}}{L^{5^{2k}}}) = \Omega(n^{1-c_\ell-2k\varepsilon-3c_\ell 5^{2k}}) = \Omega(n^{1-c_{\ell+1}})$. Set $S_{\ell+1}$ to be the set of these $y \in S'_\ell$, and note that properties (i)-(iv) are satisfied for $\ell + 1$. $\square$

2.6.2 Short paths

Our aim in this subsection is to prove the following lemma.

Lemma 2.6.6. *Let G be an F^{2k-1}-free K-almost-regular graph on n vertices with minimum degree $\delta \geq L^{100^k |V(F)||E(F)|^2 (k+1)}$. Then, provided that L is sufficiently large compared to $|V(F)|$, $|E(F)|$, k and K, the number of paths of length k that are not good is $O(\frac{n\delta^k}{L})$.*

Observe that if $s = |V(F)|$ and $t = |E(F)|$, then $H = F^{2k-1}$ is a subgraph of $K_{s,t}^{k-1}$. Hence, Lemma 2.6.6 will follow from the following result.

Lemma 2.6.7. *Let s and t be positive integers and let G be a $K_{s,t}^{k-1}$-free K-almost-regular graph on n vertices with minimum degree $\delta \geq L^{100^k st^2(k+1)}$. Then, provided that L is sufficiently large compared to s, t, k and K, the number of paths of length k that are not good is $O(\frac{n\delta^k}{L})$.*

The next definition is for notational convenience.

Definition 2.6.8. A pair of distinct vertices (x, y) in G is said to be (ℓ, L)-*bad* for some $2 \leq \ell \leq 2k$ and some L if there is an L-admissible, but not L-good, path of length ℓ from x to y.

In what follows, for $v \in V(G)$, we shall write $\Gamma_i(v)$ for the set of vertices $u \in V(G)$ for which there exists a path of length i from v to u and write $N(v) = \Gamma_1(v)$. The next lemma will be used to show that if s and t are fixed, then in a $K_{s,t}^{k-1}$-free graph there cannot be many bad pairs between $N(v) = \Gamma_1(v)$ and $\Gamma_{\ell-1}(v)$. We will take a suitable $X \subset N(v)$, $Y = \Gamma_{\ell-1}(v)$ and repeatedly apply the lemma to obtain $(i-1)$-subdivided t-stars. At the end, we piece these together to form a copy of $K_{s,t}^{k-1}$. To make sure that this is nondegenerate, the set Z of vertices that we have already used will be avoided.

Lemma 2.6.9. *Let $t \geq 1$, $2 \leq \ell \leq k$ and $1 \leq i \leq \ell$ be integers. Let G be a K-almost-regular graph on n vertices with minimum degree $\delta > 0$. Let $X, Y, Z \subset V(G)$ be such that $|Z| \leq L^{1/10}, |Y| \leq (K\delta)^{\ell-1}$ and, for any $x \in X$, the number of $y \in Y$ such that (x, y) is (ℓ, L)-bad is as at least $\frac{(K\delta)^{\ell-1}}{f(\ell-1,L)^2}$. Then, provided that L is sufficiently large compared to t, k and K, there exist an $(i-1)$-subdivided t-star in G, disjoint from Z, whose endpoints form a set $R \subset Y$, and a subset $X' \subset X$ such that $|X'| \geq |X \setminus Z|/(4f(\ell-1, L)^2)^t$ and (x', r) is (ℓ, L)-bad for every $x' \in X'$ and $r \in R$.*

Proof. After replacing X by $X \setminus Z$, we may assume $X \cap Z = \emptyset$. Let Y' be the set of those $y \in Y$ for which the number of $x \in X$ such that (x, y) is (ℓ, L)-bad is at least $\frac{|X|}{2f(\ell-1,L)^2}$. Then the number of $(x, y) \in X \times (Y \setminus Y')$ which are (ℓ, L)-bad is at most $\frac{|X||Y|}{2f(\ell-1,L)^2} \leq \frac{|X|(K\delta)^{\ell-1}}{2f(\ell-1,L)^2}$, so the number of $(x, y) \in X \times Y'$ which are (ℓ, L)-bad is at least $\frac{|X|(K\delta)^{\ell-1}}{2f(\ell-1,L)^2}$. Now there exists some $x^* \in X$ such that there are at least $\frac{(K\delta)^{\ell-1}}{2f(\ell-1,L)^2}$ choices $y \in Y'$ for which (x^*, y) is (ℓ, L)-bad. If a pair (x^*, y) is (ℓ, L)-bad, then there are at least $f(\ell, L)$ paths of length ℓ from x^* to y. Hence, there are at least $\frac{(K\delta)^{\ell-1}}{2f(\ell-1,L)^2} \cdot f(\ell, L) = \Omega(f(\ell-1, L)^3\delta^{\ell-1})$ paths of length ℓ starting at x^* and ending in Y'.

The number of such paths intersecting Z is at most $|Z|\ell(K\delta)^{\ell-1}$. Indeed, there are at most $|Z|$ choices for the element of Z in the path, at most ℓ choices for its position in the path and, given a fixed choice for these, at most $(K\delta)^{\ell-1}$ choices for the other $\ell-1$ vertices in the path. (Note that as $X \cap Z = \emptyset$, the vertex in Z is not x^*.) But $|Z|\ell(K\delta)^{\ell-1} \leq L^{1/10}\ell K^{\ell-1}\delta^{\ell-1}$, so, for L sufficiently large there are $\Omega(f(\ell-1, L)^3\delta^{\ell-1})$ paths of length ℓ starting at x^* and ending in Y' that avoid Z. Moreover, there are at most $(K\delta)^{\ell-i}$ different initial segments of length $\ell-i$ for these paths, so, by the pigeonhole principle, there exist $\Omega(f(\ell-1,L)^3\delta^{i-1})$ of them which start with the same $\ell-i$ edges. It follows that there exists some $u \in \Gamma_{\ell-i}(x^*)$ such that there are $\Omega(f(\ell-1,L)^3\delta^{i-1})$ paths of length i from u to Y', all avoiding Z.

Take now a maximal set of such paths which are pairwise vertex-disjoint apart from at u. We claim that there are $\Omega(f(\ell-1,L)^3)$ such paths. Suppose otherwise. Then all the $\Omega(f(\ell-1,L)^3\delta^{i-1})$ paths of length i from u to Y' intersect a certain set of size $o(f(\ell-1,L)^3)$ not containing u. But there are $o(f(\ell-1,L)^3)\delta^{i-1}$ such paths, which is a contradiction.

So we have $r = \Omega(f(\ell-1,L)^3)$ paths $P_1,\dots,P_r$ of length i from u to Y' which are pairwise vertex-disjoint except at u and avoid Z. Let the endpoints of these paths be $y_1,\dots,y_r$. Since $y_j \in Y'$ for all j, the number of pairs (x,y_j) with $x \in X$ which are (ℓ,L)-bad is at least $\frac{r|X|}{2f(\ell-1,L)^2}$. Therefore, by Jensen's inequality, there are at least $|X|\cdot\binom{r/(2f(\ell-1,L)^2)}{t}$ choices $x \in X$, $1 \le j_1 < j_2 < \cdots < j_t \le r$ such that $(x,y_{j_1}),\dots,(x,y_{j_t})$ are all (ℓ,L)-bad. Since $\binom{r/(2f(\ell-1,L)^2)}{t} \ge (\frac{1}{4f(\ell-1,L)^2})^t\binom{r}{t}$, there exist $1 \le j_1 < \cdots < j_t \le r$ such that the set

$$X' = \{x \in X : (x,y_{j_1}),\dots,(x,y_{j_t}) \text{ are all } (\ell,L)\text{-bad}\}$$

has size at least $|X|/(4f(\ell-1,L)^2)^t$. We can now take $R = \{y_{j_1},\dots,y_{j_t}\}$, and the union of the paths $P_{j_1},\dots,P_{j_t}$ is a suitable $(i-1)$-subdivided t-star. □

We now iterate Lemma 2.6.9, as promised, to find a copy of $K_{s,t}^{k-1}$.

Lemma 2.6.10. *Let s and t be positive integers and let G be an $K_{s,t}^{k-1}$-free K-almost-regular graph on n vertices with minimum degree $\delta \ge L^{100^k st^2(k+1)}$. Let $2 \le \ell \le k$ and $v \in V(G)$. Then, provided that L is sufficiently large compared to s, t, k and K, the number of L-admissible, but not L-good, paths of the form $v_0vv_2v_3\dots v_\ell$ is at most $\frac{2(K\delta)^\ell}{f(\ell-1,L)}$.*

Proof. Suppose otherwise. Let $Y = \Gamma_{\ell-1}(v)$ and note that $|Y| \le (K\delta)^{\ell-1}$. For any $x \in N(v)$ and any $y \in Y$, the number of L-admissible paths of the form $xvv_2\dots v_{\ell-1}y$ is at most $f(\ell-1,L)$. Indeed, in any such path, the subpath $vv_2v_3\dots v_{\ell-1}y$ is L-good, and for any fixed $y \in Y$ there are at most $f(\ell-1,L)$ such L-good paths. Hence, by assumption, the number of pairs $(x,y) \in N(v)\times Y$ such that there is an L-admissible, but not L-good, path of the form $xvv_2\dots v_{\ell-1}y$ is at least $\frac{2(K\delta)^\ell}{f(\ell-1,L)^2} \ge \frac{2|N(v)|(K\delta)^{\ell-1}}{f(\ell-1,L)^2}$. By definition, any such pair (x,y) is (ℓ,L)-bad. Let X consist of those $x \in N(v)$ for which there are at least $\frac{(K\delta)^{\ell-1}}{f(\ell-1,L)^2}$ choices of $y \in Y$ such that (x,y) is (ℓ,L)-bad. Then the number of pairs $(x,y) \in X \times Y$ which are (ℓ,L)-bad is at least $\frac{|N(v)|(K\delta)^{\ell-1}}{f(\ell-1,L)^2}$, and so $|X| \ge \frac{|N(v)|}{f(\ell-1,L)^2} \ge \frac{\delta}{f(\ell-1,L)^2}$.

Our aim now is to find a copy of $K_{s,t}^{k-1}$ in G, which will yield a contradiction. Consider first the case $\ell = k$. By Lemma 2.6.9 with $Z = \emptyset$, there exists a set $X' \subset X$ of size at least $|X|/(4f(\ell-1,L)^2)^t$ and a set $R_1 \subset Y$ of size t such that (x,y) is (ℓ,L)-bad for any

$x \in X'$ and $y \in R_1$. Note that this uses Lemma 2.6.9 in a rather weak sense since we do not need the subdivided star provided by the lemma, only its leaves. Now applying Lemma 2.6.9 with $Z = R_1$ and with X' in place of X, we find a set $X'' \subset X'$ of size at least $|X' \setminus R_1|/(4f(\ell-1,L)^2)^t$ and a set $R_2 \subset Y$ of size t, disjoint from R_1 such that (x,y) is (ℓ,L)-bad for any $x \in X''$ and $y \in R_2$. Repeat this procedure. Note that for L sufficiently large we have

$$\begin{aligned} |X| &\geq \frac{\delta}{f(\ell-1,L)^2} \geq \frac{L^{100^k st^2(k+1)}}{f(\ell-1,L)^2} \geq \frac{f(\ell-1,L)^{20st^2(k+1)}}{f(\ell-1,L)^2} \\ &\geq 2L(4f(\ell-1,L)^2)^{st^2(k+1)+st}, \end{aligned} \tag{2.6}$$

so we may apply Lemma 2.6.9 $\lceil \frac{s}{t} \rceil$ times (or even $st(k+1)+s$ times) as above to find a set $X_{\text{final}} \subset X$ of size at least t and a set $U = R_1 \cup R_2 \cup \cdots \cup R_{\lceil s/t \rceil} \subset Y$ with $|U| \geq s$ such that X_{final} and U are disjoint and (x,y) is (ℓ,L)-bad for any $x \in X_{\text{final}}$ and $y \in U$. Choose distinct vertices $x_1,\dots,x_t \in X_{\text{final}}$ and $y_1,\dots,y_s \in U$. Since (x_i,y_j) is (ℓ,L)-bad for every i,j, if L is sufficiently large, we can find pairwise internally vertex-disjoint paths of length $\ell = k$ joining x_i to y_j for every i,j. The union of these paths forms a copy of $K_{s,t}^{k-1}$.

Now assume that $\ell < k$. Write $k = j\ell + i$ with $1 \leq i \leq \ell$. Note that $i < k$. Assume first that j is odd. Using equation (2.6), we may apply Lemma 2.6.9 $st(k+1)+s$ times to find a set $X_{\text{final}} \subset X$ of size at least $st(k+1)$, $(i-1)$-subdivided t-stars $T_1,\dots,T_s$ with leaf sets $Y_1,\dots,Y_s \subset Y$ and a set $U \subset Y$ with $|U| \geq st(k+1)$ such that the sets $X_{\text{final}}, V(T_1),\dots,V(T_s),U$ are pairwise disjoint and (x,y) is (ℓ,L)-bad for any $x \in X_{\text{final}}$ and $y \in Y_1 \cup \cdots \cup Y_s \cup U$.

Label the vertices of $K_{s,t}^{k-1}$ as follows. Let the vertices in the part of size s be $u_1,\dots,u_s$, let the vertices in the part of size t be $v_1,\dots,v_t$ and, for each $1 \leq a \leq s$ and $1 \leq b \leq t$, let the path of length k connecting u_a and v_b be $u_a w_{a,b,1} w_{a,b,2} \dots w_{a,b,k-1} v_b$. We now embed $K_{s,t}^{k-1}$ in G as follows. For each $1 \leq a \leq s$, the $(i-1)$-subdivided t-star T_a will take the role of the $(i-1)$-subdivided t-star in $K_{s,t}^{k-1}$ with vertices u_a, $w_{a,1,1}$, $w_{a,1,2}$, $\dots$, $w_{a,1,i}$, $w_{a,2,1}$, $w_{a,2,2}$, $\dots$, $w_{a,2,i}$, $\dots$, $w_{a,t,1}$, $w_{a,t,2}$, $\dots$, $w_{a,t,i}$. Furthermore, the roles of $w_{a,b,i+q\ell}$ for q odd $(1 \leq a \leq s, 1 \leq b \leq t, q \geq 1)$ will be taken by vertices in X_{final} in an arbitrary injective manner and the roles of $w_{a,b,i+q\ell}$ for q even $(1 \leq a \leq s, 1 \leq b \leq t, q \geq 2)$ will be taken by vertices in U in an arbitrary injective manner. Finally, let $v_1,\dots,v_t$ be mapped to X_{final} in an injective manner avoiding all previous vertices. See Figure 2.2, which illustrates the embedding in the case $s=2$, $t=3$, $k=7$, $\ell=2$. It remains to define the vertices that correspond to $w_{a,b,c}$ with $1 \leq a \leq s, 1 \leq b \leq t, i < c \leq k-1$ and $c-i$ not divisible by ℓ. But, since the images of $w_{a,b,i+q\ell}, w_{a,b,i+(q+1)\ell}$ $(1 \leq a \leq s, 1 \leq b \leq t, 0 \leq q \leq j-1$, where $w_{a,b,k} = v_b)$ are such that one is in X_{final} and the other is in $R_1 \cup \cdots \cup R_s \cup U$, the image of any pair $(w_{a,b,i+q\ell}, w_{a,b,i+(q+1)\ell})$ is (ℓ,L)-bad. Therefore, by Lemma 2.6.4, provided that L is sufficiently large, we may join these pairs by paths of length ℓ, all internally disjoint from each other and from the previous vertices, yielding a copy of $K_{s,t}^{k-1}$.

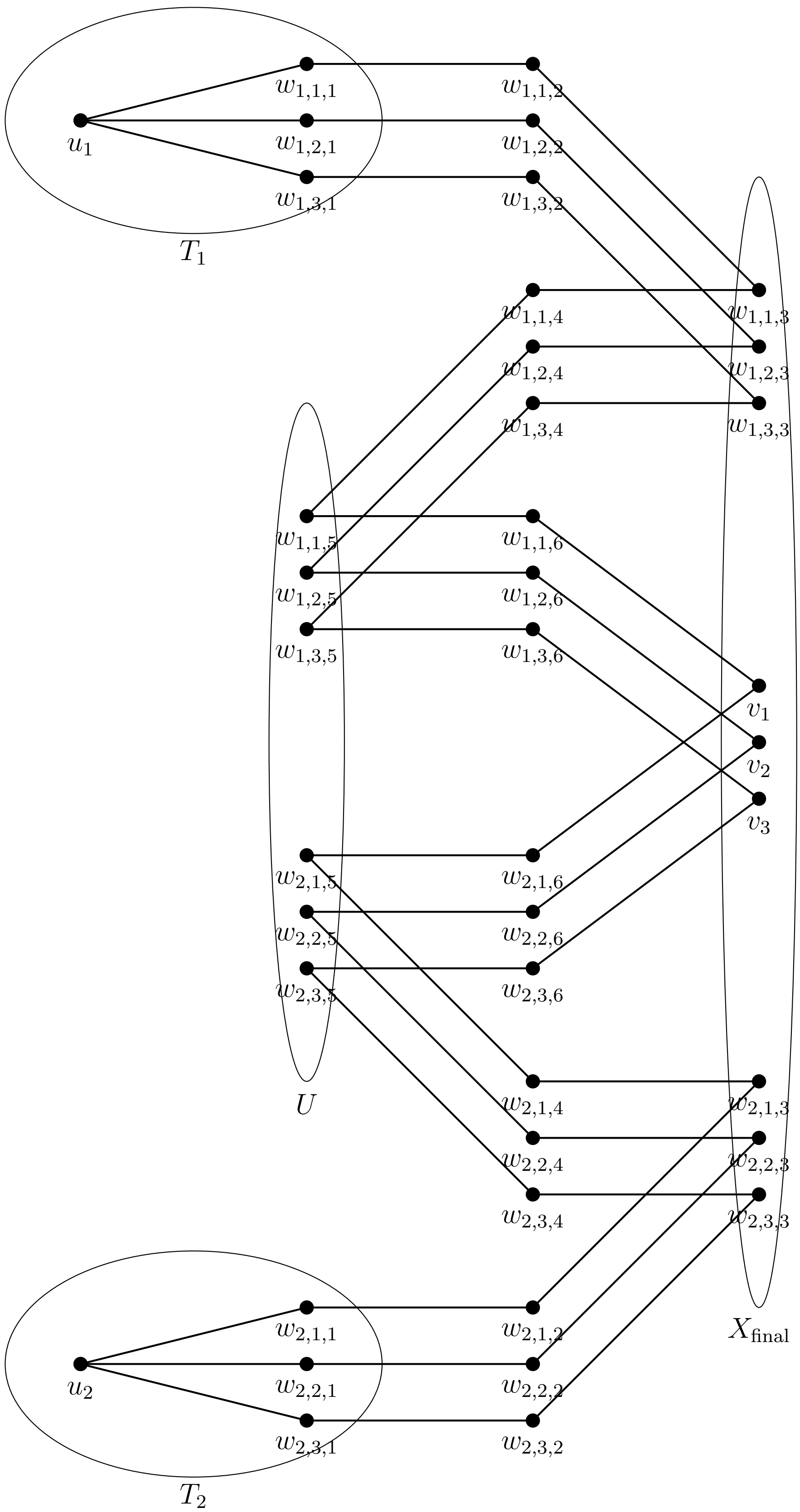

Figure 2.2: The embedding of $K_{s,t}^{k-1}$ in the case $s = 2$, $t = 3$, $k = 7$, $\ell = 2$.

The case where j is even is very similar. The only difference is that the vertices $v_1, \dots, v_t$ are mapped to U. $\square$

Corollary 2.6.11. *Let s and t be positive integers and let G be a $K_{s,t}^{k-1}$-free K-almost-regular graph on n vertices with minimum degree $\delta \geq L^{100^k st^2(k+1)}$. Then, provided that L is sufficiently large compared to s, t, k and K, for any $2 \leq \ell \leq k$, the number of L-admissible, but not L-good, paths of length ℓ is at most $n\frac{2(K\delta)^\ell}{f(\ell-1,L)}$.*

Now we are in a position to prove Lemma 2.6.7.

Proof of Lemma 2.6.7. Suppose that the path $u_0u_1 \dots u_k$ is not L-good. Take $0 \leq i < j \leq k$ with $j - i$ minimal such that $u_iu_{i+1} \dots u_j$ is not L-good. Then $u_i \dots u_j$ is L-admissible. For any fixed i, j, by Corollary 2.6.11, the number of such paths is at most $n\frac{2(K\delta)^{j-i}}{f(j-i-1,L)} \cdot 2(K\delta)^{k-(j-i)} = 4K^k\frac{n\delta^k}{f(j-i-1,L)} \leq 4K^k\frac{n\delta^k}{L}$. Using that i and j can take at most $k+1$ values each, it follows that the number of not L-good paths of length k is at most $(k+1)^2 4K^k \cdot \frac{n\delta^k}{L}$. $\square$

2.6.3 Long paths

In what follows, for a vertex $x \in V(G)$ and a nonnegative integer i, we write $\mathcal{P}_i(x)$ for the set of directed paths of length i starting at x. For an element $P \in \mathcal{P}_i(x)$, we let $v(P)$ be the endpoint of the path P. In the next definition the notion of richness also depends on the value of δ, but we do not emphasise this.

Definition 2.6.12. Let i, j be nonnegative integers with $i + j < 2k$. Call a pair (x, y) of vertices (i, j)-*rich* if $x \neq y$ and the number of pairs $(P, Q) \in \mathcal{P}_i(x) \times \mathcal{P}_j(y)$ such that there are at least $(|V(H)| + 2)(2k + 1) + 1$ pairwise internally vertex-disjoint paths of length $2k-i-j$ between $v(P)$ and $v(Q)$ is more than $(2(i+j)|V(H)|(2k+1)+2(i+1)j)(K\delta)^{i+j-1}$. Otherwise (including when $x = y$) call it (i, j)-*poor*.

Lemma 2.6.13. *Let G be a graph with maximum degree at most $K\delta$. Let $x, y \in V(G)$ and let i, j be nonnegative integers with $i + j < 2k$. If (x, y) is (i, j)-rich, then there exist $|V(H)|$ pairwise internally vertex-disjoint paths of length $2k$ between x and y.*

Proof. Choose a maximal set of pairwise internally vertex-disjoint paths $R_1, \dots, R_\alpha$ between x and y and assume that $\alpha < |V(H)|$. Let T be the set of the vertices appearing in at least one of these paths. Note that $|T| < |V(H)|(2k + 1)$.

Claim. If there is a pair $(P, Q) \in \mathcal{P}_i(x) \times \mathcal{P}_j(y)$ such that

(i) P is disjoint from $T \setminus \{x\}$,

(ii) Q is disjoint from $T \setminus \{y\}$,

(iii) P and Q are vertex-disjoint and

(iv) there are at least $(|V(H)|+2)(2k+1)+1$ pairwise internally vertex-disjoint paths of length $2k-i-j$ between $v(P)$ and $v(Q)$,

then there is a path of length $2k$ between x and y which is internally vertex-disjoint from all of $R_1, \dots, R_\alpha$.

Proof of Claim. Clearly, it suffices to find a path of length $2k-i-j$ between $v(P)$ and $v(Q)$ which is disjoint from the vertices of $R_1, \dots, R_\alpha, P, Q$, except for $v(P)$ and $v(Q)$. But such a path exists since there are at most $(\alpha+2)\cdot(2k+1) \le (|V(H)|+2)(2k+1)$ vertices in one of $R_1, \dots, R_\alpha, P, Q$ and there are at least $(|V(H)|+2)(2k+1)+1$ pairwise internally vertex-disjoint paths of length $2k-i-j$ between $v(P)$ and $v(Q)$.

A path provided by the claim would contradict the maximality of $R_1, \dots, R_\alpha$, so it suffices to prove that there are paths P, Q satisfying (i)-(iv) above.

Since the maximum degree of G is at most $K\delta$, the number of paths of length $i-1$ in G intersecting T is at most $i|T|(K\delta)^{i-1}$, so the number of $P \in \mathcal{P}_i(x)$ which have a vertex in $T \setminus \{x\}$ is at most $2i|T|(K\delta)^{i-1}$. Since $|\mathcal{P}_j(y)| \le (K\delta)^j$, the number of pairs $(P,Q) \in \mathcal{P}_i(x) \times \mathcal{P}_j(y)$ failing condition (i) above is at most $2i|T|(K\delta)^{i-1}(K\delta)^j$. Similarly, the number of pairs failing (ii) is at most $2j|T|(K\delta)^{j-1}(K\delta)^i$. Finally, for every $P \in \mathcal{P}_i(x)$, the number of paths of length $j-1$ which intersect P is at most $(i+1)j(K\delta)^{j-1}$, so the number of pairs $(P,Q) \in \mathcal{P}_i(x) \times \mathcal{P}_j(y)$ for which P and Q share a vertex other than y is at most $(K\delta)^i \cdot 2(i+1)j(K\delta)^{j-1}$. So the number of pairs which fail at least one of (i),(ii),(iii) is at most $(2(i+j)|T|+2(i+1)j)(K\delta)^{i+j-1} \le (2(i+j)|V(H)|(2k+1)+2(i+1)j)(K\delta)^{i+j-1}$. By the definition of (i,j)-richness of (x,y) it follows that there is a pair (P,Q) satisfying (i)-(iv). $\square$

Definition 2.6.14. For a vertex $v \in V(G)$ and some $1 \le \ell \le k$, define an auxiliary graph $\mathcal{G}_\ell(v)$ as follows. The vertices of $\mathcal{G}_\ell(v)$ are the $(k+1)$-tuples $(u_0, u_1, \dots, u_k) \in V(G)^{k+1}$ with $u_0 = v$ such that $u_i u_{i+1} \in E(G)$ for all i. Vertices $(u_0, \dots, u_k)$ and $(u'_0, \dots, u'_k)$ are joined by an edge if $v, u_1, u_2, \dots, u_k, u'_1, \dots, u'_k$ are distinct and there exist $0 \le i,j \le k-1$ such that the pair (u_ℓ, u'_ℓ) is (i,j)-rich. Since the vertex set of $\mathcal{G}_\ell(v)$ does not depend on ℓ, we may define $\mathcal{G}(v)$ to be the union $\bigcup_{1 \le \ell \le k} \mathcal{G}_\ell(v)$.

Lemma 2.6.15. *Let G be a graph with maximum degree at most $K\delta$ which does not contain F^{2k-1} as a subgraph. Let $t = |V(F)|$. Then for any $v \in V(G)$ and any $1 \le \ell \le k$, the graph $\mathcal{G}_\ell(v)$ is K_t-free.*

Moreover, let $r = R_k(t)$ be the k-colour Ramsey number. Then $\mathcal{G}(v)$ is K_r-free.

Proof. Suppose that $\mathcal{G}_\ell(v)$ contains K_t as a subgraph. Let the corresponding vertices be the vectors $u^1, \dots, u^t$. Let their respective $(\ell+1)$th coordinate be $u^1_\ell, \dots, u^t_\ell$. For every $a \ne b$, since $u^a u^b$ is an edge in $\mathcal{G}_\ell(v)$, it follows that u^a_ℓ and u^b_ℓ are distinct, and, by Lemma 2.6.13, there exist $|V(H)|$ pairwise internally vertex-disjoint paths of length $2k$ between them. It is not hard to see that this implies that there is a copy of $H = F^{2k-1}$ in G

in which the vertices of F are mapped to $u_\ell^1, \dots, u_\ell^t$. This is a contradiction, so $\mathcal{G}_\ell(v)$ is indeed K_t-free.

Suppose there is a copy of K_r in $\mathcal{G}(v)$. Then each edge in this K_r can be coloured with one of the colours $1, 2, \dots, k$ such that if an edge gets colour i, then it lies in $\mathcal{G}_i(v)$. By the definition of r, there exists a monochromatic K_t in this k-edge-coloured K_r, which gives a K_t in some $\mathcal{G}_\ell(v)$, contradicting the first paragraph. $\square$

The next lemma provides us a large set of walks of length $2k$ with both endpoints in S. Later, we will argue that most of them are L-good paths.

Lemma 2.6.16. *Let $r = R_k(t)$ denote the k-colour Ramsey number where $t = |V(F)|$. Let G be an F^{2k-1}-free K-almost-regular graph on n vertices with minimum degree δ and let $S \subset V(G)$ such that $|S| \geq 2nr/\delta^k$. Then there are at least $\frac{|S|^2\delta^{2k}}{4r^2n}$ vectors $(u_{-k}, \dots, u_k) \in V(G)^{2k+1}$ with the following properties*

(i) $u_{-k} \in S$, $u_k \in S$

(ii) $u_\ell u_{\ell+1} \in E(G)$ for every $-k \leq \ell \leq k-1$

(iii) $(u_{-\ell}, u_\ell)$ is (i,j)-poor for every $1 \leq \ell \leq k$ and every $0 \leq i, j \leq k-1$.

Proof. Since the minimum degree of G is δ, the number of $(k+1)$-tuples $(v_0, v_1, \dots, v_k) \in V(G)^{k+1}$ with $v_k \in S$ and $v_i v_{i+1} \in E(G)$ for every $0 \leq i \leq k-1$ is at least $|S|\delta^k$. Writing $\mathcal{T}(v_0)$ for the set of such vectors for a fixed v_0 and letting $g(v_0) = |\mathcal{T}(v_0)|$, we get that $\sum_{v_0 \in V(G)} g(v_0) \geq |S|\delta^k$. Note that $\sum_{v_0 \in V(G): g(v_0) < r} g(v_0) \leq nr \leq \frac{|S|\delta^k}{2}$, so

$$\sum_{v_0 \in V(G): g(v_0) \geq r} g(v_0) \geq \frac{|S|\delta^k}{2}. \tag{2.7}$$

Note that $\mathcal{T}(v_0) \subset V(\mathcal{G}(v_0))$. By Lemma 2.6.15, the graph $\mathcal{G}(v_0)[\mathcal{T}(v_0)]$ is K_r-free. This graph has $g(v_0)$ vertices, so if $g(v_0) \geq r$, then the number of non-edges in $\mathcal{G}(v_0)[\mathcal{T}(v_0)]$ is at least $\frac{1}{\binom{r}{2}}\binom{g(v_0)}{2} \geq \frac{g(v_0)^2}{r^2}$. But if $v = (v_0, v_1, \dots, v_k) \in \mathcal{T}(v_0)$ and $v' = (v_0, v_1', \dots, v_k') \in \mathcal{T}(v_0)$ are such that vv' is not an edge in $\mathcal{G}(v_0)$, then $(u_{-k}, \dots, u_k) = (v_k', v_{k-1}', \dots, v_1', v_0, v_1, \dots, v_k)$ satisfies all three properties in the statement of the lemma. Therefore the number of such $(2k+1)$-tuples with $u_0 = v_0$ is at least $\frac{g(v_0)^2}{r^2}$ provided that $g(v_0) \geq r$. By (2.7) and Jensen's inequality, we get $\sum_{v_0 \in V(G): g(v_0) \geq r} \frac{g(v_0)^2}{r^2} \geq \frac{|S|^2\delta^{2k}}{4r^2n}$, and the proof is complete. $\square$

The following simple lemma shows that most walks of length $2k$ are paths.

Lemma 2.6.17. *Let G be a graph on n vertices with maximum degree at most $K\delta$. Then the number of $(2k+1)$-tuples $(u_{-k}, \dots, u_k) \in V(G)^{2k+1}$ such that $u_i u_{i+1} \in E(G)$ for every i and $u_i = u_j$ for some $i \neq j$ is at most $\binom{2k+1}{2}K^{2k-1} \cdot n\delta^{2k-1}$.*

Proof. There are $\binom{2k+1}{2}$ ways to choose the pair $\{i, j\}$ and there are n ways to choose $u_i = u_j$. Given any such choices, there are at most $(K\delta)^{2k-1}$ ways to choose the vertices u_b for $b \notin \{i, j\}$ since any vertex in G has degree at most $K\delta$. $\square$

Our strategy now is to take all the paths guaranteed by Lemmas 2.6.16 and 2.6.17 and discard those which contain a subpath of length k which is not L-good. The next result shows that doing this we discard only a small proportion of the paths.

Lemma 2.6.18. *Let G be an F^{2k-1}-free K-almost-regular graph on n vertices with minimum degree $\delta \geq L^{100^k|V(F)||E(F)|^2(k+1)}$. Then, provided that L is sufficiently large compared to $|V(F)|$, $|E(F)|$, k and K, the number of paths $u_{-k}u_{-k+1}\dots u_k$ of length $2k$ in G with the property that there is some $-k \leq j \leq 0$ for which the path $u_ju_{j+1}\dots u_{j+k}$ is not L-good is $O(\frac{n\delta^{2k}}{L})$.*

Proof. By Lemma 2.6.6, there are $O(\frac{n\delta^k}{L})$ paths $u_ju_{j+1}\dots u_{j+k}$ which are not L-good, and since the maximum degree of G is at most $K\delta$, there are at most $2(K\delta)^k$ ways to extend such a path to a path $u_{-k}u_{-k+1}\dots u_k$ of length $2k$. The result follows after summing these terms for all $-k \leq j \leq 0$. $\square$

The next lemma relates the notion of L-goodness and the notion of (i, j)-richness.

Lemma 2.6.19. *Suppose that $u_{-k}u_{-k+1}\dots u_k$ is a path in G which is not L-good but each of its subpaths of length k is L-good. Then, provided that L is sufficiently large compared to $|V(F)|$, $|E(F)|$ and k, there exist $1 \leq \alpha, \beta \leq k$ with $\alpha + \beta > k$ such that there exist $(|V(H)| + 2)(2k + 1) + 1$ pairwise internally vertex-disjoint paths of length $\alpha + \beta$ between $u_{-\alpha}$ and u_β.*

Proof. Choose $-k \leq i < j \leq k$ with $j - i$ minimal such that $u_iu_{i+1}\dots u_j$ is not L-good. By the minimality of $j - i$, every proper subpath of $u_iu_{i+1}\dots u_j$ is L-good, so $u_iu_{i+1}\dots u_j$ is L-admissible. By Lemma 2.6.4 and our assumption about L, there exist $(|V(H)| + 2)(2k + 1) + 1$ pairwise internally vertex-disjoint paths of length $j - i$ between u_i and u_j.

By the assumption that every subpath of $u_{-k}u_{-k+1}\dots u_k$ of length k is L-good, we have $j - i > k$, so $i < 0$ and $j > 0$. Thus, the choices $\alpha = -i$ and $\beta = j$ satisfy the conditions described in the lemma. $\square$

The next result is the final ingredient to the proof of Lemma 2.6.5.

Lemma 2.6.20. *Let G be a graph on n vertices with maximum degree at most $K\delta$. Then there are $O(n\delta^{2k-1})$ paths $u_{-k}u_{-k+1}\dots u_k$ in G with the following two properties*

(i) $(u_{-\ell}, u_\ell)$ *is (i, j)-poor for every $1 \leq \ell \leq k$ and every $0 \leq i, j \leq k - 1$ and*

(ii) *there exist $1 \leq \alpha, \beta \leq k$ with $\alpha + \beta > k$ such that there exist $(|V(H)| + 2)(2k + 1) + 1$ pairwise internally vertex-disjoint paths of length $\alpha + \beta$ between $u_{-\alpha}$ and u_β.*

Proof. Fix a pair (α,β) with $1 \le \alpha,\beta \le k$ and $\alpha+\beta > k$. It suffices to prove that the number of paths satisfying (i) and (ii) for *this* pair (α,β) is $O(n\delta^{2k-1})$.

Let $\ell = \alpha+\beta-k$. Note that $1 \le \ell \le k$. Also, let $i = \alpha-\ell = k-\beta$ and $j = \beta-\ell = k-\alpha$. Observe that $0 \le i,j \le k-1$.

Suppose that $u_{-\ell}\dots u_\ell$ is a path such that $(u_{-\ell},u_\ell)$ is (i,j)-poor. By the definition of (i,j)-poorness, the number of pairs of paths $(u_{-\ell}u_{-\ell-1}\dots u_{-\alpha}, u_\ell u_{\ell+1}\dots u_\beta)$ such that there exist $(|V(H)|+2)(2k+1)+1$ pairwise internally vertex-disjoint paths of length $\alpha+\beta = 2k-i-j$ between $u_{-\alpha}$ and u_β is $O(\delta^{i+j-1})$. Thus, the number of ways to extend $u_{-\ell}u_{-\ell+1}\dots u_\ell$ to a path $u_{-k}u_{-k+1}\dots u_k$ possessing property (ii) with our fixed choice of α and β is $O(\delta^{i+j-1}\cdot (K\delta)^{k-\alpha+k-\beta}) = O(\delta^{2k-2\ell-1})$, where the first factor bounds the number of possible ways to extend to $u_{-\alpha}u_{-\alpha+1}\dots u_\beta$, and the second factor bounds the number of possible ways to extend that to $u_{-k}u_{-k+1}\dots u_k$. The number of possible choices for $u_{-\ell}u_{-\ell+1}\dots u_\ell$ is $O(n\delta^{2\ell})$, so the result follows. □

We are now in a position to complete the proof of Lemma 2.6.5.

Proof of Lemma 2.6.5. The condition $|S| = \omega(\frac{n}{\delta^{1/2}})$ implies that $n\delta^{2k-1} = o(\frac{|S|^2\delta^{2k}}{n})$, so by Lemmas 2.6.16 and 2.6.17, there are $\Omega(\frac{|S|^2\delta^{2k}}{n})$ paths $u_{-k}u_{-k+1}\dots u_k$ with both endpoints in S such that $(u_{-\ell},u_\ell)$ is (i,j)-poor for every $1 \le \ell \le k$ and every $0 \le i,j \le k-1$. Discard all those paths among these in which there is a subpath of length k which is not L-good. By Lemma 2.6.18, we discarded $O(\frac{n\delta^{2k}}{L})$ paths, which is $o(\frac{|S|^2\delta^{2k}}{n})$, by the condition $|S| = \omega(\frac{n}{L^{1/2}})$. Of the remaining paths, discard all those for which there exist $1 \le \alpha,\beta \le k$ with $\alpha+\beta > k$ such that there exist $(|V(H)|+2)(2k+1)+1$ pairwise internally vertex-disjoint paths of length $\alpha+\beta$ between $u_{-\alpha}$ and u_β. By Lemma 2.6.20, there are $O(n\delta^{2k-1})$ such paths, which is again $o(\frac{|S|^2\delta^{2k}}{n})$. Hence, we are left with $\Omega(\frac{|S|^2\delta^{2k}}{n})$ paths.

We claim that each such path is L-good. Suppose otherwise, and take a path $u_{-k}u_{-k+1}\dots u_k$ which is not L-good. Since each of its subpaths of length k is L-good, by Lemma 2.6.19 there exist $1 \le \alpha,\beta \le k$ with $\alpha+\beta > k$ such that there exist $(|V(H)|+2)(2k+1)+1$ pairwise internally vertex-disjoint paths of length $\alpha+\beta$ between $u_{-\alpha}$ and u_β. But we discarded these paths, which is a contradiction, and the proof is complete. □

2.7 Longer subdivisions of the complete bipartite graph

In this section we prove Theorem 2.1.17.

2.7.1 The high-level structure of the proof

Using Lemma 2.2.1, Theorem 2.1.17 reduces to the following statement.

Theorem 2.7.1. *Let $s, t, k \geq 2$ be integers. Let G be a K-almost-regular graph on n vertices with minimum degree $\delta = \omega(n^{\frac{s-1}{sk}})$. Then, for n sufficiently large, G contains $K_{s,t}^{k-1}$ as a subgraph.*

In what follows, let us fix the integers $s, t, k \geq 2$. It will be tacitly assumed throughout the section that n is sufficiently large compared to all other parameters.

The next definition is due to Jiang and Qiu [75].

Definition 2.7.2. Let $\ell_1, \ldots, \ell_s$ be positive integers. An *s-legged spider* S with length vector $(\ell_1, \ldots, \ell_s)$ consists of a vertex u, called the centre of the spider, and paths $P_1, \ldots, P_s$, called the legs of S, of lengths $\ell_1, \ldots, \ell_s$, starting at u and sharing no vertex other than u. For convenience, we define two spiders S and S' to be different if $P_i \neq P_i'$ for some $1 \leq i \leq s$, where $P_1', \ldots, P_s'$ are the legs of S'. So different spiders can form the same graph, e.g. if $\ell_1 = \ell_2$, $P_1 = P_2'$, $P_2 = P_1'$ and $P_i = P_i'$ for $i \geq 3$.

Let v_i be the endpoint of P_i different from u. Then we say that S has leaf vector $(v_1, \ldots, v_s)$.

We say that S' is a subspider of S if they have the same centre and for each $1 \leq i \leq s$, the ith leg of S' is a subpath of the ith leg of S.

In Subsection 2.6.2 we showed that (roughly speaking) if a graph has many pairs of short paths (P, P') such that P and P' are of equal length and have the same endpoints, then the graph contains $K_{s,t}^{k-1}$ as a subgraph. In Subsection 2.7.2 we shall prove an analogous statement for spiders; that is, if there are many pairs of spiders (S, S') such that S and S' have the same length vector and the same leaf vector, then the graph contains $K_{s,t}^{k-1}$ as a subgraph.

The next definition extends Definition 2.6.3 to spiders.

Definition 2.7.3. We define the notions of *L-admissible* and *L-good* spiders recursively as follows.

Every s-legged spider with length vector $(1, \ldots, 1)$ is L-admissible. Now let $1 \leq \ell_1, \ldots, \ell_s \leq k$ and assume that $\ell_i > 1$ for some i. A spider with centre u and legs $P_i = uw_{i,1} \ldots w_{i,\ell_i}$ (for $1 \leq i \leq s$) is L-admissible if the following two conditions hold:

- for any $1 \leq i \leq s$ and any $1 \leq j < \ell_i$, the s-legged spider with centre u and legs $P_1, \ldots, P_{i-1}, P_i', P_{i+1}, \ldots, P_s$ is L-good, where $P_i' = uw_{i,1} \ldots w_{i,j}$
- for any $1 \leq i \leq s$, the path P_i is L-good.

Finally, we say that a spider with length vector $(\ell_1, \ldots, \ell_s)$ and leaf vector $(v_1, \ldots, v_s)$ is L-good if it is L-admissible and the number of L-admissible spiders with length vector $(\ell_1, \ldots, \ell_s)$ and leaf vector $(v_1, \ldots, v_s)$ is at most $f(\ell, L)$, where $\ell = \ell_1 + \cdots + \ell_s$.

Remark. (1) This is well-defined since whether a spider is L-admissible or not depends only on the L-goodness of smaller spiders and paths.

(2) In this section L is always a constant not depending on n.

The next lemma follows easily from Corollary 2.6.11 from the previous section, and states that most short paths can be assumed to be good.

Lemma 2.7.4. *Let G be a $K_{s,t}^{k-1}$-free K-almost-regular graph on n vertices with minimum degree $\delta = \omega(1)$. Then for any $1 \leq j \leq k$, the number of paths of length j which are not L-good is at most $c_L n\delta^j$, where $c_L \to 0$ as $L \to \infty$.*

The main technical result of this section is the following lemma, which is the analogue of Lemma 2.7.4 for spiders.

Lemma 2.7.5. *Let G be a $K_{s,t}^{k-1}$-free K-almost-regular graph on n vertices with minimum degree $\delta = \omega(1)$ and let $1 \leq \ell_1, \dots, \ell_s \leq k$. Then the number of s-legged spiders with length vector $(\ell_1, \dots, \ell_s)$ which are L-admissible but not L-good is at most $c_L' n\delta^{\ell_1+\dots+\ell_s}$, where $c_L' \to 0$ as $L \to \infty$.*

We postpone the proof of this lemma to the next subsection and first show how it implies Theorem 2.7.1. The next lemma is an easy corollary of Lemma 2.7.5.

Lemma 2.7.6. *Let G be a $K_{s,t}^{k-1}$-free K-almost-regular graph on n vertices with minimum degree $\delta = \omega(1)$. Then the number of s-legged spiders with length vector $(k, k \dots, k)$ which are not L-good is at most $c_L'' n\delta^{sk}$, where $c_L'' \to 0$ as $L \to \infty$.*

Proof. Suppose that some s-legged spider S with length vector $(k, \dots, k)$ and legs $P_1, \dots, P_s$ is not L-good.

We distinguish two cases. First, assume that some P_i is not L-good. By Lemma 2.7.4, there are at most $c_L n\delta^k$ choices for P_i, where $c_L \to 0$ as $L \to \infty$. Since the maximum degree of G is at most $K\delta$, the number of ways to extend a given P_i to an s-legged spider with length vector $(k, \dots, k)$ is at most $(K\delta)^{(s-1)k}$. Thus, the number of s-legged spiders with length vector $(k, \dots, k)$ such that one of the legs is not L-good is at most $s \cdot c_L n\delta^k \cdot (K\delta)^{(s-1)k} = sK^{(s-1)k} c_L n\delta^{sk}$.

Now assume that all the P_i are L-good. Choose an s-legged subspider S' with the same centre and legs $P_1', \dots, P_s'$ which are subpaths of $P_1, \dots, P_s$ such that S' is minimal with respect to the condition that S' is not L-good. Let ℓ_i be the length of P_i'. Suppose that S' is not L-admissible. Since each P_i is L-good, so is every P_i'. Thus, there must be a proper s-legged subspider in S' which is not L-good. This contradicts the minimality of S'. So S' is L-admissible but not L-good. By Lemma 2.7.5, for any fixed $1 \leq \ell_1, \dots, \ell_s \leq k$, the number of s-legged spiders with length vector $(k, \dots, k)$ whose subspider with length vector $(\ell_1, \dots, \ell_s)$ is L-admissible but not L-good is at most $c_L' n\delta^{\ell_1+\dots+\ell_s} \cdot (K\delta)^{sk-\ell_1-\dots-\ell_s}$. Summing over all choices for $\ell_1, \dots, \ell_s$, we find that the number of s-legged spiders with length vector $(k, \dots, k)$ which are not L-good but whose legs are all L-good is at most $k^s \cdot K^{sk} c_L' n\delta^{sk}$. □

We are now in a position to complete the proof of Theorem 2.7.1.

Proof of Theorem 2.7.1. Choose L such that the c''_L provided by Lemma 2.7.6 satisfies $c''_L \leq 1/2$. Then by Lemma 2.7.6, for n sufficiently large, the number of L-good s-legged spiders with length vector $(k, \dots, k)$ is at least $\frac{1}{3}n\delta^{sk} > f(sk, L)n^s$. Thus, there exists an s-tuple $(v_1, \dots, v_s)$ of vertices such that the number of L-good s-legged spiders with length vector $(k, \dots, k)$ and leaf vector $(v_1, \dots, v_s)$ is greater than $f(sk, L)$. This contradicts the definition of an L-good spider. □

2.7.2 Spiders

In this subsection we prove Lemma 2.7.5, after which the proof of Theorem 2.7.1 is complete. For this subsection, we fix some $1 \leq \ell_1, \dots, \ell_s \leq k$ and write $\ell = \ell_1 + \dots + \ell_s$.

In what follows, it will be crucial to look at "spiders" some of whose legs may consist of zero edges.

Definition 2.7.7. Let $\ell'_1, \dots, \ell'_s$ be nonnegative integers. A *generalised spider* S with length vector $(\ell'_1, \dots, \ell'_s)$ consists of a vertex u (the centre of S) and paths $P_1, \dots, P_s$ (the legs of S) of lengths $\ell'_1, \dots, \ell'_s$, starting at u and sharing no vertex other than u. Let P_i have endpoints u and v_i. Then we say that S has leaf vector $(v_1, \dots, v_s)$.

The next lemma states that if there are many L-admissible but not L-good spiders with length vector $(\ell_1, \dots, \ell_s)$ in our graph, then we can find many L-admissible spiders with length vector $(\ell_1, \dots, \ell_s)$ and some useful extra properties.

Lemma 2.7.8. *Let G be a K-almost-regular graph on n vertices with minimum degree δ. Assume that L is sufficiently large compared to s, k and K and that there are at least $\frac{n\delta^{\ell_1+\dots+\ell_s}}{L}$ L-admissible but not L-good spiders with length vector $(\ell_1, \dots, \ell_s)$. Then there exists a non-empty set $\mathcal{S}$ of L-admissible spiders with length vector $(\ell_1, \dots, \ell_s)$ such that the following conditions hold.*

(i) For any $S \in \mathcal{S}$, the number of spiders $T \in \mathcal{S}$ with the same leaf vector as that of S is at least $\frac{f(\ell,L)}{2}$.

(ii) For any $S \in \mathcal{S}$, and any $\gamma_1, \dots, \gamma_s \in \{0, 1\}$, the subspider of S with length vector $(\ell_1 - \gamma_1, \dots, \ell_s - \gamma_s)$ (which is a generalised spider) is contained as a subspider in at least $\frac{\delta^{\gamma_1+\dots+\gamma_s}}{L^2}$ elements of $\mathcal{S}$.

Proof. Define a sequence of sets $\mathcal{T}_0, \mathcal{T}_1, \dots, \mathcal{T}_m$ recursively as follows. Take $\mathcal{T}_0$ be the set of all L-admissible but not L-good spiders with length vector $(\ell_1, \dots, \ell_s)$. Then, if there is some $S \in \mathcal{T}_i$ which violates condition (i), ie. the number of spiders $T \in \mathcal{T}_i$ with the same leaf vector as that of S is less than $\frac{f(\ell,L)}{2}$, then choose such an S arbitrarily and let $\mathcal{T}_{i+1} = \mathcal{T}_i \backslash \{S\}$. Also, if no such S exists, but there is some $S \in \mathcal{T}_i$ which violates condition (ii), ie. there exist some $\gamma_1, \dots, \gamma_s \in \{0, 1\}$ such that the subspider of S with length vector $(\ell_1 - \gamma_1, \dots, \ell_s - \gamma_s)$ is contained in less than $\frac{\delta^{\gamma_1+\dots+\gamma_s}}{L^2}$ elements of $\mathcal{T}_i$, then choose such an S arbitrarily and let $\mathcal{T}_{i+1} = \mathcal{T}_i \backslash \{S\}$. The process eventually terminates with some set $\mathcal{T}_m$.

Let $\mathcal{S} = \mathcal{T}_m$. It is clear that $\mathcal{S}$ satisfies conditions (i) and (ii); all we need to prove is that $\mathcal{S} \neq \emptyset$. Note that every $S \in \mathcal{T}_0$ is L-admissible but not L-good, so there are at least $f(\ell, L)$ elements $T \in \mathcal{T}_0$ with the same leaf vector as that of S. Among the set of elements of $\mathcal{T}_0$ with a fixed leaf vector, at most $\frac{f(\ell,L)}{2}$ are discarded because of violating condition (i) at some point. Thus, if $\mathcal{S} = \emptyset$, then at least half of the elements of $\mathcal{T}_0$, and so at least $\frac{n\delta^{\ell_1+\dots+\ell_s}}{2L}$ spiders are discarded because of violating condition (ii) at some point. However, any generalised spider R with length vector $(\ell_1 - \gamma_1, \dots, \ell_s - \gamma_s)$ is "responsible" for discarding at most $\frac{\delta^{\gamma_1+\dots+\gamma_s}}{L^2}$ elements, meaning that the number of elements discarded because they contain R which is contained in less than $\frac{\delta^{\gamma_1+\dots+\gamma_s}}{L^2}$ elements of some $\mathcal{T}_i$ is at most $\frac{\delta^{\gamma_1+\dots+\gamma_s}}{L^2}$. Since the number of generalised spiders with length vector $(\ell_1 - \gamma_1, \dots, \ell_s - \gamma_s)$ is at most $n(K\delta)^{(\ell_1-\gamma_1)+\dots+(\ell_s-\gamma_s)}$, the total number of elements discarded because of violating condition (ii) at some point is at most $2^s \cdot \frac{n(K\delta)^{\ell_1+\dots+\ell_s}}{L^2}$. For $L > 2^{s+1}K^{\ell_1+\dots+\ell_s}$, this is less than $\frac{n\delta^{\ell_1+\dots+\ell_s}}{2L}$, contradicting our earlier claim. Thus, $\mathcal{S} \neq \emptyset$. $\square$

The next lemma is an analogue of Lemma 2.6.4 for spiders.

Lemma 2.7.9. *Let $L \geq 1$ be sufficiently large compared to ℓ and let $v_1, \dots, v_s$ be vertices. Suppose that there is a set $\mathcal{T}$ of at least $\frac{f(\ell,L)}{2}$ L-admissible spiders with length vector $(\ell_1, \dots, \ell_s)$ and leaf vector $(v_1, \dots, v_s)$. Then, among these, there exist more than $f(\ell - 1, L)$ spiders which are pairwise vertex-disjoint apart from their leaves.*

Proof. Suppose otherwise. Take a maximal set of such spiders. By assumption, we have chosen at most $f(\ell-1, L)$ spiders. Each such spider has $\ell+1-s \leq \ell-1$ non-leaf vertices, so altogether they have at most $f(\ell - 1, L)(\ell - 1)$ non-leaf vertices. By the maximality assumption, each $S \in \mathcal{T}$ contains at least one of these vertices. Thus, by the pigeonhole principle, there exist some vertex x and a set $\mathcal{S} \subset \mathcal{T}$ of size at least $\frac{f(\ell,L)/2}{(\ell-1)\cdot f(\ell-1,L)(\ell-1)}$ such that the elements of $\mathcal{S}$ all contain the vertex x in the same non-leaf position (meaning that there exist some i and $j < \ell_i$ such that in all $S \in \mathcal{S}$, x is the jth vertex on the ith leg, where the centre of the spider is viewed as the 0th vertex on the leg). Note that $|\mathcal{S}| \geq \frac{f(\ell,L)/2}{f(\ell-1,L)(\ell-1)^2} > \max_{1\leq b\leq \ell-1} f(b, L)f(\ell - b, L)$.

We now distinguish two cases. First, let us assume that x is not the centre in the spiders in $\mathcal{S}$. Then there exists some $1 \leq i \leq s$ and some $1 \leq j < \ell_i$ such that x is the jth vertex on the ith leg in each of these spiders. Let $b = \ell_i - j$. Since $|\mathcal{S}| > f(b, L)f(\ell-b, L)$ and each element of $\mathcal{S}$ is L-admissible, either there are more than $f(b, L)$ L-good paths of length b between x and v_i or there are more than $f(\ell-b, L)$ L-good s-legged spiders with length vector $(\ell_1, \dots, \ell_{i-1}, j, \ell_{i+1}, \dots, \ell_s)$ and leaf vector $(v_1, \dots, v_{i-1}, x, v_{i+1}, \dots, v_s)$. The first case contradicts the definition of an L-good path and the second case contradicts the definition of an L-good s-legged spider.

Let us now assume that x is the centre in the spiders in $\mathcal{S}$. Note that

$$|\mathcal{S}| > f(\ell_1, L)f(\ell_2, L)\dots f(\ell_s, L).$$

Thus, there exists some $i \le s$ such that there are more than $f(\ell_i, L)$ L-good paths of length ℓ_i between x and v_i. This contradicts the definition of an L-good path. □

In the key part of the proof of Lemma 2.7.5 it will be necessary to assume that $\ell_i = 1$ holds for at most one choice of i. Accordingly, we first deal with the other case separately.

Lemma 2.7.10. *Let G be a $K_{s,t}^{k-1}$-free K-almost-regular graph on n vertices with minimum degree $\delta = \omega(1)$, and assume that $\ell_1 = \ell_2 = 1$. Then the number of s-legged spiders with length vector $(\ell_1, \dots, \ell_s)$ which are L-admissible but not L-good is at most $c'_L n\delta^{\ell_1+\dots+\ell_s}$, where $c'_L \to 0$ as $L \to \infty$.*

Proof. If $s = 2$, then the result follows from Lemma 2.7.4, since a spider with length vector $(1, 1)$ is L-good if and only if it is L-good when viewed as a path of length 2. Assume that $s \ge 3$. Note that in this case $\ell \ge 3$.

Let S be an L-admissible but not L-good spider with length vector $(\ell_1, \dots, \ell_s)$ and leaf vector $(v_1, \dots, v_s)$. By definition, there exist at least $f(\ell, L)$ L-admissible spiders with length vector $(\ell_1, \dots, \ell_s)$ and leaf vector $(v_1, \dots, v_s)$. Hence, by Lemma 2.7.9, for L sufficiently large there exist more than $f(\ell-1, L)$ L-admissible spiders with length vector $(\ell_1, \dots, \ell_s)$ and leaf vector $(v_1, \dots, v_s)$ which are pairwise vertex-disjoint apart from at their leaves. In particular, there are more than $f(\ell - 1, L) \ge f(2, L)$ paths of length 2 between v_1 and v_2. Note that any path of length 2 is L-admissible. Let u be the centre of S. Then the path v_1uv_2 is not L-good.

The number of ways to extend a path xyz to a spider with length vector $(\ell_1, \dots, \ell_s)$, centre y and first two legs yx and yz in this order is at most $(K\delta)^{\ell_3+\dots+\ell_s}$. Thus, by Lemma 2.7.4, the number of L-admissible but not L-good spiders with length vector $(\ell_1, \dots, \ell_s)$ is at most $c_L n\delta^2 \cdot 2 \cdot (K\delta)^{\ell_3+\dots+\ell_s}$ with $c_L \to 0$ as $L \to \infty$, where the factor $c_L n\delta^2$ bounds the number of not L-good paths of length 2, the factor 2 accounts for the two edges in this path that we can use as the first leg of the spider, and the factor $(K\delta)^{\ell_3+\dots+\ell_s}$ bounds the number of ways to get a spider with fixed first two legs. Since $c_L n\delta^2 \cdot 2 \cdot (K\delta)^{\ell_3+\dots+\ell_s} = 2K^{\ell_3+\dots+\ell_s} c_L n\delta^{\ell_1+\dots+\ell_s}$, the result follows. □

Using Lemma 2.7.10 and symmetry, it is enough to prove Lemma 2.7.5 in the case where $\ell_i = 1$ holds for at most one value of i.

The next result is the key step in the proof of Lemma 2.7.5, and contains the main idea of this section. It is proved in greater generality than what is necessary for Theorem 2.1.17, to allow for use in future work. Indeed, as we will see in Section 2.8, this lemma was used by Jiang and Qiu [74] to prove a generalisation of Theorem 2.1.17.

Lemma 2.7.11. *Let $\ell_i \le k_i \le k$ for each i. Assume that $\ell_i = 1$ holds for at most one value of i. Let G be a K-almost-regular graph on n vertices with minimum degree $\delta = \omega(1)$. Assume that L is sufficiently large compared to s, k and K and that there exists a set $\mathcal{S}$ of spiders satisfying the conditions in Lemma 2.7.8. For each $1 \le i \le s$, let $\gamma_{i,0} \in \{0, 1\}$ such that $k_i - \ell_i - \gamma_{i,0}$ is even. Let R_0 be the subspider with length vector*

$(\ell_1-\gamma_{1,0},\dots,\ell_s-\gamma_{s,0})$ of an arbitrary element of $\mathcal{S}$. Let R_0 have leaf vector $(v_1,\dots,v_s)$. Let $Z\subset V(G)$ be a set of size at most L, disjoint from $\{v_1,\dots,v_s\}$. Then there exist vertices $w_1,\dots,w_s$ and paths $P_1,\dots,P_s$ such that

(1) for each i, P_i is a path of length $k_i-\ell_i$ between v_i and w_i

(2) $(w_1,\dots,w_s)$ is the leaf vector of an element of $\mathcal{S}$ and

(3) the paths $P_1,\dots,P_s$ are pairwise vertex-disjoint and avoid Z.

Proof. Since $k_i-\ell_i-\gamma_{i,0}$ is an even number between 0 and k, there exist $\gamma_{i,1},\dots,\gamma_{i,k-1}\in\{0,1\}$ such that $k_i-\ell_i-\gamma_{i,0}=2\gamma_{i,1}+\dots+2\gamma_{i,k-1}$.

We now define a sequence $R_1,\dots,R_{k-1}$ of generalised spiders, and sequences $S_1,\dots,S_k$ and $T_1,\dots,T_k$ of spiders recursively.

R_0 is given as a subspider of some element of $\mathcal{S}$, so by property (ii) in Lemma 2.7.8, the number of elements of $\mathcal{S}$ containing R_0 as a subspider is at least $\frac{\delta^{\gamma_{1,0}+\dots+\gamma_{s,0}}}{L^2}$. Thus, there is some $S_1\in\mathcal{S}$ containing R_0 such that $V(S_1)\setminus V(R_0)$ is disjoint from Z. Indeed, any fixed vertex not in $V(R_0)$ is a vertex in $O(\delta^{\gamma_{1,0}+\dots+\gamma_{s,0}-1})$ elements of $\mathcal{S}$ containing R_0, so the number of elements of $\mathcal{S}$ containing R_0 and intersecting $Z\setminus V(R_0)$ is $O(\delta^{\gamma_{1,0}+\dots+\gamma_{s,0}-1})$. Hence, as $\delta=\omega(1)$ and $L=O(1)$, a suitable $S_1\in\mathcal{S}$ indeed exists.

Now choose $T_1\in\mathcal{S}$ with the same leaf vector as that of S_1 such that T_1 and S_1 are disjoint apart from their leaves. This is possible, if L is sufficiently large, by property (i) in Lemma 2.7.8 and Lemma 2.7.9. Let R_1 be the subspider of T_1 with length vector $(\ell_1-\gamma_{1,1},\dots,\ell_s-\gamma_{s,1})$.

More generally, for any $1\le j\le k$, given a generalised spider R_{j-1} with length vector $(\ell_1-\gamma_{1,j-1},\dots,\ell_s-\gamma_{s,j-1})$ which is a subspider of an element of $\mathcal{S}$, we define S_j, T_j and R_j as follows.

Choose some $S_j\in\mathcal{S}$ containing R_{j-1} such that $V(S_j)\setminus V(R_{j-1})$ is disjoint from $Z\cup(V(S_1)\cup\dots\cup V(S_{j-1}))\cup(V(T_1)\cup\dots\cup V(T_{j-1}))$. This is possible by property (ii) in Lemma 2.7.8.

Also, choose $T_j\in\mathcal{S}$ with the same leaf vector as that of S_j such that T_j is disjoint from $Z\cup(V(S_1)\cup\dots\cup V(S_j))\cup(V(T_1)\cup\dots\cup V(T_{j-1}))$ apart from its leaves. This is possible by property (i) in Lemma 2.7.8 and Lemma 2.7.9.

Finally, if $j<k$, let R_j be the subspider of T_j with length vector $(\ell_1-\gamma_{1,j},\dots,\ell_s-\gamma_{s,j})$.

Now for $1\le i\le s$ and $0\le j\le k-1$, let $x_{i,2j}$ be the endpoint of the ith leg of R_j and let $x_{i,2j+1}$ be the endpoint of the ith leg of S_{j+1}. Then, when we ignore the repetitions, the vertices $x_{i,0},x_{i,1},\dots,x_{i,2k-1}$ form a path of length $\gamma_{i,0}+2\gamma_{i,1}+\dots+2\gamma_{i,k-1}=k_i-\ell_i$. Indeed, if $\gamma_{i,0}=0$, then $x_{i,1}=x_{i,0}$ and if $\gamma_{i,0}=1$, then $x_{i,1}$ is a neighbour of $x_{i,0}$. Moreover, for any $1\le j\le k-1$, if $\gamma_{i,j}=0$, then $x_{i,2j+1}=x_{i,2j}=x_{i,2j-1}$ and if $\gamma_{i,j}=1$, then $x_{i,2j}$ is a neighbour of $x_{i,2j-1}$ and does not belong to $\{x_{p,q}:1\le p\le s,0\le q\le 2j-1\}\cup Z$, and $x_{i,2j+1}$ is a neighbour of $x_{i,2j}$ and does not belong to $\{x_{p,q}:1\le p\le s,0\le q\le 2j\}\cup Z$. Let P_i be the path formed by the vertices $x_{i,0},x_{i,1}\dots,x_{i,2k-1}$ and let $w_i=x_{i,2k-1}$.

Note that $(x_{1,0}, \ldots, x_{s,0})$ is the leaf vector of R_0, so $x_{i,0} = v_i$, therefore condition (1) in this lemma is satisfied. Moreover, $(w_1, \ldots, w_s) = (x_{1,2k-1}, \ldots, x_{s,2k-1})$ is the leaf vector of S_k, so property (2) is also satisfied.

By assumption, Z is disjoint from $\{v_1, \ldots, v_s\} = \{x_{1,0}, \ldots, x_{s,0}\}$, so it follows from the above that $P_1, \ldots, P_s$ avoid Z. Finally, it is clear by the above discussion that if $P_1, \ldots, P_s$ are not pairwise vertex-disjoint, then $x_{i,j} = x_{i',j}$ holds for some $i \neq i'$ and some $0 \leq j \leq 2k-1$. However, for each $0 \leq j \leq 2k-1$, $(x_{1,j}, \ldots, x_{s,j})$ is the leaf vector of a generalised spider whose ith leg consists of at least $\ell_i - 1$ edges, so at most one of its legs has 0 edges. Thus, the vertices $x_{1,j}, \ldots, x_{s,j}$ are distinct and condition (3) is satisfied. $\square$

It is not hard to connect the paths given by the previous lemma to form spiders with length vector $(k_1, \ldots, k_s)$.

Lemma 2.7.12. *Let $\ell_i \leq k_i \leq k$ for each i. Assume that $\ell_i = 1$ holds for at most one value of i. Let G be a K-almost-regular graph on n vertices with minimum degree $\delta = \omega(1)$. Assume that L is sufficiently large compared to s, k and K and that there exists a set $\mathcal{S}$ of spiders satisfying the conditions in Lemma 2.7.8. For each $1 \leq i \leq s$, let $\gamma_{i,0} \in \{0,1\}$ such that $k_i - \ell_i - \gamma_{i,0}$ is even. Let R_0 be the subspider with length vector $(\ell_1 - \gamma_{1,0}, \ldots, \ell_s - \gamma_{s,0})$ of an arbitrary element of $\mathcal{S}$. Let R_0 have leaf vector $(v_1, \ldots, v_s)$. Let $Z \subset V(G)$ be a set of size at most L, disjoint from $\{v_1, \ldots, v_s\}$.*

Then there exists an s-legged spider with length vector $(k_1, \ldots, k_s)$ and leaf vector $(v_1, \ldots, v_s)$ that avoids Z.

Proof. Choose vertices $w_1, w_2 \ldots, w_s$ and paths $P_1, P_2, \ldots, P_s$ as in the conclusion of Lemma 2.7.11. $(w_1, \ldots, w_s)$ is the leaf vector of an element of $\mathcal{S}$, so by condition (i) in Lemma 2.7.8 and Lemma 2.7.9, there exist at least $f(\ell - 1, L)$ spiders with length vector $(\ell_1, \ldots, \ell_s)$ and leaf vector $(w_1, \ldots, w_s)$ which are pairwise vertex-disjoint apart from at their leaves. Thus, if L is sufficiently large, then there exists a spider S with length vector $(\ell_1, \ldots, \ell_s)$ and leaf vector $(w_1, \ldots, w_s)$ such that $V(S)$ is disjoint from Z and intersects $\bigcup_{1 \leq i \leq s} V(P_i)$ only at $\{w_1, \ldots, w_s\}$. Let J_i be the ith leg of S, let u be the centre of S and let Q_i be the union of J_i and P_i. Then the spider with centre u and legs $Q_1, \ldots, Q_s$ is suitable. $\square$

The next result, together with Lemma 2.7.10, completes the proof of Lemma 2.7.5.

Lemma 2.7.13. *Assume that $\ell_i = 1$ holds for at most one value of i. Let G be a K-almost-regular graph on n vertices with minimum degree $\delta = \omega(1)$. Assume that L is sufficiently large compared to s, t, k and K and that there are at least $\frac{n\delta^{\ell_1 + \cdots + \ell_s}}{L}$ L-admissible but not L-good spiders with length vector $(\ell_1, \ldots, \ell_s)$. Then G contains $K_{s,t}^{k-1}$ as a subgraph.*

Proof. Choose a set $\mathcal{S}$ with the properties described in Lemma 2.7.8. Define $v_1, \ldots, v_s$ as in the statement of Lemma 2.7.12. We may repeatedly apply Lemma 2.7.12 to find s-legged spiders $S_1, \ldots, S_t$, each with length vector $(k, \ldots, k)$ and leaf vector $(v_1, \ldots, v_s)$

such that $V(S_j)$ is disjoint from $(\bigcup_{1\leq i\leq j-1} V(S_i)) \setminus \{v_1,\dots,v_s\}$. Then the union of these spiders is a copy of $K_{s,t}^{k-1}$. $\square$

2.8 Concluding remarks

Our main objective in this chapter was to prove upper bounds for extremal numbers. In some cases we can use a result of Bukh and Conlon to show that there is a matching lower bound.

Let F be a graph with a set of roots $R \subsetneq V(F)$. Recall from Section 2.5 that the rooted t-blowup of this rooted graph is the graph obtained by taking t vertex-disjoint copies of F and identifying the different copies of v for each $v \in R$. We denote this graph by $t * F$. For any non-empty $S \subset V(F) \setminus R$, let e_S be the number of edges in F with at least one endpoint in S. Set $\rho_F(S) = \frac{e_S}{|S|}$ and $\rho(F) = \rho_F(V(F) \setminus R)$. We say that (F, R) (or F if R is clear) is balanced if $\rho(F) \leq \rho_F(S)$ holds for every non-empty $S \subset V(F) \setminus R$. Bukh and Conlon proved the following result.

Theorem 2.8.1 (Bukh–Conlon [18]). *Let F be a balanced bipartite rooted graph with $\rho(F) > 0$. Then there is some $t_0 \in \mathbb{N}$ such that for every $t \geq t_0$, we have* $\mathrm{ex}(n, t * F) = \Omega(n^{2-\frac{1}{\rho(F)}})$.

Note that if F is the s-legged spider with length vector $(k, k, \dots, k)$ and its roots are the leaves, then F is balanced with $\rho(F) = \frac{sk}{s(k-1)+1}$. Moreover, $t * F = K_{s,t}^{k-1}$. Thus, Theorem 2.8.1 implies that for t sufficiently large in terms of s and k, we have $\mathrm{ex}(n, K_{s,t}^{k-1}) = \Omega(n^{1+\frac{s-1}{sk}})$. This shows that Corollary 2.1.18 follows from Theorem 2.1.17.

In order to prove Theorem 2.1.14 and to obtain a large family of realisable exponents, together with Conlon and Lee we established the following result, which was conjectured by Kang, Kim and Liu [79].

Theorem 2.8.2 (Conlon–Janzer–Lee [23]). *Let s, k, t be positive integers such that $s \geq 2$ and let S be the s-legged spider with length vector $(1, k, k, \dots, k)$ and the roots being the leaves. Then*

$$\mathrm{ex}(n, t * S) = O(n^{1+\frac{s-1}{(s-1)k+1}}).$$

Since S is balanced, by Theorem 2.8.1 this is tight for t sufficiently large, so it implies that $1 + \frac{s-1}{(s-1)k+1}$ is a realisable exponent for every $s \geq 2$ and $k \geq 1$.

More generally, it is not hard to see that an s-legged spider S with length vector $(k_1, \dots, k_s)$ is balanced if and only if $k_1 + \dots + k_s \geq (s-1)\max_{1\leq i\leq s} k_i$, where again the roots are the leaves. If this holds, then by Theorem 2.8.1, $\mathrm{ex}(n, t * S) = \Omega(n^{1+\frac{s-1}{k_1+\dots+k_s}})$ for t sufficiently large. The author conjectures that this is tight.

Conjecture 2.8.3 (Janzer [66]). *Let $s \geq 2$ and $1 \leq k_1 \leq k_2 \leq \dots \leq k_s$ be integers satisfying $k_1 + \dots + k_s \geq (s-1)k_s$. Let S be the rooted graph which is a spider with length* *vector* $(k_1,\dots,k_s)$ *and whose roots are the leaves. Then for any integer* $t\ge 1$*,*

$$\mathrm{ex}(n,t*S)=O(n^{1+\frac{s-1}{k_1+\dots+k_s}}).$$

Note that Theorem 2.1.17 and Theorem 2.8.2 show that the conjecture holds when $(k_1,\dots,k_s)=(k,\dots,k)$ and when $(k_1,\dots,k_s)=(1,k,\dots,k)$.

As pointed out by the author in [66], the following lemma, which follows easily from the results in the previous section, might be useful for proving Conjecture 2.8.3.

Lemma 2.8.4. *Let* $1\le \ell_i\le k_i$ *be integers for each* i*. Assume that* $\ell_i=1$ *holds for at most one value of* i*. Let* G *be a* K*-almost-regular graph on* n *vertices with minimum degree* $\delta=\omega(1)$*. Assume that* L *is sufficiently large compared to* s*,* t*,* $k_1,\dots,k_s$ *and* K *and that there are at least* $\frac{n\delta^{\ell_1+\dots+\ell_s}}{L}$ L*-admissible but not* L*-good spiders with length vector* $(\ell_1,\dots,\ell_s)$*. Let* S *be the rooted graph which is a spider with length vector* $(k_1,\dots,k_s)$ *and whose roots are the leaves. Then* G *contains* $t*S$ *as a subgraph.*

Proof. Choose a set $\mathcal{S}$ with the properties described in Lemma 2.7.8. Define $v_1,\dots,v_s$ as in the statement of Lemma 2.7.12. We may repeatedly apply Lemma 2.7.12 to find s-legged spiders $S_1,\dots,S_t$, each with length vector $(k_1,\dots,k_s)$ and leaf vector $(v_1,\dots,v_s)$ such that $V(S_j)$ is disjoint from $(\bigcup_{1\le i\le j-1}V(S_i))\setminus\{v_1,\dots,v_s\}$. Then the union of these spiders is a copy of $t*S$. □

Using Lemma 2.8.4 and some new ideas, Jiang and Qiu proved the following common generalisation of Theorem 2.1.17 and Theorem 2.8.2.

Theorem 2.8.5 (Jiang–Qiu [74])**.** *Let* s,b,k,t *be positive integers with* $s\ge 2$ *and* $b\le k$*, and let* S *be the* s*-legged spider with length vector* $(b,k,k,\dots,k)$*. Then*

$$\mathrm{ex}(n,t*S)=O(n^{1+\frac{s-1}{(s-1)k+b}}).$$

By Theorem 2.8.1, this is tight for t sufficiently large, so it implies that for every $s\ge 2$ and $b\le k$, $1+\frac{s-1}{(s-1)k+b}$ is realisable. This has the following nice corollary.

Corollary 2.8.6 (Jiang–Qiu [74])**.** *For any positive integers* p,q *with* $q>p^2$*,* $1+\frac{p}{q}$ *is realisable.*

We finish the chapter by stating a conjecture of Kang, Kim and Liu about the extremal number of the 1-subdivision of an arbitrary bipartite graph.

Conjecture 2.8.7 (Kang–Kim–Liu [79])**.** *Let* F *be a bipartite graph with* $\mathrm{ex}(n,F)=O(n^{1+\alpha})$ *for some* $\alpha>0$*. Then*

$$\mathrm{ex}(n,F')=O(n^{1+\frac{\alpha}{2}}).$$

Apart from being independently interesting, they showed that this conjecture would imply Conjecture 2.1.2 on rational exponents.

Chapter 3

The extremal number of blow-ups

3.1 Introduction

A graph F is called *r-degenerate* if each of its subgraphs has minimum degree at most r. Generalising the Kővári–Sós–Turán theorem, Erdős in 1967 proposed the following conjecture.

Conjecture 3.1.1 (Erdős [34]). *Let F be a bipartite r-degenerate graph. Then* $\mathrm{ex}(n, H) = O(n^{2-\frac{1}{r}})$.

Note that, if true, this would also greatly generalise Theorem 2.1.3.

Alon, Krivelevich and Sudakov used dependent random choice to obtain the following result.

Theorem 3.1.2 (Alon–Krivelevich–Sudakov [4]). *Let H be a bipartite r-degenerate graph. Then* $\mathrm{ex}(n, H) = O(n^{2-\frac{1}{4r}})$.

Another partial result towards Conjecture 3.1.1 is due to Füredi and West [48], who confirmed that $\mathrm{ex}(K_{s,s} \setminus K_{s-r,s-r}) = O(n^{2-1/r})$. Here the forbidden graph is obtained from the complete bipartite graph $K_{s,s}$ by deleting the edges of a complete bipartite subgraph $K_{s-r,s-r}$.

Observe that there exists a permutation of the vertices $\{v_1, v_2, \ldots, v_k\}$ of any r-degenerate graph for which every vertex v_i has at most r neighbours in the set $\{v_1, v_2, \ldots, v_{i-1}\}$. With this in mind, one can define the complexity of an r-degenerate graph as follows.

Definition 3.1.3. The graph $K_{r,r} = G(A_0, B_0)$ is considered as a graph of complexity 0 and any multiplicity. A bipartite graph $G(A, B)$ is a *complete r-degenerate bipartite graph of complexity s and multiplicity m* if it can be obtained from the complete bipartite graph $G(A', B')$ of complexity $s-1$ and multiplicity m by the addition of further $m(\binom{|A'|}{r} + \binom{|B'|}{r})$ vertices such that m new vertices are assigned to each r-set in A' and each r-set in B', and every new vertex is connected to the vertices of the r-set that it is assigned to. The *complexity* of an r-degenerate bipartite graph H is defined to be the smallest possible complexity of a complete r-degenerate bipartite graph (of arbitrary multiplicity) that contains H as a subgraph.

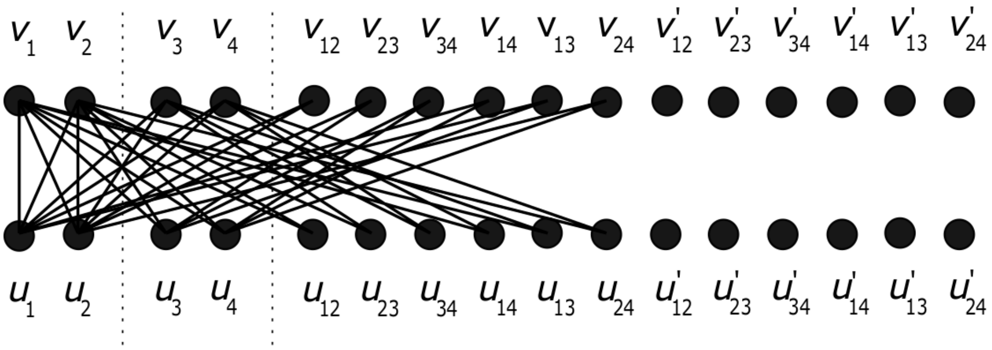

Figure 3.1: The complete 2-degenerate bipartite graph of complexity 2 and multiplicity 2. Note that the clone v' of v has the same neighbours, but we did not draw those edges in order to keep the figure transparent.

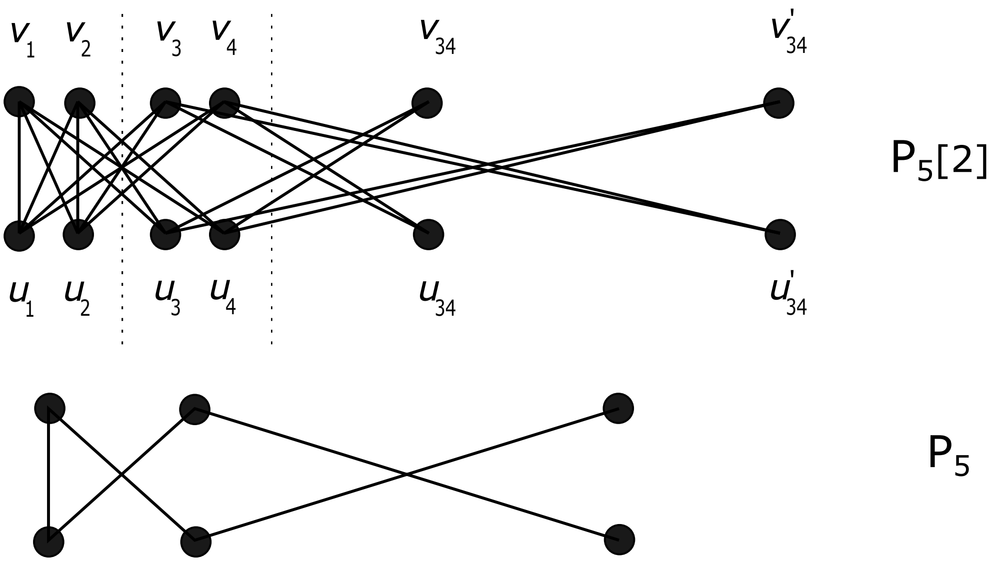

Figure 3.2: The blow-up $P_5[2]$ of the path with 5 edges, as a subgraph of the complete 2-degenerate bipartite graph of complexity 2 and multiplicity 2.

Note that the result of Füredi and West covers precisely the complexity 1 case, while Theorem 2.1.3 only applies to some r-degenerate bipartite graphs of complexity at most 2.

Our first contribution is a proof of Conjecture 3.1.1 for all graphs of complexity at most 2.

Theorem 3.1.4. *Let H be a complete r-degenerate bipartite graph of complexity* 2 *and arbitrary multiplicity. Then*

$$\mathrm{ex}(n, H) = O(n^{2-\frac{1}{r}}).$$

Our next result concerns the case where H has larger complexity but has a strong structure, namely where H is a blow-up of a tree. For a graph F and a positive integer

r, the *r-blowup* of F is the graph obtained by replacing the vertices and edges of F with independent sets of size r and copies of $K_{r,r}$, respectively. We denote this graph by $F[r]$. The 2-blowup of P_5 is shown in Figure 3.2.

Theorem 3.1.5. *Let T be a tree and let r be a positive integer. Then*

$$\mathrm{ex}(n, T[r]) = O(n^{2-\frac{1}{r}}).$$

Actually, the vertices can be replaced by sets of arbitrary sizes as long as the resulting graph is r-degenerate, and the same conclusion holds. We say that a graph H is a blow-up of the graph T if to get H from T we replace each vertex of T with an independent set (of arbitrary size) and replace each edge of T with a corresponding complete bipartite graph.

Theorem 3.1.6. *Let H be a graph that is r-degenerate and is a blow-up of a tree. Then*

$$\mathrm{ex}(n, H) = O(n^{2-\frac{1}{r}}).$$

Note that Theorem 3.1.6 is a generalisation of the result of Füredi and West [48] on the Turán number $\mathrm{ex}(n, K_{s,s} \setminus K_{s-r,s-r})$. This case corresponds to the blow-up of the path of length 3.

In fact, we prove an even more general statement from which Theorem 3.1.6 follows. To state this result, we need to introduce another definition.

Definition 3.1.7. Let $r \leq t$ and k be positive integers and let $X_1 = Y_0, Y_1, Y_2, \ldots, Y_k$ be pairwise disjoint sets with $|X_1| = r$, $|Y_1| = \ldots = |Y_k| = t$. For each $2 \leq i \leq k$, let X_i be a subset of some Y_j with $j < i$ such that $|X_i| = r$. The graph L with vertex set $Y_0 \cup Y_1 \cup \ldots \cup Y_k$ and edge set $\bigcup_{1 \leq i \leq k} \{xy : x \in X_i, y \in Y_i\}$ is called an *(r,t)-blownup tree of size k*.

See Figure 3.3 for an example of a $(2,3)$-blownup tree of size 4.

Observe that an (r,t)-blownup tree is r-degenerate. We are now ready to state our most general result.

Theorem 3.1.8. *Let L be an (r,t)-blownup tree of arbitrary size. Then*

$$\mathrm{ex}(n, L) = O(n^{2-\frac{1}{r}}).$$

Note that any bipartite graph H with maximum degree at most r on one side is a subgraph of some (r,t)-blownup tree (for a suitable t). Indeed, when the parts of H are X and Y such that every vertex in X has degree at most r, then t can be chosen to be $|Y|$. This shows that Theorem 3.1.8 generalises Theorem 2.1.3.

It is natural to ask what we can say about the extremal number of the blow-up of an arbitrary bipartite graph. We make the following conjecture.

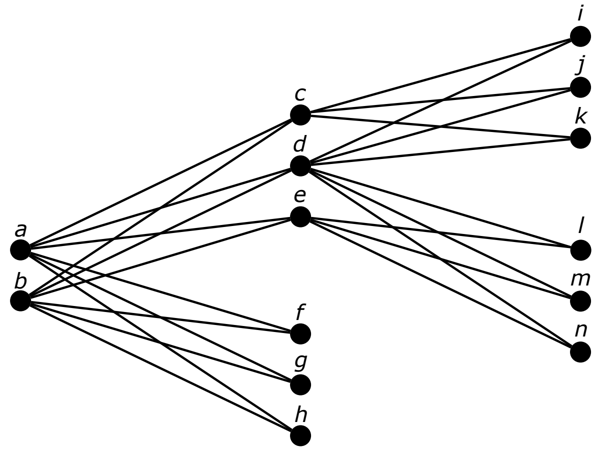

Figure 3.3: A $(2,3)$-blownup tree of size 4. Here $X_1 = Y_0 = \{a, b\}$, $X_2 = \{a, b\}$, $X_3 = \{c, d\}$, $X_4 = \{d, e\}$, $Y_1 = \{c, d, e\}$, $Y_2 = \{f, g, h\}$, $Y_3 = \{i, j, k\}$, $Y_4 = \{l, m, n\}$.

Conjecture 3.1.9. *For any* $0 \le \alpha \le 1$ *and any graph* F*, if* $\mathrm{ex}(n, F) = O(n^{2-\alpha})$*, then*

$$\mathrm{ex}(n, F[r]) = O\left(n^{2-\frac{\alpha}{r}}\right).$$

The motivation behind this conjecture is the following. Given a graph G, define an auxiliary graph $\mathcal{G}$ whose vertex set is $V(G)^{(r)}$ and in which U and V are joined by an edge if $U \cap V = \emptyset$ and $uv \in E(G)$ for every $u \in U$ and $v \in V$. Note that if the number of edges in G is $\omega\left(n^{2-\frac{\alpha}{r}}\right)$, then by supersaturation (see Lemma 3.2.1 below) there are $\omega(n^{2r-\alpha r})$ copies of $K_{r,r}$ in G, i.e. there are $\omega\left(N^{2-\alpha}\right)$ edges in $\mathcal{G}$, where $N = |V(\mathcal{G})| = \binom{n}{r}$. Therefore there exists a copy of F in $\mathcal{G}$, which provides a homomorphic copy of $F[r]$ in G. We conjecture that one can always embed F to $\mathcal{G}$ in a way that the r-sets corresponding to the vertices of F are disjoint, providing an embedding of the blow-up $F[r]$ to G.

Theorem 3.1.5 proves Conjecture 3.1.9 for trees. Note that $K_{s,t}[r] = K_{rs,rt}$, so the conjecture also holds for $F = K_{s,t}$, $\alpha = \frac{1}{s}$. It would be interesting to extend this to the family of even cycles. In this case, the conjecture can be stated as follows.

Conjecture 3.1.10. *For any* $r, k \ge 2$*,*

$$\mathrm{ex}(n, C_{2k}[r]) = O(n^{2-\frac{1}{r}+\frac{1}{rk}}).$$

We prove Conjecture 3.1.10 in the first unknown case – the 2-blowup of the hexagon.

Theorem 3.1.11.

$$\mathrm{ex}(n, C_6[2]) = O(n^{5/3}).$$

In fact, we prove a more general result about theta graphs.

Theorem 3.1.12. *For any positive integer* t*,*

$$\mathrm{ex}(n, \theta_{3,t}[2]) = O(n^{5/3}).$$

Theorem 2.8.1 from the previous chapter shows that this is tight for t sufficiently large. Indeed, if $F = P_3[2]$ with the roots being the degree 2 vertices, then $\theta_{3,t}[2]$ is the rooted t-blowup of F. (Note that rooted t-blowups are very different from t-blowups.) Since $\rho(F) = 3$, Theorem 3.1.12, combined with Theorem 2.8.1, has the following corollary.

Corollary 3.1.13. *For sufficiently large t, we have*

$$\mathrm{ex}(n, \theta_{3,t}[2]) = \Theta(n^{5/3}).$$

In the next chapter, using a different method, we give a general upper bound $\mathrm{ex}(n, C_{2k}[r]) = O(n^{2-\frac{1}{r}+\frac{1}{k+r-1}}(\log n)^{\frac{4k}{r(k+r-1)}})$ and discuss some interesting consequences.

The rest of the chapter is organised as follows. In Section 3.2 we present the proofs of Theorem 3.1.4, Theorem 3.1.6 and Theorem 3.1.8, while in Section 3.3 we prove Theorem 3.1.12.

3.2 Blow-ups of trees

For a graph G, $\overline{d}(G)$ denotes its average degree. Like in the previous chapter, the common neighbourhood of a vertex set R is denoted by $N_G(R)$ and we write $d_G(R) = |N_G(R)|$. We call a set of r vertices an r-set.

Let us briefly summarise the method we will use in this section. Roughly speaking, we prove that if we randomly and greedily try to embed an (r,t)-blownup tree L in the host graph, then with positive probability we do not get stuck. The way we choose the embedded images of the first few vertices of L is not straightforward: we make use of the stationary distribution on an auxiliary graph whose vertices are the r-sets of the original host graph. To obtain a dense enough auxiliary graph, we apply results on graph supersaturation. The embedding of the further vertices is also closely related to the usual random walk on this auxiliary graph, which allows us to prove that with high probability all r-sets that we hit in the random embedding have large enough neighbourhood.

One of the main ingredients of the proofs is a theorem on supersaturated graphs.

Lemma 3.2.1 (Erdős–Simonovits [40])**.** *For any positive integer r, there exist positive constants $c = c(r), \beta = \beta(r)$ such that any graph on n vertices with $e > cn^{2-\frac{1}{r}}$ edges contains at least $\beta\frac{e^{r^2}}{n^{2r^2-2r}}$ copies of $K_{r,r}$.*

This has the following simple corollary.

Corollary 3.2.2 (Erdős–Simonovits [34, 47])**.** *For any positive integer r and real number $\gamma > 0$ there exists a constant $c = c_r(\gamma)$ such that any graph on n vertices with $e > cn^{2-\frac{1}{r}}$ edges contains at least $\gamma\binom{n}{r}$ copies of $K_{r,r}$.*

We start with the proof of Theorem 3.1.4 which is simpler but already contains some of the ideas needed in the proof of Theorem 3.1.8.

Proof of Theorem 3.1.4. Let m be the multiplicity of H and let $\gamma = 2\binom{r+m}{r} \cdot \binom{|V(H)|}{r}$. By Corollary 3.2.2, there exists a constant $c = c_r(\gamma)$ such that any graph on n vertices with $e > c \cdot n^{2-\frac{1}{r}}$ edges contains at least $\gamma\binom{n}{r}$ copies of $K_{r,r}$.

Let G be any graph with $e > c \cdot n^{2-\frac{1}{r}}$ edges. It is not hard to see that, in order to find a copy of H in G, it suffices to find distinct vertices $u_1, u_2, \ldots, u_{r+m}$ and $v_1, v_2, \ldots, v_{r+m}$ in $V(G)$ such that

(i) $u_i v_j \in E(G)$ unless $i > r$ and $j > r$;

(ii) $d_G(\{u_{i_1}, \ldots, u_{i_r}\}) \geq |V(H)|$ and $d_G(\{v_{i_1}, \ldots, v_{i_r}\}) \geq |V(H)|$ for $1 \leq i_1 < \ldots < i_r \leq r+m$.

We assign an auxiliary graph $\mathcal{G}$ to G as follows. The vertices of $\mathcal{G}$ are the r-sets in $V(G)$, and two such r-sets U and V are joined by an edge in $\mathcal{G}$ if $uv \in E(G)$ for every $u \in U$ and $v \in V$. Clearly, we have $\bar{d}(\mathcal{G}) \geq 2\gamma$.

Let us choose a uniformly random edge of $\mathcal{G}$ and let its endpoints be X and Y in uniformly random order. Observe that for any fixed r-set $U \in V(\mathcal{G})$, we have $\mathbb{P}(X = U) = \frac{d_{\mathcal{G}}(U)}{2e(\mathcal{G})}$. Let $u_1, \ldots, u_r$ be a uniformly random listing of the elements of X and let $v_1, \ldots, v_r$ be a uniformly random listing of the elements of Y. If $d_G(X) \geq r+m$ and $d_G(Y) \geq r+m$, then let $v_{r+1}, \ldots, v_{r+m}$ be chosen uniformly at random from $N_G(X) \setminus Y$ without repetition, and similarly, let $u_{r+1}, \ldots, u_{r+m}$ be chosen uniformly at random from $N_G(Y) \setminus X$ without repetition (otherwise, let $v_{r+1}, \ldots, v_{r+m}, u_{r+1}, \ldots, u_{r+m}$ be undefined).

It is clear that if $d_G(X) \geq r+m$ and $d_G(Y) \geq r+m$, then these choices satisfy condition (i) above. It remains to be shown that with positive probability condition (ii) is also satisfied.

But note that for any $1 \leq i_1 < \ldots < i_r \leq r+m$, the set $\{v_{i_1}, \ldots, v_{i_r}\}$ is a uniformly random neighbour in $\mathcal{G}$ of X, where, as noted above, $\mathbb{P}(X = U) = \frac{d_{\mathcal{G}}(U)}{2e(\mathcal{G})}$. Hence,

$$\begin{aligned}
\mathbb{P}(\{v_{i_1}, \ldots, v_{i_r}\} = V) &= \sum_{\substack{U \sim V \\ d_G(U) \geq r+m}} \mathbb{P}\big(X = U\big) \cdot \frac{1}{d_{\mathcal{G}}(U)} \\
&\leq \sum_{U \sim V} \mathbb{P}\big(X = U\big) \cdot \frac{1}{d_{\mathcal{G}}(U)} \\
&= \sum_{U \sim V} \frac{d_{\mathcal{G}}(U)}{2e(\mathcal{G})} \cdot \frac{1}{d_{\mathcal{G}}(U)} \\
&= \frac{d_{\mathcal{G}}(V)}{2e(\mathcal{G})}, \qquad (3.1)
\end{aligned}$$

where we write $U \sim V$ if U and V are neighbours in $\mathcal{G}$.

Now let $\mathcal{S}$ consist of those $V \in V(\mathcal{G})$ for which $d_{\mathcal{G}}(V) \leq \frac{\bar{d}(\mathcal{G})}{4\binom{r+m}{r}}$. By inequality (3.1), for every $1 \leq i_1 < \ldots < i_r \leq r+m$, we have

$$\mathbb{P}(\{v_{i_1}, \ldots, v_{i_r}\} \in \mathcal{S}) \leq \frac{1}{2e(\mathcal{G})} \sum_{V \in \mathcal{S}} d_{\mathcal{G}}(V) \leq \frac{1}{4\binom{r+m}{r}}.$$

Thus, with probability at least 3/4, $\{v_{i_1}, \ldots, v_{i_r}\} \notin \mathcal{S}$ for every $1 \le i_1 < \ldots < i_r \le r+m$. Similarly, with probability at least 3/4, $\{u_{i_1}, \ldots, u_{i_r}\} \notin \mathcal{S}$ holds for every $1 \le i_1 < \ldots < i_r \le r+m$. Hence, with probability at least 1/2, we have both $\{u_{i_1}, \ldots, u_{i_r}\} \notin \mathcal{S}$ and $\{v_{i_1}, \ldots, v_{i_r}\} \notin \mathcal{S}$ for every $1 \le i_1 < \ldots < i_r \le r+m$. But if $U \notin \mathcal{S}$, then $d_{\mathcal{G}}(U) > \frac{\gamma}{2\binom{r+m}{r}} \ge \binom{|V(H)|}{r}$. Therefore $d_G(U) \ge |V(H)|$ holds for all such U. It follows that with probability at least 1/2, the vertices $u_1, \ldots, u_{r+m}, v_1, \ldots, v_{r+m}$ are well-defined and have properties (i) and (ii). □

We now turn to the proof of Theorem 3.1.8.

Proof of Theorem 3.1.8. Let k be the size of the (r,t)-blownup tree and let $\gamma = \frac{3}{2}k \cdot \binom{10k^2t^2}{r}$. By Lemma 3.2.2, there exists a constant $c = c_r(\gamma)$ such that any graph on n vertices with $e > c \cdot n^{2-\frac{1}{r}}$ edges contains at least $\gamma\binom{n}{r}$ copies of $K_{r,r}$.

Let G be any graph with $e > c \cdot n^{2-\frac{1}{r}}$ edges. Define the auxiliary graph $\mathcal{G}$ as in the proof of Theorem 3.1.4. Clearly, we have $\bar{d}(\mathcal{G}) \ge 2\gamma$.

Let us define a random function f which is a partial graph homomorphism $L \to G$, i.e., if it is defined on $S \subset V(L)$, then it is a graph homomorphism $L[S] \to G$. We define f firstly on X_1, then on Y_1, Y_2, …, and finally on Y_k.

Let $f(X_1)$ be a random vertex of $\mathcal{G}$ according to the stationary distribution, that is, $f(X_1) = U$ with probability $\frac{d_{\mathcal{G}}(U)}{2e(\mathcal{G})}$. (Once $f(X_1) = U$ is decided, each bijection $X_1 \to U$ is chosen with equal probability.) If $d_G(f(X_1)) \ge t$, then let $f(Y_1)$ be a uniformly random t-subset of $N_G(f(X_1))$. Otherwise, let f be undefined on Y_1.

More generally, for $2 \le i \le k$, choose $j < i$ such that $X_i \subset Y_j$. If f is undefined on Y_j, then declare f to be undefined on Y_i. Otherwise, let $U = f(X_i)$. If $d_G(U) < t$, then let f be undefined on Y_i, while if $d_G(U) \ge t$, then let $f(Y_i)$ be a uniformly random t-subset of $N_G(U)$.

It is clear that this produces a partial graph homomorphism $L \to G$.

The key step in our proof is the following claim.

Claim. For each $1 \le i \le k$ and each $U \in V(\mathcal{G})$,

$$\mathbb{P}(f(X_i) = U) \le \frac{d_{\mathcal{G}}(U)}{2e(\mathcal{G})}.$$

Proof of Claim. Fix $1 \le i \le k$. Observe that there is a sequence $j_1 < \ldots < j_\ell = i$ such that $X_{j_1} = X_1$ and for each $1 \le a \le \ell-1$, we have $X_{j_{a+1}} \subset Y_{j_a}$. We prove by induction on a that for each $1 \le a \le \ell$ and every $U \in V(\mathcal{G})$, we have $\mathbb{P}(f(X_{j_a}) = U) \le \frac{d_{\mathcal{G}}(U)}{2e(\mathcal{G})}$. For $a = 1$, we have $X_{j_a} = X_1$, so $\mathbb{P}(f(X_{j_a}) = U) = \frac{d_{\mathcal{G}}(U)}{2e(\mathcal{G})}$. For $a \ge 2$, observe that conditional on $f(X_{j_{a-1}}) = V$, $f(Y_{j_{a-1}})$ is defined if and only if $d_G(V) \ge t$, and if this holds, then $f(Y_{j_{a-1}})$ is a uniformly random t-set in $N_G(V)$. Therefore in this case $f(X_{j_a})$ is a uniformly random r-set in $N_G(V)$, so if $U \subset N_G(V)$ then the probability that $f(X_{j_a}) = U$ is $\frac{1}{d_{\mathcal{G}}(V)}$. Hence, we have

$$\begin{aligned}
\mathbb{P}(f(X_{j_a}) = U) &= \sum_{\substack{V \sim U \\ d_G(V) \geq t}} \mathbb{P}\big(f(X_{j_{a-1}}) = V\big) \cdot \frac{1}{d_{\mathcal{G}}(V)} \\
&\leq \sum_{V \sim U} \mathbb{P}\big(f(X_{j_{a-1}}) = V\big) \cdot \frac{1}{d_{\mathcal{G}}(V)} \\
&\leq \sum_{V \sim U} \frac{d_{\mathcal{G}}(V)}{2e(\mathcal{G})} \cdot \frac{1}{d_{\mathcal{G}}(V)} \\
&= \frac{d_{\mathcal{G}}(U)}{2e(\mathcal{G})},
\end{aligned}$$

where we write $V \sim U$ if U and V are neighbours in $\mathcal{G}$. This completes the induction step, and the case $a = \ell$ proves the claim.

Now let $\mathcal{S}$ consist of those $U \in V(\mathcal{G})$ for which $d_{\mathcal{G}}(U) \leq \frac{\bar{d}(\mathcal{G})}{3k}$. By the claim above, for every i, we have $\mathbb{P}(f(X_i) \in \mathcal{S}) \leq \frac{1}{2e(\mathcal{G})} \sum_{U \in \mathcal{S}} d_{\mathcal{G}}(U) \leq \frac{1}{3k}$. Thus, with probability at least $1/3$, $f(X_i) \notin \mathcal{S}$ for every i. Moreover, for any $U \in V(\mathcal{G}) \setminus \mathcal{S}$ we have $d_G(U) \geq t$, so if $f(X_i) \notin \mathcal{S}$ for every i, then f is defined everywhere.

Suppose that $f(X_i) = U$ for some $U \in V(\mathcal{G})$ with $d_{\mathcal{G}}(U) > \frac{\bar{d}(\mathcal{G})}{3k}$. Then $d_{\mathcal{G}}(U) > \binom{10k^2t^2}{r}$, so $d_G(U) > 10k^2t^2$. But $f(Y_i)$ is a uniformly random t-subset of $N_G(U)$, and $|f(\bigcup_{0 \leq j \leq i-1} Y_j)| \leq kt$, so the probability that $f(Y_i) \cap f(\bigcup_{0 \leq j \leq i-1} Y_j) \neq \emptyset$ is at most $\frac{1}{3k}$.

It follows that with probability at least $1/3$, f defines an injective graph homomorphism $L \to G$, thus G contains L as a subgraph. □

Given Theorem 3.1.8, it is not hard to deduce Theorem 3.1.6. Clearly, it suffices to prove that any r-degenerate blow-up of a tree is a subgraph of some (r,t)-blownup tree. We will in fact prove the following stronger statement.

Lemma 3.2.3. *Let H be a blow-up of some tree T, and suppose that H is r-degenerate. For each $u \in V(T)$, write $I(u)$ for the independent set with which the vertex u is replaced in F. Then there exists some $t = t(H)$ and an (r,t)-blownup tree L with sets $X_1, \dots, X_k$, $Y_0, \dots, Y_k$ as in Definition 3.1.7 such that there is an embedding of H in L in a way that each $I(u)$ is a subset of some Y_i for $0 \leq i \leq k$.*

Proof. The proof is by induction on the size of T. If T has one vertex, the assertion is trivial. Now assume that T has at least two vertices. The assertion is straightforward when T is a star, so let us assume that that is not the case. Let x be an arbitrary vertex of T and let u be a vertex with maximum distance from x. Clearly u is a leaf. Let v be the unique neighbour of u in T. Since T is not a star, we have $v \neq x$.

If $|I(v)| \leq r$, then by induction there exist integers t, k and an (r,t)-blownup tree L with sets $X_1, \dots, X_k$, $Y_0, \dots, Y_k$ such that there is an embedding of $H - I(u)$ in L in a way that for each $y \in V(T) \setminus \{u\}$, $I(y)$ is a subset of some Y_i. In particular, $I(v)$ is a subset of some Y_i, so we can take $X_{k+1} = I(v)$ and $Y_{k+1} = I(u)$ to get an embedding of H in an (r, t')-blownup tree L' of size $k + 1$ with $t' = \max(t, |I(u)|)$.

We may therefore assume that $|I(v)| > r$. Then

$$\sum_{w \in V(T):\ wv \in E(T)} |I(w)| \leq r,$$

for otherwise H contains $K_{r+1,r+1}$ as a subgraph and so is not r-degenerate. Let z be the unique neighbour of v on the path between v and x and let $u_1, \ldots, u_m$ be the other neighbours of v. Now $T - \{v, u_1, \ldots, u_m\}$ is a tree, so by induction there exist integers t, k and an (r, t)-blownup tree L with sets $X_1, \ldots, X_k$, $Y_0, \ldots, Y_k$ such that there is an embedding of $H - (I(v) \cup \bigcup_{j \leq m} I(u_j))$ in L in a way that for each $y \in V(T) \setminus \{v, u_1, \ldots, u_m\}$, $I(y)$ is a subset of some Y_i. In particular, $I(z)$ is a subset of some Y_i. Now if we replace Y_i with $Y_i' = Y_i \cup \bigcup_{j \leq m} I(u_j)$ and set $X_{k+1} = I(z) \cup \bigcup_{j \leq m} I(u_j) \subset Y_i'$ and $Y_{k+1} = I(v)$, then we get an embedding of F in an (r, t')-blownup tree L' of size $k + 1$ with $t' = \max(t, |Y_i'|, |I(v)|)$. $\square$

3.3 The 2-blowup of the hexagon

3.3.1 Outline of the proof

Before we get on with the proof of Theorem 3.1.12, let us give a brief sketch of the argument. First, using a standard reduction lemma, we will assume that our host graph G is nearly regular. Then we will find many copies of $P_3[2]$ in G with a fixed pair of endpoints (x_1, x_2). Here and below, $P_3[2]$ has vertices $x_1, x_2, y_1, y_2, z_1, z_2, w_1, w_2$ and edges $x_i y_j, y_i z_j, z_i w_j$. It is not hard to see that if G has minimum degree $\omega(n^{2/3})$, then for some pair $(x_1, x_2) \in V(G)^2$, $\omega(n^2)$ such copies can be found. This means that there will be $\omega(1)$ among these copies that share the same (w_1, w_2). If we take t such $P_3[2]$'s, their union is a homomorphic copy of $\theta_{3,t}[2]$. However, it may be a degenerate one, i.e. some of the internal vertices may coincide in the t copies of $P_3[2]$. In order to prevent this from happening, we will only use $P_3[2]$'s in the above argument which satisfy some extra properties. For example, we only count those $P_3[2]$'s for which $d(x_1, x_2, z_1, z_2) < 6t$ and $d(y_1, y_2, w_1, w_2) < 6t$. Lemma 3.3.2 below will show that we do not lose too many $P_3[2]$'s by doing so. We will also make sure that in all our $P_3[2]$'s, the codegree $d(z_1, z_2)$ is roughly the same. Finally, we will insist that $d(x_1, x_2, z_1)$ and $d(x_1, x_2, z_2)$ are not too large. In Lemma 3.3.4, we show that we have many $P_3[2]$'s possessing all these properties. Then we find many pairs of these $P_3[2]$'s which share the same endpoints. Using the extra properties of our $P_3[2]$'s, we can argue that (unless G contains $\theta_{3,t}[2]$) it is not possible that most pairs share an internal vertex. Hence, using these internally vertex-disjoint copies of $P_3[2]$, we get a $\theta_{3,t}[2]$ in G.

The next proof naturally splits into two main parts. In Subsection 3.3.2, we show that we have many $P_3[2]$'s with the required properties. In Subsection 3.3.3, we show that there cannot be too many pairs of these $P_3[2]$'s which share the same endpoints and an internal vertex.

It is well known that any graph with e edges contains a bipartite subgraph with at least $e/2$ edges. This observation, combined with Lemma 2.2.1, reduces Theorem 3.1.12 to the following statement.

Theorem 3.3.1. *Let K be a constant and let G be a K-almost-regular bipartite graph on n vertices with minimum degree $\delta = \omega(n^{2/3})$. Then, for n sufficiently large, G contains a copy of $\theta_{3,t}[2]$.*

3.3.2 Prescribing extra properties

As mentioned in the proof outline (Section 3.3.1), our first prescribed property is that the quadruples (x_1, x_2, z_1, z_2) and (y_1, y_2, w_1, w_2) should have few common neighbours. The next lemma will be used to achieve this.

Lemma 3.3.2. *Let G be a $\theta_{3,t}[2]$-free graph. Let x, x', y and y' be distinct vertices in G and let $R \subset N(y, y') \setminus \{x, x'\}$. Then the number of pairs of distinct vertices (z, z') in R with $d(x, x', z, z') \geq 6t$ is at most $4t|R|$.*

Proof. Take a maximal set of pairs $(z_1, z_1'), \dots, (z_s, z_s') \in R^2$ such that $z_1, z_1', \dots, z_s, z_s'$ are all distinct and $d(x, x', z_i, z_i') \geq 6t$ for every i. If $s \geq t$, then we may choose $w_1, w_1', \dots, w_t, w_t' \in V(G)$ such that x, x', y, y', z_i, z_i' $(1 \leq i \leq t)$ and w_j, w_j' $(1 \leq j \leq t)$ are all distinct, and $w_i, w_i' \in N(x, x', z_i, z_i')$ for all $1 \leq i \leq t$. Then the vertices x, x', y, y', z_i, z_i' $(1 \leq i \leq t)$ and w_j, w_j' $(1 \leq j \leq t)$ form a copy of $\theta_{3,t}$, which is a contradiction.

Thus, $s < t$. By maximality, for any $(z, z') \in R^2$ with $d(x, x', z, z') \geq 6t$ we have $\{z, z'\} \cap \{z_1, z_1', \dots, z_s, z_s'\} \neq \emptyset$. This leaves at most $2 \cdot 2s \cdot |R| < 4t|R|$ possibilities for such (z, z'). □

Roughly speaking, the next lemma will be used to find $P_3[2]$'s with the property that $d(x_1, x_2, z_1)$ and $d(x_1, x_2, z_2)$ are not too large. For a set $S \subset V(G)$, we write $d_S(v) = |N(v) \cap S|$.

Lemma 3.3.3. *Let G be a graph on n vertices with minimum degree $\delta = \omega(n^{2/3})$. Let $S \subset V(G)$ have size $s \geq n^{1/3}$. Then there exists some $\lambda = \omega(1)$ such that the number of vertices $v \in V(G)$ with $\frac{\lambda}{2} \frac{s}{n^{1/3}} < d_S(v) \leq \lambda \frac{s}{n^{1/3}}$ is at least $c \delta n^{1/3} \lambda^{-11/10}$, where $c = (\sum_{i \geq 0} 2^{-i/10})^{-1}$.*

Proof. Define $U_0 = \{v \in V(G) : d_S(v) \leq \frac{s}{n^{1/3}}\}$, and for every positive integer i, let $U_i = \{v \in V(G) : \frac{s}{n^{1/3}} 2^{i-1} < d_S(v) \leq \frac{s}{n^{1/3}} 2^i\}$.

Now we double count the number of edges between S and $V(G)$ (viewed as a bipartite graph). On the one hand, every $y \in S$ has at least δ neighbours in $V(G)$. On the other hand, any $v \in U_i$ has at most $\frac{s}{n^{1/3}}2^i$ neighbours in S. Thus,

$$\sum_{i\geq 0} |U_i| \frac{s}{n^{1/3}} 2^i \geq s\delta,$$

so

$$\sum_{i\geq 0} |U_i| 2^i \geq \delta n^{1/3}.$$

It is easy to see that then there exists some i such that $|U_i| \geq 2^{-\frac{11i}{10}} c\delta n^{1/3}$. Since $|U_i| \leq n$, we have $i = \omega(1)$. So we may take $\lambda = 2^i$. $\square$

The next lemma lists almost all properties that we require about the vertices x_1, x_2, y_1, y_2, z_1, z_2, w_1, w_2 (discussed in the proof outline). The one additional property that we will need is that $d(y_1, y_2, w_1, w_2) < 6t$.

Lemma 3.3.4. *Let K be a constant and let G be a K-almost-regular, $\theta_{3,t}[2]$-free graph on n vertices with minimum degree $\delta = \omega(n^{2/3})$. Then there exist distinct vertices x_1, x_2 in G and a set $S \subset N(x_1, x_2)$ of size at least $n^{1/3}$ as follows. Writing $s = |S|$, there exist $\lambda = \omega(1)$, $\mu = \omega(1)$, and $\Omega(\frac{s^2 n^{2/3} \lambda^{27/10}}{\mu^{11/10}})$ tuples $(y_1, y_2, z_1, z_2) \in V(G)^4$ satisfying the following properties.*

1. *$y_1, y_2 \in S$ and $y_i z_j$ are edges for every i, j.*
2. *$x_1, x_2, y_1, y_2, z_1, z_2$ are distinct.*
3. *$d_S(z_1), d_S(z_2) \leq \lambda \frac{s}{n^{1/3}}$.*
4. *$d(x_1, x_2, z_1, z_2) < 6t$.*
5. *$\mu n^{1/3} \leq d(z_1, z_2) \leq 2\mu n^{1/3}$.*

Proof. Let $c = (\sum_{i\geq 0} 2^{-i/10})^{-1}$ as in Lemma 3.3.3. For every $R \subset V(G)$ of size at least $n^{1/3}$, define $\lambda(R)$ to be the largest λ such that the number of vertices v with $\frac{\lambda}{2}\frac{|R|}{n^{1/3}} < d_R(v) \leq \lambda \frac{|R|}{n^{1/3}}$ is at least $c\delta n^{1/3} \lambda^{-11/10}$. By Lemma 3.3.3, this is well-defined and $\lambda(R) = \omega(1)$. Since G has minimum degree $\omega(n^{2/3})$, for sufficiently large n, it is easy to see that there exist distinct $u, v \in V(G)$ with $d(u, v) \geq n^{1/3}$. Choose distinct $x_1, x_2 \in V(G)$ and $S \subset N(x_1, x_2)$ such that $|S| \geq n^{1/3}$ and $\lambda(S)$ is minimal among these choices. Let $\lambda = \lambda(S)$. It remains to find μ and enough number of tuples (y_1, y_2, z_1, z_2) with properties 1.-5.

This is done in two main steps.

Step 1. We find $\Omega(s^2 n^{2/3} \lambda^{9/5})$ tuples (y_1, y_2, z_1, z_2) satisfying properties 1., 2. and 3.

Let $U = \{v \in V(G)\backslash\{x_1, x_2\} : \frac{\lambda}{2}\frac{s}{n^{1/3}} < d_S(v) \leq \lambda\frac{s}{n^{1/3}}\}$. Then $|U| \geq c\delta n^{1/3}\lambda^{-11/10} - 2 \geq n\lambda^{-11/10}$ for n sufficiently large.

Clearly, the number of triples (y_1, y_2, z) with $y_1, y_2 \in S$ distinct, $z \in U$ and $y_1z, y_2z \in E(G)$ is at least $|U|(\frac{\lambda}{2}\frac{s}{n^{1/3}})(\frac{\lambda}{2}\frac{s}{n^{1/3}} - 1) = \Omega(s^2n^{1/3}\lambda^{9/10})$. Hence, on average, for a pair $y_1, y_2 \in S$ there are $\Omega(n^{1/3}\lambda^{9/10})$ vertices $z \in N(y_1, y_2) \cap U$. By convexity, on average, for a pair $y_1, y_2 \in S$ there are $\Omega(n^{2/3}\lambda^{9/5})$ pairs of distinct vertices $z_1, z_2 \in N(y_1, y_2) \cap U$. Since any $z \in U$ has $d_S(z) \leq \lambda\frac{s}{n^{1/3}}$, this completes Step 1.

Step 2. We find $\Omega(s^2n\lambda^{27/10})$ tuples (y_1, y_2, z_1, z_2, w) satisfying properties 1., 2., 3. and 4. with the additional properties that $d(z_1, z_2) \geq n^{1/3}\lambda^{4/5}$ and $z_1w, z_2w \in E(G)$.

For $y_1, y_2 \in S$, let $N(y_1, y_2)^* = \{v \in N(y_1, y_2) \setminus \{x_1, x_2\} : d_S(v) \leq \lambda\frac{s}{n^{1/3}}\}$. The conclusion of Step 1 implies that

$$\sum_{y_1, y_2 \in S \text{ distinct}} |N(y_1, y_2)^*|^2 = \Omega(s^2n^{2/3}\lambda^{9/5}).$$

Hence,

$$\sum_{\substack{y_1, y_2 \in S \text{ distinct} \\ |N(y_1, y_2)^*| \geq n^{1/3}}} |N(y_1, y_2)^*|^2 = \Omega(s^2n^{2/3}\lambda^{9/5}). \tag{3.2}$$

We now prove that for any distinct $y_1, y_2 \in S$ with $|N(y_1, y_2)^*| \geq n^{1/3}$, the number of triples (z_1, z_2, w) of distinct vertices with $(z_1, z_2) \in N(y_1, y_2)^*$, $d(x_1, x_2, z_1, z_2) < 6t$, $d(z_1, z_2) \geq n^{1/3}\lambda^{4/5}$ and $w \in N(z_1, z_2)$ is $\Omega(|N(y_1, y_2)^*|^2n^{1/3}\lambda^{9/10})$. Using equation (3.2), this would complete Step 2.

Let some distinct $y_1, y_2 \in S$ have $|N(y_1, y_2)^*| \geq n^{1/3}$. Let $R = N(y_1, y_2)^*$. By definition, the number of vertices v with $d_R(v) > \frac{\lambda(R)}{2}\frac{|R|}{n^{1/3}}$ is at least $c\delta n^{1/3}\lambda(R)^{-11/10}$. Thus, the number of triples of distinct vertices (z_1, z_2, w) with $z_1, z_2 \in R$ and $w \in N(z_1, z_2)$ is $\Omega(|R|^2\delta n^{-1/3}\lambda(R)^{9/10}) \geq \Omega(|R|^2\delta n^{-1/3}\lambda^{9/10})$. By Lemma 3.3.2, the number of pairs of distinct vertices (z_1, z_2) in R with $d(x_1, x_2, z_1, z_2) \geq 6t$ is at most $4t|R|$. Hence, the number of triples (z_1, z_2, w) involving such pairs (z_1, z_2) is at most $4t|R|\delta$. Note that $|R| \geq n^{1/3}$ and $\lambda = \omega(1)$, so $4t|R|\delta = o(|R|^2\delta n^{-1/3}\lambda^{9/10})$. Moreover, the number of triples (z_1, z_2, w) with $z_1, z_2 \in R$, $w \in N(z_1, z_2)$ and $d(z_1, z_2) \leq n^{1/3}\lambda^{4/5}$ is clearly at most $|R|^2n^{1/3}\lambda^{4/5}$, which is again $o(|R|^2n^{1/3}\lambda^{9/10})$. Thus, the number of triples (z_1, z_2, w) of distinct vertices with $(z_1, z_2) \in R$, $d(x_1, x_2, z_1, z_2) < 6t$, $d(z_1, z_2) \geq n^{1/3}\lambda^{4/5}$ and $w \in N(z_1, z_2)$ is $\Omega(|R|^2\delta n^{-1/3}\lambda^{9/10})$. This is $\Omega(|N(y_1, y_2)^*|^2n^{1/3}\lambda^{9/10})$, as claimed.

Using the conclusion of Step 2, there exists some positive integer j such that there exist $\Omega(\frac{s^2n\lambda^{27/10}}{2^{j/10}})$ tuples (y_1, y_2, z_1, z_2, w) satisfying properties 1., 2., 3. and 4. with the additional properties $n^{1/3}\lambda^{4/5}2^{j-1} \leq d(z_1, z_2) < n^{1/3}\lambda^{4/5}2^j$ and $z_1w, z_2w \in E(G)$. Take $\mu = \lambda^{4/5}2^{j-1} = \omega(1)$. Then the number of tuples (y_1, y_2, z_1, z_2) satisfying properties 1.-5. is $\Omega(\frac{s^2n\lambda^{27/10}}{2^{j/10}n^{1/3}\lambda^{4/5}2^j})$, which is $\Omega(\frac{s^2n^{2/3}\lambda^{27/10}}{\mu^{11/10}})$. $\square$

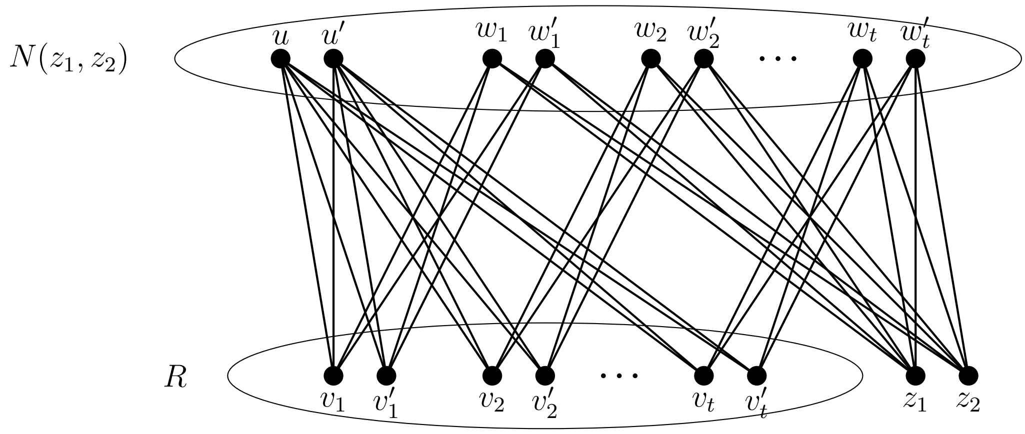

Figure 3.4: $\theta_{3,t}[2]$ in the proof of Lemma 3.3.6

3.3.3 Counting the number of pairs of $P_3[2]$'s which share an internal vertex

Lemma 3.3.5. *Let T be a tree with a special vertex v. Let G be a bipartite graph with parts X and Y of size at most n. Assume that G has $\omega(n^2)$ $K_{2,2}$'s. Then G contains a copy of $T[2]$ with the two images of v embedded in X.*

The proof of this lemma is similar to the proof of Theorem 3.1.5, so it is omitted.

Lemma 3.3.6. *Let G be a $\theta_{3,t}[2]$-free bipartite graph. Let z_1, z_2 be distinct vertices in G and let $N(z_1, z_2)$ have size $\ell = \omega(1)$. Let $q = \omega(\ell^{1/2})$. Let $R \subset \{v \in V(G) \setminus \{z_1, z_2\} : d(v, z_1, z_2) \geq q\}$. Then the number of triples (z', w_1, w_2) of distinct vertices with $z' \in R$, $w_1, w_2 \in N(z', z_1, z_2)$ is $O(\ell^2)$.*

Proof. Suppose that the number of triples (z', w_1, w_2) of distinct vertices with $z' \in R$, $w_1, w_2 \in N(z', z_1, z_2)$ is $\omega(\ell^2)$. Clearly we may assume that $|R| \leq \ell$. By assumption, on average a pair $(w_1, w_2) \in N(z_1, z_2)^2$ of distinct vertices has $\omega(1)$ common neighbours in R. Hence, there exist $\omega(\ell^2)$ many $K_{2,2}$'s in G with one part in $N(z_1, z_2)$ and the other in R. Since G is bipartite, we have $R \cap N(z_1, z_2) = \emptyset$. By Lemma 3.3.5, there exist distinct vertices $u, u', w_1, w_1', w_2, w_2', \ldots, w_t, w_t' \in N(z_1, z_2)$ and $v_1, v_1', v_2, v_2', \ldots, v_t, v_t' \in R$ such that $v_i, v_i' \in N(u, u', w_i, w_i')$ for every $1 \leq i \leq t$. Then the vertices $z_1, z_2, u, u', w_1, w_1', w_2, w_2', \ldots, w_t, w_t', v_1, v_1', v_2, v_2', \ldots, v_t, v_t'$ together form a copy of $\theta_{3,t}[2]$ (see Figure 3.4), which is a contradiction. $\square$

We are now in a position to complete the proof of Theorem 3.3.1.

Proof of Theorem 3.3.1. Assume for contradiction that G does not contain $\theta_{3,t}[2]$ as a subgraph. Choose $x_1, x_2, S, \lambda, \mu$ as in Lemma 3.3.4. Let $\mathcal{Q}$ be a set of $q = \Omega(\frac{s^2 n^{2/3} \lambda^{27/10}}{\mu^{11/10}})$ tuples (y_1, y_2, z_1, z_2) with the five properties given in Lemma 3.3.4. By property 5. and Lemma 3.3.2, any such tuple can be extended $\Theta(\mu^2 n^{2/3})$ ways to a tuple $(y_1, y_2, z_1, z_2, w_1, w_2)$ of vertices with the additional properties that w_1 and w_2 are distinct

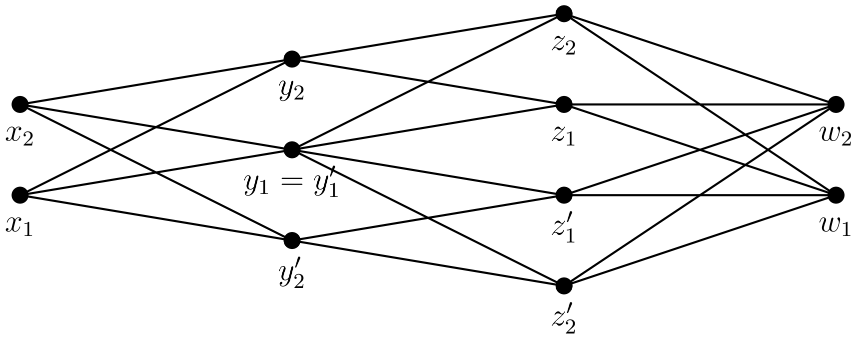

Figure 3.5: An element of $\mathcal{A}$

elements of $N(z_1, z_2) \setminus \{x_1, x_2, y_1, y_2\}$ and $d(y_1, y_2, w_1, w_2) < 6t$. Let $\mathcal{R}$ be the set of all tuples obtained this way and let $r = |\mathcal{R}|$. Note that $r = \Theta(q\mu^2 n^{2/3})$, so $r = \omega((\frac{\lambda s}{n^{1/3}})^2 n^2)$. Thus, on average a pair (w_1, w_2) of distinct vertices can be extended in $\omega((\frac{\lambda s}{n^{1/3}})^2)$ ways to a tuple $(y_1, y_2, z_1, z_2, w_1, w_2) \in \mathcal{R}$.

Assume that a pair (w_1, w_2) can be extended to $h = \omega((\frac{\lambda s}{n^{1/3}})^2)$ such tuples. Find a maximal set of disjoint tuples $(y_1^1, y_2^1, z_1^1, z_2^1), (y_1^2, y_2^2, z_1^2, z_2^2), \ldots, (y_1^k, y_2^k, z_1^k, z_2^k)$ such that $(y_1^i, y_2^i, z_1^i, z_2^i, w_1, w_2) \in \mathcal{R}$ for every $1 \leq i \leq k$. Since G is $\theta_{3,t}[2]$-free, we have $k < t$. Now for any y_1, y_2, z_1, z_2 with $(y_1, y_2, z_1, z_2, w_1, w_2) \in \mathcal{R}$, we have $\{y_1, y_2, z_1, z_2\} \cap \{y_1^1, y_2^1, z_1^1, z_2^1, y_1^2, y_2^2, z_1^2, z_2^2, \ldots, y_1^k, y_2^k, z_1^k, z_2^k\} \neq \emptyset$. By the pigeon hole principle, there exists some $v \in \{y_1^1, y_2^1, z_1^1, z_2^1, y_1^2, y_2^2, z_1^2, z_2^2, \ldots, y_1^k, y_2^k, z_1^k, z_2^k\}$ such that at least one of the following holds.

(i) There are at least $\frac{h}{16k}$ tuples $(y_1, y_2, z_1, z_2, w_1, w_2) \in \mathcal{R}$ with $y_1 = v$.

(ii) There are at least $\frac{h}{16k}$ tuples $(y_1, y_2, z_1, z_2, w_1, w_2) \in \mathcal{R}$ with $y_2 = v$.

(iii) There are at least $\frac{h}{16k}$ tuples $(y_1, y_2, z_1, z_2, w_1, w_2) \in \mathcal{R}$ with $z_1 = v$.

(iv) There are at least $\frac{h}{16k}$ tuples $(y_1, y_2, z_1, z_2, w_1, w_2) \in \mathcal{R}$ with $z_2 = v$.

If $(y_1, y_2, z_1, z_2, w_1, w_2) \in \mathcal{R}$, then by property 3. in Lemma 3.3.4, $d_S(z_1) \leq \frac{\lambda s}{n^{1/3}}$, by property 1. we have $y_1, y_2 \in S$, and finally $d(y_1, y_2, w_1, w_2) < 6t$. Thus, there are at most $(\frac{\lambda s}{n^{1/3}})^2 \cdot 6t$ ways to extend a fixed choice of z_1, w_1, w_2 to get $(y_1, y_2, z_1, z_2, w_1, w_2) \in \mathcal{R}$. In particular, (since in our case $h = \omega((\frac{\lambda s}{n^{1/3}})^2)$), case (iii) is impossible. Similarly, case (iv) is impossible. Thus, either case (i) or case (ii) holds.

Assume, without loss of generality, that case (i) holds. Since $d(y_1, y_2, w_1, w_2) < 6t$, for any $u \in V(G)$ there are at most $(6t)^2$ tuples $(y_1, y_2, z_1, z_2, w_1, w_2) \in \mathcal{R}$ with $y_1 = v, y_2 = u$. Moreover, for any $u \in V(G)$ there are at most $\frac{\lambda s}{n^{1/3}} \cdot 6t$ tuples $(y_1, y_2, z_1, z_2, w_1, w_2) \in \mathcal{R}$ with $y_1 = v, z_1 = u$, and there are at most $\frac{\lambda s}{n^{1/3}} \cdot 6t$ tuples $(y_1, y_2, z_1, z_2, w_1, w_2) \in \mathcal{R}$ with $y_1 = v, z_2 = u$. Hence, almost all pairs from our at least $\frac{h}{16k}$ tuples $(y_1, y_2, z_1, z_2, w_1, w_2) \in \mathcal{R}$ with $y_1 = v$ are disjoint apart from y_1, w_1 and w_2. Thus, for our fixed w_1, w_2, there are $\Omega(h^2)$ pairs $(y_1, y_2, z_1, z_2, w_1, w_2), (y_1', y_2', z_1', z_2', w_1, w_2) \in \mathcal{R}$ with $y_1 = y_1'$ but $\{y_2, z_1, z_2\} \cap \{y_2', z_1', z_2'\} = \emptyset$.

Summing over all pairs (w_1, w_2) and noting the symmetry of cases (i) and (ii) above, we get $\Omega(n^2 \cdot (\frac{r}{n^2})^2) = \Omega(\frac{r^2}{n^2})$ pairs $(y_1, y_2, z_1, z_2, w_1, w_2), (y_1', y_2', z_1', z_2', w_1, w_2) \in \mathcal{R}$ with $y_1 = y_1'$ but $\{y_2, z_1, z_2\} \cap \{y_2', z_1', z_2'\} = \emptyset$. Let $\mathcal{A}$ be the set of all tuples $(y_1, y_2, z_1, z_2, w_1, w_2, y_2', z_1', z_2')$ for which $(y_1, y_2, z_1, z_2, w_1, w_2), (y_1, y_2', z_1', z_2', w_1, w_2) \in \mathcal{R}$ and $\{y_2, z_1, z_2\} \cap \{y_2', z_1', z_2'\} = \emptyset$ (see Figure 3.5). Then $|\mathcal{A}| = \Omega(\frac{r^2}{n^2})$.

Note that for any $(y_1, y_2, z_1, z_2, w_1, w_2, y_2', z_1', z_2') \in \mathcal{A}$, we have $(y_1, y_2, z_1, z_2) \in \mathcal{Q}$ and $y_2' \in S$, so there are at most qs choices for y_1, y_2, z_1, z_2, y_2'. Hence, on average there are $\Omega(\frac{r^2}{n^2 qs})$ ways to extend such a choice to an element of $\mathcal{A}$. Note that $\frac{r^2}{n^2 qs} = \Theta(\frac{q\mu^4}{n^{2/3}s}) \geq \Omega(s\lambda^{27/10}\mu^{29/10}) \geq \Omega(n^{1/3}\mu^{29/10})$.

Let y_1, y_2, z_1, z_2, y_2' be vertices which extend in $g = \Omega(n^{1/3}\mu^{29/10})$ ways to an element of $\mathcal{A}$. Similarly to the pigeon hole argument above, there must exist a vertex v such that at least one of the following holds.

(i) There are at least $\frac{g}{16t}$ ways to extend y_1, y_2, z_1, z_2, y_2' to an element of $\mathcal{A}$ with $z_1' = v$.

(ii) There are at least $\frac{g}{16t}$ ways to extend y_1, y_2, z_1, z_2, y_2' to an element of $\mathcal{A}$ with $z_2' = v$.

(iii) There are at least $\frac{g}{16t}$ ways to extend y_1, y_2, z_1, z_2, y_2' to an element of $\mathcal{A}$ with $w_1 = v$.

(iv) There are at least $\frac{g}{16t}$ ways to extend y_1, y_2, z_1, z_2, y_2' to an element of $\mathcal{A}$ with $w_2 = v$.

Suppose that case (iii) holds. Then there are $\Omega(n^{1/3}\mu^{29/10})$ ways to extend $y_1, y_2, z_1, z_2, y_2', w_1$ to an element of $\mathcal{A}$. However, in any element of $\mathcal{A}$, we have $w_2 \in N(z_1, z_2)$ and, by property 5., $d(z_1, z_2) \leq 2\mu n^{1/3}$. Moreover, $z_1', z_2' \in N(y_1, y_2', w_1, w_2)$ and $d(y_1, y_2', w_1, w_2) < 6t$, so there are at most $2\mu n^{1/3} \cdot (6t)^2$ ways to extend $y_1, y_2, z_1, z_2, y_2', w_1$ to an element of $\mathcal{A}$. This contradicts $\mu = \omega(1)$, so either case (i) or case (ii) must hold. Without loss of generality, assume that (i) holds.

The number of ways to extend $y_1, y_2, z_1, z_2, y_2', z_1'$ to an element of $\mathcal{A}$ is at most $d(z_1, z_2, z_1')^2 \cdot 6t$, so we must have $d(z_1, z_2, v) \geq (\frac{g}{16t \cdot 6t})^{1/2} \geq n^{1/6}\mu^{28/20}$ when n is sufficiently large. So for our fixed choice of y_1, y_2, z_1, z_2, y_2' there are at least $\frac{g}{16t}$ ways to extend to an element of $\mathcal{A}$ such that $d(z_1, z_2, z_1') \geq n^{1/6}\mu^{28/20}$ holds. Summing over all y_1, y_2, z_1, z_2, y_2', we obtain $\Theta(|\mathcal{A}|) = \Omega(\frac{r^2}{n^2})$ elements of $\mathcal{A}$ in which $d(z_1, z_2, z_1') \geq \mu^{28/20}n^{1/6}$.

We now prove that this is impossible by counting such elements of $\mathcal{A}$ in a different way. Note that $(y_1, y_2, z_1, z_2) \in \mathcal{Q}$, so there are at most q choices for these vertices. For any such choice $\mu n^{1/3} \leq d(z_1, z_2) \leq 2\mu n^{1/3}$. Since $\mu^{28/20}n^{1/6} = \omega((2\mu n^{1/3})^{1/2})$, Lemma 3.3.6 implies that there are $O((\mu n^{1/3})^2)$ choices for (z_1', w_1, w_2). Moreover, there are at most $d_S(z_1') \leq \lambda \frac{s}{n^{1/3}}$ choices for y_2'. Finally, there are at most $d(y_1, y_2', w_1, w_2) < 6t$ choices for z_2'. Altogether, we find that there are $O(q \cdot (\mu n^{1/3})^2 \cdot \lambda \frac{s}{n^{1/3}} \cdot 6t)$ elements of $\mathcal{A}$ with $d(z_1, z_2, z_1') \geq \mu^{28/20}n^{1/6}$. But we have already seen that this number is $\Omega(\frac{r^2}{n^2}) = \Omega(q^2\mu^4 n^{-2/3})$, which is a contradiction since $q = \Omega(\frac{s^2 n^{2/3}\lambda^{27/10}}{\mu^{11/10}})$, $s \geq n^{1/3}$, $\lambda = \omega(1)$ and $\mu = \omega(1)$. $\square$

Chapter 4

The rainbow Turán number of even cycles

4.1 Introduction

In this chapter we develop a method that allows us to find cycles with suitable extra properties in graphs with sufficiently many edges. We give applications in three different areas, which are introduced in the next three subsections.

4.1.1 Rainbow Turán numbers

The following variant of the Turán number was introduced by Keevash, Mubayi, Sudakov and Verstraëte in [82]. In an edge-coloured graph, we say that a subgraph is rainbow if all its edges are of different colour. The rainbow Turán number of the graph H is then defined to be the maximum number of edges in a properly edge-coloured n-vertex graph that does not contain a rainbow H as a subgraph. This number is denoted by $\mathrm{ex}^*(n, H)$. Clearly, $\mathrm{ex}^*(n, H) \geq \mathrm{ex}(n, H)$ for every n and H. Keevash, Mubayi, Sudakov and Verstraëte proved, among other things, that for any non-bipartite graph H, we have $\mathrm{ex}^*(n, H) = (1+o(1))\mathrm{ex}(n, H)$. Hence, the most challenging case again seems to be when H is bipartite. Keevash, Mubayi, Sudakov and Verstraëte showed that $\mathrm{ex}^*(n, K_{s,t}) = O(n^{2-1/s})$, which is tight when $t > (s-1)!$ by the corresponding lower bound for $\mathrm{ex}(n, K_{s,t})$. The function has also been studied for trees (see [42, 77, 78]). About even cycles, Keevash, Mubayi, Sudakov and Verstraëte proved the following lower bound.

Theorem 4.1.1 (Keevash–Mubayi–Sudakov–Verstraëte [82]). *For any $k \geq 2$,*

$$\mathrm{ex}^*(n, C_{2k}) = \Omega(n^{1+1/k}).$$

They conjectured that this is tight.

Conjecture 4.1.2 (Keevash–Mubayi–Sudakov–Verstraëte [82]). *For any $k \geq 2$,*

$$\mathrm{ex}^*(n, C_{2k}) = \Theta(n^{1+1/k}).$$

They have verified their conjecture for $k \in \{2, 3\}$. For general k, Das, Lee and Sudakov proved the following upper bound.

Theorem 4.1.3 (Das–Lee–Sudakov [26]). *For every fixed integer $k \geq 2$,*

$$\mathrm{ex}^*(n, C_{2k}) = O\left(n^{1+\frac{(1+\varepsilon_k)\ln k}{k}}\right),$$

where $\varepsilon_k \to 0$ as $k \to \infty$.

In this chapter we prove Conjecture 4.1.2 by establishing the following result.

Theorem 4.1.4. *For any integer $k \geq 2$, we have*

$$\mathrm{ex}^*(n, C_{2k}) = O(n^{1+1/k}).$$

We remark that our proof can be easily modified to show that $\mathrm{ex}^*(n, \theta_{k,t}) = O(n^{1+1/k})$ for any fixed k and t.

Keevash, Mubayi, Sudakov and Verstraëte also asked how many edges a properly edge-coloured n-vertex graph can have if it does not contain any rainbow cycle. They constructed such graphs with $\Omega(n \log n)$ edges. Note that this is quite different from the uncoloured case, since any n-vertex acyclic graph has at most $n - 1$ edges. Das, Lee and Sudakov proved that if $\eta > 0$ and n is sufficiently large, then any properly edge-coloured n-vertex graph with at least $n \exp\left((\log n)^{\frac{1}{2}+\eta}\right)$ edges contains a rainbow cycle. We prove the following improvement.

Theorem 4.1.5. *There exists an absolute constant C such that if n is sufficiently large and G is a properly edge-coloured graph on n vertices with at least $Cn(\log n)^4$ edges, then G contains a rainbow cycle of even length.*

4.1.2 Colour-isomorphic even cycles in proper colourings

Conlon and Tyomkyn [25] have initiated the study of the following problem. We say that two subgraphs of an edge-coloured graph are colour-isomorphic if there is an isomorphism between them preserving the colours. For an integer $r \geq 2$ and a graph H, we write $f_r(n, H)$ for the smallest number C so that there is a proper edge-colouring of K_n with C colours containing no r pairwise vertex-disjoint colour-isomorphic copies of H. They proved various general results about this function, such as the following upper bound.

Theorem 4.1.6 (Conlon–Tyomkyn [25]). *For any graph H with v vertices and e edges,*

$$f_r(n, H) = O\left(\max\left(n, n^{\frac{rv-2}{(r-1)e}}\right)\right).$$

Regarding even cycles, they established the following result.

Theorem 4.1.7 (Conlon–Tyomkyn [25]). *$f_2(n, C_6) = \Omega(n^{4/3})$.*

One of the several open problems they posed is the following question.

Question 4.1.8 (Conlon–Tyomkyn [25]). *Is it true that for every $\varepsilon > 0$, there exists $k_0 = k_0(\varepsilon)$ such that, for all $k \geq k_0$, $f_2(n, C_{2k}) = \Omega(n^{2-\varepsilon})$?*

Later, Xu, Zhang, Jing and Ge made a more precise conjecture.

Conjecture 4.1.9 (Xu–Zhang–Jing–Ge [110]). *For any $k \geq 3$,*

$$f_2(n, C_{2k}) = \Omega(n^{2-\frac{2}{k}}).$$

We prove this conjecture in a more general form.

Theorem 4.1.10. *Let $k, r \geq 2$ be fixed integers. Then*

$$f_r(n, C_{2k}) = \Omega\left(n^{\frac{r}{r-1}\cdot\frac{k-1}{k}}\right).$$

4.1.3 Turán number of blow-ups of cycles

Now we return to the study of the extremal number of blow-ups of cycles, which was started in the previous chapter. In the case of forbidding all r-blowups of cycles, the following question was formulated by Jiang and Newman [73]. To state this question, we write $\mathcal{C}[r] = \{C_{2k}[r] : k \geq 2\}$.

Question 4.1.11 (Jiang–Newman [73]). *Is it true that for any positive integer r and any $\varepsilon > 0$, $\mathrm{ex}(n, \mathcal{C}[r]) = O(n^{2-\frac{1}{r}+\varepsilon})$?*

We answer this question affirmatively in a stronger form.

Theorem 4.1.12. *For any positive integer r,*

$$\mathrm{ex}(n, \mathcal{C}[r]) = O(n^{2-1/r}(\log n)^{7/r}).$$

Erdős–Rényi random graphs show that $\mathrm{ex}(n, \mathcal{C}[r]) = \Omega(n^{2-1/r})$. It would be interesting to decide whether the logarithmic factor in Theorem 4.1.12 can be removed.

We also establish an upper bound for the extremal number when only one blownup cycle is forbidden.

Theorem 4.1.13. *For any integers $r \geq 1$ and $k \geq 2$, we have*

$$\mathrm{ex}(n, C_{2k}[r]) = O\left(n^{2-\frac{1}{r}+\frac{1}{k+r-1}}(\log n)^{\frac{4k}{r(k+r-1)}}\right).$$

This is still quite a long way from the conjectured $\mathrm{ex}(n, C_{2k}[r]) = O(n^{2-\frac{1}{r}+\frac{1}{kr}})$. However, it can be used to disprove the following conjecture of Erdős and Simonovits.

Conjecture 4.1.14 (Erdős–Simonovits [31])**.** *Let H be a bipartite graph with minimum degree s. Then there exists $\varepsilon > 0$ such that* $\mathrm{ex}(n, H) = \Omega(n^{2-\frac{1}{s-1}+\varepsilon})$.

To see that this is false, note that the graph $C_{2k}[r]$ has minimum degree $2r$, but, by Theorem 4.1.13, for any $\delta > 0$, we have $\mathrm{ex}(n, C_{2k}[r]) = O(n^{2-\frac{1}{r}+\delta})$ for sufficiently large k. This means that we have, for any even s and any $\delta > 0$, a bipartite graph H with minimum degree s which has $\mathrm{ex}(n, H) = O(n^{2-\frac{2}{s}+\delta})$, disproving Conjecture 4.1.14. On the other hand, a simple application of the probabilistic method shows that if H is a bipartite graph with minimum degree $s \geq 2$, then there exists $\varepsilon > 0$ such that $\mathrm{ex}(n, H) = \Omega(n^{2-\frac{2}{s}+\varepsilon})$.

The rest of this chapter is organised as follows. In Section 4.2, we prove Theorem 4.1.4. In Section 4.3, we prove Theorem 4.1.5. In Section 4.4, we prove Theorem 4.1.10. The proofs of Theorem 4.1.12 and Theorem 4.1.13 are given in Section 4.5. We give some concluding remarks and open problems in Section 4.6.

While we see no implication relations between our results, the proofs in the three topics (rainbow Turán numbers, colour-isomorphic cycles and blow-ups of cycles) are very similar. In order to avoid repeating the same argument many times, we give the full proofs in the case of rainbow Turán problems, but we often only sketch the proofs in the sections on colour-isomorphic cycles and blow-ups of cycles. Nevertheless, we always indicate the necessary twists and in one case we give a proof in the appendix.

4.2 Rainbow cycles of length $2k$

Notation. In what follows, for graphs H and G we write $\mathrm{hom}(H, G)$ for the number of graph homomorphisms $V(H) \to V(G)$. P_k will denote the path with k edges and we use the convention $C_2 = P_1$. For vertices $x, y \in V(G)$, $\mathrm{hom}_{x,y}(P_\ell, G)$ denotes the number of walks of length ℓ in G between x and y. Moreover, $\mathrm{hom}_x(P_\ell, G)$ denotes the number of walks of length ℓ in G starting at x. Finally, we write $\delta(G)$ and $\Delta(G)$ for the minimum and maximum degree of G, respectively. Logarithms are base 2 unless stated otherwise.

4.2.1 Finding suitable short cycles

Our goal in this section is to develop a method for finding 'suitable' cycles of given length. This is done in two steps. In this subsection we prove that there exist 'suitable' cycles of length at most $2k$, while in the next subsection we push the method further to make sure that we get cycles of length exactly $2k$. We have been deliberately vague about what we mean by a 'suitable' cycle. In this section it will mean rainbow cycle, but in later sections it will have different meanings. For example, in both Section 4.4 and Section 4.5 we will work in auxiliary graphs whose vertices are sets, and a 'suitable' cycle in these auxiliary graphs will be one whose vertices are disjoint sets.

Our first key lemma is an upper bound on the number of those (homomorphic) 2ℓ-cycles which are not suitable. With a slight abuse of terminology, we call a homomorphism $H \to G$ a *homomorphic copy of H in G*. That is, a homomorphic copy of $C_{2\ell}$ is a tuple $(x_1, \ldots, x_{2\ell}) \in V(G)^{2\ell}$ such that $x_1x_2, x_2x_3, \ldots, x_{2\ell}x_1 \in E(G)$. A rainbow homomorphic copy of H is one in which the images of distinct edges of H have different colour.

Lemma 4.2.1. *Let $\ell \geq 2$ be a positive integer and let G be a properly edge-coloured graph. Then the number of homomorphic copies of $C_{2\ell}$ which are not rainbow is at most*

$$16\ell \left(\ell\Delta(G) \hom(C_{2\ell-2}, G) \hom(C_{2\ell}, G)\right)^{1/2}.$$

Proof. Let $c(e)$ be the colour of the edge $e \in E(G)$. We want to prove that the number of $(x_1, x_2, \ldots, x_{2\ell}) \in V(G)^{2\ell}$ with $x_1x_2, x_2x_3, \ldots, x_{2\ell}x_1 \in E(G)$ such that $c(x_1x_2), c(x_2x_3), \ldots, c(x_{2\ell}x_1)$ are not all distinct is at most $16\ell \left(\ell\Delta(G) \hom(C_{2\ell-2}, G) \hom(C_{2\ell}, G)\right)^{1/2}$. By symmetry, it suffices to prove that the number of $(x_1, x_2, \ldots, x_{2\ell}) \in V(G)^{2\ell}$ with $x_1x_2, x_2x_3, \ldots, x_{2\ell}x_1 \in E(G)$ such that $c(x_1x_2) = c(x_ix_{i+1})$ for some $2 \leq i \leq \ell + 1$ is at most $8 \left(\ell\Delta(G) \hom(C_{2\ell-2}, G) \hom(C_{2\ell}, G)\right)^{1/2}$.

For a positive integer s, let α_s be the number of walks of length $\ell - 1$ in G whose endpoints y and z have $2^{s-1} \leq \hom_{y,z}(P_{\ell-1}, G) < 2^s$ and let β_s be the number of walks of length ℓ in G whose endpoints y and z have $2^{s-1} \leq \hom_{y,z}(P_\ell, G) < 2^s$. Clearly,

$$\sum_{s \geq 1} \alpha_s 2^{s-1} \leq \hom(C_{2\ell-2}, G) \tag{4.1}$$

and

$$\sum_{s \geq 1} \beta_s 2^{s-1} \leq \hom(C_{2\ell}, G). \tag{4.2}$$

For positive integers s and t, write $\gamma_{s,t}$ for the number of homomorphic copies $x_1x_2 \ldots x_{2\ell}x_1$ of $C_{2\ell}$ such that $c(x_1x_2) = c(x_ix_{i+1})$ for some $2 \leq i \leq \ell + 1$, $2^{s-1} \leq \hom_{x_1,x_{\ell+2}}(P_{\ell-1}, G) < 2^s$ and $2^{t-1} \leq \hom_{x_2,x_{\ell+2}}(P_\ell, G) < 2^t$. Observe that $\gamma_{s,t} \leq \alpha_s \cdot \Delta(G) \cdot 2^t$. Indeed, if $x_1x_2 \ldots x_{2\ell}x_1$ is a homomorphic $C_{2\ell}$ with $2^{s-1} \leq \hom_{x_1,x_{\ell+2}}(P_{\ell-1}, G) < 2^s$ and $2^{t-1} \leq \hom_{x_2,x_{\ell+2}}(P_\ell, G) < 2^t$, then there are at most α_s ways to choose $(x_{\ell+2}, x_{\ell+3}, \ldots, x_{2\ell}, x_1)$, given such a choice there are at most $\Delta(G)$ choices for x_2, and given these there are at most 2^t choices for $(x_3, \ldots, x_{\ell+1})$. On the other hand, $\gamma_{s,t} \leq \beta_t \cdot \ell \cdot 2^s$. Indeed, there are at most β_t ways to choose $(x_2, \ldots, x_{\ell+2})$. Given such a choice, there are at most ℓ possibilities for x_1, since $c(x_1x_2) = c(x_ix_{i+1})$ for some $2 \leq i \leq \ell + 1$, the edges $x_2x_3, \ldots, x_{\ell+1}x_{\ell+2}$ are already fixed and c is a proper colouring. Finally, there are at most 2^s ways to complete this to a suitable homomorphic copy of $C_{2\ell}$.

Clearly, the total number of homomorphic copies $x_1x_2 \ldots x_{2\ell}x_1$ of $C_{2\ell}$ with $c(x_1x_2) = c(x_ix_{i+1})$ for some $2 \leq i \leq \ell + 1$ is $\sum_{s,t \geq 1} \gamma_{s,t}$. We give an upper bound for this sum as follows. Let q be the integer for which $(\frac{\ell \hom(C_{2\ell},G)}{\Delta(G)\hom(C_{2\ell-2},G)})^{1/2} \leq 2^q < 2(\frac{\ell \hom(C_{2\ell},G)}{\Delta(G)\hom(C_{2\ell-2},G)})^{1/2}$. Now, using $\gamma_{s,t} \leq \beta_t \cdot \ell \cdot 2^s$ and equation (4.2),

$$\begin{aligned}\sum_{s,t:s\leq t-q} \gamma_{s,t} &\leq \ell \sum_{s,t:s\leq t-q} 2^s \beta_t \leq \ell \cdot \sum_{t\geq 1} 2^{t-q+1}\beta_t \leq \ell \cdot 2^{-q+2}\hom(C_{2\ell},G)\\ &\leq 4(\ell\Delta(G)\hom(C_{2\ell-2},G)\hom(C_{2\ell},G))^{1/2}.\end{aligned}$$

Also, using $\gamma_{s,t} \leq \alpha_s \cdot \Delta(G) \cdot 2^t$ and equation (4.1),

$$\begin{aligned}\sum_{s,t:s>t-q} \gamma_{s,t} &\leq \Delta(G) \sum_{s,t:s>t-q} 2^t \alpha_s \leq \Delta(G)\sum_{s\geq 1} 2^{s+q}\alpha_s \leq \Delta(G)2^{q+1}\hom(C_{2\ell-2},G)\\ &\leq 4(\ell\Delta(G)\hom(C_{2\ell-2},G)\hom(C_{2\ell},G))^{1/2}.\end{aligned}$$

Thus,

$$\sum_{s,t\geq 1} \gamma_{s,t} \leq 8(\ell\Delta(G)\hom(C_{2\ell-2},G)\hom(C_{2\ell},G))^{1/2}.$$

This completes the proof. □

Corollary 4.2.2. *Let $k \geq 2$ be an integer and let G be a properly edge-coloured non-empty graph on n vertices with* $\hom(C_{2k},G) \geq 2^{8k}k^{3k}n\Delta(G)^k$. *Then G contains a rainbow cycle of length at most $2k$.*

Proof. Let ℓ be the smallest positive integer satisfying

$$\hom(C_{2\ell},G) \geq 2^{8\ell}k^{3\ell}n\Delta(G)^\ell.$$

This is well-defined and $\ell \leq k$ by the condition of the lemma. Since $\hom(C_2,G) = 2e(G) \leq n\Delta(G)$, we have $\ell \geq 2$.

Note that

$$\hom(C_{2\ell-2},G) < 2^{8(\ell-1)}k^{3(\ell-1)}n\Delta(G)^{\ell-1} \leq \frac{\hom(C_{2\ell},G)}{2^8k^3\Delta(G)} \leq \frac{\hom(C_{2\ell},G)}{2^8\ell^3\Delta(G)},$$

so by Lemma 4.2.1, the number of homomorphic copies of $C_{2\ell}$ which are not rainbow is less than $\hom(C_{2\ell},G)$.

Hence, there is at least one homomorphic copy of $C_{2\ell}$ in G which is rainbow. This implies the existence of a rainbow cycle. Indeed, the homomorphic $C_{2\ell}$ uses every edge of G at most once (since it is rainbow), so it is a circuit. Thus, it has a subgraph which is a cycle. Clearly, this is a rainbow cycle. □

The next lemma is another instance of an upper bound for the number of certain kind of non-suitable homomorphic copies of $C_{2\ell}$, namely non-injective ones. In what follows, an *injectively homomorphic* copy of $C_{2\ell}$ is a homomorphic copy $(x_1, x_2, \ldots, x_{2\ell})$ of $C_{2\ell}$ where the vertices $x_1, \ldots, x_{2\ell}$ are distinct. That is, it is a labelled genuine $C_{2\ell}$.

Lemma 4.2.3. *Let $\ell \geq 2$ be a positive integer and let G be a graph. Then the number of homomorphic, but not injective copies of $C_{2\ell}$ in G is at most*

$$16\ell \left(\ell\Delta(G)\,\mathrm{hom}(C_{2\ell-2}, G)\,\mathrm{hom}(C_{2\ell}, G)\right)^{1/2}.$$

Proof. The proof is almost identical to the proof of Lemma 4.2.1. The only difference is that instead of bounding those homomorphic copies $(x_1, \ldots, x_{2\ell})$ with $c(x_1x_2) = c(x_ix_{i+1})$ for some $2 \leq i \leq \ell+1$, we bound those with $x_1 = x_i$ for some $2 \leq i \leq \ell+1$. All details go through exactly the same way. □

4.2.2 Finding a cycle of given length

In this subsection we develop the necessary tools to find a suitable cycle of length exactly $2k$ (rather than length at most $2k$ as in Corollary 4.2.2).

We will need the following lemma.

Lemma 4.2.4. *Let H be a bipartite graph and suppose that it does not contain a non-empty subgraph with minimum degree at least d. Then the largest eigenvalue of H is at most $2\sqrt{d\Delta(H)}$.*

We defer its simple proof until the next subsection and proceed with the main part of the argument. The next lemma is an easy corollary of Lemma 4.2.4. It will be used to compare $\mathrm{hom}_x(C_{2\ell-2}, G)$ with $\mathrm{hom}_x(C_{2\ell}, G)$, where $\mathrm{hom}_x(C_{2j}, G)$ denotes the number of homomorphic copies $(x_1, x_2, \ldots, x_{2j})$ of C_{2j} with $x_1 = x$.

Lemma 4.2.5. *Let H be a bipartite graph with parts Y and Z. Let $f : Y \to \mathbb{R}$ be a function and let $g(z) = \sum_{y \in N_H(z)} f(y)$ for every $z \in Z$. Suppose that H does not contain a non-empty subgraph with minimum degree at least d. Then*

$$\sum_{y \in Y} f(y)^2 \geq \frac{1}{4d\Delta(H)} \sum_{z \in Z} g(z)^2.$$

The next lemma is one of our key results.

Lemma 4.2.6. *Let k be a fixed positive integer and let G be a properly edge-coloured non-empty graph on n vertices. Suppose that for some $2 \leq \ell \leq k$ we have*

$$\mathrm{hom}(C_{2\ell}, G) \geq c_k \Delta(G) \mathrm{hom}(C_{2\ell-2}, G),$$

where $c_k = 2^{18}k^7$. Then G contains a rainbow C_{2k}.

Proof. Call a pair $(x_1, x_{\ell+1})$ of vertices *nice* if the number of rainbow injectively homomorphic copies of $C_{2\ell}$ of the form $x_1x_2 \ldots x_{2\ell}x_1$ is greater than $\left(1 - \frac{1}{\binom{4k}{2}}\right) \left(\mathrm{hom}_{x_1,x_{\ell+1}}(P_\ell, G)\right)^2$. Observe that the total number of homomorphic copies of $C_{2\ell}$ of the form $x_1x_2\ldots x_{2\ell}x_1$ is $\hom_{x_1,x_{\ell+1}}(P_\ell,G)^2$, so this means that the proportion of those which are not injective or not rainbow is less than $\frac{1}{\binom{4k}{2}}$. Hence, if we choose two walks of length ℓ between x_1 and $x_{\ell+1}$ randomly with replacement, then the probability that their concatenation is a non-injective or non-rainbow homomorphic copy of $C_{2\ell}$ is less than $\frac{1}{\binom{4k}{2}}$. In particular, if we choose $4k$ random walks of length ℓ between x_1 and $x_{\ell+1}$ with replacement, then with positive probability any two of these walks form a rainbow, injectively homomorphic copy of $C_{2\ell}$. Hence, there exist at least $4k$ pairwise internally vertex-disjoint paths between x_1 and $x_{\ell+1}$ such that no colour appears more than once on these paths.

By Lemmas 4.2.1 and 4.2.3, the number of non-rainbow or non-injective homomorphic copies of $C_{2\ell}$ in G is at most

$$32\ell^{3/2}(\Delta(G)\hom(C_{2\ell-2},G)\hom(C_{2\ell},G))^{1/2} \leq \frac{32\ell^{3/2}}{c_k^{1/2}}\hom(C_{2\ell},G).$$

Hence,

$$\sum_{(x_1,x_{\ell+1})\text{ not nice}} \frac{1}{\binom{4k}{2}}\hom_{x_1,x_{\ell+1}}(P_\ell,G)^2 \leq \frac{32\ell^{3/2}}{c_k^{1/2}}\hom(C_{2\ell},G),$$

so, using $\sum_{x_1,x_{\ell+1}\in V(G)}\hom_{x_1,x_{\ell+1}}(P_\ell,G)^2 = \hom(C_{2\ell},G)$, we have

$$\begin{aligned}\sum_{(x_1,x_{\ell+1})\text{ nice}} \hom_{x_1,x_{\ell+1}}(P_\ell,G)^2 &\geq \left(1-\binom{4k}{2}\frac{32\ell^{3/2}}{c_k^{1/2}}\right)\hom(C_{2\ell},G) > \frac{1}{2}\hom(C_{2\ell},G)\\ &\geq \frac{c_k}{2}\Delta(G)\hom(C_{2\ell-2},G).\end{aligned}$$

Thus, there exists some $x\in V(G)$ such that

$$\sum_{z\in V(G):(x,z)\text{ is nice}} \hom_{x,z}(P_\ell,G)^2 > \frac{c_k}{2}\Delta(G)\hom_x(C_{2\ell-2},G), \tag{4.3}$$

where $\hom_x(C_{2\ell-2},G)$ denotes the number of homomorphic copies $(x_1,\ldots,x_{2\ell-2})$ of $C_{2\ell-2}$ with $x_1=x$. Let $Z=\{z\in V(G):(x,z)\text{ is nice}\}$ and let $Y=V(G)$. Consider the bipartite graph H with parts Y and Z, defined by G. (We view Y and Z as disjoint sets even though they overlap as subsets of $V(G)$.)

Suppose that H does not contain a subgraph with minimum degree at least $4k$. Let $f(y)=\hom_{x,y}(P_{\ell-1},G)$ for every $y\in Y=V(G)$ and define g as in Lemma 4.2.5. By that lemma with $d=4k$,

$$\sum_{y\in Y} f(y)^2 \geq \frac{1}{16k\Delta(H)}\sum_{z\in Z} g(z)^2 \geq \frac{1}{16k\Delta(G)}\sum_{z\in Z} g(z)^2.$$

However, $g(z) = \sum_{y \in N_G(z)} \hom_{x,y}(P_{\ell-1}, G) = \hom_{x,z}(P_\ell, G)$, so, using equation (4.3),

$$\sum_{y \in Y} f(y)^2 \geq \frac{1}{16k\Delta(G)} \sum_{z \in Z} \hom_{x,z}(P_\ell, G)^2 > \frac{c_k}{32k} \hom_x(C_{2\ell-2}, G).$$

However, $\sum_{y \in Y} f(y)^2 = \hom_x(C_{2\ell-2}, G)$, which is a contradiction.

Thus, H contains a subgraph with minimum degree at least $4k$. Then we can greedily find a rainbow path of length $2k - 2\ell$ in G which avoids x and which have both endpoints in Z. Let this path be Q with endpoints z_1 and z_2. Since (x, z_1) is a nice pair, there exist at least $4k$ pairwise internally vertex-disjoint paths of length ℓ between x and z_1 such that any colour appears at most once on these paths. Thus, by avoiding the vertices and colours on Q, we can choose a path Q_1 of length ℓ between x and z_1 in a way that the concatenation of Q_1 and Q is a rainbow path of length $2k - \ell$. Moreover, since (x, z_2) is a nice pair, we can extend this path to a rainbow cycle of length $2k$. □

Corollary 4.2.7. *Let k be a fixed positive integer and let G be a properly edge-coloured non-empty graph on n vertices. Suppose that for some $2 \leq j \leq k$ we have*

$$\hom(C_{2j}, G) = \omega\left(n\Delta(G)^j\right).$$

Then, for n sufficiently large, G contains a rainbow C_{2k}.

Proof. Choose $L = \omega(1)$ such that $\hom(C_{2j}, G) \geq L^j n\Delta(G)^j$. Let ℓ be the smallest positive integer satisfying $\hom(C_{2\ell}, G) \geq L^\ell n\Delta(G)^\ell$. Clearly $\ell \leq j \leq k$, and since $\hom(C_2, G) \leq n\Delta(G)$, we have $\ell \geq 2$. Now $\hom(C_{2\ell}, G) \geq L\Delta(G) \hom(C_{2\ell-2}, G)$, so by Lemma 4.2.6, G contains a rainbow C_{2k}. □

Corollary 4.2.7 shows in particular that if we have many homomorphic cycles of length $2k$ and the maximum degree is not too large, then there exists a rainbow C_{2k}. Using Lemma 2.2.1, we can pass to a suitable almost regular subgraph in which we can apply Corollary 4.2.7 to find a rainbow C_{2k}.

Proof of Theorem 4.1.4. By Lemma 2.2.1, it suffices to prove that for any fixed K, if G' is a properly edge-coloured K-almost regular graph on m vertices with minimum degree $\delta = \omega(m^{1/k})$, then, for m sufficiently large, G' contains a rainbow C_{2k}.

It is well known that C_{2k} satisfies Sidorenko's conjecture, so

$$\hom(C_{2k}, G') \geq \frac{\hom(K_2, G')^{2k}}{m^{2k}} \geq \delta^{2k} \geq \frac{\delta^k}{mK^k} m\Delta(G')^k.$$

Then $\hom(C_{2k}, G') = \omega(m\Delta(G')^k)$, so by Corollary 4.2.7, G' contains a rainbow C_{2k}. □

4.2.3 The proof of Lemma 4.2.4

It remains to prove Lemma 4.2.4.

Lemma 4.2.8. *Let H be a bipartite graph with parts Y and Z. Suppose that H does not contain a non-empty subgraph with minimum degree at least d. Then there exist bipartite graphs H_1, H_2 both with parts Y and Z such that $E(H)$ is the disjoint union of $E(H_1)$ and $E(H_2)$, every vertex in Y has degree less than d in H_1 and every vertex in Z has degree less than d in H_2.*

Proof. Since H has minimum degree less than d, there is a vertex u in H which has degree less than d. If $u \in Y$, let every edge in H of the form uv belong to H_1, otherwise let every edge of the form uv belong to H_2. Set $H' = H - u$.

Since H' has minimum degree less than d, there is a vertex u' in H' which has degree less than d. If $u' \in Y$, let every edge in H' of the form $u'v$ belong to H_1, otherwise let every edge of the form $u'v$ belong to H_2. Set $H'' = H' - u'$.

Continue this procedure until all edges are placed in H_1 or H_2. It is easy to see that these graphs are suitable. $\square$

The next two lemmas are well known.

Lemma 4.2.9. *Let H be a bipartite graph with parts Y and Z so that every vertex in Y has degree at most D_1 and every vertex in Z has degree at most D_2. Then the largest eigenvalue of H is at most $\sqrt{D_1 D_2}$.*

Lemma 4.2.10. *Let A and B be symmetric real matrices with largest eigenvalues λ and μ. Then the largest eigenvalue of $A + B$ is at most $\lambda + \mu$.*

Proof of Lemma 4.2.4. Define graphs H_1 and H_2 as in Lemma 4.2.8. By Lemma 4.2.9, both H_1 and H_2 have largest eigenvalue at most $\sqrt{d\Delta(H)}$. Hence, by Lemma 4.2.10, the largest eigenvalue of H is at most $2\sqrt{d\Delta(H)}$. $\square$

4.3 Rainbow cycles of arbitrary length

In this section we prove Theorem 4.1.5. We will use Corollary 4.2.2, but we first have to find a 'regular enough' subgraph. Using Corollary 4.2.2, one can show that there exists a constant C such that any C-almost regular graph on n vertices with at least $Cn(\log n)^3$ edges contains a rainbow cycle. Unfortunately, we think that it is not possible to find a $O(1)$-almost regular subgraph on $m = \omega(1)$ vertices with $\omega(m(\log m)^3)$ edges in an arbitrary n-vertex graph with $\omega(n(\log n)^3)$ edges. The next two lemmas give us a suitable subgraph for which Corollary 4.2.2 is applied, but we lose a $\log n$ factor on the way, that is why we need $Cn(\log n)^4$ edges in Theorem 4.1.5.

Lemma 4.3.1. *Let d be sufficiently large and let G be a graph on n vertices with average degree d. Then there exists a non-empty bipartite subgraph G' of G with parts X and Y such that $e(G') \geq |X| \cdot \frac{\Delta(G')}{80}$ and $e(G') \geq |Y| \cdot \frac{d}{10 \log n}$.*

Proof. By passing to a suitable subgraph, we may, without loss of generality, assume that every subgraph of G has average degree at most d.

Let A be the set consisting of the $\lceil n/2 \rceil$ largest degree vertices in G (we break ties arbitrarily) and let $B = V(G) \setminus A$.

Suppose first that $e(G[B]) \geq \frac{e(G)}{10}$. Then we may partition B into sets X and Y such that $e(G[X,Y]) \geq \frac{e(G)}{20} = \frac{nd}{40}$. Let $G' = G[X,Y]$. Any vertex in B has degree at most $\frac{2e(G)}{\lceil n/2 \rceil} = \frac{nd}{\lceil n/2 \rceil} \leq 2d$ in G, so $\Delta(G') \leq 2d$. Since $|X|, |Y| \leq n/2$, G' satisfies the conditions in the lemma.

Hence, we may assume that $e(G[B]) < \frac{e(G)}{10}$. Suppose that $e(G[A]) \geq \frac{6e(G)}{10}$. Then $G[A]$ has larger average degree than G, which is a contradiction. Thus, $e(G[A]) < \frac{6e(G)}{10}$ and so $e(G[A,B]) \geq \frac{3e(G)}{10}$.

Let $A_{\text{low}} = \{x \in X : |N_G(x) \cap B| \leq \frac{d}{20}\}$ and let $A' = A \setminus A_{\text{low}}$. Clearly, $e(G[A_{\text{low}}, B]) \leq n\frac{d}{20} = \frac{e(G)}{10}$, so $e(G[A',B]) \geq \frac{e(G)}{5}$. For $0 \leq i \leq \lfloor \log n \rfloor$, let $A_i = \{x \in A' : 2^i \leq |N_G(x) \cap B| < 2^{i+1}\}$. The sets A_i partition A', so there exists some i such that $e(G[A_i, B]) \geq \frac{e(G[A',B])}{\log n + 1} \geq \frac{e(G)}{10 \log n} = \frac{nd}{20 \log n} \geq |B| \cdot \frac{d}{10 \log n}$.

Let $X = A_i$, $Y = B$ and $G' = G[X,Y]$. The last inequality from the previous paragraph gives that $e(G') \geq |Y| \cdot \frac{d}{10 \log n}$. Since every $x \in A_i$ has $\frac{d}{20} < d_{G'}(x) < 2^{i+1}$, we have $\frac{d}{20} < 2^{i+1}$. But every $y \in B$ has $d_{G'}(y) \leq d_G(y) \leq 2d$, so $\Delta(G') \leq 40 \cdot 2^{i+1}$. However, for every $x \in A_i$, we have $d_{G'}(x) \geq 2^i$, so $e(G') \geq |X| \cdot 2^i \geq |X| \cdot \frac{\Delta(G')}{80}$. □

Lemma 4.3.2. *Let d be sufficiently large and let G be a graph on n vertices with average degree d. Then there exists a non-empty bipartite subgraph G'' of G with parts X and Y such that for every $x \in X$, we have $d_{G''}(x) \geq \frac{\Delta(G'')}{160}$ and for every $y \in Y$, we have $d_{G''}(y) \geq \frac{d}{20 \log n}$.*

Proof. By Lemma 4.3.1, we may choose a non-empty bipartite subgraph G' with parts X' and Y' such that $e(G') \geq |X'| \cdot \frac{\Delta(G')}{80}$ and $e(G') \geq |Y'| \cdot \frac{d}{10 \log n}$. Now perform the following simple algorithm: as long as there is a vertex in X' which has degree less than $\frac{\Delta(G')}{160}$ in the current graph, or there is a vertex in Y' which has degree less than $\frac{d}{20 \log n}$ in the current graph, then discard one such vertex. Let the final graph be G'' and let its parts be X and Y. Clearly we have $d_{G''}(x) \geq \frac{\Delta(G')}{160} \geq \frac{\Delta(G'')}{160}$ for every $x \in X$ and $d_{G''}(y) \geq \frac{d}{20 \log n}$ for every $y \in Y$. Finally, G'' is non-empty since the number of edges discarded by the algorithm is less than $|X| \cdot \frac{\Delta(G')}{160} + |Y| \cdot \frac{d}{20 \log n} \leq e(G')$. □

Now we prove that the subgraph we find by Lemma 4.3.2 has many homomorphic C_{2k}'s.

Lemma 4.3.3. *Let G be a bipartite graph with parts X and Y such that $d(x) \geq s$ for every $x \in X$ and $d(y) \geq t$ for every $y \in Y$. Then, for every positive integer k,*

$$\hom(C_{2k}, G) \geq s^k t^k.$$

Proof. If k is even, then $\hom(P_k, G) \geq |X| s^{k/2} t^{k/2}$. Hence,

$$\begin{aligned}\hom(C_{2k}, G) &\geq \sum_{x,x' \in X} \hom_{x,x'}(P_k, G)^2 \geq \frac{1}{|X|^2} \left(\sum_{x,x' \in X} \hom_{x,x'}(P_k, G) \right)^2 \\ &\geq \left(\frac{\hom(P_k, G)}{|X|} \right)^2 \geq s^k t^k.\end{aligned}$$

Now suppose that k is odd. Without loss of generality, we may assume that $|X|s \geq |Y|t$. Note that $\hom(P_k, G) \geq |X| s^{\frac{k+1}{2}} t^{\frac{k-1}{2}}$. Hence,

$$\begin{aligned}\hom(C_{2k}, G) &\geq \sum_{x \in X, y \in Y} \hom_{x,y}(P_k, G)^2 \geq \frac{1}{|X||Y|} \left(\sum_{x \in X, y \in Y} \hom_{x,y}(P_k, G) \right)^2 \\ &\geq \frac{\hom(P_k, G)^2}{|X||Y|} \geq \frac{|X|}{|Y|} s^{k+1} t^{k-1} \geq s^k t^k.\end{aligned}$$

□

Lemma 4.3.4. *Let d be sufficiently large and let G be a graph on n vertices with average degree d. Then there exists a non-empty bipartite subgraph G'' of G such that for every positive integer k,*

$$\hom(C_{2k}, G'') \geq \left(\frac{d}{20 \log n} \right)^k \left(\frac{\Delta(G'')}{160} \right)^k.$$

Proof. This follows immediately from Lemma 4.3.2 and Lemma 4.3.3. □

Proof of Theorem 4.1.5. Let n be sufficiently large and let G be a properly edge-coloured graph on n vertices with at least $Cn(\log n)^4$ edges, where $C = 2^{100}$. Let $k = \lfloor \log n \rfloor$.

By Lemma 4.3.4, G has a non-empty bipartite subgraph G'' such that

$$\hom(C_{2k}, G'') \geq \left(\frac{C}{10} (\log n)^3 \right)^k \left(\frac{\Delta(G'')}{160} \right)^k \geq 2^{50k} k^{3k} \Delta(G'')^k \geq 2^{8k} k^{3k} n \Delta(G'')^k.$$

Then, by Corollary 4.2.2, G'' contains a rainbow cycle. It has even length because G'' is bipartite. □

4.4 Colour-isomorphic cycles

In this section we prove Theorem 4.1.10. Throughout the section, let k and r be fixed.

Definition 4.4.1. Given an edge-colouring of K_n, define an auxiliary graph $\mathcal{G}_0$ as follows. Let the vertex set of $\mathcal{G}_0$ be the set of r-vertex subsets of $V(K_n)$, i.e. let $V(\mathcal{G}_0) = V(K_n)^{(r)}$. Now let U and V be joined by an edge if $U \cap V = \emptyset$ and there is a monochromatic matching between U and V.

We will prove that if K_n is coloured with $o(n^{\frac{r}{r-1}\cdot\frac{k-1}{k}})$ colours, then there exists a copy of $\theta_{k,r!+1}$ in $\mathcal{G}_0$ in which the vertices are pairwise disjoint as subsets of $V(K_n)$. This implies that there exist r colour-isomorphic, pairwise vertex-disjoint copies of C_{2k}. Indeed, let X, $Y_{i,j}$ for $1 \le i \le k-1$, $1 \le j \le r!+1$ and Z be pairwise disjoint r-subsets of $V(K_n)$ with X joined to $Y_{1,j}$ in $\mathcal{G}_0$ for $1 \le j \le r!+1$, $Y_{i,j}$ joined to $Y_{i+1,j}$ for every $1 \le i \le k-2$ and every $1 \le j \le r!+1$ and $Y_{k-1,j}$ joined to Z for every $1 \le j \le r!+1$. For each $1 \le j \le r!+1$, pair each vertex in X with the vertex in Z that we get to if we follow the edges in the monochromatic matchings between $X, Y_{1,j}, Y_{2,j}, \dots, Y_{k-1,j}, Z$. This gives, for each $1 \le j \le r!+1$, a bijection between X and Z. Since there are $r!$ bijections between two sets of size r, two of these bijections must be identical, say the one corresponding to j_1 and the one corresponding to j_2. Then $X, Y_{1,j_1}, \dots, Y_{k-1,j_1}, Z, Y_{k-1,j_2}, \dots, Y_{1,j_2}$ and the monochromatic matchings between them provide r colour-isomorphic, pairwise vertex-disjoint copies of C_{2k}.

Lemma 4.4.2. *If K_n is properly edge-coloured with $o(n^{\frac{r}{r-1}\cdot\frac{k-1}{k}})$ colours, then $e(\mathcal{G}_0) = \omega(n^{r+r/k})$.*

Proof. By the convexity of the function $\binom{x}{r}$, the number of monochromatic r-matchings in K_n is $\omega\left(n^{\frac{r}{r-1}\cdot\frac{k-1}{k}} \cdot (n^{2-\frac{r}{r-1}\cdot\frac{k-1}{k}})^r\right) = \omega(n^{r+r/k})$. Any monochromatic r-matching gives rise to an edge in $\mathcal{G}_0$ and any edge in $\mathcal{G}_0$ is counted at most r times, so the statement of the lemma follows. $\square$

For the rest of the proof, we fix a proper edge-colouring of K_n with $o(n^{\frac{r}{r-1}\cdot\frac{k-1}{k}})$ colours and define $\mathcal{G}_0$ as above. Since $\mathcal{G}_0$ has $N := \binom{n}{r}$ vertices and $\mathrm{ex}(N, \theta_{k,r!+1}) = O(N^{1+1/k})$ (see [44]), it is already clear by Lemma 4.4.2 that $\mathcal{G}_0$ contains a copy of $\theta_{k,r!+1}$. What we will prove is that this $\theta_{k,r!+1}$ can be chosen in a way that the vertices are pairwise disjoint sets.

The following simple lemma will be useful for making sure that the vertices are disjoint sets.

Lemma 4.4.3. *Let $x, y \in V(\mathcal{G}_0)$. Then the number of $z \in V(\mathcal{G}_0)$ such that $xz \in E(\mathcal{G}_0)$ and $z \cap y \neq \emptyset$ is at most r^2.*

Proof. Since y is a set of size r, there are r ways to specify which element $v \in y$ will be contained in z. Given this choice, there are r ways to choose the colour of the monochromatic matching between x and z since it must be the colour of uv for some $u \in x$. Given these two choices, z is uniquely determined (if exists) since the colouring of K_n is proper. $\square$

The next lemma is analogous to Lemma 4.2.1.

Lemma 4.4.4. *Let $\ell \geq 2$ be a positive integer and let $\mathcal{G}$ be a subgraph of $\mathcal{G}_0$. Then the number of homomorphic copies of $C_{2\ell}$ in $\mathcal{G}$ in which the vertices are not pairwise disjoint (as subsets of $V(K_n)$) is at most*

$$16\ell \left(r^2 \ell \Delta(\mathcal{G}) \hom(C_{2\ell-2}, \mathcal{G}) \hom(C_{2\ell}, \mathcal{G})\right)^{1/2}.$$

The proof is nearly identical to that of Lemma 4.2.1, so it is only briefly sketched here. As in Lemma 4.2.1, we count the number of $(x_1, \ldots, x_{2\ell}) \in V(\mathcal{G})^{2\ell}$ with $x_1x_2, \ldots, x_{2\ell}x_1 \in E(\mathcal{G})$ such that $x_1 \cap x_i \neq \emptyset$ for some $2 \leq i \leq \ell + 1$. The only minor difference is that given $x_2, \ldots, x_{\ell+2}$, there are at most $r^2\ell$, rather than ℓ ways to choose x_1. Indeed, there are ℓ ways to choose i such that $x_1 \cap x_i \neq \emptyset$, and, given any such choice, by Lemma 4.4.3, there are at most r^2 ways to choose x_1.

The next lemma is analogous to Lemma 4.2.6.

Lemma 4.4.5. *Let $\mathcal{G}$ be a non-empty subgraph of $\mathcal{G}_0$ and suppose that for some $2 \leq \ell \leq k$ we have*

$$\hom(C_{2\ell}, \mathcal{G}) = \omega\left(\Delta(\mathcal{G}) \cdot \hom(C_{2\ell-2}, \mathcal{G})\right).$$

Then, for n sufficiently large, $\mathcal{G}$ contains a $\theta_{k,r!+1}$ in which the vertices are pairwise disjoint sets.

The proof of this lemma is very similar to that of Lemma 4.2.6 and is given in the appendix, but let us list here the three minor differences.

First, whenever in the proof of Lemma 4.2.6 we said 'rainbow, injectively homomorphic copy of $C_{2\ell}$', we now say 'homomorphic copy of $C_{2\ell}$ in which the vertices are pairwise disjoint sets'.

We very slightly modify the definition of a 'nice pair' such that between any nice pair of vertices in $\mathcal{G}$ we find $r|V(\theta_{k,r!+1})|$ paths of length ℓ, such that the vertices of $\mathcal{G}$ involved in these paths are pairwise disjoint sets in $V(K_n)$.

The last difference is that we now find a subgraph of H with sufficiently large minimum degree so that (using Lemma 4.4.3) we can greedily embed a spider with $r! + 1$ legs of length $k - \ell$ in H whose vertices are pairwise disjoint sets, and such that all the legs have endpoints which form nice pairs with x. (A spider with t legs of length s is the union of t paths of length s which share one endpoint but are pairwise vertex-disjoint apart from that.) Then we can extend this spider to a copy of $\theta_{k,r!+1}$ in $\mathcal{G}$ in which the vertices are pairwise disjoint sets.

Corollary 4.4.6. *Let $\mathcal{G}$ be a subgraph of $\mathcal{G}_0$ on m vertices and suppose that for some $2 \leq j \leq k$ we have*

$$\hom(C_{2j}, \mathcal{G}) = \omega\left(m\Delta(\mathcal{G})^j\right).$$

Then, for n sufficiently large, $\mathcal{G}$ contains a $\theta_{k,r!+1}$ in which the vertices are pairwise disjoint sets.

The proof of this is identical to that of Corollary 4.2.7.

We are now in a position to prove Theorem 4.1.10. Suppose that K_n is properly edge-coloured with $o(n^{\frac{r}{r-1}\cdot\frac{k-1}{k}})$ colours. By Lemma 4.4.2, we have $e(\mathcal{G}_0) = \omega(N^{1+1/k})$, where $N = |V(\mathcal{G}_0)| = \binom{n}{r}$. By Lemma 2.2.1, $\mathcal{G}_0$ has a K-almost regular subgraph $\mathcal{G}$ on $m = \omega(1)$ vertices with minimum degree $\delta = \omega(m^{1/k})$ such that $K = O(1)$. Now $\hom(C_{2k}, \mathcal{G}) \geq \delta^{2k} = \omega(m\Delta(\mathcal{G})^k)$, so by Corollary 4.4.6, $\mathcal{G}_0$ contains a $\theta_{k,r!+1}$ in which the vertices are pairwise disjoint sets. As we have discussed after Definition 4.4.1, this guarantees the existence of r colour-isomorphic, pairwise vertex-disjoint copies of C_{2k}.

4.5 Blow-ups of cycles

In this section we prove Theorem 4.1.12 and Theorem 4.1.13.

Definition 4.5.1. Given a graph G, define an auxiliary graph $\mathcal{G}_0$ as follows. Let the vertex set of $\mathcal{G}_0$ be the set of r-vertex subsets of $V(G)$, i.e. let $V(\mathcal{G}_0) = V(G)^{(r)}$. Now let U and V be joined by an edge if $U \cap V = \emptyset$ and $uv \in E(G)$ for every $u \in U$ and $v \in V$.

For the rest of the proof, we fix a positive integer r and a graph G, and define $\mathcal{G}_0$ as above. In order to find a copy of $C_{2k}[r]$ in G, we need to find a copy of C_{2k} in $\mathcal{G}_0$ in which the vertices are disjoint as subsets of $V(G)$. The next lemma will be useful for making sure that the vertices in our cycles are disjoint sets, and it plays the role of Lemma 4.4.3 from the previous section.

Lemma 4.5.2. *Let $x, y \in V(\mathcal{G}_0)$. Then the number of $z \in V(\mathcal{G}_0)$ such that $xz \in E(\mathcal{G}_0)$ and $z \cap y \neq \emptyset$ is at most $r^{r+1} d_{\mathcal{G}_0}(x)^{1-1/r}$.*

Proof. There are r ways to choose the element of y that should belong to z, so it suffices to prove that for any $v \in V(G)$, the number of neighbours of x in $\mathcal{G}_0$ that contain v is at most $r^r d_{\mathcal{G}_0}(x)^{1-1/r}$. Let d be the size of the common neighbourhood (in G) of the vertices in x. There are $\binom{d-1}{r-1}$ ways to choose the $r-1$ vertices in z that are different from v. Since $\binom{d-1}{r-1} \leq r^r \binom{d}{r}^{1-1/r} = r^r d_{\mathcal{G}_0}(x)^{1-1/r}$, the proof is complete. □

The next lemma is analogous to Lemma 4.2.1 and Lemma 4.4.4.

Lemma 4.5.3. *Let $\ell \geq 2$ be a positive integer and let $\mathcal{G}$ be a bipartite subgraph of $\mathcal{G}_0$ with parts X_1 and X_2 such that every $x \in X_1$ has $d_{\mathcal{G}_0}(x) \leq D_1$ and every $x \in X_2$ has $d_{\mathcal{G}_0}(x) \leq D_2$, where $D_1 \leq D_2$. Then the number of homomorphic copies of $C_{2\ell}$ in $\mathcal{G}$ in which the vertices are not pairwise disjoint (as subsets of $V(G)$) is at most*

$$32\ell \left(r^{r+1} \ell D_1^{1-1/r} D_2 \hom(C_{2\ell-2}, \mathcal{G}) \hom(C_{2\ell}, \mathcal{G}) \right)^{1/2}.$$

The proof of this lemma is similar to that of Lemma 4.2.1, but not quite identical, so we give a sketch of the proof.

Sketch of proof. We want to prove that the number of $(x_1, x_2, \ldots, x_{2\ell}) \in V(\mathcal{G})^{2\ell}$ with $x_1x_2, \ldots, x_{2\ell}x_1 \in E(\mathcal{G})$ such that $x_1, x_2, \ldots, x_{2\ell}$ are not all disjoint is at most $32\ell \left(r^{r+1}\ell D_1^{1-1/r} D_2 \hom(C_{2\ell-2}, \mathcal{G}) \hom(C_{2\ell}, \mathcal{G})\right)^{1/2}$. By symmetry, it suffices to prove that the number of $(x_1, x_2, \ldots, x_{2\ell}) \in V(\mathcal{G})^{2\ell}$ with $x_1x_2, \ldots, x_{2\ell}x_1 \in E(\mathcal{G})$ such that $x_1 \cap x_i \neq \emptyset$ for some $2 \leq i \leq \ell+1$ is at most $16 \left(r^{r+1}\ell D_1^{1-1/r} D_2 \hom(C_{2\ell-2}, \mathcal{G}) \hom(C_{2\ell}, \mathcal{G})\right)^{1/2}$.

For a positive integer s, let α_s be the number of walks of length $\ell - 1$ in $\mathcal{G}$ whose endpoints y and z have $2^{s-1} \leq \hom_{y,z}(P_{\ell-1}, \mathcal{G}) < 2^s$ and let β_s be the number of walks of length ℓ in $\mathcal{G}$ whose endpoints y and z have $2^{s-1} \leq \hom_{y,z}(P_\ell, \mathcal{G}) < 2^s$.

For positive integers s and t, write $\gamma_{s,t}$ for the number of homomorphic copies $x_1x_2 \ldots x_{2\ell}x_1$ of $C_{2\ell}$ such that $x_1 \in X_1$, $x_1 \cap x_i \neq \emptyset$ for some $2 \leq i \leq \ell+1$, $2^{s-1} \leq \hom_{x_1,x_{\ell+2}}(P_{\ell-1}, \mathcal{G}) < 2^s$ and $2^{t-1} \leq \hom_{x_2,x_{\ell+2}}(P_\ell, \mathcal{G}) < 2^t$ and write $\gamma'_{s,t}$ for the number of homomorphic copies $x_1x_2 \ldots x_{2\ell}x_1$ of $C_{2\ell}$ such that $x_1 \in X_2$, $x_1 \cap x_i \neq \emptyset$ for some $2 \leq i \leq \ell+1$, $2^{s-1} \leq \hom_{x_1,x_{\ell+2}}(P_{\ell-1}, \mathcal{G}) < 2^s$ and $2^{t-1} \leq \hom_{x_2,x_{\ell+2}}(P_\ell, \mathcal{G}) < 2^t$

Here comes the main difference compared to the proof of Lemma 4.2.1. Observe that $\gamma_{s,t} \leq \alpha_s \cdot D_1 \cdot 2^t$. Indeed, if $x_1x_2 \ldots x_{2\ell}x_1$ is a homomorphic $C_{2\ell}$ with $x_1 \in X_1$, $2^{s-1} \leq \hom_{x_1,x_{\ell+2}}(P_{\ell-1}, \mathcal{G}) < 2^s$ and $2^{t-1} \leq \hom_{x_2,x_{\ell+2}}(P_\ell, \mathcal{G}) < 2^t$, then there are at most α_s ways to choose $(x_{\ell+2}, x_{\ell+3}, \ldots, x_{2\ell}, x_1)$, given such a choice, as $x_1 \in X_1$, there are at most D_1 choices for x_2, and given these there are at most 2^t choices for $(x_3, \ldots, x_{\ell+1})$. On the other hand, $\gamma_{s,t} \leq \beta_t \cdot \ell r^{r+1} D_2^{1-1/r} \cdot 2^s$. Indeed, there are at most β_t ways to choose $(x_2, \ldots, x_{\ell+2})$. By Lemma 4.5.2, given such a choice, there are at most $\ell r^{r+1} d_{\mathcal{G}_0}(x_2)^{1-1/r} \leq \ell r^{r+1} D_2^{1-1/r}$ possibilities for x_1 (since $x_1 \cap x_i \neq \emptyset$ for some $2 \leq i \leq \ell+1$). Finally, there are at most 2^s ways to complete this to a suitable homomorphic copy of $C_{2\ell}$. Similarly, $\gamma'_{s,t} \leq \alpha_s \cdot D_2 \cdot 2^t$ and $\gamma'_{s,t} \leq \beta_t \cdot \ell r^{r+1} D_1^{1-1/r} \cdot 2^s$.

Now similarly to the proof of Lemma 4.2.1, we can prove that

$$\sum_{s,t \geq 1} \gamma_{s,t} \leq 8 \left(r^{r+1}\ell D_1 D_2^{1-1/r} \hom(C_{2\ell-2}, \mathcal{G}) \hom(C_{2\ell}, \mathcal{G})\right)^{1/2}$$

and

$$\sum_{s,t \geq 1} \gamma'_{s,t} \leq 8 \left(r^{r+1}\ell D_1^{1-1/r} D_2 \hom(C_{2\ell-2}, \mathcal{G}) \hom(C_{2\ell}, \mathcal{G})\right)^{1/2}.$$

Hence, the total number of homomorphic copies of $C_{2\ell}$ in $\mathcal{G}$ in which the vertices are not pairwise disjoint is

$$\sum_{s,t \geq 1} \gamma_{s,t} + \gamma'_{s,t} \leq 16 \left(r^{r+1}\ell D_1^{1-1/r} D_2 \hom(C_{2\ell-2}, \mathcal{G}) \hom(C_{2\ell}, \mathcal{G})\right)^{1/2}. \qquad \square$$

Now we want to find a bipartite subgraph $\mathcal{G}$ in $\mathcal{G}_0$ which has many homomorphic cycles but whose vertices have not too large degree in $\mathcal{G}_0$.

Lemma 4.5.4. *Let $\mathcal{G}_0$ have average degree $d > 0$. Then there exist $D_1, D_2 \geq \frac{d}{4}$ and a non-empty bipartite subgraph $\mathcal{G}$ in $\mathcal{G}_0$ with parts X_1 and X_2 such that for every $x \in X_1$, we have*

$d_{\mathcal{G}}(x) \geq \frac{D_1}{256r^2(\log n)^2}$ *and* $d_{\mathcal{G}_0}(x) \leq D_1$*, and for every* $x \in X_2$*, we have* $d_{\mathcal{G}}(x) \geq \frac{D_2}{256r^2(\log n)^2}$ *and* $d_{\mathcal{G}_0}(x) \leq D_2$*.*

Proof. Let N and e denote the number of vertices and edges in $\mathcal{G}_0$, respectively. Observe that the number of edges in $\mathcal{G}_0$ incident to vertices of degree at most $d/4$ is at most $Nd/4 = e/2$. Hence, a random partitioning of all vertices with degree at least $d/4$ shows that there exist disjoint sets A and B in $V(\mathcal{G}_0)$ such that for every $v \in A \cup B$ we have $d_{\mathcal{G}_0}(v) \geq d/4$ and the number of edges in $\mathcal{G}_0[A, B]$ is at least $e/4$. For each $1 \leq i \leq \lceil r \log n \rceil$, let $A_i = \{v \in A : 2^{i-1} \leq d_{\mathcal{G}_0}(v) < 2^i\}$ and let $B_i = \{v \in B : 2^{i-1} \leq d_{\mathcal{G}_0}(v) < 2^i\}$. Note that the A_i's partition A. Indeed, $\Delta(\mathcal{G}_0) \leq \binom{n}{r} \leq n^r$. Similarly, the B_i's partition B. Hence, there exist i, j such that $e(\mathcal{G}_0[A_i, B_j]) \geq \frac{e}{4\lceil r \log n \rceil^2} \geq \frac{e}{16r^2(\log n)^2}$.

Note that $|A_i|2^{i-1} \leq 2e(\mathcal{G}_0) = 2e$, so $|A_i| \leq \frac{2e}{2^{i-1}}$. Thus, the average degree of the vertices in A_i in the graph $\mathcal{G}_0[A_i, B_j]$ is at least $\frac{2^{i-1}}{32r^2(\log n)^2}$. Similarly, the average degree of the vertices in B_j in the same graph is at least $\frac{2^{j-1}}{32r^2(\log n)^2}$. Thus, by a standard vertex removal argument, there exist non-empty $X_1 \subset A_i$ and $X_2 \subset B_j$ such that for $\mathcal{G} = \mathcal{G}_0[X_1, X_2]$, we have $d_{\mathcal{G}}(x) \geq \frac{2^{i-1}}{128r^2(\log n)^2}$ for every $x \in X_1$ and $d_{\mathcal{G}}(x) \geq \frac{2^{j-1}}{128r^2(\log n)^2}$ for every $x \in X_2$. Take $D_1 = 2^i$ and $D_2 = 2^j$. Since $d/4 \leq d_{\mathcal{G}_0}(v) < 2^i$ holds for every $v \in X_1 \subset A$, we have $D_1 > d/4$. Similarly, $D_2 > d/4$. $\square$

We are now ready to prove Theorem 4.1.12.

Proof of Theorem 4.1.12. Let G be an n-vertex graph with $\omega(n^{2-1/r}(\log n)^{7/r})$ edges. We will prove that if n is sufficiently large, then G contains an r-blowup cycle. By the supersaturation of $K_{r,r}$ (Lemma 3.2.1), $\mathcal{G}_0$ has $\omega(n^r(\log n)^{7r})$ edges, so it has average degree $\omega((\log n)^{7r})$. By Lemma 4.5.4, there exist $D_1, D_2 = \omega((\log n)^{7r})$ and a non-empty bipartite subgraph $\mathcal{G}$ in $\mathcal{G}_0$ with parts X_1 and X_2 such that for every $x \in X_1$, we have $d_{\mathcal{G}}(x) \geq \frac{D_1}{256r^2(\log n)^2}$ and $d_{\mathcal{G}_0}(x) \leq D_1$, and for every $x \in X_2$, we have $d_{\mathcal{G}}(x) \geq \frac{D_2}{256r^2(\log n)^2}$ and $d_{\mathcal{G}_0}(x) \leq D_2$. Without loss of generality, we may assume that $D_1 \leq D_2$.

By Lemma 4.3.3, for every positive integer k we have

$$\begin{aligned} \hom(C_{2k}, \mathcal{G}) &\geq \left(\frac{D_1}{256r^2(\log n)^2}\right)^k \left(\frac{D_2}{256r^2(\log n)^2}\right)^k \\ &= \left(\frac{D_1^{1/r}}{2^{16}r^4(\log n)^4}\right)^k (D_1^{1-1/r}D_2)^k. \end{aligned}$$

Let $k = \lfloor \log n \rfloor$. Since $D_1 = \omega((\log n)^{7r})$, we have

$$\left(\frac{D_1^{1/r}}{2^{16}r^4(\log n)^4}\right)^k \geq \binom{n}{r}(L(\log n)^3)^k$$

for some $L = \omega(1)$. Then

$$\hom(C_{2k}, \mathcal{G}) \geq \binom{n}{r}(L(\log n)^3 D_1^{1-1/r} D_2)^k.$$

Let ℓ be the smallest positive integer such that

$$\hom(C_{2\ell},\mathcal{G}) \geq \binom{n}{r}(L(\log n)^3 D_1^{1-1/r} D_2)^{\ell}.$$

Clearly, $\ell \leq k$. Moreover, since $\mathcal{G}$ has at most $\binom{n}{r}$ vertices and maximum degree at most D_2, we have $\ell \geq 2$. Now note that

$$\hom(C_{2\ell-2},\mathcal{G}) < \frac{\hom(C_{2\ell},\mathcal{G})}{L(\log n)^3 D_1^{1-1/r} D_2}.$$

Hence, by Lemma 4.5.3, the number of homomorphic copies of $C_{2\ell}$ in $\mathcal{G}$ in which the vertices are not pairwise disjoint is less than

$$\frac{32 r^{\frac{r+1}{2}} \ell^{3/2}}{L^{1/2}(\log n)^{3/2}} \hom(C_{2\ell},\mathcal{G}).$$

Since $\ell \leq k \leq \log n$ and $L = \omega(1)$, this is less than $\hom(C_{2\ell},\mathcal{G})$ provided that n is sufficiently large. Thus, there exists a homomorphic copy of $C_{2\ell}$ in $\mathcal{G}$ in which the vertices are pairwise disjoint subsets of $V(G)$. This gives a $C_{2\ell}[r]$ in G. $\square$

We will now prove Theorem 4.1.13. The key step is the following lemma, which is similar to Lemma 4.2.6 and Lemma 4.4.5 from the previous sections, but very slightly more involved.

Lemma 4.5.5. *Let $\ell \geq 2$ and $k \geq \ell$ be fixed integers and let $\mathcal{G}$ be a bipartite subgraph of $\mathcal{G}_0$ with parts X_1 and X_2 such that every $x \in X_1$ has $d_{\mathcal{G}_0}(x) \leq D_1$ and every $x \in X_2$ has $d_{\mathcal{G}_0}(x) \leq D_2$, where $D_1 \leq D_2$. Assume that*

$$\hom(C_{2\ell},\mathcal{G}) = \omega\left(D_1^{1-1/r} D_2 \hom(C_{2\ell-2},\mathcal{G})\right).$$

Then, for n sufficiently large, $\mathcal{G}$ contains a copy of C_{2k} in which the vertices are pairwise disjoint subsets of $V(G)$. In particular, G contains a copy of $C_{2k}[r]$.

To prove this lemma, we need the following strengthening of Lemma 4.2.5.

Lemma 4.5.6. *Let H be a bipartite graph with parts Y and Z. Let $f : Y \to \mathbb{R}$ be a function and let $g(z) = \sum_{y \in N_H(z)} f(y)$ for every $z \in Z$. Assume that $d_H(y) \leq D_1$ for every $y \in Y$ and that $d_H(z) \leq D_2$ for every $z \in Z$. Also suppose that H does not contain a subgraph H' with parts $Y' \subset Y$ and $Z' \subset Z$ such that for every $y \in Y'$, we have $d_{H'}(y) \geq d_1$ and for every $z \in Z$, we have $d_{H'}(z) \geq d_2$. Then*

$$\sum_{y \in Y} f(y)^2 \geq \min\left(\frac{1}{4d_1 D_2}, \frac{1}{4D_1 d_2}\right) \sum_{z \in Z} g(z)^2.$$

The proof of Lemma 4.5.6 is similar to the proof of Lemma 4.2.5 and is omitted.

Let us briefly sketch the proof of Lemma 4.5.5. It is nearly identical to the proof of Lemma 4.2.6 up to the definition of H, the only difference is that we replace each 'rainbow, injectively homomorphic copy of $C_{2\ell}$' by '$C_{2\ell}$ in which the vertices are disjoint sets'. Let us define the parts of H very slightly differently: let H have parts Y and Z where $Z = \{z \in V(\mathcal{G}) : (x, z) \text{ is nice}\}$ and let Y be the set of vertices in $\mathcal{G}$ which have a neighbour in the set Z. Since there is a walk of length ℓ from x to any element of Z, and $\mathcal{G}$ is bipartite, we have either $Y \subset X_1$ and $Z \subset X_2$ or $Y \subset X_2$ and $Z \subset X_1$. In the former case we use Lemma 4.5.6 to find a subgraph of H with parts $Y' \subset Y$ and $Z' \subset Z$ such that every $y \in Y'$ has $d_{H'}(y) = \omega(D_1^{1-1/r})$ and every $z \in Z'$ has $d_{H'}(z) = \omega(D_2^{1-1/r})$. In the latter case we use Lemma 4.5.6 to find a subgraph of H with parts $Y' \subset Y$ and $Z' \subset Z$ such that every $y \in Y'$ has $d_{H'}(y) = \omega(D_2^{1-1/r})$ and every $z \in Z'$ has $d_{H'}(z) = \omega(D_1^{1-1/r})$. Then, using Lemma 4.5.2, we can greedily find a path of length $2k - 2\ell$ in which the vertices are disjoint from each other and from x and which has endpoints in Z. Then we can extend this to a cycle of length $2k$ through x in which the vertices are disjoint sets.

Proof of Theorem 4.1.13. Let G be a graph with $\omega\left(n^{2-\frac{1}{r}+\frac{1}{k+r-1}}(\log n)^{\frac{4k}{r(k+r-1)}}\right)$ edges. By Lemma 3.2.1, $\mathcal{G}_0$ has average degree $\omega\left(n^{\frac{r^2}{k+r-1}}(\log n)^{\frac{4kr}{k+r-1}}\right)$. By Lemma 4.5.4, $\mathcal{G}_0$ has a bipartite subgraph $\mathcal{G}$ with parts X_1 and X_2 such that for every $x \in X_i$ we have $d_{\mathcal{G}}(x) \geq \frac{D_i}{256r^2(\log n)^2}$ and $d_{\mathcal{G}_0}(x) \leq D_i$, where $D_i = \omega\left(n^{\frac{r^2}{k+r-1}}(\log n)^{\frac{4kr}{k+r-1}}\right)$. Using Lemma 4.3.3, we have $\hom(C_{2k}, \mathcal{G}) \geq \Omega\left(\frac{D_1^k D_2^k}{(\log n)^{4k}}\right) \geq \omega\left((D_1^{1-1/r} D_2)^{k-1}\binom{n}{r} D_2\right)$. So there exists some $2 \leq \ell \leq k$ with $\hom(C_{2\ell}, \mathcal{G}) = \omega\left((D_1^{1-1/r} D_2) \hom(C_{2\ell-2}, \mathcal{G})\right)$. By Lemma 4.5.5, G contains $C_{2k}[r]$ as a subgraph. $\square$

4.6 Concluding remarks

Rainbow cycles. We have shown that for a sufficiently large constant C, any properly edge-coloured n-vertex graph with at least $Cn(\log n)^4$ edges contains a rainbow cycle. However, the best known construction of a graph without a rainbow cycle has only $\Theta(n \log n)$ edges. One such example, found by Keevash, Mubayi, Sudakov and Verstraëte [82], is the m-dimensional cube whose vertices are the subsets of $\{1, 2, \dots, m\}$ where A is joined to $A \setminus \{i\}$ for every $i \in A$. The colour of the edge between A and $A \setminus \{i\}$ is i. This graph has 2^m vertices and $\frac{1}{2}m2^m$ edges and it has no rainbow cycle. Examples with more than $0.58n \log n$ edges were also found by Keevash, Mubayi, Sudakov and Verstraëte.

Colour-isomorphic cycles. Recall that $f_r(n, H)$ is the smallest number C so that there is a proper edge-colouring of K_n with C colours containing no r vertex-disjoint colour-isomorphic copies of H. We have shown that $f_r(n, C_{2k}) = \Omega\left(n^{\frac{r}{r-1}\cdot\frac{k-1}{k}}\right)$. Note that our result becomes trivial when $r \geq k$ since $f_r(n, H) \geq n - 1$ holds for any r and H (as any proper colouring of K_n must use at least $n - 1$ colours).

The best general upper bound comes from the probabilistic construction that is used in Theorem 4.1.6 and says that $f_r(n, C_{2k}) = O\left(n^{\frac{r}{r-1} - \frac{1}{(r-1)k}}\right)$. Another result of Conlon and Tyomkyn [25, Theorem 1.4], proved by a variant of Bukh's random algebraic method [17], states that if H contains a cycle, then there exists r such that $f_r(n, H) = O(n)$. It would be interesting to decide what the smallest such r is when $H = C_{2k}$. Our result shows that we must have $r \geq k$. This question was studied in the case $H = C_4$ by Xu, Zhang, Jing and Ge [110], who showed that $f_r(n, C_4) = \Theta(n)$ for any $r \geq 3$.

The Erdős–Gyárfás function. For positive integers n, p and $2 \leq q \leq \binom{p}{2}$, the Erdős-Gyárfás function $g(n, p, q)$ is defined to be the smallest C such that there exists a (not necessarily proper) colouring of the edges of K_n with C colours such that every induced subgraph on p vertices receives at least q colours. A variant of our Theorem 4.1.10 can be used to give a good lower bound for this function when q is close to $\binom{p}{2}$. Indeed, assume that $p = 2kr$ and $q = \binom{p}{2} - (r-1)2k + 1$ for some $r, k \geq 2$. If we can find r vertex-disjoint colour-isomorphic cycles of length $2k$, then the p vertices of these cycles induce at most $\binom{p}{2} - (r-1)2k < q$ colours. Note that the proof of Theorem 4.1.10 can be adapted to the case where the edge-colouring is not necessarily proper, but every vertex is incident to at most $O(1)$ edges of any given colour. Now if we have an arbitrary edge-colouring of K_n, then either every vertex is incident to at most $2kr - 2$ edges of any given colour, or we can choose vertices $u_0, u_1, \ldots, u_{2kr-1}$ such that the edges u_0u_i are of the same colour for every $1 \leq i \leq 2kr - 1$. In this latter case, we have p vertices which induce at most $\binom{p}{2} - p + 2 < q$ colours. In the former case, we can use the strengthened version of Theorem 4.1.10. We obtain the following result.

Theorem 4.6.1. *For any integers $r, k \geq 2$,*

$$g\left(n, 2kr, \binom{2kr}{2} - (r-1)2k + 1\right) = \Omega(n^{\frac{r}{r-1} \cdot \frac{k-1}{k}}).$$

This generalises a recent result of Fish, Pohoata and Sheffer [45, Theorem 1.1], which is Theorem 4.6.1 in the special case $r = 2$.

Blow-ups of cycles. We have shown that $\mathrm{ex}(n, \mathcal{C}[r]) = O(n^{2-1/r}(\log n)^{7/r})$. On the other hand, a random graph with edge probabilities $p = \frac{n^{-1/r}}{2}$ contains no r-blownup cycles with probability at least 1/2, so $\mathrm{ex}(n, \mathcal{C}[r]) = \Omega(n^{2-1/r})$. We pose the following question.

Question 4.6.2. *Let $r \geq 2$. Is it true that $\mathrm{ex}(n, \mathcal{C}[r]) = \Theta(n^{2-1/r})$?*

Finally, regarding a single forbidden blownup cycle, we reiterate our conjecture that $\mathrm{ex}(n, C_{2k}[r]) = O(n^{2-\frac{1}{r}+\frac{1}{kr}})$.

4.A Appendix

Proof of Lemma 4.4.5. Let $s = r|V(\theta_{k,r!+1})|$. For a graph F, call a homomorphic copy of F in $\mathcal{G}$ *good* if the images of the vertices of F are disjoint sets (as subsets of $V(K_n)$). In particular, any such copy is an injectively homomorphic copy of F. Call a pair $(x_1, x_{\ell+1})$ of vertices in $\mathcal{G}$ *nice* if the number of good copies of $C_{2\ell}$ of the form $x_1x_2 \dots x_{2\ell}x_1$ is greater than $(1 - \frac{1}{\binom{s}{2}}) \left(\hom_{x_1,x_{\ell+1}}(P_\ell, \mathcal{G})\right)^2$. Observe that the total number of homomorphic copies of $C_{2\ell}$ of the form $x_1x_2 \dots x_{2\ell}x_1$ in $\mathcal{G}$ is $\hom_{x_1,x_{\ell+1}}(P_\ell, \mathcal{G})^2$, so this means that the proportion of those which are not good is less than $\frac{1}{\binom{s}{2}}$. In particular, if we choose s random walks of length ℓ between x_1 and $x_{\ell+1}$ with replacement, then with positive probability any two of these walks form a good copy of $C_{2\ell}$. Hence, there exist at least s pairwise internally vertex-disjoint paths between x_1 and $x_{\ell+1}$ such that the vertices involved in these paths are pairwise disjoint sets in $V(K_n)$.

By Lemma 4.4.4, the number of non-good copies of $C_{2\ell}$ in $\mathcal{G}$ is

$$O_{r,\ell}\left((\Delta(\mathcal{G}) \hom(C_{2\ell-2}, \mathcal{G}) \hom(C_{2\ell}, \mathcal{G}))^{1/2}\right) \leq o(\hom(C_{2\ell}, \mathcal{G})).$$

Hence,

$$\sum_{(x_1,x_{\ell+1}) \text{ not nice}} \frac{1}{\binom{s}{2}} \hom_{x_1,x_{\ell+1}}(P_\ell, \mathcal{G})^2 = o(\hom(C_{2\ell}, \mathcal{G})),$$

so, using $\sum_{x_1,x_{\ell+1} \in V(\mathcal{G})} \hom_{x_1,x_{\ell+1}}(P_\ell, \mathcal{G})^2 = \hom(C_{2\ell}, \mathcal{G})$, we have

$$\begin{aligned}\sum_{(x_1,x_{\ell+1}) \text{ nice}} \hom_{x_1,x_{\ell+1}}(P_\ell, \mathcal{G})^2 &\geq (1 - o(1)) \hom(C_{2\ell}, \mathcal{G}) \\ &> (1 - o(1)) L\Delta(\mathcal{G}) \hom(C_{2\ell-2}, \mathcal{G})\end{aligned}$$

for some $L = \omega(1)$.

Thus, there exists some $x \in V(\mathcal{G})$ such that

$$\sum_{z \in V(\mathcal{G}) : (x,z) \text{ is nice}} \hom_{x,z}(P_\ell, \mathcal{G})^2 > (1 - o(1)) L\Delta(\mathcal{G}) \hom_x(C_{2\ell-2}, \mathcal{G}). \tag{4.4}$$

Let $Z = \{z \in V(\mathcal{G}) : (x, z) \text{ is nice}\}$ and let $Y = V(\mathcal{G})$. Consider the bipartite graph H with parts Y and Z, defined by $\mathcal{G}$. (We view Y and Z as disjoint sets even though they overlap as subsets of $V(\mathcal{G})$.)

Suppose that H does not contain a subgraph with minimum degree at least $r^2k(r!+1)$. Let $f(y) = \hom_{x,y}(P_{\ell-1}, \mathcal{G})$ for every $y \in Y = V(\mathcal{G})$ and define g as in Lemma 4.2.5. By that lemma with $d = r^2k(r! + 1)$,

$$\sum_{y \in Y} f(y)^2 \geq \frac{1}{4d\Delta(H)} \sum_{z \in Z} g(z)^2 \geq \frac{1}{4d\Delta(\mathcal{G})} \sum_{z \in Z} g(z)^2.$$

However, $g(z) = \sum_{y \in N_{\mathcal{G}}(z)} \hom_{x,y}(P_{\ell-1}, \mathcal{G}) = \hom_{x,z}(P_\ell, \mathcal{G})$, so, using equation (4.4),

$$\sum_{y \in Y} f(y)^2 \geq \frac{1}{4d\Delta(\mathcal{G})} \sum_{z \in Z} \hom_{x,z}(P_\ell, \mathcal{G})^2 > \frac{(1-o(1))L}{4d} \hom_x(C_{2\ell-2}, \mathcal{G}).$$

However, $\sum_{y \in Y} f(y)^2 = \hom_x(C_{2\ell-2}, \mathcal{G})$, which is a contradiction, as $L = \omega(1)$ and n is sufficiently large.

Thus, H contains a subgraph with minimum degree at least $r^2k(r!+1)$. Then, by Lemma 4.4.3 we can greedily find in H a spider whose vertices are disjoint (as subsets of $V(K_n)$) from x and from each other and which has $r!+1$ legs of length $k-\ell$ such that the endpoints of these legs are in Z. Let this spider be S with endpoints $z_1, z_2, \ldots, z_{r!+1}$. Since for every i, (x, z_i) is a nice pair, there exist at least $s = r|V(\theta_{k,r!+1})|$ paths of length ℓ between x and z_i such that all the internal vertices in these paths are distinct and pairwise disjoint from each other. Hence, we can choose paths of length ℓ between x and z_i for every $1 \leq i \leq r!+1$ such that all the vertices involved are disjoint from the vertices of S and from each other. Then the union of these paths with S gives a suitable $\theta_{k,r!+1}$. $\square$

Chapter 5

Improved bounds for the Erdős-Rogers function

5.1 Introduction

Let G be a graph with n vertices that contains no K_4. How large a triangle-free induced subgraph must G have? The standard proof of Ramsey's theorem implies that G contains an independent set of size $n^{1/3}$, but can we do better?

A simple argument shows that the answer is yes. Indeed, each vertex in G has a triangle-free neighbourhood, and either there is a vertex of degree $n^{1/2}$ or one can find an independent set of size roughly $n^{1/2}$ by repeatedly choosing vertices and discarding their neighbours.

This stronger argument still feels a little wasteful, because in the second case one finds an independent set rather than a triangle-free subgraph. Moreover, there is no obvious example that yields a matching upper bound, so it is not immediately clear whether 1/2 is the correct exponent.

The problem above is an example of a general problem that was first considered by Erdős and Rogers. Given positive integers $1 < s < t$ and $n > 2$, define $f_{s,t}(n)$ to be the minimum over all K_t-free graphs G with n vertices of the order of the largest induced K_s-free subgraph of G. We have just been discussing the function $f_{3,4}$. The function $f_{s,t}$ is known as the *Erdős-Rogers function*. It has been studied by several authors: for a detailed survey covering many of the known results on the subject, see [29]. For a more recent exposition, see also section 3.5.2 of [22].

The first bounds were obtained by Erdős and Rogers [37] who showed that for every s there exists a positive constant $\epsilon(s)$ such that $f_{s,s+1}(n) \leq n^{1-\epsilon(s)}$. About 30 years later, Bollobás and Hind [13] improved the estimate for $\epsilon(s)$ and established the lower bound $f_{s,t}(n) \geq n^{1/(t-s+1)}$. In particular, $f_{s,s+1}(n) \geq n^{1/2}$ (by the obvious generalization of the argument for $f_{3,4}$ above).

Subsequently, Krivelevich [90, 91] improved these lower bounds by a small power of $\log n$ and also gave a new general upper bound, which is

$$f_{s,t}(n) \leq O(n^{\frac{s}{t+1}}(\log n)^{\frac{1}{s-1}}). \tag{5.1}$$

Later, the lower bound was significantly improved by Sudakov [105, 106]. He showed that if $t > s+1$, then $f_{s,t}(n) \geq \Omega(n^{a_{s,t}})$ where $a_{s,t}$ is defined recursively. In particular, when s is fixed and $t \to \infty$, he obtained the bound

$$f_{s,t}(n) \geq \Omega(n^{\frac{s}{2t}+O(1/t^2)}).$$

We remark that if $t \geq 2s$ then (5.1) is the best known upper bound, while Sudakov's lower bound is the best known for every $t > s+1$. In particular, the upper bound is roughly the square of the lower bound in the range $t \geq 2s$.

Recently, there has been quite a lot of progress on the case $t = s+1$. First, Dudek and Rödl [28] showed that $f_{s,s+1}(n) \leq O(n^{2/3})$. Then Wolfovitz [109] proved that for sufficiently large n we have $f_{3,4}(n) \leq n^{1/2}(\log n)^{120}$, yielding the slightly surprising fact that the exponent $1/2$ is indeed the right one in that case. Finally, Dudek, Retter and Rödl [27], generalizing Wolfovitz's construction, showed that for any $s \geq 3$ there exist constants c_1 and c_2 such that

$$f_{s,s+1}(n) \leq c_1 n^{1/2}(\log n)^{c_2}$$

so the exponent $1/2$ is correct for all $f_{s,s+1}$. However, the problem of finding the correct exponent of n for general s, t remains open.

A particularly important case is when $t = s+2$ since $f_{s,t}(n) \leq f_{s,s+2}(n)$ for any $t \geq s+2$. Sudakov's lower bound gives $f_{s,s+2}(n) = \Omega(n^{\beta_s})$ where $\beta_s = 1/2 - \frac{1}{6s-6}$. Dudek, Retter and Rödl in [27] showed that for any $s \geq 4$ there exists a constant c depending only on s such that

$$f_{s,s+2}(n) \leq cn^{1/2}.$$

Note that the exponent $1/2$ follows from the bound for $f_{s,s+1}$, so this improves it by removing the log factor. Having established this, Dudek, Retter and Rödl asked the following question.

Question 5.1.1. *Does there exist $s \geq 3$ such that $f_{s,s+2}(n) = o(n^{1/2})$?*

Another central open problem in the area is the following question of Erdős [36].

Question 5.1.2. *Is it true that*

$$\lim_{n\to\infty} \frac{f_{s+1,t}(n)}{f_{s,t}(n)} = \infty \tag{5.2}$$

for every $t > s+1$?

The answer has been shown to be yes when $t = s + 2 \geq 6$ and when (s,t) is one of the pairs $(2,4),(2,5),(2,6),(2,7),(2,8)$ or $(3,6)$.

5.1.1 Our results

In this chapter, we prove that the answer to the first question above is yes. We also establish (5.2) for the families of pairs $t = s+3 \geq 7$ and $t = s+2 \geq 5$. We obtain these results by proving a significant improvement for the upper bound on $f_{s,t}$ when $s+2 \leq t \leq 2s-1$. The previous best upper bound for these parameters appeared in [27] and was $f_{s,t}(n) \leq cn^{1/2}$ (except for the pair $s=3, t=5$, where this bound was not established). We do not just obtain bounds of the form $o(n^{1/2})$, but we improve the exponents throughout the range. Our construction is probabilistic, and has some similarities to the constructions that established the previous best upper bounds. However, an important difference is that we do not make use of algebraic objects such as projective planes.

To state the bound that comes out of our argument takes a small amount of preparation. Let $s \geq 3$ and $s+2 \leq t \leq 2s-1$. Call (s,t) *regular* if $s \geq 11$ and $s+3 \leq t \leq 2s-4$ or if $(s,t) \in \{(10,14),(10,15)\}$ and call it *exceptional* otherwise. Let

$$\alpha = \alpha_{s,t} = \begin{cases} \alpha(1) = \frac{(s-2)(t-s)(t+s-1)+2t-2s}{(2s-3)(t-s)(t+s-1)-2s+4}, & \text{if } (s,t) \text{ is regular} \\ \alpha(2) = \frac{(s-2)(t-s)(s-1)+s-1}{(2s-3)(t-s)(s-1)+2s-t}, & \text{if } (s,t) \text{ is exceptional} \end{cases}$$

We will prove the following theorem.

Theorem 5.1.3. *For any $s \geq 3, s+2 \leq t \leq 2s-1$, there exists some constant $c = c(s,t)$ such that*

$$f_{s,t}(n) \leq n^{\alpha}(\log n)^c.$$

It is not hard to check that $\alpha < 1/2$ for all pairs (s,t) in the given range. Thus, as mentioned above, we obtain a strong answer to the question of Dudek, Retter and Rödl.

Corollary 5.1.4. *For every $s \geq 3$, we have $f_{s,s+2}(n) = o(n^{1/2})$.*

The simplest case where our result is new is the case $s=3, t=5$. There we obtain an upper bound of $n^{6/13}(\log n)^c$. For comparison, Sudakov's lower bound is $cn^{5/12}$.

Since the exponent when $t = s+1$ is $1/2$, our result also implies a positive answer to the question of Erdős in the following family of cases.

Corollary 5.1.5.

$$\lim_{n\to\infty} \frac{f_{s+1,s+2}(n)}{f_{s,s+2}(n)} \to \infty$$

That is, (5.2) holds for $t = s+2 \geq 5$.

If $t = s+3$, then

$$\alpha = \begin{cases} \frac{3s^2-3s-3}{6s^2-4s-7}, & \text{if } s \geq 11 \\ \frac{3s^2-8s+5}{6s^2-14s+6}, & \text{if } 4 \leq s \leq 10 \end{cases}$$

Comparing this with Sudakov's lower bound $f_{s+1,s+3}(n) \geq \Omega(n^{\beta_{s+1}})$, where $\beta_{s+1} = \frac{3s-1}{6s}$, we get the following additional result.

Corollary 5.1.6.

$$\lim_{n\to\infty} \frac{f_{s+1,s+3}(n)}{f_{s,s+3}(n)} \to \infty$$

That is, (5.2) holds for $t = s + 3 \geq 7$.

In the following table, we compare the exponent of n in the best known lower bound with that in our new upper bound (both rounded to three decimal places).

	Our new upper bound	Best known lower bound
$s = 3, t = 5$	0.462	0.417
$s = 4, t = 6$	0.467	0.444
$s = 4, t = 7$	0.457	0.375
$s = 5, t = 7$	0.475	0.458
$s = 5, t = 8$	0.465	0.404
$s = 5, t = 9$	0.460	0.351

In the case $t = s + 2$, our bound is $f_{s,s+2}(n) \leq n^{\alpha+o(1)}$ for $\alpha = 1/2 - \frac{s-2}{8s^2-18s+8} \approx 1/2 - \frac{1}{8s}$ while Sudakov's lower bound is $f_{s,s+2}(n) \geq n^{\beta+o(1)}$ for $\beta = 1/2 - \frac{1}{6s-6} \approx 1/2 - \frac{1}{6s}$. It would be very interesting to know whether either of these two estimates reflects the true asymptotics of $f_{s,s+2}$. It would be particularly interesting to know whether either of the exponents 5/12 or 6/13 is the correct one for $(s, t) = (3, 5)$. We have made some effort to optimize our construction, whereas there appear to be places where Sudakov's argument is potentially throwing information away, so our guess is that 6/13 is correct, but this guess is very tentative and could easily turn out to be wrong.

5.1.2 An overview of the argument

We will now sketch the key steps in our argument. For simplicity, we will focus on the $s = 3, t = 5$ case. Then, as mentioned above, Theorem 5.1.3 says that $f_{3,5}(n) \leq n^{6/13}(\log n)^c$. That is, we construct a K_5-free graph G in which every subset of size roughly $n^{6/13}$ induces a triangle.

The basic idea is very simple. We are looking for a graph that contains "triangles everywhere" but does not contain any K_5s. The obvious way to create a large number of triangles without creating K_5s is to take a complete tripartite graph. Of course, this on its own does nothing, since a complete tripartite graph has a huge independent set, but we can use it as a building block by taking a union of many complete tripartite graphs. In previous constructions, such as Wolfovitz's graph that gives an upper bound for $f_{3,4}(n)$, the vertex sets of these tripartite graphs are chosen algebraically – in Wolfovitz's case they are the lines of a projective plane. The main difference in our approach is that we simply choose them at random, where the number we choose and the size of each one are parameters that we optimize at the end of the argument. This creates difficulties that are not present in the earlier approaches, but in the end allows us to prove stronger bounds.

Thus, we begin by taking a graph G_0, which is a union of roughly $n^{9/13}$ complete tripartite graphs with parts having size roughly $n^{6/13}$ each, these parts being randomly chosen subsets of $V(G_0)$. It is not hard to prove that G_0 contains a triangle in every set of vertices of size roughly $n^{6/13}$.

However, G_0 also contains many K_5s, so we have to delete some edges. It is here that the proof becomes less simple: while random constructions followed by edge deletions are very standard, in this case we need rather delicate arguments in order to prove that it can be done without removing all the triangles from a set of size around $n^{6/13}$.

First, let us check that every set of size roughly $n^{6/13}$ does indeed induce a triangle in G_0. Let A be a subset of $V(G_0)$ of size $n^{6/13}$. A given tripartite copy will intersect A in at least 3 vertices with probability roughly $n^{-3/13}$. Thus, as we place $n^{9/13}$ tripartites (we write "tripartite" as a shorthand for our complete tripartite graphs with parts of size roughly $n^{6/13}$), the expected number of those tripartites that give a triangle in A is roughly $n^{6/13}$. Hence, by the Chernoff bound, the probability that A does not contain a triangle is roughly $e^{-n^{6/13}}$. But the number of subsets of $V(G_0)$ of size $n^{6/13}$ is *very* roughly $n^{n^{6/13}}$. Modifying the parameters by $\log n$ factors suitably, a union bound shows that almost surely every subset A of size roughly $n^{6/13}$ will contain a triangle. In fact, a slightly more careful examination of this argument reveals that almost surely every such subset will contain at least $n^{6/13}$ triangles, each coming from a single tripartite graph such that the tripartites corresponding to different triangles are all distinct.

Now let us specify which edges get deleted. We shall delete them in two stages. The first stage consists of what we call *Type 1* deletions. Given any two of our random tripartite graphs, with vertex sets $A = A_0 \cup A_1 \cup A_2$ and $B = B_0 \cup B_1 \cup B_2$, we remove all edges xy such that $x, y \in A \cap B$. We do not insist that xy is an edge of both tripartite graphs: if, for example, $x, y \in A_0$, $x \in B_0$ and $y \in B_1$, then the edge xy will be removed. Let G_1 be the resulting graph when all such edges have been deleted. The reason for these deletions is that each of our tripartite graphs contains many copies of $K_{3,1,1}$, which are somewhat "dangerous" for us, since all it takes to convert a $K_{3,1,1}$ into a K_5 is the addition of a further triangle. If we do not do Type 1 deletions, then we will obtain K_5s in this way too frequently, with the result that most edges in the graph are contained in a K_5. Indeed, the expected number of edges in G_0 is roughly $n^{9/13}(n^{6/13})^2 = n^{21/13}$ and the expected number of K_5s of the above form is roughly $n^5(n^{9/13})^2(n^{-7/13})^8 = n^{27/13}$.

Type 1 deletion is feasible in the sense that it destroys only a small proportion of the edges of G_0. That is because it is significantly less likely for a pair of vertices to be contained in two tripartite copies than for it to be contained in one tripartite copy.

Thanks to Type 1 deletions, it has become "difficult" for K_5s to appear in G_1, since now none of our random tripartite graphs can intersect a K_5 in more than 3 vertices. Indeed, if one of them intersects a K_5 in say 4 vertices, then there exist two of those vertices between which this tripartite does not provide an edge, and if one of the other tripartites gives an edge in G_0 between those two vertices, that edge is deleted.

Thus, it is easy to check that if a K_5 appears in G_1, then it has to do so in one of the following ways.

(i) All 10 edges of the K_5 come from distinct tripartites.

(ii) There is one tripartite giving a triangle in the K_5 but all the other 7 edges come from distinct tripartites.

(iii) There are two tripartites that each give a triangle in the K_5, these two triangles sharing a single vertex, and all the other 4 edges come from distinct tripartites.

We now delete at least one edge from each of these remaining K_5s. This will be done probabilistically and the precise method will be explained later. The deletions in this second round we call *Type 2* deletions. Once they have been performed, the resulting graph is our final graph G.

The graph G is K_5-free, by definition, but we now have to show that we have not inadvertently destroyed all the triangles in some set of $n^{6/13}(\log n)^c$ vertices. We begin by checking the more basic requirement that the Type 2 deletions destroy only a small proportion of the edges. That is, we check that the expected number of K_5s in G_1 is less than the expected number of edges (which is already computed to be $n^{21/13}$). To do this, we split into the three cases mentioned above. To calculate the expected number of K_5s of type (i), observe that there are at most n^5 choices for the vertex set, and $(n^{9/13})^{10}$ choices for the copies of tripartites giving an edge (since there are $n^{9/13}$ tripartites to choose from and we need 10 of them), and the probability that the vertices of the K_5 are in these tripartites as prescribed is $(n^{-7/13})^{20}$ (since the probability that a given vertex is in a given tripartite is $n^{-7/13}$), giving that the expected number of these K_5 is $n^{15/13}$. Similarly, the expected number of K_5s of type (ii) is $n^5(n^{9/13})^8(n^{-7/13})^{17} = n^{18/13}$. Finally, the expected number of K_5s of type (iii) is $n^5(n^{9/13})^6(n^{-7/13})^{14} = n^{21/13}$. This last number is roughly equal to the expected number of edges, therefore we will need to modify the parameters by $\log n$ factors. However, the main point is that after this second round of deletions, most edges of the original graph are still present.

In order to finish off the proof, there are two main difficulties to overcome. The first one is that even though we have made sure that globally not too many edges are deleted, this is, as we have already mentioned, just a necessary condition for the argument to have a chance of working. What we actually need is the stronger statement that every induced subset of size $n^{6/13}(\log n)^c$ still contains a triangle. We can hope that the small set of edges we have removed is "sufficiently random" for this to be the case, but actually proving that takes some work. Let us sketch how we do it. From now on, it will be convenient to think of each tripartite as having a colour: accordingly, we call the tripartites "colour classes". If a vertex belongs to, say, the red tripartite, then we say that that vertex is red.

Let us now fix a set A of size $n^{6/13}(\log n)^c$. As shown above, we can take it for granted that G_0 contains a big set $\mathcal{T}$ of triangles in A, all coming from different colour classes. Moreover, these triangles will be uniformly distributed over A. Let $T_C \in \mathcal{T}$ be a triangle coming from the colour class C. (Note that not every colour gives a triangle, and not every triangle in A comes from just one colour class.) Let us first deal with Type 1 deletions. An edge of some T_C gets deleted by the Type 1 deletions if the endpoints of this edge share a colour other than C. So intuitively we can imagine that G_0 has already been constructed, and then we place these triangles T_C randomly inside A and hope that most triangles will not have any edge contained in another colour class. It is not too hard to show, under suitable assumptions, that with very high probability the density of pairs of vertices in A sharing a colour is fairly low (this essentially comes from the fact that the typical sizes of the tripartites are smaller - after adjusting the parameters by suitable log factors - than the size of A). Therefore for a fixed T_C it is indeed true that with fairly high probability its edges will not be deleted by Type 1 deletions. However, these events are not independent for different colours C. To overcome this difficulty, we define a set Π of roughly $\log n$ partitions with the property that for any pair of distinct colours C, D there is a $\pi \in \Pi$ such that D is in the first part of π and C is in the second part. We now define a π-*dangerous pair* to be a pair of vertices that share a colour from the first part of π. If an edge xy of a T_C gets deleted (by Type 1 deletions) then x and y share a colour $D \neq C$ and there is some $\pi \in \Pi$ such that D is in the first part of π and C is in the second part of π and therefore (x, y) is a π-dangerous pair. But note that, as indicated above, the density of π-dangerous pairs will be fairly low, so the probability that an edge of T_C is deleted because of a colour in the first part of π is low, and, conditional on the outcome of colours in the first part of π, these events are now independent for all C in the second part of π. We can therefore conclude that only a small proportion of these T_Cs will lose an edge thanks to colours in the first part of π. Thus, since Π is small, we deduce that most triangles T_C will not lose an edge. That is, we can find many triangles in A even after the Type 1 deletions.

Now let us define Type 2 deletions. Given the graph G_1, we order its edges randomly and keep each edge provided that it does not form a K_5 when combined with the edges that we have already decided to keep. We remark that this construction is a variant of the so called K_5-free process. The edges we keep will form our final graph G.

To be more precise, we note here that in fact we keep an edge only if it does not form a so called *core* of a K_5 of G_1 when combined with the edges that we have already decided to keep. The core is a certain subgraph of a K_5 defined in terms of the colours of its edges. The reader is encouraged to think of the core of a K_5 as the K_5 itself (especially as we can prove that the core of any K_t is itself, but the proof of this fact is very long and we do not include it here).

As shown above, the number of K_5s in G_1 is less than the number of edges, that is, on average an edge is contained in less than one K_5. In fact, one can show that almost surely *every* edge will be contained in a relatively small number of K_5s. It is not hard to see that this means that any triangle in G_1 is also present in G with probability not very close to 0. Since the number of triangles in $G_1[A]$ is large, standard concentration inequalities will imply that with very high probability $G[A]$ still contains a triangle. Using the union bound over all A (of size roughly $n^{6/13}$), we conclude that almost surely every $G[A]$ contains a triangle, finishing the proof.

Let us briefly discuss how we determined the parameters of our construction. Let n^δ be the number of tripartite copies placed, let n^β be the size of each part of each of these copies, and let n^α be the set size that will guarantee an induced triangle. The parameters δ, β have been chosen to optimize the result: that is, to allow α to be as small as possible. There are three main conditions that we need to impose on these parameters.

The first one is that we need enough triangles in G_0 inside every A of size n^α. It is not hard to see that this condition is equivalent to

$$\delta + 3(\alpha + \beta - 1) \geq \alpha. \tag{5.3}$$

The second one comes from the fact that the parts of the tripartites will not contain a triangle in G (since every edge inside a part of a tripartite gets deleted by Type 1 deletions), so we trivially need

$$\alpha \geq \beta. \tag{5.4}$$

Finally, we want the expected number of K_5s in G_1 to be less than the expected number of edges in G_1 which gives (only considering those K_5s which are type (iii) in the sense described a few paragraphs above)

$$\delta + 2\beta \geq 5 + 6\delta + 14(\beta - 1). \tag{5.5}$$

It is not hard to see that these conditions force $\alpha \geq 6/13$ and that equality is achieved by taking $\delta = 9/13, \beta = 6/13$.

This leads us to the other main difficulty, which arises only when we consider more general values of s, t. While (5.3) is essentially the same but with 3 replaced by s, and (5.4) is exactly the same, (5.5) becomes completely different. Indeed, it will be crucial to analyse all possible ways that a K_t can occur in G_1 in some systematic way, rather than writing down the three possibilities (i),(ii),(iii) as we did above in the $s = 3, t = 5$ case, since in general there are many ways that a K_t can be formed from the contributions of the various s-partite graphs. Analysing these decompositions of K_t, which we shall refer to as colour schemes (again by imagining that each s-partite graph has its own colour), is necessary to determine the best parameters δ, β, and also to prove Theorem 5.1.3 for these parameters. The complicated formula for α is obtained by solving the system of inequalities (5.3),(5.4),(5.5) that we obtain in the general case.

The organization of this chapter is as follows. In Section 5.2 we present our construction. In Section 5.3 we give the main part of the proof conditional on three lemmas. These lemmas are proved in Section 5.4. The first one, which asserts that each edge in G_1 is contained in a small number of (cores of) K_ts, is proved in Subsection 5.4.1, conditional on a lemma about colour schemes that is proved in Subsection 5.4.3. The result that says that $G_1[A]$ contains many K_ss is proved in Subsection 5.4.2. Finally, there is an appendix that contains some tedious computations and the source code of a program relevant to some results in Subsection 5.4.3.

5.2 The precise construction and the main result

Remark. Logarithms throughout the chapter are to base e. We will not be concerned with floor signs, divisibility, and so on. Also, we will tacitly assume that n is sufficiently large whenever this is needed. Moreover, throughout the rest of the chapter, it is to be understood that $s \geq 3$ and that $s+2 \leq t \leq 2s-1$. Recall that a pair (s,t) is regular if $s \geq 11$ and $s+3 \leq t \leq 2s-4$ or if $(s,t) \in \{(10,14),(10,15)\}$, and otherwise it is exceptional.

Let

$$\delta = s-(2s-1)\alpha = \begin{cases} \delta(1) = \frac{(2s-2)(t-s)(t+s-1)+2s^2-4st+2t+2s}{(2s-3)(t-s)(t+s-1)-2s+4}, & \text{if } (s,t) \text{ is regular} \\ \delta(2) = \frac{(2s-2)(t-s)(s-1)-st+3s-1}{(2s-3)(t-s)(s-1)+2s-t}, & \text{if } (s,t) \text{ is exceptional} \end{cases}$$

Lemma 5.2.1. $\delta < 2\alpha < 1$.

Proof. If (s,t) is regular, then

$$\begin{aligned} 2\alpha - \delta &= \frac{4st-2s^2+2t-6s-2(t-s)(t+s+1)}{(2s-3)(t-s)(t+s-1)-2s+4} \\ &= \frac{4st-2t^2-4s}{(2s-3)(t-s)(t+s-1)-2s+4} > 0, \end{aligned}$$

since $s+1 \leq t \leq 2s-2$. If (s,t) is exceptional, then

$$2\alpha-\delta = \frac{st-s-1-2(t-s)(s-1)}{(2s-3)(t-s)(s-1)+2s-t} = \frac{2s^2-st+2t-3s-1}{(2s-3)(t-s)(s-1)+2s-t} > 0,$$

since $s+1 \leq t \leq 2s-1$.

By Lemma 5.A.2 (e) from the appendix, we have $\delta > 2/3 > 1/2$, which implies that $\alpha < 1/2$. □

Remark. Intuitively, one can think of α as $1/2-\epsilon$ for ϵ quite small and $\delta = 1/2+(2s-1)\epsilon$. This makes δ significantly greater than 1/2 but less than 1. Also, it may be helpful to bear in mind the case $s=3, t=5$, where, as we have seen, $\delta = 9/13$ and $\alpha = 6/13$.

Let

$$m = n^{\delta}(\log n)^{-c_1}$$

$$\gamma = n^{\alpha-1}(\log n)^{-c_2}$$

$$a = n^{\alpha}(\log n)^{c_3}$$

where c_1, c_2, c_3 are positive constants, to be specified, that depend on s and t. (In fact, c_1 can be taken to be 0. All we need are that c_2 is suitably large and that c_3 is sufficiently larger than c_1, c_2.)

The following estimates will be used several times later in the chapter.

Lemma 5.2.2. *$m\gamma > 1$ and $m\gamma^2 < 1$.*

Proof. Note that $\delta + (\alpha - 1) = (s-1) - (2s-2)\alpha > 0$ since $\alpha < 1/2$. This implies that $m\gamma > 1$.

Also, $\delta + 2(\alpha - 1) < 4\alpha - 2 < 0$, by Lemma 5.2.1. This implies that $m\gamma^2 < 1$. □

We construct the graph G_0 as follows. Let $V = V(G_0) = \{1, 2, ..., n\}$. Define independent random subsets $S_1, ..., S_m$ of V in such a way that each S_i contains each $v \in V$ independently with probability γ. We call S_i the ith *colour class*. If $v \in S_i$, we say that v *has colour i*. Now randomly partition each S_i into s sets, $S_{i1}, S_{i2}, ..., S_{is}$ by placing each element of S_i independently at random in one of these parts, and use these sets to define a complete s-partite graph. Let G_0 be the union of these s-partite graphs. We say that a pair of vertices has colour i if both its members have colour i. We do not require the pair to form an edge in G_0. Remove all edges of G_0 that have at least two colours to obtain the subgraph G_1. Again, we do not require both colours to give an edge. Another way to state the condition is that if xy is an edge of colour i and x and y both have colour j for some $j \neq i$, then we remove the edge xy even if x and y belong to the same set S_{jr}. Finally, for every K_t in G_1 we randomly remove a certain edge, which we shall specify in a moment. The resulting graph is called G.

The graph G is obviously K_t-free. We shall prove that for suitable choices for the constants c_1, c_2, c_3, we have the following result, which is our main theorem.

Theorem 5.2.3. *For n sufficiently large, there is a positive probability that every subset A of G with $|A| = a$ contains a K_s.*

Obviously Theorem 5.2.3 implies Theorem 5.1.3.

Let us now specify which edges are removed from G_1. Suppose that $x_1, ..., x_t$ form a K_t in G_1. Then necessarily any two distinct vertices x_i and x_j share precisely one colour. Indeed, they must share at least one colour since $x_ix_j \in E(G_0)$ but they cannot share more than one since then x_ix_j would have been removed from G_0 during the first round of deletions.

Definition 5.2.4. A *colour scheme* for K_t with parameter s, or *scheme* for short, is a set X of t nodes and a set $\mathcal{D}$ of subsets of X, which we call *colours*, or *blocks*, such that

(i) For any $x, y \in X$, there is a unique $D \in \mathcal{D}$ such that x, y both belong to D.

(ii) Every colour appears on at least two nodes.

(iii) Every colour appears at most s times.

A pair of nodes is called an *edge* and the *colour* of an edge is the unique colour that contains both endpoints. (Note that a node may have several colours.) If a node x belongs to a colour D, we shall say that D *labels* x. We also define a *label* to be a pair (x, D) such that x is a node and D labels x. The number of labels in a scheme is thus the sum of the sizes of all the colours.

If $X = \{x_1, ..., x_t\}$ forms a K_t in G_1, then there is set of (at most $\binom{t}{2}$) colours such that X is a colour scheme with respect to those colours, and no other colour labels more than one vertex in X. Indeed, we have already observed that property (i) holds. Choosing the colours suitably, (ii) can clearly be achieved. For property (iii), observe that if some colour D labels at least $s+1$ vertices, then there must exist distinct vertices x_i and x_j that belong to the same part of the complete s-partite graph of colour D. Then D does not provide an edge between x_i and x_j, so some other colour must, but then x_i and x_j share at least two colours, which contradicts (i).

Thus, any K_t in G_1 can be viewed as a scheme in a natural way. A simple upper bound for the expected number of K_ts associated with a scheme Q is $n^t m^b \gamma^l$, where l is the number of labels of Q and b is the number of colours of Q. Indeed, the number of ways choosing the t nodes is at most n^t, the number of ways of choosing the b colours (from the m colours used to construct G_1) is at most m^b, and the probability that any given choice of nodes and colours realizes the scheme is γ^l, since for each label the probability that the given node receives the given colour is γ, and all these events are independent.

Now $n^t m^b \gamma^l = n^{t+b\delta+l(\alpha-1)}(\log n)^f$ for some $f = f(s,t,b,l)$. Also, once we know that a certain pair u, v of vertices have a colour in common, the expected number of K_ts associated with Q that contain u and v becomes at most roughly $n^{t-2}m^{b-1}\gamma^{l-2} = n^{t-2+(b-1)\delta+(l-2)(\alpha-1)}(\log n)^{f'}$. This motivates the following definition.

Definition 5.2.5. The *value* of a scheme Q with b colours and l labels, denoted $v(Q)$, is given by the formula

$$v(Q) = t - 2 + (b-1)\delta + (l-2)(\alpha - 1).$$

Thus, roughly speaking, the expected number of K_ts associated with a scheme Q that contain a given edge in G_1 is at most $n^{v(Q)}$ up to log factors. The following lemma – proved in Subsection 5.4.3 – shows that this number is small.

Lemma 5.2.6. *Let Q be a scheme. Then $v(Q) \leq 0$.*

We shall also need a generalization of the notion of a scheme where a pair of nodes does not need to have a colour, if it does have a colour then that colour does not have to be unique, and a colour is allowed to label more than s nodes.

Definition 5.2.7. A *colour configuration* consists of a set of nodes and a set of colours labelling the nodes such that every colour appears on at least two nodes.

Given a colour configuration W and a subset S of its nodes, we define the *subconfiguration induced by* S to be the configuration whose nodes are the elements of S and whose colours are the colours of W that appear at least twice on S (which then label the nodes in S that they labelled in W).

The *value* of a configuration W is defined to be

$$v(W) = h - 2 + (b-1)\delta + (l-2)(\alpha - 1),$$

where h is the number of nodes, b is the number of colours and l is the number of labels in W (where a label is again a pair (x, D) where x is a node labelled by the colour D).

The same argument as for schemes shows that, once we condition on the event that u and v are both coloured red, the expected number of occurrences of a colour configuration W that contain both u and v is at most $n^{v(W)}$ up to log factors. (In fact, it is smaller unless u and v share a colour in W.)

Definition 5.2.8. The *core* of a scheme Q, denoted $C(Q)$, is the induced subconfiguration S on at least two nodes for which $v(S)$ is minimal. If several subconfigurations have the same value then the core is the one with the maximum number of nodes. If this is still not unique, then we simply pick an arbitrary one with the given properties.

Remark. We can in fact prove that $C(Q) = Q$ for every scheme Q. Although using that fact would simplify the argument in this chapter slightly, this gain does not compensate for the extra work needed to establish it, so we shall avoid using it. Nevertheless, the reader is encouraged to think of a core just as a scheme: that is, as a K_t in the graph G_1 with the colours given by the s-partite graphs with vertex sets that contain at least two of its vertices.

Lemma 5.2.9. *Let Q be a scheme. Then $C(Q)$ has at least 3 nodes, $v(C(Q)) \leq 0$, and $v(S) \geq v(C(Q))$ for every induced subconfiguration S of $C(Q)$ with at least two nodes.*

Proof. The first two assertions follow from Lemma 5.2.6, since an induced subconfiguration of Q with two nodes has value 0. The third assertion follows immediately from the definition of the core. □

We can now define G precisely. Following an idea in [109], we assign independently to each edge e of G_1 a birthtime β_e, chosen uniformly randomly from $[0, 1]$. Equivalently, we order the edges of G_1 uniformly at random from all the possible orderings. To define the edge set $E(G)$, which will be a subset of $E(G_1)$, we recursively decide for each $e \in E(G_1)$ whether $e \in E(G)$, as follows. Suppose that the decision has been made for every $e' \in E(G_1)$ with $\beta_{e'} < \beta_e$. Then let $e \in E(G)$ unless there is a K_t in G_1, which we view as a scheme Q, for which the edges of $C(Q)$ all have birthtime at most β_e and they all (apart from e) already belong to $E(G)$.

For any K_t in G_1 there is an edge in the core of that K_t that is not an edge of G, since if all the edges in the core apart from the last one are chosen to belong to $E(G)$, then the last one is not. Thus, G is K_t-free. It remains to prove that with positive probability every set of a vertices still contains a K_s, which was Theorem 5.2.3 above.

5.3 The proof of Theorem 5.2.3

In this section, we shall prove Theorem 5.2.3 conditional on two lemmas, which we shall prove in Section 5.4 and which are where most of the work will be. The first one says, roughly speaking, that for any A of size a, the induced subgraph $G_1[A]$ of G_1 contains many copies of K_s.

Lemma 5.3.1. *Almost surely, for every A of size a there is a set of $\Omega(ma^s\gamma^s)$ monochromatic copies of K_s inside $G_1[A]$, each with a different colour.*

The second tells us that any edge in G_1 is contained in few cores. Here, and in what follows, we use the word "core" to refer to the core of a K_t in G_1.

Lemma 5.3.2. *Almost surely, any edge in G_1 is contained in at most $(\log n)^{2t}$ cores.*

We shall use McDiarmid's inequality [95] in the next proof, which for convenience we recall here. Let $Y_1, \dots, Y_N$ be independent random variables, taking values in a set S, and let $X = g(Y_1, \dots, Y_N)$ for some $g : S^N \to \mathbb{R}$ with the property that if $y, y' \in S^N$ only differ in their ith coordinate, then $|g(y) - g(y')| \le c_i$. Then the inequality states that

$$\mathbb{P}\big[|X - \mathbb{E}[X]| \ge r\big] \le 2\exp\left(\frac{-2r^2}{\sum_i c_i^2}\right).$$

The following lemma, together with Lemmas 5.3.1 and 5.3.2 and a union bound, implies Theorem 5.2.3.

Lemma 5.3.3. *Suppose that G_1 is such that any edge in G_1 is contained in at most $(\log n)^{2t}$ cores. Let A be a set of vertices of size a such that the induced subgraph $G_1[A]$ contains $\Omega(ma^s\gamma^s)$ monochromatic copies of K_s, each with a different colour. Then the probability, conditional on the graph G_1, that $G[A]$ does not contain any K_s is $o\big(\frac{1}{\binom{n}{a}}\big)$.*

Proof. Choose $\Omega(ma^s\gamma^s)$ monochromatic copies of K_s in $G_1[A]$, all of distinct colours. Let the set of these copies be $\mathcal{T}$. Then by the definition of the first deletion process, the elements of $\mathcal{T}$ are edge disjoint. Let $T \in \mathcal{T}$. Let E_T be the set of all edges of cores that have at least one edge that belongs to T, together with the edges of T itself. Clearly, $|E_T| \le \binom{s}{2} + \binom{s}{2}(\log n)^{2t}\binom{t}{2} \le (\log n)^{3t}$. Let B_T be the event that the birthtimes of the edges of T precede the birthtimes of all other edges in E_T. If B_T occurs, then the only way an edge of T could be deleted from G_1 and therefore fail to be present in G is if T itself contains a core of some K_t. But note that there is no colour that labels every vertex in a core C. Indeed, if there is such a colour, then since all edges in a core belong to G_1, there is no other colour appearing at least twice on the node set of C, therefore C, considered as a colour configuration, has value $h-2+(h-2)(\alpha-1)=(h-2)\alpha$ (where h is the number of nodes in C), which contradicts Lemma 5.2.9. It follows that if B_T occurs, then every edge of T is present in G.

For a fixed G_1, let X be the number of events B_T that occur over all $T \in \mathcal{T}$. Then X is a random variable with the property that if $X \neq 0$, then there is some $T \in \mathcal{T}$ that belongs to $G[A]$. It therefore suffices to prove that $\mathbb{P}[X=0]=o\left(\frac{1}{\binom{n}{a}}\right)$.

To do this, we apply McDiarmid's inequality when Y_i is the birthtime of the ith edge. Since the $T \in \mathcal{T}$ are edge disjoint, and any edge e in G_1 is contained in at most $(\log n)^{2t}$ cores, it follows that e is contained in at most $1+(\log n)^{2t}\binom{t}{2} \leq (\log n)^{3t}$ of the graphs E_T. Hence, changing the birthtime β_e of e influences at most $(\log n)^{3t}$ of the events B_T. Also, if $e \notin \cup_{T\in\mathcal{T}} E_T$, then β_e does not influence *any* event B_T. Thus, by McDiarmid's inequality (with some $N \leq |\mathcal{T}|(\log n)^{3t}$), we get

$$\mathbb{P}[X=0] \leq 2\exp\left(\frac{-2(\mathbb{E}[X])^2}{|\mathcal{T}|(\log n)^{3t}((\log n)^{3t})^2}\right).$$

Now note that $\mathbb{P}[B_T] \geq |E_T|^{-\binom{s}{2}} \geq (\log n)^{-3s^2t}$, so $\mathbb{E}[X] \geq |\mathcal{T}|(\log n)^{-3s^2t}$, and

$$\mathbb{P}[X=0] \leq 2\exp\left(\frac{-2|\mathcal{T}|}{(\log n)^{6s^2t+9t}}\right).$$

Finally, note that $\binom{n}{a} \leq n^a = \exp(a\log n)$. To finish the proof we just need to verify that $\frac{|\mathcal{T}|}{(\log n)^{6s^2t+9t}} = \omega(a\log n)$. Since

$$\frac{|\mathcal{T}|}{a} = \Omega(ma^{s-1}\gamma^s) = n^{\delta+(s-1)\alpha+s(\alpha-1)}(\log n)^{-c_1+(s-1)c_3-sc_2} = (\log n)^{-c_1+(s-1)c_3-sc_2},$$

we are done provided that $(s-1)c_3 - sc_2 - c_1 > 6s^2t+9t+1$. □

5.4 The proofs of the auxiliary lemmas

In this section we shall prove Lemmas 5.2.6, 5.3.1 and 5.3.2, which are the results we used in the proof of Theorem 5.2.3 but have not yet proved.

5.4.1 The proof of Lemma 5.3.2

Let e be an edge in G_1. We would like to show that it belongs to at most $(\log n)^{2t}$ cores. Any core that contains e can be viewed as a core in a scheme that contains e, and as such it has nonpositive value. But for any colour configuration W (with more than two labels), the expected number of occurrences of that colour configuration in G_0 containing a fixed edge in G_1 is at most $n^{v(W)}(\log n)^{-c_2}$ (as we remarked slightly less precisely after Definition 5.2.7), which is at most $(\log n)^{-c_2}$ if $v(W) \leq 0$. In particular, the probability that an edge e in G_1 is contained in r cores that are pairwise disjoint apart from their intersection on e is at most $(\log n)^{-rc_2}$. If $r = \log n$ then this is much less than $1/n^2$, and therefore almost surely no edge is contained in $\log n$ cores of the above form.

In general, the cores containing e need not be disjoint. This adds a complication, and we need to introduce a few definitions to handle it, but the main reason Lemma 5.3.2 holds is the one given in the previous paragraph. The next definition describes the kind of colour configuration which – if it occurs in G_0 – can produce many cores in G_1 (that is, cores of K_ts in G_1 that we view as schemes) that contain a given edge xy. Soon we shall argue that almost surely no such large configuration occurs in G_0.

Definition 5.4.1. An *abstract core container* W is a colour configuration whose nodes are $\{x\} \cup \{y\} \cup Z$ and in which every $z \in Z$ is contained in at least one abstract core, where an abstract core is defined as follows.

An *abstract core* in a core container is an induced subconfiguration S consisting of at most t nodes and containing x and y such that for any induced subconfiguration $S' \subset S$ containing x, y, we have $v(S') \geq v(S)$ and such that for any two distinct $u, v \in S$ there is a unique colour that labels both u and v.

The *size* of a core container is the number of nodes it contains.

A core container is *irreducible* if it is not possible to remove a label or colour and still have a core container.

Remark. Assume for a moment that we know that the core of a scheme is the scheme itself (see the remark after Definition 5.2.8). Then Lemma 5.3.2 just asserts that each edge in G_1 is contained in few K_ts. Then we can replace the technical notion of abstract core container with the notion of *abstract scheme container* instead. What we mean by that is a colour configuration whose nodes are $\{x\} \cup \{y\} \cup Z$ and in which every $z \in Z$ is contained in at least one colour scheme containing x and y as well. This is a configuration that is dangerous to us since if it occurs in G_0, then the edge xy is contained in many K_ts (corresponding to the various schemes in the configuration).

Note that as the vertices of G_0 are coloured, we can naturally talk about G_0 containing various colour configurations. We shall now establish that:

1. If an edge in G_1 is contained in many cores, then there is a large irreducible core container in G_0.

2. There are not too many irreducible abstract core containers of fixed size.

3. The expected number of occurrences in G_0 of any large abstract core container is small.

The last two points will imply that almost surely there is no large irreducible core container in G_0, which in turn implies that there is no edge in G_1 that is contained in many cores.

Note that for the second point it is important that we count only irreducible core containers because otherwise the number of colours in the core container could be arbitrarily large.

Lemma 5.4.2. *If the edge $e = uv$ is contained in at least r cores of K_ts in G_1, then there is an irreducible core container W in G_0 with $x = u, y = v$ (as in Definition 5.4.1) and with size between $\frac{1}{2}r^{1/t}$ and tr.*

Proof. Define a colour configuration W_0 as follows. Arbitrarily pick r cores that contain e. The set of nodes of W_0 is the set of vertices of G_1 that are in one of these r cores. The set of colours is the set of those colours in G_0 that appear at least twice on this set of nodes. This does indeed define a core container, since any core of a K_t in G_1 that contains e satisfies the two properties required of an abstract core in W_0: the minimality of v follows from the definition of a core, and the condition about the colours follows from the fact that the K_t belongs to G_1.

How many nodes does W_0 have? Any core consists of between 2 and t nodes, so if the number of nodes of W_0 is h, then $r \leq \sum_{2\leq j\leq t} \binom{h}{j} \leq (2h)^t$. Thus, $h \geq \frac{1}{2}r^{1/t}$. On the other hand, $h \leq rt$, since the vertex set of W_0 is a union of r cores. Now remove labels or colours as long as we still get a core container; the object we end up with is an irreducible core container of the required size. □

Lemma 5.4.3. *The number of distinct irreducible abstract core containers of size h is at most $ht^2 \cdot 2^{ht^2} \cdot h^{ht^2}$.*

Proof. First we shall prove that the number of labels in an irreducible core container of size h is at most $2h\binom{t}{2} \leq ht^2$. For any occurrence of a colour D at some node u (that is, for any label (u, D)), there must exist $v \in \{x\} \cup \{y\} \cup Z$ such that every abstract core containing v contains u and the colour D, or else we could remove the occurrence of D at u and still have a core configuration. But for any v, there are at most $2\binom{t}{2}$ such pairs (u, D), since u must belong to the intersection of the vertex sets of the abstract cores containing v, and in a given abstract core there are at most $2\binom{t}{2}$ labels. Indeed, an abstract core is an induced subconfiguration so each of its colours labels at least two nodes. Now if an abstract core has q colours and they label $d_1, \ldots, d_q$ nodes, then $\sum_{i\leq q} \binom{d_i}{2} \leq \binom{t}{2}$ because the abstract core has at most $\binom{t}{2}$ pairs of nodes. Since $d_i \geq 2$ for each i, it follows that $\sum_{i\leq q} d_i \leq 2\binom{t}{2}$.

So there are at most ht^2 choices for the total number of labels. Since the partition function $p(k)$ is at most 2^k, it follows that for each possibility for the number of labels, there are at most 2^{ht^2} choices for the number of occurrences for each colour class. Suppose we have b colours and the numbers of times that they occur are $l_1, \ldots, l_b$. Then the number of choices for the vertices labelled by these colours is at most $\binom{h}{l_1}\binom{h}{l_2}\dots\binom{h}{l_b} \leq h^{l_1+\dots+l_b} \leq h^{ht^2}$. $\square$

Next, we shall investigate how many copies we expect to have in G_0 of a given abstract core container. Let W be more generally any colour configuration with h nodes, b colours and l labels. Then the expected number of occurrences of such a configuration is at most $n^h m^b \gamma^l$. Indeed, the number of ways of choosing the h nodes is at most n^h. The number of ways of choosing the b colours is at most m^b. And for each label, the probability that the given node receives the given colour is γ, and all these events are independent, so the probability that any given choice of nodes and colours realizes the scheme is γ^l.

Definition 5.4.4. We call $n^h m^b \gamma^l$ the *frequency* of the configuration W and denote it by $\omega(W)$.

Lemma 5.4.5. *Let W be an abstract core container of size h. Then*

$$\omega(W) \leq n^2(\log n)^{-\frac{h-2}{t}c_2}.$$

To prove this result, we will kill some of the nodes and colours and remove some of the labels of the core container in steps. To keep track of which nodes and colours have been killed, we introduce the following definition.

Definition 5.4.6. A *partial configuration* P consists of four pairwise disjoint sets $\{x\}$, $\{y\}$, Z_0 and Z_1 of nodes, and two disjoint sets $\mathcal{B}_0, \mathcal{B}_1$ of colours that label those nodes in such a way that any $B \in \mathcal{B}_1$ labels at least two nodes. We write $\mathcal{B}$ for $\mathcal{B}_0 \cup \mathcal{B}_1$ and Z for $Z_0 \cup Z_1$.

We now generalize the notion of frequency to this setting, which can be thought of as the expected number of occurrences of the colour configuration for given choices of the nodes in Z_0 and colours in $\mathcal{B}_0$, which represent the nodes and colours that have already been killed. Thus, we let $r = |\{x\} \cup \{y\} \cup Z_1|$ be the number of nodes yet to choose, we let $g = |\mathcal{B}_1|$ be the number of colours yet to choose, and we let u be the total number of labels, including the labels on nodes in Z_0 and of colours in $\mathcal{B}_0$. Then we can choose the remaining nodes in at most n^r ways and the remaining colours in at most m^g ways, and for each label there is a probability γ that the given node receives the given colour. So we define the frequency $\omega(P)$ to be $n^r m^g \gamma^u$.

Proof of Lemma 5.4.5. We shall define a sequence $P_0, \dots, P_k$ of partial configurations such that $\omega(P_0) = \omega(W)$, $\omega(P_k) \leq n^2$, $k \geq \frac{h-2}{t}$ and $\omega(P_j) \geq \omega(P_{j-1})(\log n)^{c_2}$. Clearly, this suffices to prove the lemma.

We shall define the P_j recursively. In what follows we use the notation of Definition 5.4.1 and Definition 5.4.6. When there is ambiguity, we will write $Z_0(P)$ to mean Z_0 in the partial configuration P, and similarly for $Z_1, \mathcal{B}_0, \mathcal{B}_1$. The set of all nodes (respectively, colours) for every P_j will be the same as the set of all nodes (respectively, colours) of W, namely $\{x\} \cup \{y\} \cup Z$ (respectively, $\mathcal{B}$). However, $\mathcal{B}_0, \mathcal{B}_1, Z_0, Z_1$ and the labels will be different for the various P_j.

Let us define P_0 to be the partial configuration whose nodes, colours and labels are the same as those of W and which has $Z_0 = \mathcal{B}_0 = \emptyset$. Then $\omega(P_0) = \omega(W)$.

Given P_{j-1} with $Z_1(P_{j-1}) \neq \emptyset$, we define P_j as follows. Pick some $z \in Z_1(P_{j-1})$ arbitrarily. As W is a core container, we can choose an abstract core S in W that contains z. Let $S_1 = S \cap Z_1(P_{j-1})$. Let $\mathcal{D}$ be the set of those colours $B \in \mathcal{B}_1(P_{j-1})$ that occur at least twice on S in P_{j-1}. Then let the sets of nodes of P_j be $Z_0(P_j) = Z_0(P_{j-1}) \cup S_1$ and $Z_1(P_j) = Z_1(P_{j-1}) \setminus S_1$, and let the sets of colours be $\mathcal{B}_0(P_j) = \mathcal{B}_0(P_{j-1}) \cup \mathcal{D}$ and $\mathcal{B}_1(P_j) = \mathcal{B}_1(P_{j-1}) \setminus \mathcal{D}$. The labels of P_j are those of P_{j-1} except that all occurrences of colours in $\mathcal{B}_0(P_j)$ are removed from S. It is clear that P_j is a partial configuration.

We want to prove that $\omega(P_j) \geq \omega(P_{j-1})(\log n)^{c_2}$.

Claim. $\frac{\omega(P_j)}{\omega(P_{j-1})} \geq \frac{\omega(S \setminus S_1)}{\omega(S)}$, where S and $S \setminus S_1$ are identified with their induced subconfigurations from W.

Proof of Claim. The contribution of the nodes is (a factor of) $n^{-|S_1|}$ to both $\frac{\omega(P_j)}{\omega(P_{j-1})}$ and $\frac{\omega(S \setminus S_1)}{\omega(S)}$. Hence it suffices to prove that the contribution of any colour (and its labels) to $\frac{\omega(P_j)}{\omega(P_{j-1})}$ is at least as much as its contribution to $\frac{\omega(S \setminus S_1)}{\omega(S)}$. There are two cases to consider.

Case 1. If B is a colour that occurs at most once on S in W, then its contribution to $\frac{\omega(S \setminus S_1)}{\omega(S)}$ is 1, whereas its contribution to $\frac{\omega(P_j)}{\omega(P_{j-1})}$ is at least 1. (Indeed, since $m\gamma^2 < 1$, the contribution of *any* colour to $\frac{\omega(P_j)}{\omega(P_{j-1})}$ is at least 1.)

Case 2. Suppose, then, that B is a colour that occurs at least twice on S in W.

Case 2a. If $B \in \mathcal{B}_0(P_{j-1})$, then let d be the number of occurrences of B on S_1 in W. The contribution of B to $\frac{\omega(S \setminus S_1)}{\omega(S)}$ is at most γ^{-d}. Indeed, this is clear unless B occurs exactly once on $S \setminus S_1$ in W. But if this is the case, then the contribution of B is precisely $m^{-1}\gamma^{-(d+1)}$, which is at most γ^{-d}, by Lemma 5.2.2.

Note that any node in S_1 (and in fact more generally in $Z_1(P_{j-1})$) that is labelled by B in W is also labelled by B in P_{j-1}. Therefore, the contribution of B to $\frac{\omega(P_j)}{\omega(P_{j-1})}$ is at least γ^{-d}.

Case 2b. If $B \in \mathcal{B}_1(P_{j-1})$, then let d be the number of occurrences of B on S in W. The contribution of B to $\frac{\omega(S \setminus S_1)}{\omega(S)}$ is at most $m^{-1}\gamma^{-d}$. Indeed, this is clear unless B occurs at least twice on $S \setminus S_1$ in W. But in this case the contribution of B is at most $\gamma^{-(d-2)}$, which is at most $m^{-1}\gamma^{-d}$, by Lemma 5.2.2.

Note that any node that is labelled by B in W is also labelled by B in P_{j-1}. Therefore, $B \in \mathcal{D}$ and the contribution of B to $\frac{\omega(P_j)}{\omega(P_{j-1})}$ is precisely $m^{-1}\gamma^{-d}$.

This completes the proof of the claim.

Since S is an abstract core in W, we have $v(S) \leq v(S \setminus S_1)$, by the minimality of S. Because $S_1 \neq \emptyset$, and every node in a core has a label on it, it follows that, considering S and $S \setminus S_1$ as induced subconfigurations of W, we have $\omega(S \setminus S_1) \geq \omega(S)(\log n)^{c_2}$. Using the claim above, the inequality $\omega(P_j) \geq \omega(P_{j-1})(\log n)^{c_2}$ follows.

Eventually we obtain a partial configuration P_j with $Z_1(P_j) = \emptyset$. When this happens, we set $k = j$. By definition, we have in that case that $\omega(P_k) = n^2 m^g \gamma^u$ where $g = |\mathcal{B}_1(P_k)|$ and u is the number of labels in P_k. Since any $B \in \mathcal{B}_1(P_k)$ labels at least two nodes in P_k and $m\gamma^2 \leq 1$, we find that $\omega(P_k) \leq n^2$. Also note that $|Z_1(P_j)| \geq |Z_1(P_{j-1})| - t$ for any j, and $|Z_1(P_0)| = |Z| = h - 2$, so $k \geq \frac{h-2}{t}$. □

We are now in a position to complete the proof of Lemma 5.3.2.

Proof of Lemma 5.3.2. By Lemma 5.4.2, it suffices to prove that in G_0 the expected number of irreducible core containers of size between $\log n$ and $(\log n)^{3t}$ is $o(1)$.
Claim. If $\log n \leq h \leq (\log n)^{3t}$, then the expected number of irreducible core containers of size h in G_0 is at most $n^2 (\log n)^{-t^3 h}$.
Proof of Claim. By Lemmas 5.4.3 and 5.4.5, the expected number of irreducible core containers of size h in G_0 is at most $ht^2 2^{ht^2} h^{ht^2} n^2 (\log n)^{-\frac{h-2}{t} c_2} \leq h^{3ht^2} n^2 (\log n)^{-\frac{h-2}{t} c_2}$. If $c_2 \geq 11t^4$, then this is at most $h^{3ht^2} n^2 (\log n)^{-11(h-2)t^3} \leq h^{3ht^2} n^2 (\log n)^{-10ht^3} \leq n^2 (\log n)^{-ht^3}$ so the claim is proved.

But $\sum_{h \geq \log n} n^2 (\log n)^{-t^3 h} = o(1)$, and the proof is complete. □

5.4.2 The proof of Lemma 5.3.1

Our proof is based on the following two observations.

1. For any set of vertices A of size a, $G_0[A]$ contains many monochromatic s-cliques with pairwise distinct colours.

2. If a monochromatic s-clique is present in G_0, then it is present also in G_1 with high probability, and, crucially, the events that various s-cliques are preserved are "sufficiently independent".

First, we shall construct a small set of bipartitions of the set of colours with a suitable property. In a moment it will become clear why we need this. We will refer to the two parts of a bipartition as the *first part/first half* and the *second part/second half.*

Lemma 5.4.7. *There exists a constant c and a set Π of $c \log n$ partitions of the set of m colours, each into two sets of size $m/2$, such that for any two distinct colours C and D there is a $\pi \in \Pi$ such that D is contained in the first part of π and C is contained in the second part of π.*

Proof. Take $l = c \log n$ random partitions. For any C, D, the probability that none of the partitions is suitable is less than $(1 - \frac{1}{5})^l = n^{-c \log(5/4)}$. For c sufficiently large this is less than n^{-2}, which is in turn less than m^{-2} and the result follows from the union bound over all choices of C, D. □

Let xy be an edge in G_0. Recall that it is not an edge in G_1 if x, y have at least two colours in common. Suppose that this is the case. Then there exists some $\pi \in \Pi$ such that x and y have a colour in common from the first half of π and also a colour in common from the second half of π.

Remark. From now on, when we say "the first $m/2$ colours", we will mean "the $m/2$ colours in the first part of π" provided it is clear which π we are talking about.

Definition 5.4.8. A pair (x, y) of vertices is π-*dangerous* for some $\pi \in \Pi$ if there is a colour class among the first $m/2$ colours that contains both x and y.

Fix a set A of vertices with $|A| = a$. Let $\mathcal{D}$ be the collection of colours D such that at least one K_s inside A is entirely coloured with colour D in G_0. (We require that every edge is given by this colour: that is, the vertices of the K_s are in different parts of the complete s-partite graph with colour D.) For each $\pi \in \Pi$, let $\mathcal{D}_\pi$ be the set of all $D \in \mathcal{D}$ such that D is one of the last $m/2$ colours.

To make sense of the statement of the next lemma, the reader should recall that $a\gamma$ is significantly less than 1. (See the beginning of Section 5.2 for their precise values.)

Lemma 5.4.9. *With probability* $1 - o(\frac{1}{\binom{n}{a}})$, $|\mathcal{D}_\pi| = \Omega(ma^s\gamma^s)$ *for every* $\pi \in \Pi$.

Proof. Let C be any colour class. The probability that C intersects A in exactly s elements is

$$\mathbb{P}[\mathrm{Bin}(a, \gamma) = s] = \binom{a}{s}\gamma^s(1-\gamma)^{a-s} = \Omega(a^s\gamma^s(1-\gamma)^a) = \Omega(a^s\gamma^s(e^{-2\gamma})^a) = \Omega(a^s\gamma^s),$$

where the last inequality follows from the fact that $a\gamma = n^{2\alpha-1}(\log n)^{c_3-c_2} = o(1)$.

Hence $\mathbb{P}[C \in \mathcal{D}] = \Omega(a^s\gamma^s)$. Moreover, the events $\{C \in \mathcal{D}\}$ are independent. Thus, for any π, by the Chernoff bound we get $\mathbb{P}\big[|\mathcal{D}_\pi| = o(ma^s\gamma^s)\big] \le e^{-\Omega(ma^s\gamma^s)}$. Therefore, using the union bound over all $\pi \in \Pi$, it suffices to prove that $(\log n)e^{-\Omega(ma^s\gamma^s)} = o(\frac{1}{\binom{n}{a}})$.

But $\binom{n}{a} \le n^a = e^{a\log n}$. Hence, we need $(\log n)e^{-\Omega(ma^s\gamma^s)} = o(e^{-a\log n})$. For this, it is enough to prove that $a\log n = o(ma^s\gamma^s)$, ie. $\log n = o(ma^{s-1}\gamma^s)$. Since

$$ma^{s-1}\gamma^s = n^{\delta+(s-1)\alpha+s(\alpha-1)}(\log n)^{-c_1+(s-1)c_3-sc_2} = (\log n)^{-c_1+(s-1)c_3-sc_2}, \tag{5.6}$$

we are done provided that $(s-1)c_3 - sc_2 - c_1 > 1$. □

Therefore, using the union bound over all sets A of size a, we may assume that $|\mathcal{D}_\pi| = \Omega(ma^s\gamma^s)$ for every $\pi \in \Pi$ and every such set A.

Lemma 5.4.10. *With probability* $1 - o(1)$ *the following holds. For every* A *of size* a *and for every* $\pi \in \Pi$, *the density of* π-*dangerous pairs in* A *is* $o(\frac{1}{\log n})$.

This result, which we shall prove later, allows us to assume for our fixed set A that the following statement holds.

($\star$) For any $\pi \in \Pi$, the density of π-dangerous pairs in A is $o(\frac{1}{\log n})$.

For each $C \in \mathcal{D}$, pick a K_s uniformly at random in $G_0[A]$ of colour C, and call it T_C. We can now prove that with sufficiently high probability, most T_C will be present in G_1.

Lemma 5.4.11. *Let $\pi \in \Pi$. Then with probability $1 - o(\frac{1}{(\log n)\binom{n}{a}})$, the number of colours $C \in \mathcal{D}_\pi$ for which T_C has a π-dangerous pair of vertices is $o(\frac{|\mathcal{D}_\pi|}{\log n})$.*

Proof. We condition everything on the already chosen first $m/2$ colour classes. Now let $C \in \mathcal{D}_\pi$. (Recall that this means that there is a K_s in A in the graph G_0 with all its edges of colour C, and moreover that C is one of the last $m/2$ colours with respect to π.) Label the vertices of T_C by $1, 2, ..., s$. Note that any pair of vertices in A is chosen with equal probability and, by condition ($\star$), at most $o(\frac{|A|^2}{\log n})$ of them are π-dangerous. So the probability that the first two vertices of T_C form a π-dangerous pair is $o(\frac{1}{\log n})$. Hence, for any $C \in \mathcal{D}_\pi$, the probability that T_C has a pair of vertices which form a π-dangerous pair is bounded above by some $p = o(\frac{1}{\log n})$. Moreover, this holds for all such C *independently* of the others. Thus, the probability that T_C contains a π-dangerous pair for more than $\Omega(\frac{|\mathcal{D}_\pi|}{\log n})$ choices of $C \in \mathcal{D}_\pi$ is at most $\mathbb{P}\big[\operatorname{Bin}(|\mathcal{D}_\pi|, p) = \Omega(\frac{|\mathcal{D}_\pi|}{\log n})\big]$. But this is $e^{-\Omega(\frac{|\mathcal{D}_\pi|}{\log n})}$. So it remains to show that $(\log n)\binom{n}{a} = o(e^{\Omega(\frac{|\mathcal{D}_\pi|}{\log n})})$. Since $\binom{n}{a} \le n^a = e^{a \log n}$, it suffices to prove that $a \log n = o(\frac{|\mathcal{D}_\pi|}{\log n})$. But $|\mathcal{D}_\pi| = \Omega(ma^s\gamma^s)$ so it is enough to prove that $(\log n)^2 = o(ma^{s-1}\gamma^s)$. By equation (5.6), this holds provided that $(s-1)c_3 - sc_2 - c_1 > 2$. $\square$

Corollary 5.4.12. *With probability $1 - o(\frac{1}{\binom{n}{a}})$, for all but $o(|\mathcal{D}|)$ colours $C \in \mathcal{D}$, all the edges of T_C are present in G_1.*

Proof. Suppose that $C \in \mathcal{D}$ and T_C has an edge e which is not present in G_1. Then there exists some $\pi \in \Pi$ such that C is in the second half of π (so $C \in \mathcal{D}_\pi$) and e is π-dangerous. But by the previous lemma, with probability $1 - o(\frac{1}{\binom{n}{a}})$ the number of such colours C is $o(|\Pi| \cdot \frac{|\mathcal{D}|}{\log n}) = o(|\mathcal{D}|)$. $\square$

Using Lemma 5.4.9 and the union bound over all A, Lemma 5.3.1 follows.

We now return to proving Lemma 5.4.10. Recall that we want to show that almost surely for every A and every π, the density of π-dangerous pairs in A is $o(\frac{1}{\log n})$. This is essentially best possible, since if we choose A to contain one of our colour classes entirely (for a colour chosen from the first part of π), then the pairs of vertices in that colour class will all be π-dangerous. Moreover, as the typical size of a colour class is $n\gamma = n^\alpha(\log n)^{-c_2} = a(\log n)^{-c_2-c_3}$, the set of these pairs will have density roughly $(\log n)^{-2c_2-2c_3}$.

Accordingly, the next lemma is to make sure that no colour class is exceptionally large.

Lemma 5.4.13. *With probability $1 - o(1)$, the size of every colour class is at most $2n\gamma$.*

Proof. $\mathbb{P}[\text{Bin}(n,\gamma) > 2n\gamma] = e^{-\Omega(n\gamma)} = o(\frac{1}{m})$. The result follows from the union bound over all colours. $\square$

So we may assume that all colour classes have size at most $2n\gamma$.

After applying the union bound over all $\pi \in \Pi$ and A, the next result completes the proof of Lemma 5.3.1.

Lemma 5.4.14. *Fix $\pi \in \Pi$ and a set A of size a. With probability $1 - o(\frac{1}{(\log n)\binom{n}{a}})$, the number of pairs in A which are π-dangerous is at most $4\frac{a^2}{(\log n)^2}$.*

Proof. The number of π-dangerous pairs in A is at most

$$\sum_{i=1}^{m} \big(\min\{\text{Bin}(a,\gamma), 2n\gamma\}\big)^2. \tag{5.7}$$

Let $h = \frac{a}{m^{1/2}\log n}$. Note that $\log h = (\alpha - \frac{1}{2}\delta)\log n + O(\log\log n)$ and recall that $\alpha > \frac{1}{2}\delta$. Now let $p = \mathbb{P}(\text{Bin}(a,\gamma) \geq h) \leq \binom{a}{h}\gamma^h \leq (a\gamma)^h \leq e^{-\Omega(h\log n)}$.

Pick some tiny positive $\rho > 0$. Note that

$$\begin{aligned}\mathbb{P}[\text{Bin}(m,p) \geq m^{1/2+\rho}] &\leq \binom{m}{m^{1/2+\rho}} p^{m^{1/2+\rho}} \leq (mp)^{m^{1/2+\rho}} = e^{-\Omega(m^{1/2+\rho}h\log n)} \\ &= e^{-\Omega(am^{\rho})} = o\left(\frac{1}{(\log n)\binom{n}{a}}\right).\end{aligned}$$

Therefore we may assume that at most $m^{1/2+\rho}$ of the random variables $\text{Bin}(a,\gamma)$ take value more than h.

The total contribution to (5.7) of the terms with $\text{Bin}(a,\gamma) \leq h$ is at most $mh^2 = \frac{a^2}{(\log n)^2}$. The random variable $X \sim \text{Bin}(a,\gamma)$, conditional on $X \geq h$, is bounded above by $h + X'$ where X' is an independent instance of $\text{Bin}(a,\gamma)$. As we assume that all colour classes have size at most $2n\gamma$, it follows that the total contribution to (5.7) of the terms with $\text{Bin}(a,\gamma) \geq h$ is bounded above by

$$\sum_{i=1}^{m^{1/2+\rho}} \Big(h + \min\{\text{Bin}(a,\gamma), 2n\gamma\}\Big)^2 \tag{5.8}$$

and we just need to show that this sum is less than $3\frac{a^2}{(\log n)^2}$ with probability $1 - o(\frac{1}{\binom{m}{m^{1/2+\rho}}(\log n)\binom{n}{a}})$.

The sum in (5.8) is at most $m^{1/2+\rho}h^2 + (2h + 2n\gamma)\sum_{i=1}^{m^{1/2+\rho}} \text{Bin}(a,\gamma)$. The first term is at most $\frac{a^2}{(\log n)^2}$. Also, $\log(n\gamma) = \alpha\log n + O(\log\log n)$ and therefore $n\gamma \geq h$, so we just need to show that $\sum_{i=1}^{m^{1/2+\rho}} \text{Bin}(a,\gamma) \leq \frac{a^2}{2n\gamma(\log n)^2}$ with the required probability. But the left-hand side is $\text{Bin}(m^{1/2+\rho}a,\gamma)$ and $\mathbb{P}\big[\text{Bin}(m^{1/2+\rho}a,\gamma) \geq \frac{a^2}{2n\gamma(\log n)^2}\big] = e^{-\Omega(\frac{a^2}{2n\gamma(\log n)^2})}$

since $m^{1/2+\rho}a\gamma = o(\frac{a^2}{2n\gamma(\log n)^2})$. This last inequality holds because

$$\log(m^{1/2+\rho}a\gamma) = \big((1/2+\rho)\delta + \alpha + (\alpha-1)\big)\log n + O(\log\log n)$$

and

$$\log(\frac{a^2}{2n\gamma(\log n)^2}) = \big(2\alpha - 1 - (\alpha-1)\big)\log n + O(\log\log n),$$

and $(1/2+\rho)\delta + \alpha < 1$ for ρ sufficiently small (since $\delta < 1$ and $\alpha < 1/2$).

Finally, $\binom{m}{m^{1/2+\rho}}(\log n)\binom{n}{a} = e^{O(a\log n)}$ because $m^{1/2+\rho} = o(a)$ for ρ sufficiently small (as $\delta < 2\alpha$). But $a\log n = o(\frac{a^2}{n\gamma(\log n)^2})$ provided that $c_3 + c_2 > 3$, so we are done. □

5.4.3 The proof of Lemma 5.2.6

It is convenient to introduce the parameter

$$\eta = 2(1-\alpha) - \delta = \begin{cases} \eta(1) = \frac{-2s^2+4st-2s-6t+8}{(2s-3)(t-s)(t+s-1)-2s+4}, & \text{if } (s,t) \text{ is regular} \\ \eta(2) = \frac{st-s-2t+3}{(2s-3)(t-s)(s-1)+2s-t}, & \text{if } (s,t) \text{ is exceptional} \end{cases}$$

Remark. $-\eta$ is the contribution of a block of size two to the value of a scheme. By Lemma 5.2.1, we have $\eta > 2 - 4\alpha > 0$.

The next lemma follows easily from Definition 5.2.5 and is a convenient way to look at the value of a scheme.

Lemma 5.4.15. *Let Q be a scheme. Then*

$$v(Q) = t + \sum_{D\in\mathcal{D}}(\delta + |D|(\alpha-1)) - (\delta + 2\alpha)$$

where $\mathcal{D}$ is the set of colours in Q and $|D|$ is the number of nodes in Q that are coloured with D.

We shall now identify a scheme for which equality in Lemma 5.2.6 will hold: the value of α was chosen so that the value of this scheme would be 0. This is the (in)equality that generalizes equation (5.5) from the introduction. This "extremal scheme" turns out to be different in the regular and the exceptional case, which is why the formula for α also differs in the two cases.

Definition 5.4.16. Let Q_1 be the scheme where one colour gives a block of size s and the rest of the edges are given by pairwise distinct colours.

Let Q_2 be the scheme where one colour gives a block of size s, another gives a block of size $t-s+1$ sharing a single vertex with the previous block and the rest of the edges are given by pairwise distinct colours.

Lemma 5.4.17. *(a) If (s,t) is regular, then $v(Q_1) = 0$.*

(b) If (s,t) is exceptional, then $v(Q_2) = 0$.

(c) If (s,t) is regular, then $v(Q_2) \le 0$.

(d) If (s,t) is exceptional, then $v(Q_1) \le 0$.

Proof. We have

$$v(Q_1) = t + (\delta + s(\alpha - 1)) + (\binom{t}{2} - \binom{s}{2})(\delta + 2(\alpha - 1)) - (\delta + 2\alpha),$$

and (a) follows by direct substitution.

We also have

$$\begin{aligned} v(Q_2) = {} & t + (\delta + s(\alpha - 1)) + (\delta + (t - s + 1)(\alpha - 1)) \\ & + \left(\binom{t}{2} - \binom{s}{2} - \binom{t - s + 1}{2}\right)(\delta + 2(\alpha - 1)) - (\delta + 2\alpha), \end{aligned}$$

and (b) follows by direct substitution.

The difference between Q_1 and Q_2 is that the former contains $\binom{t-s+1}{2}$ edges of distinct colours where the latter contains a block of size $t - s + 1$. Using Lemmas 5.A.1 and 5.A.2 (a) from the appendix, we obtain statements (c) and (d). □

Definition 5.4.18. We call a block in a scheme *large* if it has size at least 3 and *small* otherwise. We call it an *s-block* if it has size s.

We shall begin by proving Lemma 5.2.6 in the special case when there is an s-block in the scheme.

Lemma 5.4.19. *If Q is a scheme and it has an s-block then $v(Q) \le 0$.*

Proof. Assume that Q is such that $v(Q)$ is maximal. It is enough to show that $Q = Q_1$ or $Q = Q_2$. Since Q has an s-block, any other block must have size at most $t - s + 1$. By Lemmas 5.A.1 and 5.A.2 (c) from the appendix, any large block of size smaller than $t - s$ gives a smaller contribution to the value than one obtains if the corresponding edges have pairwise distinct colours. Therefore, we may assume that Q has no such block. So every block in Q, other than the one of size s, has size $2, t - s$ or $t - s + 1$. If there is a block of size $t - s + 1$, then $Q = Q_2$. If there are no large blocks, then $Q = Q_1$. Otherwise, there is a block of size $t - s \ge 3$.

If there are no other large blocks, then we claim that $v(Q) \le v(Q_1)$ or $v(Q) \le v(Q_2)$. Indeed, the $(t-s)$-block can be modified to become a $(t-s+1)$-block (and Q then becomes Q_2) and this increases the value provided that $(\alpha - 1) \ge (t - s)(-\eta)$, or equivalently $(t-s)\eta \ge (1-\alpha)$. So we may assume that $(t-s)\eta < (1-\alpha)$. But $\delta = s-(2s-1)\alpha > 1-\alpha$, since $\alpha < 1/2$. Hence, $(t-s)\eta < \delta$, but then $v(Q) \leq v(Q_1)$ by Lemma 5.A.1 from the appendix.

We may therefore assume that there are at least two large blocks other than the one of size s, and that both have size $t-s$. This forces $t-s$ to equal 3. Moreover, by Lemmas 5.A.1 and 5.A.2 (b), we have that $t = 2s-1$. It follows that $s = 4$ and $t = 7$. So Q consists of a 4-block and several 3-blocks (there can be at most 3) and the rest of the edges are given by distinct colours. It is easy to check that in this case $v(Q) \leq 0$. □

Using the previous result, to prove Lemma 5.2.6, it is sufficient to prove the following statement.

Lemma 5.4.20. *Suppose that Q is a scheme with $v(Q)$ as large as possible. Assume also that Q does not contain a block of size s. Then $v(Q) \leq 0$.*

To prove Lemma 5.4.20, we shall introduce the following definition.

Definition 5.4.21. Let P be a node in a scheme. The *local value* at P, which we denote by $v(P)$, is defined by the formula

$$v(P) = 1 + \sum_{D:P\in D} (\delta/|D| + (\alpha - 1)),$$

where the summation is over all blocks containing P.

Example. If P is in a block of size 2 and two blocks of size 4, then

$$v(P) = 1 + 3(\alpha - 1) + \delta/2 + 2 \cdot \delta/4.$$

Lemma 5.4.22. *For any scheme Q, we have*

$$\sum_P v(P) = v(Q) + (\delta + 2\alpha)$$

where the summation is over all nodes of Q.

Proof. This statement follows easily from Lemma 5.4.15. □

The next result is the key part in the proof of Lemma 5.2.6.

Lemma 5.4.23. *Suppose that Q is a scheme such that $v(Q)$ is maximal. Let P be a node and assume that every block containing P has size less than $t/2$. Then $v(P) < 2\delta/t$.*

Proof. Let the blocks of Q that contain P have sizes $r_1, ..., r_u$. Then $\sum_i r_i = t + u - 1$. Let k be the minimal integer greater than 2 that is equal to some r_i (or, if no such integer exists, then let k be large enough that $\delta/k - \delta/(k+1) < \eta/2$). Let $R = \lfloor \frac{t-1}{2} \rfloor$. By assumption, $r_i \leq R$ for all i. Moreover, by the maximality of $v(Q)$ and Lemma 5.A.1, we have the inequality $k\eta \geq \delta$ and therefore $\delta/k - \delta/(k+1) = \frac{\delta}{k(k+1)} \leq \frac{\eta}{k+1} < \eta/2$.

Claim 1. There exist positive integers w and $q_1, ..., q_w$ such that

(i) $2 \leq q_j \leq R$ for all j

(ii) $\sum_j q_j = t + w - 1$

(iii) There is at most one j for which $2 < q_j < k$ and if there is any i with $q_i = 2$, then there is no j with $2 < q_j < k$.

(iv) $v(P) \leq 1 + \sum_j (\delta/q_j + (\alpha - 1))$

(v) Either all but at most one q_j are equal to R or else $q_j \in \{2, R\}$ for all j

Proof of Claim 1. Note that $v(P) = 1 + \sum_i (\delta/r_i + (\alpha - 1))$. Define $w, q_1, q_2, ...q_w$ to be the integers that maximize the quantity $1 + \sum_j (\delta/q_j + (\alpha - 1))$ subject to the conditions (i),(ii) and (iii). Since the r_i satisfy (i),(ii),(iii), we get $v(P) \leq 1 + \sum_j (\delta/q_j + (\alpha - 1))$. We are left to prove (v), so let us suppose that it does not hold. There are two cases to consider.

Case 1. If there exists some i with $q_i = 2$, then there is a j such that $q_j \notin \{2, R\}$ and by (iii) we have $q_j \geq k$. Hence, $\delta/q_j - \delta/(q_j + 1) < \eta/2$. After relabelling, we may assume that $j = w - 1, i = w$. Now set $w' = w - 1$, $q'_h = q_h$ for all $h \leq w - 2$ and $q'_{w-1} = q_{w-1} + 1$. Then $q'_1, ..., q'_{w'}$ satisfy (i),(ii),(iii) and

$$1 + \sum_{h \leq w} (\delta/q_h + (\alpha - 1)) < 1 + \sum_{h \leq w'} (\delta/q'_h + (\alpha - 1)),$$

which is a contradiction.

Case 2. If there is no i with $q_i = 2$, then since (v) is assumed to fail, there must exist $i \neq j$ with $2 < q_i \leq q_j < R$. Moreover, we may assume that q_i is minimal among all q_hs. Without loss of generality, $i = w - 1, j = w$. Now define $q'_h = q_h$ for all $h \leq w - 2$, $q'_{w-1} = q_{w-1} - 1$ and $q'_w = q_w + 1$. Then $q'_1, ..., q'_w$ satisfy (i),(ii),(iii) and

$$1 + \sum_{h \leq w} (\delta/q_h + (\alpha - 1)) < 1 + \sum_{h \leq w} (\delta/q'_h + (\alpha - 1)),$$

which is a contradiction.

This completes the proof of Claim 1.

Claim 2. If $q_1, ..., q_w$ satisfy the conditions (i),(ii),(v) in Claim 1, then

$$1 + \sum_{h \leq w} (\delta/q_h + (\alpha - 1)) < 2\delta/t.$$

Proof of Claim 2. For $t \leq 13$, this is a straightforward check, which we performed using a computer program, since it would have taken inordinately long to do it by hand. (The code, written in Matlab, can be found at the end of the appendix.) So we shall assume that $t \geq 14$. Then $3R \geq 3 \cdot \frac{t-2}{2} > t + 2$, so there are at most two q_js with $q_j = R$. Using (v), this leaves the following cases.

Case 1: $q_j = 2$ for all j
Case 2: $q_1 = R$ and $q_j = 2$ for all $j \geq 2$
Case 3a: $q_1 = q_2 = R = \frac{t-2}{2}$ and $q_3 = q_4 = q_5 = 2$ ($w = 5$)
Case 3b: $q_1 = q_2 = R = \frac{t-1}{2}$ and $q_3 = q_4 = 2$ ($w = 4$)
Case 4a: $q_1 = q_2 = R = \frac{t-2}{2}, q_3 = 4$ ($w = 3$)
Case 4b: $q_1 = q_2 = R = \frac{t-1}{2}, q_3 = 3$ ($w = 3$)

By Lemmas 5.A.1 and 5.A.2 (d) we have

$$(l-1)(\delta/2 + (\alpha - 1)) < (\delta/l + (\alpha - 1))$$

when $l = \frac{t-1}{2}$. Moreover, we have the inequality

$$\left(\delta/\left(\frac{t-2}{2}\right) + (\alpha - 1)\right) + \frac{1}{2}\left(\delta/2 + (\alpha - 1)\right) < \left(\delta/\left(\frac{t-1}{2}\right) + (\alpha - 1)\right),$$

since this is equivalent to $\frac{2\delta}{(t-1)(t-2)} < \eta/4$, which holds because $(t-1)\eta \geq 2\delta$ and $t-2 > 4$. It is not hard to see that these two observations allow us to deduce all Cases 1-3 from Case 3b. To prove Case 3b, we need the inequality

$$1 + 2(\delta/(\frac{t-1}{2}) + (\alpha - 1)) - \eta < 2\delta/t,$$

which is given in Lemma 5.A.2 (f).

Clearly, Case 4a follows from Case 4b. To prove Case 4b, we need

$$1 + 2(\delta/(\frac{t-1}{2}) + (\alpha - 1)) + (\delta/3 + (\alpha - 1)) < 2\delta/t.$$

Using $\alpha < 1/2$ and $\delta < 1$, it suffices to prove that $4/(t-1) - 2/t \leq 1/6$, which holds for $t \geq 14$.

This completes the proof of Claim 2, and the two claims imply the lemma. □

Lemma 5.4.24. *Suppose that Q is a scheme such that its $v(Q)$ is as large as possible and such that the largest block D of Q has size at least $t/2$. Then D has size s.*

Proof. Suppose not. Pick a node P with $P \notin D$. Let D have size $k \geq t/2$. Suppose that P is contained in exactly r large blocks. Define a scheme Q' as follows. Q' has the same blocks as Q except that

- P is removed from all large blocks,
- all small blocks containing P and a node in D are deleted,
- P is added to D,
- the missing edges are now provided by distinct colours.

We now compare the values $v(Q)$ and $v(Q')$. The node P is in only one large block in Q' while it is in r large blocks in Q. The number of small blocks containing P is precisely $t-k-1$ in Q' while it is at least $k-r$ in Q. That is because any large block containing P contains at most one element of D. So

$$\begin{aligned} v(Q') - v(Q) &\geq (r-1)(1-\alpha) + ((t-k-1)-(k-r))(\delta + 2(\alpha - 1)) \\ &= (r-1)(1-\alpha) - (t-2k+(r-1))\eta \geq (r-1)(1-\alpha-\eta). \end{aligned}$$

But $1-\alpha-\eta = \delta - (1-\alpha) = 1/2 + (2s-1)\epsilon - (1/2+\epsilon) > 0$. This contradicts the maximality of $v(Q)$ if $r \geq 2$.

If $r = 1$, then let the unique large block containing P have size l. By assumption, $l \leq k$. Hence,

$$v(Q') - v(Q) = ((t-k-1)-(t-l))(\delta + 2(\alpha-1)) = (k+1-l)\eta > 0,$$

a contradiction.

If $r = 0$, then

$$v(Q') - v(Q) = -(1-\alpha) - k(\delta + 2(\alpha - 1)) = k\eta - (1-\alpha) \geq \frac{t}{2}\eta - (1-\alpha).$$

But by Lemma 5.A.2 (d), this is at least $\delta - (1-\alpha) > 0$. This is a contradiction and the lemma is proved. $\square$

We are ready to complete the proof of Lemma 5.2.6.

Proof of Lemma 5.2.6. We may assume that $v(Q)$ is maximal possible among all schemes Q. If Q has a block of size s, then we are done by Lemma 5.4.19. Otherwise, by Lemma 5.4.24, there is no block of size greater than or equal to $t/2$. But then Lemma 5.4.22 and Lemma 5.4.23 together imply that $v(Q) \leq t\frac{2\delta}{t} - (\delta + 2\alpha) = \delta - 2\alpha < 0$. $\square$

5.A Appendix

Lemma 5.A.1. *For any $k > 2$, we have*

$$\binom{k}{2}(\delta + 2(\alpha - 1)) > \delta + k(\alpha - 1) \iff k\eta < \delta$$

Proof.

$$\begin{aligned}\binom{k}{2}(\delta+2(\alpha-1)) &> \delta+k(\alpha-1)\\ &\iff (k-1)(\delta+2(\alpha-1)) > 2\delta/k+2(\alpha-1)\\ &\iff (k-1)\eta < 2(1-\alpha)-2\delta/k = \eta+\delta(1-2/k)\\ &\iff (k-2)\eta < \delta(k-2)/k\\ &\iff k\eta < \delta.\end{aligned}$$

□

Lemma 5.A.2. *(a)* $(t-s+1)\eta < \delta$ *if and only if* (s,t) *is regular.*

(b) $(t-s)\eta < \delta$ *unless* $t = 2s-1$

(c) $(t-s-1)\eta < \delta$

(d) $(t-1)\eta > 2\delta$.

(e) $\delta > 2/3$.

(f) $1+2(\delta/\binom{t-1}{2}+(\alpha-1))-\eta < 2\delta/t$

Proof. Assume first that (s,t) is regular. Then after some tedious calculations, one finds that (a) is equivalent to the inequality

$$(s-2)(t-s-2)(2s-t-3)-t-3s+8>0$$

The left hand side is a quadratic in t with negative leading coefficient so it is enough to check that the inequality holds when $t = s+3$ and when $t = 2s-4$.

For $t = s+3$ we require $(s-2)(s-6)-4s+5>0$, which holds for $s \geq 11$, and for $t = 2s-4$ we require $(s-2)(s-6)-5s+12>0$, which holds for $s \geq 11$. It therefore suffices to check the inequality for the pairs $(s,t) = (10,14)$ and $(s,t) = (10,15)$. This can be done by direct substitution. So (a) is proved (when (s,t) is regular) which immediately implies (b) and (c).

Now let us assume that (s,t) is exceptional. Then the inequality $(t-s+c)\eta < \delta$ is equivalent to the inequality

$$(s-2)(t-s-c-1)(2s-t-2c-1)+(-2c^2-2c+1)s-t+4c^2+3c+1>0 \quad (5.9)$$

When $c = 1$, this says that $(s-2)(t-s-2)(2s-t-3)-3s-t+8>0$, so in order to prove (a) we need to show that this does not hold. For $t \in \{s+2, 2s-3, 2s-2, 2s-1\}$ that is clear, since $(s-2)(t-s-2)(2s-t-3) \leq 0$. We are left to check that the inequality fails for the pairs $(7,10), (8,11), (8,12), (9,12), (9,13), (9,14), (10,13)$, and $(10,16)$. If $t = s+3$, then we need $s^2-12s+17 \leq 0$ which indeed holds for $7 \leq s \leq 10$. If $t = 2s-4$, then we need $s^2 - 13s + 24 \leq 0$ which indeed holds for $7 \leq s \leq 10$. We have only $(s,t) = (9,13)$ left to check. That is done by direct substitution.

When $c = 0$, then (5.9) says that $(s-2)(t-s-1)(2s-t-1) + s - t + 1 > 0$. But if $s + 2 \leq t \leq 2s - 2$, then the left hand side is minimal at $t = 2s - 2$ and there it takes value $(s-3)^2 > 0$. (Note that $s > 3$ in this case.) This proves (b).

When $c = -1$ in (5.9), then it says that $(s-2)(t-s)(2s-t+1) + s - t + 2 > 0$. But the left hand side is minimal when $t = 2s - 1$, and then it is $2s^2 - 7s + 7 > 0$. This proves (c).

(d) In the regular case the statement is equivalent to the inequality

$$2s^3 - s^2t - 5s^2 + 3st - t^2 + s + 3t - 4 > 0$$

But in the regular case we have $2s - t \geq 4$, so $2s^3 - s^2t \geq 4s^2$. Since $-s^2 + 3st - t^2 \geq 0$ and $s + 3t - 4 > 0$, the statement follows.

In the exceptional case, (d) is equivalent to the inequality

$$(s-2)(2s-t)^2 + (t-s-1) > 0$$

which is clear.

(e) In the regular case, the statement is equivalent to the inequality

$$2s(t+s-5)(t-s-2) + 2s^2 - 10s + 6t - 8 > 0,$$

which is easily seen to hold.

In the exceptional case, it is equivalent to the inequality

$$(2s^2 - 5s)(t-s-2) + s^2 - 5s + 2t - 3 > 0,$$

which again clearly holds.

(f) Since (by (d)) we have $\delta/\binom{t-1}{2} < \eta$, this inequality reduces to

$$1 + 2(\alpha - 1) + 2\delta/(t-1) < 2\delta/t,$$

or, equivalently, to

$$2\alpha < 1 - \frac{2\delta}{t(t-1)}.$$

Expressing α in terms of δ and performing some routine algebraic manipulations, we find that we need to prove that

$$\frac{2(2s-1)}{t(t-1)}\delta < 2\delta - 1.$$

Since $\delta < 1$, the left hand side of this inequality is less than $4/t < 1/3$ while the right hand side is greater than 1/3, by part (e), so the proof is complete. □

Below we present the Matlab code that we used to perform the case check in the proof of Lemma 5.4.23.

```
% go through all pairs (s,t)
for t=5:13
for s=(floor(t/2)+1):(t-2)
% these pairs are all exceptional
alpha=((s-2)*(t-s)*(s-1)+s-1)/((2*s-3)*(t-s)*(s-1)+2*s-t);
delta=s-(2*s-1)*alpha;
eta=2*(1-alpha)-delta;
% bad will be changed to 1 if the inequality that we want
% to prove fails
bad=0;
R=floor((t-1)/2);
% j will count the number of q_h which are equal to R
for j=0:4
a=(t-1)-j*(R-1);
if 0<=a
% in the following case every q_h is 2 or R
v=1+a*(delta/2+alpha-1)+j*(delta/R+alpha-1);
% check that our inequality holds with a suitably large
% difference which can't be due to rounding errors
if v>2*delta/(t)-10^(-3)
bad=1;
end
end
if (2<=a+1) && (a+1<=R)
% in the following case there is only one q_h that
% is not equal to R
v=1+(delta/(a+1)+alpha-1)+j*(delta/R+alpha-1);
if v>2*delta/(t)-10^(-3)
bad=1;
end
end
end
% tabulate the result: for each pair (s,t) we print
% whether the inequality failed (1) or not (0)
fprintf('%5d %5d %5d \n',s,t,bad)
end
end
```

Chapter 6

Polynomial bound for the partition rank vs the analytic rank of tensors

6.1 Introduction

6.1.1 Bias and rank of polynomials

For a finite field $\mathbb{F}$ and a polynomial $P : \mathbb{F}^n \to \mathbb{F}$, we say that P is unbiased if the distribution of the values $P(x)$ is close to the uniform distribution on $\mathbb{F}$; otherwise we say that P is biased. It is an important direction of research in higher order Fourier analysis to understand the structure of biased polynomials.

Note that a generic degree d polynomial should be unbiased. In fact, as we will see below, if a degree d polynomial is biased, then it can be written as a function of not too many polynomials of degree at most $d-1$. Let us now make this discussion more precise.

Definition 6.1.1. Let $\mathbb{F}$ be a finite field and let χ be a nontrivial character of $\mathbb{F}$. The *bias* of a function $f : \mathbb{F}^n \to \mathbb{F}$ with respect to χ is defined to be $\text{bias}_\chi(f) = \mathbb{E}_{x \in \mathbb{F}^n}[\chi(f(x))]$. (Here and elsewhere in the chapter $\mathbb{E}_{x \in G} h(x)$ denotes $\frac{1}{|G|}\sum_{x \in G} h(x)$.)

Remark. Most of the previous work is on the case $\mathbb{F} = \mathbb{F}_p$ with p a prime, in which case the standard definition of bias is $\text{bias}(f) = \mathbb{E}_{x \in \mathbb{F}^n}\omega^{f(x)}$ where $\omega = e^{\frac{2\pi i}{p}}$.

Definition 6.1.2. Let P be a polynomial $\mathbb{F}^n \to \mathbb{F}$ of degree d. The *rank* of P (denoted $\text{rank}(P)$) is defined to be the smallest integer r such that there exist polynomials $Q_1, \dots, Q_r : \mathbb{F}^n \to \mathbb{F}$ of degree at most $d-1$ and a function $f : \mathbb{F}^r \to \mathbb{F}$ such that $P = f(Q_1, \dots, Q_r)$.

As discussed above, it is known that if a polynomial has large bias, then it has low rank. The first result in this direction was proved by Green and Tao [56] who showed that if $\mathbb{F}$ is a field of prime order and $P : \mathbb{F}^n \to \mathbb{F}$ is a polynomial of degree d with $d < |\mathbb{F}|$ and $\text{bias}(P) \geq \delta > 0$, then $\text{rank}(P) \leq c(\mathbb{F}, \delta, d)$. Kaufman and Lovett [80] proved that the condition $d < |\mathbb{F}|$ can be omitted. In both results, c has Ackermann-type dependence on its parameters. Finally, Bhowmick and Lovett [9] proved that if $d < \mathrm{char}(\mathbb{F})$ and $\mathrm{bias}(P) \geq |\mathbb{F}|^{-s}$, then $\mathrm{rank}(P) \leq c'(d, s)$. The novelty of this result is that c' does not depend on $\mathbb{F}$. However, it still has Ackermann-type dependence on d and s.

One of our main results is the following theorem, which improves the result of Bhowmick and Lovett, unless $|\mathbb{F}|$ is very large.

Theorem 6.1.3. *Let $\mathbb{F}$ be a finite field and let χ be a nontrivial character of $\mathbb{F}$. Let P be a polynomial $\mathbb{F}^n \to \mathbb{F}$ of degree $d <$ char$(\mathbb{F})$. Suppose that* $\mathrm{bias}_\chi(P) \geq \varepsilon > 0$ *where $\varepsilon \leq 1/|\mathbb{F}|$. Then*

$$\mathrm{rank}(P) \leq (c \cdot 2^d \cdot \log(1/\varepsilon))^{c'(d)} + 1$$

where c is an absolute constant and $c'(d) = 4^{d^d}$.

Recall that if G is an Abelian group and d is a positive integer, then the Gowers U^d norm (which is only a seminorm for $d = 1$) of $f : G \to \mathbb{C}$ is defined to be

$$\|f\|_{U^d} = \big|\mathbb{E}_{x,y_1,\ldots,y_d \in G} \prod_{S \subset [d]} \mathcal{C}^{d-|S|} f(x + \sum_{i \in S} y_i)\big|^{1/2^d},$$

where $\mathcal{C}$ is the conjugation operator. It is a major area of research to understand the structure of functions f whose U^d norm is large. Our next theorem is a result in this direction.

Theorem 6.1.4. *Let $\mathbb{F}$ be a finite field and let χ be a nontrivial character of $\mathbb{F}$. Let P be a polynomial $\mathbb{F}^n \to \mathbb{F}$ of degree $d <$ char$(\mathbb{F})$. Let $f(x) = \chi(P(x))$ and assume that $\|f\|_{U^d} \geq \varepsilon > 0$ where $\varepsilon \leq 1/|\mathbb{F}|$. Then*

$$\mathrm{rank}(P) \leq (c \cdot 2^d \cdot \log(1/\varepsilon))^{c'(d)} + 1$$

where c is an absolute constant and $c'(d) = 4^{d^d}$.

Our result implies a similar improvement to the bounds for the quantitative inverse theorem for Gowers norms for polynomial phase functions of degree d.

Theorem 6.1.5. *Let $\mathbb{F}$ be a field of prime order and let P be a polynomial $\mathbb{F}^n \to \mathbb{F}$ of degree $d <$ char$(\mathbb{F})$. Let $f(x) = \omega^{P(x)}$ where $\omega = e^{\frac{2\pi i}{|\mathbb{F}|}}$ and assume that $\|f\|_{U^d} \geq \varepsilon > 0$ where $\varepsilon \leq 1/|\mathbb{F}|$. Then there exists a polynomial $Q : \mathbb{F}^n \to \mathbb{F}$ of degree at most $d - 1$ such that*

$$|\mathbb{E}_{x \in \mathbb{F}^n} \omega^{P(x)} \overline{\omega^{Q(x)}}| \geq |\mathbb{F}|^{-(c \cdot 2^d \cdot \log(1/\varepsilon))^{c'(d)} - 1}$$

where c is an absolute constant and $c'(d) = 4^{d^d}$.

Theorems 6.1.3 and 6.1.5 easily follow from Theorem 6.1.4.

Proof of Theorem 6.1.3. Note that when $f(x) = \chi(P(x))$, then $\|f\|_{U^1}^2 = |\mathbb{E}_{x,y \in \mathbb{F}^n} \overline{f(x)} f(x + y)| = |\mathbb{E}_{x \in \mathbb{F}^n} f(x)|^2$, so $\|f\|_{U^1} = |E_{x \in \mathbb{F}^n} f(x)| = |\mathrm{bias}_\chi(P)|$. However,

$\|f\|_{U^k}$ is increasing in k (see eg. Claim 6.2.2 in [61]), therefore $\|f\|_{U^d} \geq |\text{bias}_\chi(P)| \geq \varepsilon$. The result is now immediate from Theorem 6.1.4.

Proof of Theorem 6.1.5. By Theorem 6.1.4, there exists a set of $r \leq (c \cdot 2^d \cdot \log(1/\varepsilon))^{c'(d)} + 1$ polynomials $Q_1, \dots, Q_r$ such that $P(x)$ is a function of $Q_1(x), \dots, Q_r(x)$. Then $\omega^{P(x)} = g(Q_1(x), \dots, Q_r(x))$ for some function $g : \mathbb{F}^r \to \mathbb{C}$. Let $G = \mathbb{F}^r$. Note that $|g(y)| = 1$ for all $y \in G$, therefore $|\hat{g}(\chi)| \leq 1$ for every character $\chi \in \hat{G}$. Now $\omega^{P(x)} = \sum_{\chi \in \hat{G}} \hat{g}(\chi)\, \chi((Q_1(x), \dots, Q_r(x))$, so

$$1 = \mathbb{E}_{x \in \mathbb{F}^n} |\omega^{P(x)}|^2 = \sum_{\chi \in \hat{G}} \overline{\hat{g}(\chi)} \left(\mathbb{E}_{x \in \mathbb{F}^n} \omega^{P(x)} \overline{\chi(Q_1(x), \dots, Q_r(x))} \right).$$

Thus, there exists some $\chi \in \hat{G}$ with $|\mathbb{E}_{x \in \mathbb{F}^n} \omega^{P(x)} \overline{\chi(Q_1(x), \dots, Q_r(x))}| \geq 1/|G| = 1/|\mathbb{F}|^r$. But χ is of the form $\chi(y_1, \dots, y_r) = \omega^{\sum_{i \leq r} \alpha_i y_i}$ for some $\alpha_i \in \mathbb{F}$. Then $\chi(Q_1(x), \dots, Q_r(x)) = \omega^{Q_\alpha(x)}$, where Q_α is the degree $d-1$ polynomial $Q_\alpha(x) = \sum_{i \leq r} \alpha_i Q_i(x)$. So $Q = Q_\alpha$ is a suitable choice.

6.1.2 Analytic rank and partition rank of tensors

Related to the bias and rank of polynomials are the notions of *analytic rank* and *partition rank* of tensors. Recall that if $\mathbb{F}$ is a field and $V_1, \dots, V_d$ are finite dimensional vector spaces over $\mathbb{F}$, then an order d tensor is a multilinear map $T : V_1 \times \dots \times V_d \to \mathbb{F}$. (In this subsection, assume that $d \geq 2$.) Each V_k can be identified with $\mathbb{F}^{n_k}$ for some n_k, and then there exist $t_{i_1, \dots, i_d} \in \mathbb{F}$ for all $i_1 \leq n_1, \dots, i_d \leq n_d$ such that $T(v^1, \dots, v^d) = \sum_{i_1 \leq n_1, \dots, i_d \leq n_d} t_{i_1, \dots, i_d} v^1_{i_1} \dots v^d_{i_d}$ for every $v^1 \in \mathbb{F}^{n_1}, \dots, v^d \in \mathbb{F}^{n_d}$ (where v_k is the kth coordinate of the vector v). Indeed, $t_{i_1, \dots, i_d}$ is just $T(e^{i_1}, \dots, e^{i_d})$, where e^i is the ith standard basis vector.

The following notion was introduced by Gowers and Wolf [54].

Definition 6.1.6. Let $\mathbb{F}$ be a finite field, let $V_1, \dots, V_d$ be finite dimensional vector spaces over $\mathbb{F}$ and let $T : V_1 \times \dots \times V_d \to \mathbb{F}$ be an order d tensor. Then the *analytic rank* of T is defined to be $\text{arank}(T) = -\log_{|\mathbb{F}|} \text{bias}(T)$, where $\text{bias}(T) = \mathbb{E}_{v^1 \in V_1, \dots, v^d \in V_d} [\chi(T(v^1, \dots, v^d))]$ for any nontrivial character χ of $\mathbb{F}$.

Remark. This is well-defined. Indeed, if χ is a nontrivial character of $\mathbb{F}$, then

$$\begin{aligned} \mathbb{E}_{v^1 \in V_1, \dots, v^d \in V_d} [\chi(T(v^1, \dots, v^d))] &= \mathbb{E}_{v^1 \in V_1, \dots, v^{d-1} \in V_{d-1}} [\mathbb{E}_{v^d \in V_d} \chi(T(v^1, \dots, v^d))] \\ &= \mathbb{P}_{v^1 \in V_1, \dots, v^{d-1} \in V_{d-1}} [T(v^1, \dots, v^{d-1}, x) \equiv 0], \end{aligned}$$

where $T(v^1, \dots, v^{d-1}, x)$ is viewed as a function in x. The second equality holds because $\mathbb{E}_{v^d \in V_d} \chi(T(v^1, \dots, v^d)) = 0$ unless $T(v^1, \dots, v^{d-1}, x) \equiv 0$, in which case it is 1.

Thus, $\mathbb{E}_{v^1 \in V_1, \dots, v^d \in V_d} [\chi(T(v^1, \dots, v^d))]$ does not depend on χ, and is always positive. Moreover, it is at most 1, therefore the analytic rank is always nonnegative.

A different notion of rank was defined by Naslund [100].

Definition 6.1.7. Let $T : V_1 \times \cdots \times V_d \to \mathbb{F}$ be a (non-zero) order d tensor. We say that T has *partition rank* 1 if there is some $S \subset [d]$ with $S \neq \emptyset, S \neq [d]$ such that $T(v^1, \ldots, v^d) = T_1(v^i : i \in S)T_2(v^i : i \notin S)$ where $T_1 : \prod_{i \in S} V_i \to \mathbb{F}, T_2 : \prod_{i \notin S} V_i \to \mathbb{F}$ are tensors. In general, the partition rank of T is the smallest r such that T can be written as the sum of r tensors of partition rank 1. This number is denoted $\mathrm{prank}(T)$.

Kazhdan and Ziegler [81] and Lovett [93] proved that $\mathrm{arank}(T) \leq \mathrm{prank}(T)$. In the other direction, it follows from the work of Bhowmick and Lovett [9] that if an order d tensor T has $\mathrm{arank}(T) \leq r$, then $\mathrm{prank}(T) \leq f(r, d)$ for some function f. Note that f does not depend on $|\mathbb{F}|$ or the dimension of the vector spaces V_k. However, f has an Ackermann-type dependence on d and r. For $d = 3, 4$, better bounds were established by Haramaty and Shpilka [60]. They proved that for $d = 3$ we have $\mathrm{prank}(T) = O(r^4)$, and that for $d = 4$ we have $\mathrm{prank}(T) = \exp(O(r))$.

Our main result is a polynomial upper bound, which holds for general d.

Theorem 6.1.8. *Let $T : V_1 \times \cdots \times V_d \to \mathbb{F}$ be an order d tensor with* $\mathrm{arank}(T) \leq r$ *and assume that $r \geq 1$. Then*

$$\mathrm{prank}(T) \leq (c \cdot \log |\mathbb{F}|)^{c'(d)} \cdot r^{c'(d)}$$

for some absolute constant c, and $c'(d) = 4^{d^d}$.

We remark that a very similar result was obtained independently and simultaneously by Milićević [96]. Moreover, in the special case $d = 4$, a similar bound was proved independently by Lampert [92].

It is not hard to see that Theorem 6.1.8 implies Theorem 6.1.4. Indeed, let P be a polynomial $\mathbb{F}^n \to \mathbb{F}$ of degree $d < \mathrm{char}(\mathbb{F})$, let $f(x) = \chi(P(x))$ and assume that $\|f\|_{U^d} \geq \varepsilon > 0$, where $\varepsilon \leq 1/|\mathbb{F}|$. Define $T : (\mathbb{F}^n)^d \to \mathbb{F}$ by $T(y_1, \ldots, y_d) = \sum_{S \subset [d]} (-1)^{d-|S|} P(\sum_{i \in S} y_i)$. By Lemma 2.4 from [54], T is a tensor of order d. Moreover, by the same lemma, we have $T(y_1, \ldots, y_d) = \sum_{S \subset [d]} (-1)^{d-|S|} P(x + \sum_{i \in S} y_i)$ for any $x \in \mathbb{F}^n$. Thus,

$$\mathrm{bias}(T) = \mathbb{E}_{y_1,\ldots,y_d \in \mathbb{F}^n} \chi(T(y_1, \ldots, y_d)) = \mathbb{E}_{y_1,\ldots,y_d \in \mathbb{F}^n} \prod_{S \subset [d]} \mathcal{C}^{d-|S|} f(x + \sum_{i \in S} y_i)$$

for any $x \in \mathbb{F}^n$. By averaging over all $x \in \mathbb{F}^n$, it follows that $\mathrm{bias}(T) = \|f\|_{U^d}^{2^d} \geq \varepsilon^{2^d}$. Thus, $\mathrm{arank}(T) \leq 2^d \log_{|\mathbb{F}|}(1/\varepsilon)$. Note that $2^d \log_{|\mathbb{F}|}(1/\varepsilon) \geq 1$. Therefore, by Theorem 6.1.8 with $r = 2^d \log_{|\mathbb{F}|}(1/\varepsilon)$, we get

$$\mathrm{prank}(T) \leq (c \cdot 2^d \cdot \log(1/\varepsilon))^{c'(d)}. \tag{6.1}$$

We claim that $d!P(x) - T(x, \ldots, x)$ is a polynomial of degree at most $d - 1$. Clearly it is a polynomial of degree at most d, so it suffices to check that the coefficient of x^d is the same in $d!P(x)$ and in $T(x,\dots,x)$. Note that $T(x,\dots,x)=\sum_{S\subset[d]}(-1)^{d-|S|}P(|S|x)$, so if the coefficient of x^d in $P(x)$ is c, then in $T(x,\dots,x)$ it is $c\sum_{i=0}^{d}\binom{d}{i}(-1)^{d-i}i^d$. By the inclusion-exclusion principle, the sum $\sum_{i=0}^{d}\binom{d}{i}(-1)^{d-i}i^d$ is equal to the number of surjective functions $[d]\to[d]$. Hence, the coefficient of x^d in $T(x,\dots,x)$ is $c\cdot d!$.

Thus, $d!P(x)-T(x,\dots,x)$ indeed has degree at most $d-1$. Since $d<\mathrm{char}(\mathbb{F})$, we can let $W(x)=P(x)-\frac{1}{d!}T(x,\dots,x)$. By equation (6.1), T can be written as a sum of at most $(c\cdot 2^d\cdot\log(1/\varepsilon))^{c'(d)}$ tensors of partition rank 1. Hence, $\frac{1}{d!}T(x,\dots,x)$ can be written as a sum of at most $(c\cdot 2^d\cdot\log(1/\varepsilon))^{c'(d)}$ expressions of the form $Q(x)R(x)$ where Q,R are polynomials of degree at most $d-1$ each. Thus, $P-W$ has rank at most $(c\cdot 2^d\cdot\log(1/\varepsilon))^{c'(d)}$, and therefore P has rank at most

$$(c\cdot 2^d\cdot\log(1/\varepsilon))^{c'(d)}+1.$$

6.2 The proof of Theorem 6.1.8

6.2.1 Notation and preliminaries

In the rest of the chapter, we identify V_i with $\mathbb{F}^{n_i}$. Thus, the set of all tensors $V_1\times\cdots\times V_d\to\mathbb{F}$ is the tensor product $\mathbb{F}^{n_1}\otimes\cdots\otimes\mathbb{F}^{n_d}$, which will be denoted by $\mathcal{G}$ throughout this section. Also, $\mathcal{B}$ will always stand for the multiset $\{u_1\otimes\cdots\otimes u_d : u_i\in\mathbb{F}^{n_i}\text{ for all }i\}$. The elements of $\mathcal{B}$ will be called *pure tensors*. Note that $\mathcal{G}=\mathbb{F}^{n_1}\otimes\cdots\otimes\mathbb{F}^{n_d}$ can be viewed as the set of d-dimensional $(n_1,\dots,n_d)$-arrays over $\mathbb{F}$ which in turn can be viewed as $\mathbb{F}^{n_1n_2\dots n_d}$, equipped with the entry-wise dot product.

For $I\subset[d]$, we write $\mathbb{F}^I$ for $\bigotimes_{i\in I}\mathbb{F}^{n_i}$ so that we naturally have $\mathcal{G}=\mathbb{F}^I\otimes\mathbb{F}^{I^c}$, where I^c always denotes $[d]\setminus I$.

If $r\in\mathbb{F}^{[d]}=\mathcal{G}$ and $s\in\mathbb{F}^{[k]}$ (for some $k\le d$), then we define rs to be the tensor in $\mathbb{F}^{[k+1,d]}$ with coordinates $(rs)_{i_{k+1},\dots,i_d}=\sum_{i_1\le n_1,\dots,i_k\le n_k}r_{i_1,\dots,i_d}s_{i_1,\dots,i_k}$. If $k=d$, then rs is the same as the entry-wise dot product $r.s$. Also, note that viewing r as a d-multilinear map $R:\mathbb{F}^{n_1}\times\cdots\times\mathbb{F}^{n_d}\to\mathbb{F}$, we have $R(v^1,\dots,v^d)=\sum_{i_1\le n_i,\dots,i_d\le n_d}r_{i_1,\dots,i_d}v^1_{i_1}\dots v^d_{i_d}=r(v^1\otimes\cdots\otimes v^d)$.

Finally, we use a non-standard notation and write kB to mean the set of elements of $\mathcal{G}$ which can be written as a sum of *at most* k elements of B, where B is some fixed (multi)subset of $\mathcal{G}$, and similarly, we write $kB-lB$ for the set of elements that can be obtained by adding at most k members and subtracting at most l members of B.

We will use the next result several times in our proofs. It is a version of Bogolyubov's lemma, due to Sanders.

Lemma 6.2.1 (Sanders [102]). *There is an absolute constant C with the following property. Let A be a subset of $\mathbb{F}^n$ with $|A|\ge\delta|\mathbb{F}^n|$. Then $2A-2A$ contains a subspace of $\mathbb{F}^n$ of codimension at most $C(\log(1/\delta))^4$.*

Throughout the chapter, C stands for the constant appearing in the previous lemma.

Clearly we may assume that $C \geq 1$. Logarithms are base 2.

6.2.2 The main lemma and some consequences

Theorem 6.1.8 will follow easily from the next lemma, which is the main technical result of this chapter. See Section 7.2 for another application of this lemma.

Lemma 6.2.2. *Let $d \geq 1$ be an integer and let $\delta \leq 1/2$. Let $f_1(d) = 2^{3^{d+3}}$, $f_2(d) = 2^{-3^{d+3}}$ and $G(d, \delta, \mathbb{F}) = ((\log |\mathbb{F}|)c_1(d)(\log 1/\delta))^{c_2(d)}$ where $c_1(d) = C \cdot 2^{3^{d+6}}$ and $c_2(d) = 4^{d^d}$. If $\mathcal{B}' \subset \mathcal{B}$ is a multiset such that $|\mathcal{B}'| \geq \delta|\mathcal{B}|$, then there exists a multiset Q whose elements are pure tensors chosen from $f_1(d)\mathcal{B}' - f_1(d)\mathcal{B}'$ (but with arbitrary multiplicity) with the following property. The set of arrays $r \in \mathcal{G}$ with $r.q = 0$ for at least $(1 - f_2(d))|Q|$ choices $q \in Q$ is contained in $\sum_{I \subset [d], I \neq \emptyset} V_I \otimes \mathbb{F}^{I^c}$ for subspaces $V_I \subset \mathbb{F}^I$ of dimension at most $G(d, \delta, \mathbb{F})$.*

Throughout the chapter, the functions G, c_1, c_2 will refer to the functions introduced in the previous lemma. In fact, as $\mathbb{F}$ is fixed, we will write $G(d, \delta)$ to mean $G(d, \delta, \mathbb{F})$.

In this subsection we deduce Theorem 6.1.8 from Lemma 6.2.2.

The notion introduced in the next definition is closely related to the partition rank, but will be somewhat more convenient to work with.

Definition 6.2.3. Let k be a positive integer. We say that $r \in \mathcal{G}$ is *k-degenerate* if for every $I \subset [d], I \neq \emptyset, I \neq [d]$, there exists a subspace $H_I \subset \mathbb{F}^I$ of dimension at most k such that $r \in \sum_{I \subset [d-1], I \neq \emptyset} H_I \otimes H_{I^c}$.

If $r \in H_I \otimes \mathbb{F}^{I^c}$ with $\dim(H_I) \leq k$, then $r \in H_I \otimes H_{I^c}$ for some $H_{I^c} \subset \mathbb{F}^{I^c}$ of dimension at most k. (This follows by writing r as $\sum_{j \leq m} s_j \otimes t_j$ with $\{s_j\}$ a basis for H_I and letting H_{I^c} be the span of all the t_j.) Thus, r is k-degenerate if and only if $r \in \sum_{I \subset [d-1], I \neq \emptyset} H_I \otimes \mathbb{F}^{I^c}$ for some $H_I \subset \mathbb{F}^I$ of dimension at most k, or equivalently, if and only if $r \in \sum_{I \subset [d-1], I \neq \emptyset} \mathbb{F}^I \otimes H_{I^c}$ for some $H_{I^c} \subset \mathbb{F}^{I^c}$ of dimension at most k. Moreover, note that if r is k-degenerate, then $\text{prank}(r) \leq 2^{d-1}k$. This is because if $I \neq \emptyset, I \subset [d-1]$ and $w \in H_I \otimes H_{I^c}$ for subspaces $H_I \subset \mathbb{F}^I$ and $H_{I^c} \subset \mathbb{F}^{I^c}$ of dimension at most k, then $w = \sum_{i \leq k} s_i \otimes t_i$ for some $s_i \in H_I$, $t_i \in H_{I^c}$. But clearly, $s_i \otimes t_i$ has partition rank 1.

Lemma 6.2.4. *Let $\delta \leq 1/2$ and $d \geq 2$. Suppose that Lemma 6.2.2 has been proved for $d' = d - 1$. Let $r \in \mathcal{G}$ be such that $r(v_1 \otimes \cdots \otimes v_{d-1}) = 0 \in \mathbb{F}^{n_d}$ for at least $\delta|\mathbb{F}|^{n_1 \cdots n_{d-1}}$ choices $v_1 \in \mathbb{F}^{n_1}, \dots, v_{d-1} \in \mathbb{F}^{n_{d-1}}$. Then r is f-degenerate for $f = G(d-1, \delta)$.*

Proof. Write $r = \sum_i s_i \otimes t_i$ where $s_i \in \mathbb{F}^{[d-1]}$ and $\{t_i\}_i$ is a basis for $\mathbb{F}^{n_d}$. Let $\mathcal{D}$ be the multiset $\{u_1 \otimes \cdots \otimes u_{d-1} : u_1 \in \mathbb{F}^{n_1}, \dots, u_{d-1} \in \mathbb{F}^{n_{d-1}}\}$ and let $\mathcal{D}' = \{w \in \mathcal{D} : rw = 0\}$. Since $|\mathcal{D}'| \geq \delta|\mathcal{D}|$, by Lemma 6.2.2 there is a multiset Q with elements from $2^{3^{d+2}}\mathcal{D}' - 2^{3^{d+2}}\mathcal{D}'$ such that the set of arrays $r' \in \mathbb{F}^{[d-1]}$ with $r'.q = 0$ for all choices $q \in Q$ is contained in some $\sum_{I \subset [d-1], I \neq \emptyset} V_I \otimes \mathbb{F}^{[d-1] \setminus I}$, where $\dim(V_I) \leq G(d-1, \delta)$. Note that for every i we have $s_i.w = 0$ for all $w \in \mathcal{D}'$ and so also $s_i.q = 0$ for all $q \in Q$. Thus, $r \in \sum_{I \subset [d-1], I \neq \emptyset} V_I \otimes \mathbb{F}^{I^c}$. $\square$

Now we are in a position to prove Theorem 6.1.8 conditional on Lemma 6.2.2.

Proof of Theorem 6.1.8. Let $T : \mathbb{F}^{n_1} \times \cdots \times \mathbb{F}^{n_d} \to \mathbb{F}$ be an order d tensor with $\text{arank}(T) \leq r$. By Remark 6.1.2, we have $\mathbb{P}_{v_1 \in \mathbb{F}^{n_1}, \ldots, v_{d-1} \in \mathbb{F}^{n_{d-1}}}[T(v_1, \ldots, v_{d-1}, x) \equiv 0] \geq |\mathbb{F}|^{-r}$. Writing t for the element in $\mathcal{G}$ corresponding to T, we get that $t(v_1 \otimes \cdots \otimes v_{d-1} \otimes x) \equiv 0$ as a function of x for at least $\delta |\mathbb{F}|^{n_1 \ldots n_d}$ choices $v_1 \in \mathbb{F}^{n_1}, \ldots, v_{d-1} \in \mathbb{F}^{n_{d-1}}$, where $\delta = |\mathbb{F}|^{-r}$. But $t(v_1 \otimes \cdots \otimes v_{d-1} \otimes x) = \big(t(v_1 \otimes \cdots \otimes v_{d-1})\big).x$, so we have $t(v_1 \otimes \cdots \otimes v_{d-1}) = 0$ for all these choices of v_i. The condition $r \geq 1$ implies $\delta \leq 1/2$, therefore by Lemma 6.2.4, t is f-degenerate for $f = G(d-1, \delta)$. Hence,

$$\begin{aligned}
\text{prank}(T) &\leq 2^{d-1} G(d-1, \delta) \\
&= 2^{d-1}((\log |\mathbb{F}|) \cdot c_1(d-1) \cdot \log(|\mathbb{F}|^r))^{c_2(d-1)} \\
&= 2^{d-1}((\log |\mathbb{F}|)^2 \cdot c_1(d-1) \cdot r)^{c_2(d-1)} \\
&\leq ((\log |\mathbb{F}|)^2 \cdot c_1(d) \cdot r)^{c_2(d-1)}
\end{aligned}$$

But there exists some absolute constant c such that $c_1(d)^{c_2(d-1)} \leq c^{c_2(d)}$ holds for all d. Moreover, $2c_2(d-1) \leq c_2(d)$. Thus, $\text{prank}(T) \leq (c \cdot \log |\mathbb{F}|)^{c_2(d)} \cdot r^{c_2(d)} = (c \cdot \log |\mathbb{F}|)^{c'(d)} \cdot r^{c'(d)}$. $\square$

6.2.3 The overview of the proof of Lemma 6.2.2

The proof of the lemma goes by induction on d. In what follows, we shall prove results conditional on the assumption that Lemma 6.2.2 has been verified for all $d' < d$. Eventually, we will use these results to prove the induction step.

In this subsection, we give a detailed sketch of the proof in the $d = 3$ case. At the end of the subsection, we also briefly sketch the $d > 3$ case.

6.2.3.1 The high-level outline in the case $d = 3$

We assume that Lemma 6.2.2 has been proven for $d \leq 2$ and use this assumption to show that it holds for $d = 3$. We will take $Q = Q_{\{1,2,3\}} \cup Q_{\{1\}} \cup Q_{\{2\}} \cup Q_{\{3\}}$ with elements chosen from $2^{3^{d+3}}\mathcal{B}' - 2^{3^{d+3}}\mathcal{B}'$ such that the Q_I have roughly equal size. This implies that if for some $r \in \mathcal{G}$ we have $r.q = 0$ for almost all $q \in Q$, then $r.q = 0$ holds for almost all $q \in Q_I$ for every $I = \{1\}, \{2\}, \{3\}, \{1,2,3\}$. We define $Q_{\{1,2,3\}}$ first, in a way that if $r.q = 0$ for almost all $q \in Q_{\{1,2,3\}}$, then $r = x + y$ where $x \in V_{\{1,2,3\}}$ for a vector space $V_{\{1,2,3\}}$ which is independent of r and have small dimension, and y has small partition rank. This already implies that any array $r \in \mathcal{G}$ with $r.q = 0$ for almost all $q \in Q$ is contained in $V_{\{1,2,3\}} + \mathbb{F}^{n_1} \otimes H_{\{2,3\}}(r) + \mathbb{F}^{n_2} \otimes H_{\{1,3\}}(r) + \mathbb{F}^{n_3} \otimes H_{\{1,2\}}(r)$ for some subspaces

$H_I(r) \subset \mathbb{F}^I$ depending on r and of small dimension. We then find $Q_{\{1\}}$ such that if $r \in V_{\{1,2,3\}} + \mathbb{F}^{n_1} \otimes H_{\{2,3\}}(r) + \mathbb{F}^{n_2} \otimes H_{\{1,3\}}(r) + \mathbb{F}^{n_3} \otimes H_{\{1,2\}}(r)$ has $r.q = 0$ for almost all $q \in Q_{\{1\}}$, then $r \in V_{\{1,2,3\}} + V_{\{1\}} \otimes \mathbb{F}^{\{2,3\}} + \mathbb{F}^{n_1} \otimes V_{\{2,3\}} + \mathbb{F}^{n_2} \otimes K_{\{1,3\}}(r) + \mathbb{F}^{n_3} \otimes K_{\{1,2\}}(r)$, where $V_{\{1\}} \subset \mathbb{F}^{n_1}$ and $V_{\{2,3\}} \subset \mathbb{F}^{\{2,3\}}$ are subspaces independent of r and have small dimension, and $K_I(r) \subset \mathbb{F}^I$ are subspaces of small dimension (although quite a bit larger than $\dim(H_I(r))$). Then we find $Q_{\{2\}}$ such that if $r \in V_{\{1,2,3\}} + V_{\{1\}} \otimes \mathbb{F}^{\{2,3\}} + \mathbb{F}^{n_1} \otimes V_{\{2,3\}} + \mathbb{F}^{n_2} \otimes K_{\{1,3\}}(r) + \mathbb{F}^{n_3} \otimes K_{\{1,2\}}(r)$ has $r.q = 0$ for almost all $q \in Q_{\{2\}}$, then $r \in V_{\{1,2,3\}} + V_{\{1\}} \otimes \mathbb{F}^{\{2,3\}} + \mathbb{F}^{n_1} \otimes V_{\{2,3\}} + V_{\{2\}} \otimes \mathbb{F}^{\{1,3\}} + \mathbb{F}^{n_2} \otimes V_{\{1,3\}} + \mathbb{F}^{n_3} \otimes L_{\{1,2\}}(r)$, where $V_{\{2\}} \subset \mathbb{F}^{n_2}$ and $V_{\{1,3\}} \subset \mathbb{F}^{\{1,3\}}$ are subspaces independent of r and have small dimension, and $L_{\{1,2\}}(r) \subset \mathbb{F}^{\{1,2\}}$ is a subspace of small dimension. Finally, we find $Q_{\{3\}}$ such that if $r \in V_{\{1,2,3\}} + V_{\{1\}} \otimes \mathbb{F}^{\{2,3\}} + \mathbb{F}^{n_1} \otimes V_{\{2,3\}} + V_{\{2\}} \otimes \mathbb{F}^{\{1,3\}} + \mathbb{F}^{n_2} \otimes V_{\{1,3\}} + \mathbb{F}^{n_3} \otimes L_{\{1,2\}}(r)$ has $r.q = 0$ for almost all $q \in Q_{\{3\}}$, then $r \in V_{\{1,2,3\}} + V_{\{1\}} \otimes \mathbb{F}^{\{2,3\}} + \mathbb{F}^{n_1} \otimes V_{\{2,3\}} + V_{\{2\}} \otimes \mathbb{F}^{\{1,3\}} + \mathbb{F}^{n_2} \otimes V_{\{1,3\}} + V_{\{3\}} \otimes \mathbb{F}^{\{1,2\}} + \mathbb{F}^{n_3} \otimes V_{\{1,2\}}$, where $V_{\{3\}} \subset \mathbb{F}^{n_3}$ and $V_{\{1,2\}} \subset \mathbb{F}^{\{1,2\}}$ are subspaces independent of r and have small dimension.

How will we find $Q_{\{1,2,3\}}, Q_{\{1\}}, Q_{\{2\}}$ and $Q_{\{3\}}$? In this outline we will only explain how to find $Q_{\{2\}}$ (but finding $Q_{\{1\}}$ and $Q_{\{3\}}$ is very similar). We take $Q_{\{2\}} = \bigcup_{u \in U} u \otimes Q_u$ where $U \subset \mathbb{F}^{n_2}$ is a subspace of low codimension, and for each $u \in U$, $Q_u \subset \mathbb{F}^{\{1,3\}}$ is a multiset consisting of pure tensors such that if for some $x \in \mathbb{F}^{\{1,3\}}$ we have $x.t = 0$ for almost all $t \in Q_u$, then $x \in W_{\{1,3\}}(u) + \mathbb{F}^{n_1} \otimes W_{\{3\}}(u) + W_{\{1\}}(u) \otimes \mathbb{F}^{n_3}$ for some subspaces $W_I(u) \subset \mathbb{F}^I$ not depending on x and of small dimension. Let us call a Q_u with this property *forcing*. We will also make sure that all the Q_u have roughly the same size.

6.2.3.2 Why does this $Q_{\{2\}}$ work?

In what follows, we will sketch why this choice is suitable. We remark that in the general case this is done in Lemma 6.2.15. Let R consist of those

$$r \in V_{\{1,2,3\}} + V_{\{1\}} \otimes \mathbb{F}^{\{2,3\}} + \mathbb{F}^{n_1} \otimes V_{\{2,3\}} + \mathbb{F}^{n_2} \otimes K_{\{1,3\}}(r) + \mathbb{F}^{n_3} \otimes K_{\{1,2\}}(r)$$

such that $r.q = 0$ for almost all $q \in Q_{\{2\}}$. Let $r \in R$. Write $r = r_2 + r_3 + r_4$ where

$$r_2 \in V_{\{1\}} \otimes \mathbb{F}^{\{2,3\}} + \mathbb{F}^{n_1} \otimes V_{\{2,3\}} + \mathbb{F}^{n_3} \otimes K_{\{1,2\}}(r), \quad r_3 \in V_{\{1,2,3\}}, \quad r_4 \in \mathbb{F}^{n_2} \otimes K_{\{1,3\}}(r).$$

It is enough to prove that

$$r_4 \in V_{\{2\}} \otimes \mathbb{F}^{\{1,3\}} + \mathbb{F}^{n_2} \otimes V_{\{1,3\}} + \mathbb{F}^{n_3} \otimes L'_{\{1,2\}}(r) \tag{6.2}$$

for some small subspaces $V_{\{2\}} \subset \mathbb{F}^{n_2}$, $V_{\{1,3\}} \subset \mathbb{F}^{\{1,3\}}$ and $L'_{\{1,2\}}(r) \subset \mathbb{F}^{\{1,2\}}$ (in fact, we will be able to take $V_{\{2\}} = U^{\perp}$).

First note that r_2u has small (partition) rank for every $u \in U$. Indeed, $r_2u \in V_{\{1\}} \otimes \mathbb{F}^{n_3} + \mathbb{F}^{n_1} \otimes V_{\{2,3\}}u + \mathbb{F}^{n_3} \otimes K_{\{1,2\}}(r)u$, where, for a vector space L of tensors, Lu denotes the space $\{su : s \in L\}$.

Moreover, since the Q_u all have roughly the same size, for almost every $u \in U$ we have that $r.(u \otimes t) = 0$ holds for almost every $t \in Q_u$. But $r.(u \otimes t) = (ru).t$, therefore as Q_u is forcing, it follows that for any such u

$$ru \in W_{\{1,3\}}(u) + \mathbb{F}^{n_1} \otimes W_{\{3\}}(u) + W_{\{1\}}(u) \otimes \mathbb{F}^{n_3}$$

for some subspaces $W_I(u) \subset \mathbb{F}^I$ not depending on r and of small dimension. Since any element of $\mathbb{F}^{n_1} \otimes W_{\{3\}}(u) + W_{\{1\}}(u) \otimes \mathbb{F}^{n_3}$ has small partition rank, it follows that for almost every $u \in U$,

$$r_4 u = ru - r_2 u - r_3 u \in W_{\{1,3\}}(u) + V_{\{1,2,3\}} u + s(u) \tag{6.3}$$

where $s(u)$ is a tensor of small partition rank.

Define a sequence $0 = Z(0) \subset Z(1) \subset \ldots \subset Z(m) \subset \mathbb{F}^{\{1,3\}}$ of subspaces recursively as follows. Given $Z(j)$, if there is some $r \in R$ such that $r_4 u$ is far from $Z(j)$ for many $u \in U$, then set $Z(j+1) = Z(j) + K_{1,3}(r)$. What we mean by $r_4 u$ being far from $Z(j)$ is that there is no $z \in Z(j)$ such that $r_4 u - z$ has small partition rank. For suitably chosen parameters, one can show that this procedure cannot go on for too long, ie. that for some not too large m we have that for every $r \in R$, for almost all $u \in U$ there is some $z \in Z(m)$ with $r_4 u - z$ having small partition rank.

Now let $r \in R$. Let $X(r)$ be the set consisting of those $x \in K_{\{1,3\}}(r)$ which are close to $Z(m)$. Then $r_4 u \in X(r)$ for almost every $u \in U$. Let $t_1, \ldots, t_\alpha$ be a maximal linearly independent subset of $X(r)$ and extend it to a basis $t_1, \ldots, t_\alpha, t'_1, \ldots, t'_\beta$ for $K_{\{1,3\}}(r)$. Now if a linear combination of $t_1, \ldots, t_\alpha, t'_1, \ldots, t'_\beta$ is in $X(r)$, then the coefficients of $t'_1, \ldots, t'_\beta$ are all zero. Write $r_4 = \sum_{i \le \alpha} s_i \otimes t_i + \sum_{j \le \beta} s'_j \otimes t'_j$ for some $s_i, s'_j \in \mathbb{F}^{n_2}$. Since $r_4 u \in X(r)$ for almost all $u \in U$, we have, for all j, that $s'_j.u = 0$ for almost all $u \in U$. Since these hold for more than half of $u \in U$, we obtain $s'_j \in U^\perp$ for every j, therefore $\sum_{j \le \beta} s'_j \otimes t'_j \in U^\perp \otimes \mathbb{F}^{\{1,3\}}$.

Since $t_i \in X(r)$ for every i, we may choose $z_i \in Z(m)$ such that $t_i = z_i + y_i$ where $y_i \in \mathbb{F}^{\{1,3\}}$ has small partition rank. Now $\sum_{i \le \alpha} s_i \otimes t_i \in \mathbb{F}^{n_2} \otimes Z(m) + \sum_{i \le \alpha} s_i \otimes y_i$. Moreover, as α is small and each y_i has small partition rank, we have $\sum_{i \le \alpha} s_i \otimes y_i \in L'_{\{1,2\}}(r) \otimes \mathbb{F}^{n_3}$ for some small $L'_{\{1,2\}}(r) \subset \mathbb{F}^{\{1,2\}}$. So we have proved (6.2) with $V_{\{2\}} = U^\perp$ and $V_{\{1,3\}} = Z(m)$.

6.2.3.3 Why can we find such a $Q_{\{2\}}$ inside $2^{3^{d+3}}\mathcal{B}' - 2^{3^{d+3}}\mathcal{B}'$?

Now we describe why there must exist $Q_{\{2\}}$ with elements chosen from $2^{3^{3+3}}\mathcal{B}' - 2^{3^{3+3}}\mathcal{B}'$ and having the required properties. We remark that in the general case this is done in Lemma 6.2.14. We want to find a subspace $U \subset \mathbb{F}^{n_2}$ of low codimension, and forcing multisets $Q_u \subset \mathbb{F}^{\{1,3\}}$ ($u \in U$) consisting of pure tensors such that for every $u \in U$, $u \otimes Q_u \subset 2^{3^{3+3}}\mathcal{B}' - 2^{3^{3+3}}\mathcal{B}'$. Let $\mathcal{D}$ be the multiset $\{v \otimes w : v \in \mathbb{F}^{n_1}, w \in \mathbb{F}^{n_3}\}$. Notice that if some set R is dense in $\mathcal{D}$, then by the induction hypothesis we can find a forcing set in $2^{3^{2+3}}R - 2^{3^{2+3}}R$ consisting of pure tensors. Therefore it is enough to find a low codimensional subspace U and dense sets $R_u \subset \mathcal{D}$ (for every $u \in U$) such that $u \otimes R_u \subset 32\mathcal{B}' - 32\mathcal{B}'$. As $\mathcal{B}'$ is dense in $\mathcal{B}$, we have a dense subset $S \subset \mathbb{F}^{n_2}$ and dense subsets $T_s \subset \mathcal{D}$ ($s \in S$) such that $s \otimes T_s \subset \mathcal{B}'$ for every $s \in S$. By Bogolyubov's lemma (Lemma 6.2.1), there is a low codimensional subspace U contained in $2S - 2S$. To establish the existence of a dense $R_u \subset \mathcal{D}$ with $u \otimes R_u \subset 32\mathcal{B}' - 32\mathcal{B}'$ for every $u \in U$, it is enough to prove the following lemma.

Lemma 6.2.5. *Let T_1, T_2, T_3, T_4 be dense subsets of $\mathcal{D}$. Then $\mathcal{D} \cap \bigcap_{i \leq 4}(8T_i - 8T_i)$ is dense in $\mathcal{D}$.*

Indeed, once we have this lemma, it follows that for any $s_1, s_2, s_3, s_4 \in S$, the set $\mathcal{D} \cap \bigcap_{i \leq 4}(8T_{s_i} - 8T_{s_i})$ is dense in $\mathcal{D}$. But if $u \in U$, then we can write $u = s_1 + s_2 - s_3 - s_4$ for some $s_i \in S$, and then $u \otimes \bigcap_{i \leq 4}(8T_{s_i} - 8T_{s_i}) \subset s_1 \otimes \bigcap_{i \leq 4}(8T_{s_i} - 8T_{s_i}) + s_2 \otimes \bigcap_{i \leq 4}(8T_{s_i} - 8T_{s_i}) - s_3 \otimes \bigcap_{i \leq 4}(8T_{s_i} - 8T_{s_i}) - s_4 \otimes \bigcap_{i \leq 4}(8T_{s_i} - 8T_{s_i}) \subset 32\mathcal{B}' - 32\mathcal{B}'$.

Lemma 6.2.5 follows easily from the next two lemmas.

Lemma 6.2.6. *Let A be a dense subset of $\mathcal{D}$. Then there exist a dense subspace $V \subset \mathbb{F}^{n_1}$ and for each $v \in V$ a dense subspace $W_v \subset \mathbb{F}^{n_3}$ such that $v \otimes W_v \subset 8A - 8A$ for every $v \in V$.*

Proof. There exist a dense subset $B \subset \mathbb{F}^{n_1}$ and dense subsets $C_b \subset \mathbb{F}^{n_3}$ for each $b \in B$ such that $b \otimes C_b \subset A$. By Bogolyubov's lemma, $2B - 2B$ contains a dense subspace $V \subset \mathbb{F}^{n_1}$, and for every $b \in B$, $2C_b - 2C_b$ contains a dense subspace $L_b \subset \mathbb{F}^{n_3}$. For any $v \in V$, choose $b_1, b_2, b_3, b_4 \in B$ with $v = b_1 + b_2 - b_3 - b_4$ and set $W_v = \bigcap_{i \leq 4} L_{b_i}$. Note that $b_i \otimes w \in 2A - 2A$ for every $i \leq 4$ and $w \in W_v$, therefore $v \otimes w \in 8A - 8A$. $\square$

Lemma 6.2.7. *Suppose that we have dense subspaces $V, V' \subset \mathbb{F}^{n_1}$, for each $v \in V$ a dense subspace $W_v \subset \mathbb{F}^{n_3}$, and for each $v' \in V'$ a dense subspace $W'_{v'} \subset \mathbb{F}^{n_3}$. Then $(\bigcup_{v \in V} v \otimes W_v) \cap (\bigcup_{v' \in V'} v' \otimes W'_{v'}) = \bigcup_{v \in V \cap V'} v \otimes (W_v \cap W'_v)$. In particular, this intersection is a dense subset of $\mathcal{D}$.*

Proof. The identity is trivial. Since the subspaces $V \cap V'$ and $W_v \cap W'_v$ are dense, the second assertion follows. $\square$

6.2.3.4 How can this be extended to $d > 3$?

Now we briefly sketch what the main difficulties are in the $d > 3$ case and how we can address them. The underlying strategy is similar: we take an ordering $\prec$ of the set of non-empty subsets $I \subset [d-1]$, and for each such I we choose Q_I such that any array

$$r \in W_{[d]} + \sum_{J \prec I}(W_J \otimes \mathbb{F}^{J^c} + \mathbb{F}^J \otimes W_{J^c}) + \sum_{J \succeq I} \mathbb{F}^J \otimes H_{J^c}(r) \tag{6.4}$$

with $r.q = 0$ for almost all $q \in Q_I$ has

$$r \in W_{[d]} + \sum_{J \preceq I} (U_J \otimes \mathbb{F}^{J^c} + \mathbb{F}^J \otimes U_{J^c}) + \sum_{J \succ I} \mathbb{F}^J \otimes K_{J^c}(r)$$

where $U_J, U_{J^c}, K_{J^c}(r)$ can have dimension slightly larger than those of W_J, W_{J^c} and H_{J^c}, but they are still low dimensional. In the $d = 3$ case, we have made use of a decomposition $r = r_2 + r_3 + r_4$ where $r_4 \in \mathbb{F}^I \otimes H_{I^c}(r)$, $r_2 u$ has small partition rank and $r_3 u$ is in a small subspace independent of r for every $u \in \mathbb{F}^I$. In general, such a decomposition need not exist. For example, when $d = 4$ and $I = \{1, 2\}$, then an array in $W_{\{1\}} \otimes \mathbb{F}^{\{2,3,4\}}$ (or in $\mathbb{F}^{n_1} \otimes H_{\{2,3,4\}}(r)$ if we were to take $\{1, 2\} \prec \{1\}$), when multiplied by some pure tensor $u \in \mathbb{F}^{\{1,2\}}$, yields a tensor which need not have small partition rank and need not lie a small space independent of r. However, by restricting the possible choices for u, we can make sure that the product is always zero. So we will take a decomposition $r = r_1 + r_2 + r_3 + r_4$ such that $r_4 \in \mathbb{F}^I \otimes H_{I^c}(r)$; for every pure tensor $u \in \mathbb{F}^I$, $r_2 u$ has small partition rank and $r_3 u$ lies in a small space depending only on u; and crucially, for every $q \in Q_I$, $r_1.q = 0$. To achieve this, we need to insist that $J \prec I$ whenever $J \subsetneq I$ and that Q_I is orthogonal to certain subspaces. To see this, note that in the above example where $d = 4$ and $I = \{1, 2\}$ we need that $\{1\} \prec \{1, 2\}$ and $Q_{\{1,2\}}$ is orthogonal to $W_{\{1\}} \otimes \mathbb{F}^{\{2,3,4\}}$. (If we had $\{1, 2\} \prec \{1\}$, then in (6.4) we would have a term $\mathbb{F}^{n_1} \otimes H_{\{2,3,4\}}(r)$ rather than $W_{\{1\}} \otimes \mathbb{F}^{\{2,3,4\}}$, which we could not control.)

We also need to generalise Lemma 6.2.5 to the case $d > 3$. Instead of using $\bigcup_{v \in V} v \otimes W_v$ as in Lemma 6.2.6, we need to define an object in $\mathcal{B}$ such that

1. an instance of the object can be found in $k\mathcal{B}' - k\mathcal{B}'$ for some small k whenever $\mathcal{B}'$ is dense in $\mathcal{B}$ (generalising Lemma 6.2.6)

2. the intersection of few instances of this object is a dense subset of $\mathcal{B}$ (generalising Lemma 6.2.7)

In the next subsection we describe this object and show that it has the required properties.

6.2.4 Construction of some auxiliary sets

Definition 6.2.8. Suppose that we have a collection of vector spaces as follows. The first one is $U \subset \mathbb{F}^{n_1}$, of codimension at most l. Then, for every $u_1 \in U$, there is some $U_{u_1} \subset \mathbb{F}^{n_2}$. In general, for every $2 \leq k \leq d$ and every $u_1 \in U, u_2 \in U_{u_1}, \ldots, u_{k-1} \in U_{u_1,\ldots,u_{k-2}}$, there is a subspace $U_{u_1,\ldots,u_{k-1}} \subset \mathbb{F}^{n_k}$. Assume, in addition, that the codimension of $U_{u_1,\ldots,u_{k-1}}$ in $\mathbb{F}^{n_k}$ is at most l for every $u_1 \in U, \ldots, u_{k-1} \in U_{u_1,\ldots,u_{k-2}}$. Then the multiset $Q = \{u_1 \otimes \cdots \otimes u_d : u_1 \in U, \ldots, u_d \in U_{u_1,\ldots,u_{d-1}}\}$ is called an l-system.

The next lemma is the generalisation of Lemma 6.2.7 from the previous subsection.

Lemma 6.2.9. *Let Q be an l-system and let Q' be an l'-system. Then $Q \cap Q'$ contains an $(l+l')$-system.*

Proof. Let Q have spaces as in Definition 6.2.8 and let Q' have spaces $U'_{u'_1,\dots,u'_{k-1}}$. We define an $(l+l')$-system P contained in $Q \cap Q'$ as follows. Let $V = U \cap U'$. Suppose we have defined $V_{v_1,\dots,v_{j-1}}$ for all $j \le k$. Let $v_1 \in V, v_2 \in V_{v_1}, \dots, v_{k-1} \in V_{v_1,\dots,v_{k-2}}$. We let $V_{v_1\dots,v_{k-1}} = U_{v_1\dots,v_{k-1}} \cap U'_{v_1\dots,v_{k-1}}$. This is well-defined and has codimension at most $l+l'$ in $\mathbb{F}^{n_k}$. Let P be the $(l+l')$-system with spaces $V_{v_1,\dots,v_{k-1}}$. □

The next lemma is the generalisation of Lemma 6.2.6 from the previous subsection.

Lemma 6.2.10. *Let $\mathcal{B}' \subset \mathcal{B}$ be a multiset such that $|\mathcal{B}'| \ge \delta|\mathcal{B}|$. Then there exists an f_1-system whose elements are chosen from $f_2\mathcal{B}' - f_2\mathcal{B}'$ with $f_1 = C \cdot 4^d(\log(2^d/\delta))^4$ and $f_2 = 4^d$.*

Proof. The proof is by induction on d. The case $d = 1$ is a direct consequence of Lemma 6.2.1. Suppose that the lemma has been proved for all $d' < d$ and let $\mathcal{B}' \subset \mathcal{B}$ be a multiset such that $|\mathcal{B}'| \ge \delta|\mathcal{B}|$. Let $\mathcal{D}$ be the multiset $\{v_2 \otimes \dots \otimes v_d : v_2 \in \mathbb{F}^{n_2}, \dots, v_d \in \mathbb{F}^{n_d}\}$. For each $u \in \mathbb{F}^{n_1}$, let $\mathcal{B}'_u = \{s \in \mathcal{D} : u \otimes s \in \mathcal{B}'\}$ and let $T = \{u \in \mathbb{F}^{n_1} : |\mathcal{B}'_u| \ge \frac{\delta}{2}|\mathcal{D}|\}$. By averaging, we have that $|T| \ge \frac{\delta}{2}|\mathbb{F}^{n_1}|$. Now by the induction hypothesis, for every $t \in T$, there exists a g_1-system in $\mathbb{F}^{n_2} \otimes \dots \otimes \mathbb{F}^{n_d}$ (whose definition is analogous to the definition of a system in $\mathbb{F}^{n_1} \otimes \dots \otimes \mathbb{F}^{n_d}$), called P_t, contained in $g_2\mathcal{B}'_t - g_2\mathcal{B}'_t$ where $g_1 = C \cdot 4^{d-1}(\log(2^d/\delta))^4$ and $g_2 = 4^{d-1}$. By Lemma 6.2.1, $2T - 2T$ contains a subspace $U \subset \mathbb{F}^{n_1}$ of codimension at most $C(\log(2/\delta))^4$. For each $u \in U$, write $u = t_1 + t_2 - t_3 - t_4$ arbitrarily with $t_i \in T$, and let $Q_u = P_{t_1} \cap P_{t_2} \cap P_{t_3} \cap P_{t_4}$, which is a g_3-system with $g_3 = 4g_1 = C \cdot 4^d(\log(2^d/\delta))^4$, by Lemma 6.2.9. Thus, $Q = \bigcup_{u \in U}(u \otimes Q_u)$ is indeed an f_1-system. Moreover, for any $u \in U, s \in Q_u$, we have $u \otimes s = t_1 \otimes s + t_2 \otimes s - t_3 \otimes s - t_4 \otimes s$ for some $t_i \in T$ and $s \in \bigcap_{i \le 4} P_{t_i}$. Then $t_i \otimes s \in g_2\mathcal{B}' - g_2\mathcal{B}'$, therefore $u \otimes s \in 4g_2\mathcal{B}' - 4g_2\mathcal{B}'$, so the elements of Q are indeed chosen from $f_2\mathcal{B}' - f_2\mathcal{B}'$. □

The next lemma describes a property of systems which was not needed for us in the $d = 3$ case, but is crucial in the general case. It is required for finding a suitable decomposition $r = r_1 + r_2 + r_3 + r_4$ described at the end of the previous subsection. Indeed, we need a set Q_I which is orthogonal to certain spaces of the form $W_J \otimes \mathbb{F}^{J^c}$ (ie. is contained in $W_J^{\perp} \otimes \mathbb{F}^{J^c}$) to make sure that $r_1.q = 0$ for every $q \in Q_I$. We will use the following lemma to guarantee the existence of such a set Q_I.

Lemma 6.2.11. *Let Q be a k-system and for every non-empty $I \subset [d]$, let $L_I \subset \mathbb{F}^I$ be a subspace of codimension at most l. Let $T = \bigcap_I (L_I \otimes \mathbb{F}^{I^c})$. Then $Q \cap T$ contains an f-system for $f = k + 2^d l$.*

Proof. Let the spaces of Q be $U_{u_1,\dots,u_{j-1}}$. It suffices to prove that for every $1 \le j \le d$, and every $u_1 \in U, \dots, u_{j-1} \in U_{u_1,\dots,u_{j-2}}$, the codimension of $(u_1 \otimes \dots \otimes u_{j-1} \otimes U_{u_1,\dots,u_{j-1}}) \cap \bigcap_{I \subset [j], j \in I} (L_I \otimes \mathbb{F}^{[j]\setminus I})$ in $u_1 \otimes \dots \otimes u_{j-1} \otimes U_{u_1,\dots,u_{j-1}}$ is at most $2^d l$. Thus, it suffices to prove that for every $I \subset [j]$ with $j \in I$, the codimension of $(u_1 \otimes \dots \otimes u_{j-1} \otimes U_{u_1,\dots,u_{j-1}}) \cap (L_I \otimes \mathbb{F}^{[j]\setminus I})$ in $u_1 \otimes \dots \otimes u_{j-1} \otimes U_{u_1,\dots,u_{j-1}}$ is at most l. But this is equivalent to the statement that $\big((\bigotimes_{i \in I\setminus\{j\}} u_i) \otimes U_{u_1,\dots,u_{j-1}}\big) \cap L_I$ has codimension at most l in $(\bigotimes_{i \in I\setminus\{j\}} u_i) \otimes U_{u_1,\dots,u_{j-1}}$, which clearly holds. □

6.2.5 The proof of Lemma 6.2.2

We now turn to the proof of Lemma 6.2.2. As described in the outline, the first step is to find a $Q_{[d]}$ such that if $r.q = 0$ for almost all $q \in Q_{[d]}$, then $r = x + y$ where $x \in V_{[d]}$ for a small space $V_{[d]}$ independent of r, and y has low partition rank.

Lemma 6.2.12. *Let $d \ge 2$ and suppose that Lemma 6.2.2 has been proved for $d' = d-1$. Let $\mathcal{B}' \subset \mathcal{B}$ be such that $|\mathcal{B}'| \ge \delta|\mathcal{B}|$ for some $\delta > 0$. Then there exist some $Q \subset 2\mathcal{B}' - 2\mathcal{B}'$ consisting of pure tensors and a subspace $V_{[d]} \subset \mathbb{F}^{[d]}$ of dimension at most $4C(\log(2/\delta))^4$ with the following property. Any array r with $r.q = 0$ for at least $\frac{7}{8}|Q|$ choices $q \in Q$ can be written as $r = x + y$ where $x \in V_{[d]}$ and y is f-degenerate for $f = G(d-1, \frac{\delta}{4|\mathbb{F}|^{4C(\log 2/\delta)^4}})$.*

Proof. Let $\mathcal{D}$ be the multiset $\{u_1 \otimes \dots \otimes u_{d-1} : u_1 \in \mathbb{F}^{n_1}, \dots, u_{d-1} \in \mathbb{F}^{n_{d-1}}\}$ and let $\mathcal{D}' = \{t \in \mathcal{D} : t \otimes u \in \mathcal{B}'$ for at least $\frac{\delta}{2}|\mathbb{F}|^{n_d}$ choices $u \in \mathbb{F}^{n_d}\}$. Clearly, we have $|\mathcal{D}'| \ge \frac{\delta}{2}|\mathcal{D}|$. Moreover, by Lemma 6.2.1, for every $t \in \mathcal{D}'$, there exists a subspace $U_t \subset \mathbb{F}^{n_d}$ of codimension at most $C(\log(2/\delta))^4$ such that $t \otimes U_t \subset 2\mathcal{B}' - 2\mathcal{B}'$. After passing to suitable subspaces, we may assume that all U_t have the same codimension $k \le C(\log(2/\delta))^4$. Now let $Q = \cup_{t \in \mathcal{D}'}(t \otimes U_t)$.

Write R for the set of arrays r with $r.q = 0$ for at least $\frac{7}{8}|Q|$ choices $q \in Q$.

We now define a sequence of subspaces $0 = V(0) \subset V(1) \subset \dots \subset V(m) \subset \mathbb{F}^{[d]}$ recursively as follows.

Given $V(j)$, if for every $r \in R$ there are at least $\frac{|\mathcal{D}'|}{2}$ choices $t \in \mathcal{D}'$ with $rt \in V(j)t$, then we set $m = j$ and terminate. (Here and below, for a subspace $L \subset \mathcal{G}$ and an array $s \in \mathbb{F}^I$, we write Ls for the subspace $\{rs : r \in L\} \subset \mathbb{F}^{I^c}$.)

Else, we choose some $r \in R$ such that there are at most $\frac{|\mathcal{D}'|}{2}$ choices $t \in \mathcal{D}'$ with $rt \in V(j)t$. We set $V(j+1) = V(j) + \operatorname{span}(r)$. Note that $r.(t \otimes s) = (rt).s$ for every $s \in U_t$. If $rt \notin U_t^\perp$, then $(rt).s = 0$ holds for only a proportion $1/|\mathbb{F}| \le 1/2$ of all $s \in U_t$. Thus, as $r \in R$, we have $rt \in U_t^\perp$ for at least $\frac{3}{4}|\mathcal{D}'|$ choices $t \in \mathcal{D}'$. Moreover, since $rt \in V(j)t$ holds for at most $\frac{|\mathcal{D}'|}{2}$ choices $t \in \mathcal{D}'$, it follows that for at least $\frac{|\mathcal{D}'|}{4}$ choices $t \in \mathcal{D}'$ we have $rt \in U_t^\perp \setminus V(j)t$. Thus, we have $\dim(U_t^\perp \cap V(j+1)t) > \dim(U_t^\perp \cap V(j)t)$ for at least $\frac{|\mathcal{D}'|}{4}$ choices $t \in \mathcal{D}'$.

However, for any j we have $\sum_{t \in \mathcal{D}'} \dim(U_t^\perp \cap V(j)t) \le \sum_{t \in \mathcal{D}'} \dim U_t^\perp \le C|\mathcal{D}'|(\log(2/\delta))^4$. Thus, we get $m \le 4C(\log(2/\delta))^4$. Set $V_{[d]} = V(m)$. Then $\dim V_{[d]} \le 4C(\log(2/\delta))^4$, as claimed.

Now let $r \in R$ be arbitrary. By definition, there are at least $|\mathcal{D}'|/2$ choices $t \in \mathcal{D}'$ with $rt \in V_{[d]}t$. Then there is some $v \in V_{[d]}$ such that $rt = vt$ for at least $\frac{|\mathcal{D}'|}{2|V_{[d]}|}$ choices $t \in \mathcal{D}'$, and hence also for at least $\frac{\delta|\mathcal{D}|}{4|V_{[d]}|}$ choices $t \in \mathcal{D}$. Note that $\frac{\delta}{4|V_{[d]}|} \geq \frac{\delta}{4|\mathbb{F}|^{4C(\log 2/\delta)^4}}$, therefore by Lemma 6.2.4, $r - v$ is f-degenerate. $\square$

Definition 6.2.13. Let k be a positive integer and let $0 \leq \alpha \leq 1$. Let Q be a multiset with elements chosen from $\mathcal{G}$ (with arbitrary multiplicity). We say that Q is (k, α)-forcing if the set of all arrays $r \in \mathcal{G}$ with $r.q = 0$ for at least $\alpha|Q|$ choices $q \in Q$ is contained in a set of the from $\sum_{I \subset [d], I \neq \emptyset} V_I \otimes \mathbb{F}^{I^c}$ for some $V_I \subset \mathbb{F}^I$ of dimension at most k.

We now turn to the main part of the proof of Lemma 6.2.2. For each non-empty $I \subset [d-1]$ we will construct Q_I as defined in the next result, and (roughly) we will take $Q = Q_{[d]} \cup \bigcup_{I \subset [d-1], I \neq \emptyset} Q_I$, where $Q_{[d]}$ is provided by Lemma 6.2.12. The properties that Q_I has are generalisations of the properties that $Q_{\{2\}}$ had in Subsection 6.2.3. Accordingly, the next lemma is the generalisation of the discussion in Subsubsection 6.2.3.3.

Lemma 6.2.14. *Let $d \geq 2$ and suppose that Lemma 6.2.2 has been proved for every $d' < d$. Let $\mathcal{B}' \subset \mathcal{B}$ have $|\mathcal{B}'| \geq \delta|\mathcal{B}|$ for some $0 < \delta \leq 1/2$. Let $k \geq G(d-1, \delta)$ be arbitrary, let $I \subset [d-1], I \neq \emptyset$, and let $W_J \subset \mathbb{F}^J$ be subspaces of dimension at most k for every $J \subset I, J \neq I, J \neq \emptyset$. Then there exist a multiset Q', and a multiset Q_s for each $s \in Q'$ with the following properties.*

(1) The elements of Q' are pure tensors chosen from $\bigcap_{J \subset I, J \neq I, J \neq \emptyset} (W_J^{\perp} \otimes \mathbb{F}^{I \setminus J}) \subset \mathbb{F}^I$

(2) Q' is $(f_1, 1 - f_2)$-forcing with $f_1 = G(|I|, |\mathbb{F}|^{-2^{d+1}dk})$, $f_2 = 2^{-3^{d+2}}$

(3) For each $s \in Q'$, the elements of Q_s are pure tensors chosen from $\mathbb{F}^{I^c}$

(4) For each $s \in Q'$, Q_s is $(f_3, 1 - f_4)$-forcing with
$f_3 = G\left(d - |I|, |\mathbb{F}|^{-2^{3^{d+4}}C(\log(2^{d-1}/\delta))^4}\right)$, $f_4 = 2^{-3^{d+2}}$

(5) $\max_{s \in Q'} |Q_s| \leq 2 \min_{s \in Q'} |Q_s|$

(6) The elements of the multiset $Q_I := \{s \otimes t : s \in Q', t \in Q_s\} = \bigcup_{s \in Q'} (s \otimes Q_s)$ are chosen from $f_5\mathcal{B}' - f_5\mathcal{B}'$ with $f_5 = 2^{3^{d+3}}$.

Proof. By symmetry, we may assume that $I = [a]$ for some $1 \leq a \leq d-1$. Let $\mathcal{C}$ be the multiset $\{u_1 \otimes \cdots \otimes u_a : u_i \in \mathbb{F}^{n_i}\}$ and let $\mathcal{D}$ be the multiset $\{u_{a+1} \otimes \cdots \otimes u_d : u_i \in \mathbb{F}^{n_i}\}$. For each $s \in \mathcal{C}$, let $\mathcal{D}_s = \{t \in \mathcal{D} : s \otimes t \in \mathcal{B}'\}$. Also, let $\mathcal{C}' = \{s \in \mathcal{C} : |\mathcal{D}_s| \geq \frac{\delta}{2}|\mathcal{D}|\}$. Clearly, $|\mathcal{C}'| \geq \frac{\delta}{2}|\mathcal{C}|$. By Lemma 6.2.10, there exists a g_1-system R (with respect to $\mathbb{F}^I$) with elements chosen from $g_2\mathcal{C}' - g_2\mathcal{C}'$ with $g_1 = C \cdot 4^d(\log(2^{d-1}/\delta))^4$ and $g_2 = 4^d$. By Lemma 6.2.11, $R \cap \bigcap_{J \subset I, J \neq I, J \neq \emptyset} (W_J^{\perp} \otimes \mathbb{F}^{I \setminus J})$ contains a g_3-system T' for $g_3 = C \cdot 4^d(\log(2^{d-1}/\delta))^4 + 2^d k$. Now $|T'| \geq |\mathbb{F}|^{-dg_3}|\mathcal{C}|$. By Lemma 6.2.2 (applied to a in place of d), it follows that there exists a multiset Q' whose elements are pure tensors chosen from

$g_4T' - g_4T'$ and which is $(g_5, 1-g_6)$-forcing for $g_4 = 2^{3^{a+3}} \leq 2^{3^{d+2}}$, $g_5 = G(a, |\mathbb{F}|^{-dg_3})$ and $g_6 = 2^{-3^{a+3}} \geq 2^{-3^{d+2}}$. Note that since $\delta \leq 1/2$, we have $C \cdot 4^d(\log(2^{d-1}/\delta))^4 = C \cdot 4^d(d-1+\log(1/\delta))^4 \leq C \cdot 4^d(d\log(1/\delta))^4$. But this is at most as $G(d-1,\delta) \leq k$, so $g_3 \leq 2 \cdot 2^d k$, therefore Q' satisfies (1) and (2) in the statement of this lemma.

By Lemma 6.2.10, for each $s \in \mathcal{C}'$ there exists a g_7-system R_s (with respect to $\mathbb{F}^{I^c}$) contained in $g_8\mathcal{D}_s - g_8\mathcal{D}_s$, where $g_7 = C \cdot 4^d(\log(2^{d-1}/\delta))^4$ and $g_8 = 4^d$. For every $s \in Q'$, choose $s_1, \dots, s_{l+l'} \in \mathcal{C}'$ with $l, l' \leq 2^{3^{d+3}}$ such that $s = s_1 + \dots + s_l - s_{l+1} - \dots - s_{l+l'}$ (this is possible, since the elements of Q' are chosen from $2g_2g_4\mathcal{C}' - 2g_2g_4\mathcal{C}'$ and $2g_2g_4 \leq 2^{3^{d+3}}$), and let $P_s = \bigcap_{i \leq l+l'} R_s$. By Lemma 6.2.9, P_s contains a g_9-system with $g_9 = 2 \cdot 2^{3^{d+3}} \cdot C \cdot 4^d(\log(2^{d-1}/\delta))^4$, therefore $|P_s| \geq g_{10}|\mathcal{D}|$ for $g_{10} = |\mathbb{F}|^{-dg_9} \geq |\mathbb{F}|^{-2^{3^{d+4}}C(\log(2^{d-1}/\delta))^4}$. By Lemma 6.2.2 (applied to $d-a$ in place of d), for every $s \in Q'$ there exists a multiset Q_s consisting of pure tensors with elements chosen from $g_{11}P_s - g_{11}P_s$ which is $(g_{12}, 1-g_{13})$-forcing for $g_{11} = 2^{3^{d-a+3}} \leq 2^{3^{d+2}}$, $g_{12} = G(d-a, |\mathbb{F}|^{-dg_9}) \leq G(d-a, |\mathbb{F}|^{-2^{3^{d+4}}C(\log(2^{d-1}/\delta))^4})$ and $g_{13} = 2^{-3^{d-a+3}} \geq 2^{-3^{d+2}}$. Notice that if we repeat every element of Q_s the same number of times, then the multiset obtained is still $(g_{12}, 1-g_{13})$-forcing, so we may assume that $\max_{s \in Q'} |Q_s| \leq 2\min_{s \in Q'} |Q_s|$. Thus, the Q_s satisfy (3), (4) and (5).

Define $Q_I = \{s \otimes t : s \in Q', t \in Q_s\} = \bigcup_{s \in Q'}(s \otimes Q_s)$. Note that as $R_s \subset g_8\mathcal{D}_s - g_8\mathcal{D}_s$ for all $s \in \mathcal{C}'$, we have $s \otimes R_s \subset g_8\mathcal{B}' - g_8\mathcal{B}'$ for all $s \in \mathcal{C}'$. But the elements of Q' are chosen from $2g_2g_4\mathcal{C}' - 2g_2g_4\mathcal{C}'$, so $s \otimes P_s \subset 4g_2g_4g_8\mathcal{B}' - 4g_2g_4g_8\mathcal{B}'$ for all $s \in Q'$. Finally, the elements of Q_s are chosen from $g_{11}P_s - g_{11}P_s$, so the elements of $s \otimes Q_s$ are chosen from $8g_2g_4g_8g_{11}\mathcal{B}' - 8g_2g_4g_8g_{11}\mathcal{B}'$ for every $s \in Q'$. Since $8g_2g_4g_8g_{11} \leq 8 \cdot (4^d)^2 \cdot (2^{3^{d+2}})^2 = 2^{3+4d+2\cdot 3^{d+2}} \leq 2^{3^{d+3}}$, property (6) is satisfied. $\square$

The next lemma is the last ingredient of the proof. It is a generalisation of the discussion in Subsubsection 6.2.3.2. Given a tensor $r \in V_{[d]} + \sum_{I \subset [d-1], I \neq \emptyset} \mathbb{F}^I \otimes H_{I^c}(r)$, we turn the terms $\mathbb{F}^I \otimes H_{I^c}(r)$ one by one into terms $V_I \otimes \mathbb{F}^{I^c} + \mathbb{F}^I \otimes V_{I^c}$ where V_J are small and do not depend on r. (Note that this is not quite the same as our approach to the case $d = 3$.) As briefly explained in Subsubsection 6.2.3.4, the order in which the various I are considered is important: we define $\prec$ to be any total order on the set of non-empty subsets of $[d-1]$ such that if $J \subsetneq I$ then $J \prec I$. It is worth noting that unlike in the $d = 3$ case, the subspaces V_J, V_{J^c} with $J \prec I$ are allowed to change when V_I and V_{I^c} get defined (although in fact the V_{J^c} will not change, and the V_J change only for $J \subsetneq I$). All we require is that they do not become much larger.

Lemma 6.2.15. *Let $d \geq 2$, $0 < \delta \leq 1/2$ and $k \geq G(d-1,\delta)^2$. Let $I \subset [d-1], I \neq \emptyset$ and let $W_J \subset \mathbb{F}^J, W_{J^c} \subset \mathbb{F}^{J^c}$ be subspaces of dimension at most k for every $J \prec I$. Moreover, let $W_{[d]} \subset \mathbb{F}^{[d]}$ have dimension at most k. Suppose that Q', Q_s (and Q_I) have the six properties described in Lemma 6.2.14. Then any array*

$$r \in W_{[d]} + \sum_{J \prec I}(W_J \otimes \mathbb{F}^{J^c} + \mathbb{F}^J \otimes W_{J^c}) + \sum_{J \succeq I} \mathbb{F}^J \otimes H_{J^c}(r)$$

with $\dim(H_{J^c}(r)) \le k$ *and the property that* $r.q = 0$ *for at least* $(1-\frac{1}{4}(2^{-3^{d+2}})^2)|Q_I|$ *choices* $q \in Q_I$ *is contained in*

$$W_{[d]} + \sum_{J \preceq I}(U_J \otimes \mathbb{F}^{J^c} + \mathbb{F}^J \otimes U_{J^c}) + \sum_{J \succ I} \mathbb{F}^J \otimes K_{J^c}(r)$$

for some $U_J \subset \mathbb{F}^J, U_{J^c} \subset \mathbb{F}^{J^c}$ *not depending on* r *and some* $K_{J^c}(r) \subset \mathbb{F}^{J^c}$ *possibly depending on* r*, all of dimension at most* $k^{2c_2(|I|)}$.

Proof. By (4) in Lemma 6.2.14, for every $s \in Q'$ there exist subspaces $V_J(s) \subset \mathbb{F}^J$ for every $J \subset I^c$, $J \neq \emptyset$, with dimension at most $g_1 = G(d-1, |\mathbb{F}|^{-2^{3^{d+4}}C(\log 2^{d-1}/\delta)^4})$ such that the set of arrays $t \in \mathbb{F}^{I^c}$ with $t.q = 0$ for at least $(1-g_2)|Q_s|$ choices $q \in Q_s$ is contained in $\sum_{J \subset I^c, J \neq \emptyset} V_J(s) \otimes \mathbb{F}^{I^c \setminus J}$, where $g_2 = 2^{-3^{d+2}}$. Note, for future reference, that

$$\begin{aligned} g_1 &= G(d-1, |\mathbb{F}|^{-2^{3^{d+4}}C(\log 2^{d-1}/\delta)^4}) = ((\log|\mathbb{F}|)^2 c_1(d-1)2^{3^{d+4}}C(\log 2^{d-1}/\delta)^4)^{c_2(d-1)} \\ &\le ((\log|\mathbb{F}|)^2 c_1(d-1)2^{3^{d+4}}C(d\log 1/\delta)^4)^{c_2(d-1)} \\ &\le ((\log|\mathbb{F}|)^2 (c_1(d-1))^2(\log 1/\delta)^4)^{c_2(d-1)} \le G(d-1,\delta)^4 \le k^2. \end{aligned}$$

Let R consist of the set of arrays with $r \in W_{[d]} + \sum_{J \prec I}(W_J \otimes \mathbb{F}^{J^c} + \mathbb{F}^J \otimes W_{J^c}) + \sum_{J \succeq I} \mathbb{F}^J \otimes H_{J^c}(r)$ with $\dim(H_{J^c}(r)) \le k$ and the property that $r.q = 0$ for at least $(1-\frac{1}{4}(2^{-3^{d+2}})^2)|Q_I|$ choices $q \in Q_I$.

Let $r \in R$. Then by averaging and using (5) from Lemma 6.2.14, for at least $(1-g_3)|Q'|$ choices $s \in Q'$ we have $r.(s \otimes t) = 0$ for at least $(1-g_2)|Q_s|$ choices $t \in Q_s$, where $g_3 = \frac{1}{2}2^{-3^{d+2}}$. Thus, (noting that $r.(s \otimes t) = (rs).t$), $rs \in \sum_{J \subset I^c, J \neq \emptyset} V_J(s) \otimes \mathbb{F}^{I^c \setminus J}$ holds for at least $(1-g_3)|Q'|$ choices $s \in Q'$. Let $Q'(r)$ be the submultiset of Q' consisting of those $s \in Q'$ for which $rs \in \sum_{J \subset I^c, J \neq \emptyset} V_J(s) \otimes \mathbb{F}^{I^c \setminus J}$. Then we have $|Q'(r)| \ge (1-g_3)|Q'|$.

Note that we can write $r = r_1 + r_2 + r_3 + r_4$ where

$$r_1 \in \sum_{J \subset I, J \neq I, J \neq \emptyset} W_J \otimes \mathbb{F}^{J^c},$$

$$r_2 \in \sum_{J \prec I, J \not\subset I} (W_J \otimes \mathbb{F}^{J^c} + \mathbb{F}^J \otimes W_{J^c}) + \sum_{J \succ I} \mathbb{F}^J \otimes H_{J^c}(r),$$

$$r_3 \in W_{[d]} + \sum_{J \subset I, J \neq I, J \neq \emptyset} \mathbb{F}^J \otimes W_{J^c},$$

$$r_4 \in \mathbb{F}^I \otimes H_{I^c}(r).$$

By (1) in Lemma 6.2.14, the elements of Q' belong to $\bigcap_{J \subset I, J \neq I, J \neq \emptyset}(W_J^\perp \otimes \mathbb{F}^{I \setminus J})$, so we have $r_1 s = 0$ for every $s \in Q'$.

Note that for every pure tensor $s \in \mathbb{F}^I$, $r_2 s$ is $2^d k$-degenerate. Indeed, for any $J \subset [d-1]$ with $J \not\subset I$ there are some $s_1 \in \mathbb{F}^{I \cap J}, s_2 \in \mathbb{F}^{I \cap J^c}$ with $s = s_1 \otimes s_2$. Then $(W_J \otimes \mathbb{F}^{J^c})s \subset (W_J s_1) \otimes \mathbb{F}^{I^c \setminus J}$. Since $\dim(W_J s_1) \le k$, $J \not\subset I$ and $d \in I^c \setminus J$, any tensor in

$(W_J s_1) \otimes \mathbb{F}^{I^c \setminus J}$ is k-degenerate. Similarly, any tensor in $(\mathbb{F}^J \otimes W_{J^c})s$ or $(\mathbb{F}^J \otimes H_{J^c}(r))s$ is also k-degenerate, so $r_2 s$ is indeed $2^d k$-degenerate. Since Q' consists of pure tensors, this holds for every $s \in Q'$.

Also, $r_3 s \in \sum_{J \subset I, J \neq I}((\mathbb{F}^J \otimes W_{J^c})s)$. It follows that for every $s \in Q'(r)$, there exists some $t(s) \in V_{I^c}(s) + \sum_{J \subset I, J \neq I}((\mathbb{F}^J \otimes W_{J^c})s)$ such that $r_4 s - t(s)$ is g_4-degenerate for $g_4 = g_1 + 2^d k$ (we have used that $\dim(V_J(s)) \leq g_1$). To ease the notation, write $T(s)$ for the space $V_{I^c}(s) + \sum_{J \subset I, J \neq I}((\mathbb{F}^J \otimes W_{J^c})s)$. We claim that the dimension of $T(s)$ is at most $g_4 = g_1 + 2^d k$. Indeed, $\dim(V_{I^c}) \leq g_1$, so it suffices to prove that $\dim((\mathbb{F}^J \otimes W_{J^c})s) \leq k$ for every $J \subset I, J \neq I$. Since $s \in Q'$, s is a pure tensor, so for any such J we have $s = s_1 \otimes s_2$ for some $s_1 \in \mathbb{F}^J, s_2 \in \mathbb{F}^{I \setminus J}$. But then $(\mathbb{F}^J \otimes W_{J^c})s \subset W_{J^c} s_2$, which has dimension at most $\dim(W_{J^c}) \leq k$.

Let us define a sequence of subspaces $0 = Z(0) \subset Z(1) \subset \ldots \subset Z(m) \subset \mathbb{F}^{I^c}$ recursively as follows. Given $Z(j)$, if for all $r \in R$ we have that for all but at most $2g_3|Q'|$ choices $s \in Q'$ there is some $z \in Z(j)$ such that $r_4 s - z$ is $(g_4+1)g_4$-degenerate, then set $m = j$ and terminate.

Else, choose some $r \in R$ such that for at least $2g_3|Q'|$ choices $s \in Q'$ there is no $z \in Z(j)$ such that $r_4 s - z$ is $(g_4+1)g_4$-degenerate, and set $Z(j+1) = Z(j) + H_{I^c}(r)$. Recall that for every $s \in Q'(r)$, and in particular, for at least $(1-g_3)|Q'|$ choices $s \in Q'$, there exists some $t(s) \in T(s)$ such that $r_4 s - t(s)$ is g_4-degenerate. So for at least $g_3|Q'|$ choices $s \in Q'$ there is some $t(s) \in T(s)$ such that $r_4 s - t(s)$ is g_4-degenerate, but there is no $z \in Z(j)$ such that $r_4 s - z$ is $(g_4+1)g_4$-degenerate. In this case there is no $z \in Z(j)$ such that $z - t(s)$ is g_4^2-degenerate. On the other hand, since $r_4 s \in H_{I^c}(r) \subset Z(j+1)$, there is some $z \in Z(j+1)$ such that $z - t(s)$ is g_4-degenerate. For any i, let $K(i,s)$ be the subspace of $T(s)$ spanned by those $t \in T(s)$ for which there is some $z \in Z(i)$ with $z - t$ being g_4-degenerate. Since the dimension of $T(s)$ is at most g_4, we have $t(s) \notin K(j,s)$, else there would exist some $z \in Z(j)$ such that $z - t(s)$ is g_4^2-degenerate. On the other hand, $t(s) \in K(j+1,s)$. Thus, $\dim K(j+1,s) > \dim K(j,s)$. This holds for at least $g_3|Q'|$ choices $s \in Q'$, so

$$\sum_{s \in Q'} \dim K(j+1,s) \geq g_3|Q'| + \sum_{s \in Q'} \dim K(j,s).$$

Since $K(m,s) \subset T(s)$, we have $\dim K(m,s) \leq g_4$. Thus,

$$|Q'|g_4 \geq \sum_{s \in Q'} \dim K(m,s) \geq m g_3 |Q'|,$$

so $m \leq \frac{g_4}{g_3}$ and $\dim Z(m) \leq \frac{kg_4}{g_3}$. Write $Z = Z(m)$.

Now let $r \in R$. Let $X(r)$ be the set consisting of those $x \in H_{I^c}(r)$ for which there is some $z \in Z$ with $x - z$ being $(g_4+1)g_4$-degenerate. Then $r_4 s \in X(r)$ apart from at most $2g_3|Q'|$ choices $s \in Q'$. Let $t_1, \ldots, t_\alpha$ be a maximal linearly independent subset of $X(r)$ and extend it to a basis $t_1, \ldots, t_\alpha, t'_1, \ldots, t'_\beta$ for $H_{I^c}(r)$. Now if a linear combination of $t_1, \dots, t_\alpha, t'_1, \dots, t'_\beta$ is in $X(r)$, then the coefficients of $t'_1, \dots, t'_\beta$ are all zero. Write $r_4 = \sum_{i \le \alpha} s_i \otimes t_i + \sum_{j \le \beta} s'_j \otimes t'_j$ for some $s_i, s'_j \in \mathbb{F}^I$. Since $r_4 q \in X(r)$ for at least $(1 - 2g_3)|Q'| = (1 - 2^{-3^{d+2}})|Q'|$ choices $q \in Q'$, we have, for all j, that $s'_j.q = 0$ for at least $(1 - 2^{-3^{d+2}})|Q'|$ choices $q \in Q'$. Thus, by (2) in Lemma 6.2.14 there exist subspaces $L_J \subset \mathbb{F}^J$ ($J \subset I, J \neq \emptyset$) not depending on r, and of dimension at most $G(|I|, |\mathbb{F}|^{-2^{d+1}dk})$ such that $s'_j \in \sum_{J \subset I, J \neq \emptyset} L_J \otimes \mathbb{F}^{I \setminus J}$ for all j. Thus, $r_4 \in \sum_{i \le \alpha} s_i \otimes t_i + \sum_{J \subset I, J \neq \emptyset} L_J \otimes \mathbb{F}^{J^c}$. Moreover, for every $i \le \alpha$, we have $t_i \in X(r)$, so there exist $z_i \in Z$ such that $t_i - z_i$ is $(g_4 + 1)g_4$-degenerate. It follows that $r_4 \in \mathbb{F}^I \otimes Z + \sum_{J \supset I, J \neq I, J \subset [d-1]} \mathbb{F}^J \otimes K'_{J^c}(r) + \sum_{J \subset I, J \neq \emptyset} L_J \otimes \mathbb{F}^{J^c}$ for some $K'_{J^c}(r) \subset \mathbb{F}^{J^c}$ of dimension at most $\alpha \cdot (g_4 + 1)g_4 \le k \cdot (g_4 + 1)g_4$.

We claim that $\dim(Z), \dim(K'_{J^c})$ and $\dim(L_J)$ are all bounded by $k^{2c_2(|I|)} - k$.

Firstly, note that $g_4 = g_1 + 2^d k \le k^2 + 2^d k \le 2k^2$.

Now $\dim(K'_{J^c}) \le k(g_4 + 1)g_4 \le k^6 \le k^{2c_2(|I|)} - k$. Also, $\dim(Z) \le \frac{kg_4}{g_3} \le k^4 \le k^{2c_2(|I|)} - k$. Finally,

$$\begin{aligned}\dim(L_J) &\le G(|I|, |\mathbb{F}|^{-2^{d+1}dk}) = ((\log|\mathbb{F}|)^2 c_1(|I|)(2^{d+1}dk))^{c_2(|I|)} \\ &\le ((\log|\mathbb{F}|)^2 c_1(d-1)^2 k)^{c_2(|I|)} \le G(d-1, \delta)^2 k^{c_2(|I|)} \\ &\le k^{c_2(|I|)+1} \le k^{2c_2(|I|)} - k\end{aligned}$$

This completes the proof of the claim and the lemma. □

Proof of Lemma 6.2.2. As stated earlier, the proof goes by induction on d. For $d = 1$, by Lemma 6.2.1 there is a subspace $U \subset \mathbb{F}^{n_1}$ of codimension at most $C(\log 1/\delta)^4$ contained in $2\mathcal{B}' - 2\mathcal{B}'$. Choose $Q = U$. Now if $r.q = 0$ for at least $(1 - 2^{-3^4})|Q|$ choices $q \in Q$ then the same holds for all $q \in Q$, therefore $r \in U^\perp$, but $\dim(U^\perp) \le C(\log 1/\delta)^4$, so the case $d = 1$ is proved.

Now let us assume that $d \ge 2$. Extend the total order $\prec$ defined above such that it now contains $\emptyset$ which has $\emptyset \prec I$ for every non-empty $I \subset [d-1]$. Say $\emptyset = I_0 \prec I_1 \prec I_2 \prec \dots \prec I_{2^{d-1}-1}$ where $\{I_0, \dots, I_{2^{d-1}-1}\} = P([d-1])$.

Claim. For every $0 \le i \le 2^{d-1} - 1$ there exists a multiset Q_{I_i} of pure tensors with elements chosen from $2^{3^{d+3}}\mathcal{B}' - 2^{3^{d+3}}\mathcal{B}'$, and subspaces $W_{I_j}(i) \subset \mathbb{F}^{I_j}$, $W_{(I_j)^c}(i) \subset \mathbb{F}^{(I_j)^c}$ for every $j \le i$ (for $j = 0$, we only require $W_{[d]}(i)$ and not $W_\emptyset(i)$) with the following properties. The dimension of each of these spaces is at most $g_1(i) = G(d-1, \delta)^{\alpha(i)}$, where $\alpha(i) = 4 \cdot \Pi_{1 \le j \le i} 2c_2(|I_j|)$. Moreover, if $r \in \mathcal{G}$ has $r.q = 0$ for at least $(1 - \frac{1}{4}(2^{-3^{d+2}})^2)|Q_{I_j}|$ choices $q \in Q_{I_j}$ for all $j \le i$, then $r \in W_{[d]}(i) + \sum_{1 \le j \le i}(W_{I_j}(i) \otimes \mathbb{F}^{(I_j)^c} + \mathbb{F}^{I_j} \otimes W_{(I_j)^c}(i)) + \sum_{j > i} \mathbb{F}^{I_j} \otimes H_{(I_j)^c}(i, r)$ holds for some $H_{(I_j)^c}(i, r)$ possibly depending on r and of dimension at most $g_1(i)$.

Proof of Claim. This is proved by induction on i. For $i = 0$, by Lemma 6.2.12, there exist $Q_\emptyset \subset 2\mathcal{B}' - 2\mathcal{B}'$ consisting of pure tensors and $V_{[d]} \subset \mathbb{F}^{[d]}$ of dimension at most $4C(\log(2/\delta))^4 \le 4C(2\log(1/\delta))^4 \le G(d-1, \delta)^4$ such that if $r.q = 0$ for at least $\frac{7}{8}|Q_\emptyset|$ choices $q \in Q_\emptyset$, then r can be written as $r = x + y$ where $x \in V_{[d]}$ and y is g_2-degenerate for $g_2 = G(d-1, \frac{\delta}{4|\mathbb{F}|^{4C(\log 2/\delta)^4}})$. Since

$$g_2 \leq G(d-1, |\mathbb{F}|^{-5C(\log 2/\delta)^4}) = ((\log|\mathbb{F}|)^2 c_1(d-1)5C(\log(2/\delta))^4)^{c_2(d-1)}$$
$$\leq ((\log|\mathbb{F}|)^2 c_1(d-1)5C(2\log(1/\delta))^4)^{c_2(d-1)} \leq G(d-1,\delta)^4,$$

we can take $W_{[d]}(0) = V_{[d]}$.

Once we have found suitable sets $W_{I_j}(i-1)$ and $W_{(I_j)^c}(i-1)$ for all $j \leq i-1$, we can apply Lemmas 6.2.14 and 6.2.15 with $I = I_i$ and $k = g_1(i-1)$ to find a suitable Q_{I_i}, $W_{I_j}(i)$ and $W_{(I_j)^c}(i)$ for all $j \leq i$, and the claim is proved, since $g_1(i) = g_1(i-1)^{2c_2(|I_i|)}$.

Now, after taking several copies of each Q_I, we may assume that additionally $\max_I |Q_I| \leq 2\min_I |Q_I|$. Let $Q = \bigcup_{I\subset[d-1]} Q_I$ and suppose that $r.q = 0$ for at least $(1-2^{-3^{d+3}})|Q|$ choices $q \in Q$. Since $2^{-3^{d+3}} \leq \frac{1}{2\cdot 2^{d-1}} \cdot \frac{1}{4}(2^{-3^{d+2}})^2$, it follows that for every $I \subset [d-1]$ we have $r.q = 0$ for at least $(1-\frac{1}{4}(2^{-3^{d+2}})^2)|Q_I|$ choices $q \in Q_I$. By the Claim with $i = 2^{d-1}-1$, we get that $r \in \sum_{I\subset[d], I\neq\emptyset} V_I \otimes \mathbb{F}^{I^c}$ for some $V_I \subset \mathbb{F}^I$ not depending on r, and of dimension at most $g_1(2^{d-1}-1) = G(d-1,\delta)^{\alpha(2^{d-1}-1)}$. Note that

$$\alpha(2^{d-1}-1) = 4 \cdot 2^{2^{d-1}-1} \cdot \Pi_{1\leq i\leq d-1} c_2(i)^{\binom{d-1}{i}}.$$

But

$$\Pi_{1\leq i\leq d-1} c_2(i)^{\binom{d-1}{i}} = 4^{\sum_{1\leq i\leq d-1}\binom{d-1}{i} i^i} \leq 4^{\sum_{1\leq i\leq d-1}\binom{d-1}{i}(d-1)^i} \leq 4^{(d-1+1)^{d-1}} = 4^{d^{d-1}}.$$

Thus, $\alpha(2^{d-1}-1) \leq 4^{d^d}$. This completes the proof of the lemma. □

Chapter 7

Subsets of Cayley graphs that induce many edges

7.1 Introduction

The Unique Games Conjecture, formulated by Khot [83] in 2002, is a central conjecture in theoretical computer science. If true, it implies that for a wide class of natural problems it is NP-hard to find even a very crude approximate solution in polynomial time. Recently, a weakening of the conjecture known as the 2-to-2 Games Conjecture, where the approximation is required to be less crude (so it is easier to prove hardness) was proved by Khot, Minzer and Safra [84], a result that is considered as a major step towards the Unique Games Conjecture itself. More precisely, after work by various authors, the problem had been reduced to a statement about a certain Cayley graph, and Khot, Minzer and Safra proved that statement.

The Cayley graph Γ in question has as its vertex set the set of all $m \times n$ matrices over $\mathbb{F}_2$, with two vertices joined by an edge if their difference has rank 1. Let us say that a subset $\mathcal{A} \subset \mathrm{M}_{m,n}(\mathbb{F}_2)$ is *η-closed* if the probability that $A + B \in \mathcal{A}$, when A is chosen uniformly from $\mathcal{A}$ and B is chosen uniformly from all rank-1 matrices, is at least η. In graph terms, this is the probability that a random neighbour of a random point in $\mathcal{A}$ is itself in $\mathcal{A}$.

A simple example of an η-closed set is the set $\{A \in \mathrm{M}_{m,n}(\mathbb{F}_2) : Ax = y\}$, for some pair of vectors $x \in \mathbb{F}_2^n, y \in \mathbb{F}_2^m$. Indeed, if $Ax = y$ and B is a random matrix of rank 1, then $x \in \ker B$ with probability roughly $1/2$. But if $x \in \ker B$, then $(A+B)x = y$, so $A+B \in \mathcal{A}$ as well. A very similar, but distinct, example is the set $\{A \in \mathrm{M}_{m,n}(\mathbb{F}_2) : A^T x = y\}$. Let us call sets of one of these two kinds *basic sets*.

We can form further examples by taking intersections of a small number of basic sets. For example, if $x_1, \dots, x_k$ are linearly independent and we take a set of the form

$$\{A \in \mathrm{M}_{m,n}(\mathbb{F}_2) : Ax_1 = y_1, \dots, Ax_k = y_k\},$$

then with probability approximately 2^{-k} each x_i belongs to $\ker B$, so for any A in the set, $A+B$ belongs to the set with probability approximately 2^{-k}. The result of Khot, Minzer and Safra is the following.

Theorem 7.1.1. *For every $\eta > 0$ there exist $\delta > 0$ and a positive integer k such that if $\mathcal{A}$ is any η-closed subset of $\mathrm{M}_{m,n}(\mathbb{F}_2)$, then there exists an intersection $\mathcal{C}$ of at most k basic sets such that $|\mathcal{A} \cap \mathcal{C}| \geq \delta|\mathcal{C}|$ and $\mathcal{C} \neq \emptyset$.*

In other words, every closed set is dense inside some intersection of a small number of basic sets.

It is well known and not hard to see that this in fact leads to a characterization (at least qualitatively) of closed sets. Indeed, observe first that if $\mathcal{A}$ is η-closed, then the subgraph induced by $\mathcal{A}$ has average degree at least $\eta|\mathcal{B}|$, where $\mathcal{B}$ is the set of rank-1 matrices, and maximal degree at most $|\mathcal{B}|$. Therefore, any subset of $\mathcal{A}$ of size at least $(1-\eta/4)|\mathcal{A}|$ has average degree at least $\eta|\mathcal{B}|/2$. It follows from this observation and Theorem 7.1.1 that we can find disjoint subsets $\mathcal{A}_1, \dots, \mathcal{A}_r$ of $\mathcal{A}$, subsets $\mathcal{C}_1, \dots, \mathcal{C}_r$ of $\mathrm{M}_{m,n}(\mathbb{F}_2)$, a positive real number $\delta = \delta(\eta)$ and a positive integer $k = k(\eta)$ with the following properties.

1. The sets $\mathcal{A}_i$ are disjoint.
2. Each $\mathcal{C}_i$ is an intersection of at most k basic sets.
3. For each i, $|\mathcal{A}_i \cap \mathcal{C}_i| \geq \delta|\mathcal{C}_i|$.
4. $|\bigcup_i \mathcal{A}_i| \geq \eta|\mathcal{A}|/4$.

Conversely, if such sets exist, then the probability that a random matrix $A \in \mathcal{A}$ belongs to some $\mathcal{A}_i$ is at least $\eta/4$. If it belongs to $\mathcal{A}_i$, then we can use the following lemma. We write $u \otimes v$ for the rank-1 matrix M with $M_{ij} = u_i v_j$, which sends a vector x to the vector $\langle x, v\rangle u$. Note also that $(u \otimes v)^T$ sends x to $\langle x, u\rangle v$.

Lemma 7.1.2. *Let $\mathcal{C}$ be an intersection of at most k basic sets and let $\mathcal{A} \subset \mathcal{C}$ be a subset of relative density at least δ. Then $\mathcal{A}$ is $2^{-k}(\delta - 2^{-(m-k)})$-closed.*

Proof. Let us set $\mathcal{C}(x,y) = \{A \in \mathrm{M}_{m,n}(\mathbb{F}_2) : Ax = y\}$, and $\mathcal{C}'(x,y) = \{A \in \mathrm{M}_{m,n}(\mathbb{F}_2) : A^T x = y\}$. Let $x_1, \dots, x_k, y_1, \dots, y_k$ be non-zero vectors such that $\mathcal{C} = \bigcap_{i=1}^r \mathcal{C}(x_i, y_i) \cap \bigcap_{i=r+1}^k \mathcal{C}'(x_i, y_i)$.

Let $u \otimes v$ be a rank-1 matrix. If there exists $i \leq r$ such that $\langle x_i, v\rangle \neq 0$, then $(A + u \otimes v)(x_i) = y_i + u$, so $A + u \otimes v \notin \mathcal{C}(x_i, y_i)$ and hence $A + u \otimes v \notin \mathcal{C}$. Similarly, if there exists $i > r$ such that $\langle x_i, u\rangle \neq 0$, then $(A + u \otimes v)^T(x_i) = y_i + v$ and again $A + u \otimes v \notin \mathcal{C}$.

We shall now bound from below the probability that $A + u \otimes v \in \mathcal{A}$ given that $A \in \mathcal{A}$ and that $\langle x_i, v\rangle = 0$ for every $i \leq r$ and $\langle x_i, u\rangle = 0$ for every $r < i \leq k$, noting that the condition on $u \otimes v$ states that $(u,v) \in U \times V$ for a pair of subspaces U and V with codimensions that add up to at most k, a condition that occurs with probability 2^{-k}.

Let us now condition further on the choice of $v \in V$. That means that we fix v, choose a random $u \in U$, and add $u \otimes v$ to A. If we allow u to take the value 0, then the resulting matrix is uniformly distributed in the affine subspace $A + U \otimes v$, so the probability that it is in $\mathcal{A}$ is equal to the relative density of $\mathcal{A}$ inside this affine subspace.

The translates of $U \otimes v$ by matrices in $\mathcal{C}$ partition $\mathcal{C}$. Let us write them as $\mathcal{W}_1, \dots, \mathcal{W}_s$, and let the relative density of $\mathcal{A}$ inside $\mathcal{W}_i$ be δ_i. Then, still fixing v, we have that

$$\mathbb{P}[A + u \otimes v \in \mathcal{A}] = \sum_i \mathbb{P}[A \in \mathcal{W}_i]\, \mathbb{P}[A + u \otimes v \in \mathcal{A} | A \in \mathcal{W}_i] = \sum_i \frac{\delta_i^2}{s\delta} \geq \delta.$$

This statement is true regardless of v, so we deduce that the probability that $A + u \otimes v \in \mathcal{A}$ given that $A \in \mathcal{A}$ and $(u, v) \in U \times (V \setminus \{0\})$ is at least δ. If we now insist that $u \neq 0$, we reduce this probability by at most $2^{-(m-k)}$, so the result is proved. $\square$

Let $B \in \mathcal{B}$ be chosen uniformly at random. Given the lemma above, applied to the sets $\mathcal{A}_i$ and $\mathcal{C}_i$, we deduce that the conditional probability that $A + B \in \mathcal{A}_i$ given that $A \in \mathcal{A}_i$ is at least some $c(\delta, k) > 0$, and from that it follows that $\mathcal{A}$ is $c(\delta, k)\eta/4$-closed.

Thus, a set $\mathcal{A}$ is η-closed for some not too small η if and only if an appreciable fraction of $\mathcal{A}$ is efficiently covered by disjoint intersections of few basic sets.

Barak, Kothari and Steurer suggest in [7] that establishing a higher dimensional analogue of Theorem 7.1.1 may be a useful step in obtaining a proof of the full Unique Games Conjecture, though they do not actually provide a formal reduction. The main purpose of this chapter is to formulate a suitable conjecture and prove some partial results towards it. We say that $\mathcal{A} \subset \mathbb{F}_2^{n_1} \otimes \cdots \otimes \mathbb{F}_2^{n_d}$ is η-closed if with probability at least η, we have $A + u_1 \otimes \cdots \otimes u_d \in \mathcal{A}$, when $A \in \mathcal{A}$ and vectors $u_i \in \mathbb{F}_2^{n_i} \setminus \{0\}$ are chosen independently and uniformly at random.

Problem 7.1.3. Give a qualitative description of η-closed sets $\mathcal{A} \subset \mathbb{F}_2^{n_1} \otimes \cdots \otimes \mathbb{F}_2^{n_d}$.

To see that this is indeed a generalization of the problem about matrices considered above, we identify $\mathrm{M}_{m,n}(\mathbb{F}_2)$ with $\mathbb{F}_2^m \otimes \mathbb{F}_2^n$ in the usual way, which leads to a slight reformulation of Theorem 7.1.1 in terms of tensor products. Note first that under this identification, the set $\{M \in \mathrm{M}_{m,n}(\mathbb{F}_2) : Mx_1 = \cdots = Mx_a = M^T y_1 = \cdots = M^T y_b = 0\}$ becomes the set $H \otimes K$, where $H = \langle y_1, \dots, y_b \rangle^\perp$ and $K = \langle x_1, \dots, x_a \rangle^\perp$. It follows that an intersection of at most k basic sets is either empty or a translate of $H \otimes K$ for some pair of subspaces $H \subset \mathbb{F}_2^m, K \subset \mathbb{F}_2^n$ with $\operatorname{codim}(H) + \operatorname{codim}(K) \leq k$.

In the higher-dimensional case, there is a richer class of sets $\mathcal{A} \subset \mathbb{F}_2^{n_1} \otimes \cdots \otimes \mathbb{F}_2^{n_d}$ that are η-closed. To describe them, we recall the notation defined in Subsection 6.2.1: given a non-empty subset $I \subset [d]$, we write $\mathbb{F}_2^I$ for $\bigotimes_{i \in I} \mathbb{F}_2^{n_i}$.

Now we say that $\mathcal{C}$ is *k-simple* if there exists a collection of subspaces $H_I \subset \mathbb{F}_2^I$ of codimension at most k, one for each non-empty subset $I \subset [d]$, such that $\mathcal{C}$ is a translate of the set $\bigcap_{I \subset [d], I \neq \emptyset} (H_I \otimes \mathbb{F}_2^{I^c})$ (where $H_{[d]} \otimes \mathbb{F}_2^{\emptyset}$ is just $H_{[d]}$), which is a subspace of $\mathbb{F}_2^{n_1} \otimes \cdots \otimes \mathbb{F}_2^{n_d}$. It is not hard to see that this subspace contains at least a proportion

$c(d,k) > 0$ of all rank-1 tensors $u_1 \otimes \cdots \otimes u_d$ (provided that $n_1, \ldots, n_d$ are sufficiently large), so it is $c(d,k)$-closed. It follows that any translate of it is $c(d,k)$-closed too.

We now make the following conjecture.

Conjecture 7.1.4. *For any $\eta > 0$ and any positive integer d, there exist $k = k(d,\eta)$ and $\rho = \rho(d,\eta) > 0$ with the following property. Let $\mathcal{A} \subset \mathbb{F}_2^{n_1} \otimes \cdots \otimes \mathbb{F}_2^{n_d}$ be η-closed. Then there is a k-simple set $\mathcal{C}$ such that $|\mathcal{A} \cap \mathcal{C}| \geq \rho|\mathcal{C}|$.*

Note that in the $d = 2$ case, we allow translates of sets $(H_{\{1\}} \otimes H_{\{2\}}) \cap H_{\{1,2\}}$ rather than just translates of $H_{\{1\}} \otimes H_{\{2\}}$, so Conjecture 7.1.4 might seem to be weaker than Theorem 7.1.1. However, this actually makes no difference, since when intersecting with $H_{\{1,2\}}$, the cardinality of the set drops by a factor at most 2^k.

The main result of this chapter, stated later in this section, is a proof of Conjecture 7.1.4 in an important special case.

7.1.1 What can be said about more general Cayley graphs?

It is tempting to try to prove Conjecture 7.1.4 by identifying and proving a statement that applies to a much wider class of Cayley graphs, of which Conjecture 7.1.4 would be a special case. We would begin with an Abelian (or even non-Abelian) group G and a pair of subsets $A, B \subset G$, where we think of B as the set of generators, satisfying the hypothesis that $|\{(a,b) \in A \times B : a + b \in A\}| \geq \eta|A||B|$. We shall say in this situation that A is (B,η)-*closed* (in G).

Another way of writing the condition is

$$\langle \mathbb{1}_A * \mu_B, \mathbb{1}_A \rangle \geq \eta\alpha,$$

where α is the density of A, μ_B is the characteristic measure of B (that is, the function that takes the value $|G|/|B|$ on B and 0 elsewhere) and we define $f * g(x)$ to be $\mathbb{E}_{y+z=x} f(y)g(z)$. By the Cauchy-Schwarz inequality the left-hand side is at most $\|\mathbb{1}_A * \mu_B\|_2 \|\mathbb{1}_A\|_2 = \|\mathbb{1}_A * \mu_B\|_2 \alpha^{1/2}$, where inner products and L_p norms are defined using expectations, so our hypothesis implies that $\|\mathbb{1}_A * \mu_B\|_2^2 \geq \eta^2\alpha$. It is easy to see that this "mixed energy" $\|\mathbb{1}_A * \mu_B\|_2^2$ can be at most α, with equality if and only if $a + b - b' \in A$ for every $a \in A, b, b' \in B$.

At this point let us recall the so-called asymmetric Balog-Szemerédi-Gowers theorem, which can be found in [108] as Theorem 2.35. (For a useful alternative presentation of the theorem, see also [55].) The main assumption of the theorem is that A, B are two finite subsets of an Abelian group, with densities α and β, such that $\|\mathbb{1}_A * \mathbb{1}_B\|_2^2 \geq \eta\alpha\beta^2$ (which is equivalent to saying that $\|\mathbb{1}_A * \mu_B\|_2^2 \geq \eta\alpha$), but there is also an assumption that A is not too much bigger than B. The precise statement is as follows.

Theorem 7.1.5. *For every $\varepsilon > 0$ there exists a constant $C = C(\varepsilon)$ with the following property. Let G be a finite Abelian group, let $L \geq 1$, let $0 < \eta \leq 1$ and let A and B be* *finite subsets of G with densities α and β, such that $\alpha \leq L\beta$ and $\|\mathbb{1}_A * \mathbb{1}_B\|_2^2 \geq 2\eta\alpha\beta^2$. Then there exist a subset $H \subset G$ such that $|H+H| \leq C\eta^{-C}L^{\varepsilon}|H|$, a subset $X \subset G$ of size at most $C\eta^{-C}L^{\varepsilon}|A|/|H|$ such that $|A \cap (X+H)| \geq C^{-1}\eta^{C}L^{-\varepsilon}|A|$, and some $x \in G$ such that $|B \cap (x+H)| \geq C^{-1}\eta^{C}L^{-\varepsilon}|B|$.*

More qualitatively speaking, if A is not too much larger than B and $\|\mathbb{1}_A * \mathbb{1}_B\|_2^2$ is within a constant of its largest possible value, then there is a set H of small doubling such that a small number of translates of H cover a substantial proportion of A, and some translate of H covers a substantial proportion of B. It is not hard to see that the converse holds as well.

This theorem cannot be used to prove Conjecture 7.1.4 because of the condition that $\alpha \leq L\beta$, which does not apply here since the set $\mathcal{A}$ in Conjecture 7.1.4 can be much bigger than the set $\mathcal{B}$. That raises the following question, which generalizes Problem 7.1.3.

Question 7.1.6. *Let G be a finite Abelian group, let $\eta > 0$, and let $A, B \subset G$ be subsets such that A is (B,η)-closed in G. What can be said about A, B and the relationship between them?*

A similar question can of course be asked with the slightly weaker hypothesis that $\|\mathbb{1}_A * \mathbb{1}_B\|_2^2 \geq \eta^2\alpha\beta^2$, but we shall concentrate on the question as stated, since it is more closely related to Conjecture 7.1.4.

An immediate observation is that we cannot hope to say anything interesting about the structure of B, even if $\eta = 1$. For example, $\eta = 1$ if $A = G$ and B is an arbitrary subset of G. For a more general example, one can let A be an arbitrary union of cosets of some subgroup H and let B be an arbitrary subset of H. For a slightly different example, let $G = \mathbb{F}_2^n$, let B be the set $\{e_1, \ldots, e_n\}$ of standard basis vectors, and let A be a union of $n/3$-dimensional affine subspaces V_i, such that each V_i is a random translate of the subspace generated by $n/3$ randomly chosen e_j. Then if $x \in V_i$ and $b \in B$, the probability that $x + b \in V_i$ is $1/3$, so A is $1/3$-closed.

Any general statement will have to be weak enough to allow for examples like these. The last example shows that we cannot hope to find a single set H of small doubling and cover a large portion of A efficiently with translates of H, unless H is of constant size, in which case the conclusion becomes trivial. To sketch briefly why not, observe first that by Freiman's theorem we can assume that H is a subspace. Next, note that for each vector x, the probability that it belongs to the span of a random $n/3$ standard basis vectors is exponentially small in the size of the support of x. We can also use the following simple lemma.

Lemma 7.1.7. *Let H be a subspace of dimension d and let $k \leq d$. Then the number of vectors in H of support size at most k is at most the number of vectors of support size at most k in the subspace generated by the first d standard basis vectors $e_1, \ldots, e_d$, namely $\sum_{i=0}^{k} \binom{d}{i}$.*

Proof. Let $u_1, \ldots, u_d$ be a basis for d. By Gaussian elimination, we can convert $u_1, \ldots, u_d$ into a basis $v_1, \ldots, v_d$ and find coordinates $t_1, \ldots, t_d$ such that $v_i(t_j) = \delta_{ij}$. Then the support size of $\sum_i \lambda_i v_i$ is at least the number of non-zero λ_i, which proves the result. □

When d is large, it follows that the proportion of vectors in H of small support is very small. Combining these observations, one can show that for every η there exists d such that if H is a d-dimensional subspace, then the probability that a random subspace V of dimension $n/3$ is $(H, \eta/2)$-closed is at most $\eta/2$. This in turn can be used to prove that with high probability the set A described above (for a suitable number of V_i) is not (H, η)-closed for any H of dimension d or above.

However, these examples do not rule out a weakening along the following lines.

Question 7.1.8. *Let G be a finite Abelian group, let $\eta > 0$, and let A, B be subsets of G such that A is (B, η)-closed in G. Does it follow that A has a non-empty subset A' such that A' is (B, η')-closed in G, and $|A' + A'| \le C|A'|$, where $\eta' > 0$ and C are constants that depend on η only?*

An argument similar to the one we mentioned just after the statement of Theorem 7.1.1 shows that if the answer is yes, then we can find a collection of disjoint subsets $A_1, \ldots, A_m$ that cover a substantial proportion of A, each one with small doubling and each one (B, η')-closed (with a slightly smaller η'). Thus, we would be able to obtain a conclusion similar to that of Theorem 7.1.5 but without the requirement that the structured sets are all translates of one another.

A positive answer would also imply Conjecture 7.1.4. Indeed, by Freiman's theorem A_i is contained in a subspace V_i not much larger than A_i. This reduces the conjecture to the case where $\mathcal{A}$ is a subspace. In that case, a very simple corollary of our main result, Corollary 7.1.12 (stated later) proves the conjecture.

However, the answer to Question 7.1.8 is easily seen to be negative (which implies that it is also negative if we assume the weaker mixed-energy hypothesis instead). The example we are about to give was communicated to us privately by Boaz Barak as a counterexample to a related but slightly different statement.

For convenience let n be odd, let $A \subset \mathbb{F}_2^n$ be the set of all vectors with $(n \pm 1)/2$ coordinates equal to 1, and let B be the set of standard basis vectors. Then it is easy to see that A is (B, η)-closed for $\eta = (n + 1)/2n \approx 1/2$. Suppose now that we could find a subset $A' \subset A$ such that $|A' + A'| \le C|A'|$, and A' is (B, η')-closed. By Freiman's theorem, A' is contained in a subspace V that is not much bigger than A', which implies that V is (B, c)-closed for some positive constant $c = c(\eta)$. That implies that at least cn of the standard basis vectors belong to V. Let W be the subspace spanned by these basis vectors. The maximum number of elements of A that can belong to a translate $x + W$ of W is $2(cn)^{-1/2}|W|$, and therefore $|A'| \le 2(cn)^{-1/2}|V|$. This contradicts the fact that V is not much bigger than A'.

In this chapter we formulate a yet weaker conjecture and prove that it still implies Conjecture 7.1.4. Unfortunately, we also give a counterexample to the weaker conjecture.

The counterexample does not make the implication vacuous, however, because the implication depends on a non-trivial theorem that is true and of some interest: it is just that for a general Cayley graph (on a finite Abelian group) one cannot deduce the hypotheses of the theorem from the assumption that a set is η-closed. It is conceivable that one might be able to prove Conjecture 7.1.4 (and thereby also give a different proof of the theorem of Khot, Minzer and Safra) by using additional properties of the particular Cayley graph that that conjecture is about.

How, then, might one try to find a conjecture that would not be contradicted by the "two-layers" example just discussed? One observation that suggests a possible way forward is the following. Suppose that we extend the set by adding a few more layers. If, say, we take not just the middle two layers but the middle ε^{-1} layers (or thereabouts), then we obtain a new set inside which the first set has relative density approximately 2ε, and this new set is $(1-2\varepsilon)$-closed, since a random element of the set will be in one of the interior layers with probability approximately (and in fact slightly bigger than) $1-2\varepsilon$, and adding an arbitrary basis vector to such an element will give another element of the set.

So perhaps we could hope that if A is (B,η)-closed, then there is a set C that is $(B,1-\varepsilon)$-closed such that $|A\cap C|\geq\delta|C|$ for some δ that depends on η and ε only.

However, simple modifications of the example show that this is too much to ask. For instance, we can take as our set A the set of all $x\in\mathbb{F}_2^n$ such that m or $m+1$ coordinates are equal to 1 and all but the first $2m$ coordinates are zero. If m is around $n/4$, say, then the resulting set is $(B,1/4)$-closed, but there is no prospect of A living densely in a set that is almost perfectly closed, because of the need to add basis vectors corresponding to coordinates beyond $2m$.

A further example to consider is the set of all $x\in\mathbb{F}_2^n$ such that at most $n/3$ coordinates are equal to 1. This set is $(B,1/3)$-closed (at least – in fact it is more like $(B,2/3)$-closed because the probability that a random element of the set has exactly $\lfloor n/3\rfloor$ coordinates equal to 1 is approximately 1/2), but for similar reasons to the previous example, one cannot find an almost perfectly closed set with a significant proportion of its elements in the set.

However, the picture changes if we ask for slightly less. Let us informally call a set C *good* if there is a proportional-sized subset $B'\subset B$ such that C is $(B',1-\varepsilon)$-good for some small constant ε. Thus, now we ask only that C should be almost closed for a *large subset* of B rather than for the whole set.

It is not immediately clear how to use this definition, because the statement that $|A\cap C|\geq\delta|C|$ for a good set C can be true for uninteresting reasons. For example, we could take C to be the union of a subspace V generated by $n/5$ basis vectors together with an arbitrary subset of A of cardinality $2\delta|V|$. To remedy this, we insist that C is "related to A" in the graph in a different sense from that of A being dense in C.

Here, then, is a question that replaces Question 7.1.8.

Question 7.1.9. *Is it true that for every $\eta, \varepsilon > 0$ there exist $c > 0, \delta > 0$ and positive integer l with the following property? Let G be a finite Abelian group and let $A, B \subset G$ be subsets such that A is (B, η)-closed. Then there is a subset $B' \subset B$ and a non-empty subset $C \subset G$ with the following properties.*

1. $|B'| \geq \delta |B|$.

2. *C is $(B', 1 - \varepsilon)$-closed.*

3. $C \subset \big\{x \in G : \mathbb{1}_A * \mu_B * \cdots * \mu_B * \mu_{-B} * \cdots * \mu_{-B}(x) \geq c\big\}$ *where the number of μ_Bs and μ_{-B}s in the convolution is l.*

Condition (3) is saying that for any $x \in C$, the probability that $x - b_1 - \cdots - b_l + b_{l+1} + \cdots + b_{2l} \in A$, when the b_i are chosen uniformly and independently at random from B, is at least c. When the group G is $\mathbb{F}_2^n$ for some n, we can and will simplify it, since $B = -B$.

To see that this question improves on Question 7.1.8, let us consider the two problematic examples for that question. If m is odd and $A \subset \mathbb{F}_2^n$ consists of all sequences with $(m \pm 1)/2$ 1s and with no 1s after the mth coordinate, then let C be the set of all sequences with no 1s after the mth coordinate that have between $(m-1)/2 - \varepsilon^{-1}$ and $(m+1)/2 + \varepsilon^{-1}$ 1s. If $l = \varepsilon^{-1}$, then for any $x \in C$, the probability that $x - b_1 - \cdots - b_l \in A$ is at least $(\frac{m}{2n})^l$. (This is because conditional on $b_i \in \{e_1, \ldots, e_m\}$ this probability is at least $\frac{1}{2^l}$.) Moreover, if $B' = \{e_1, \ldots, e_m\}$, then for every $b \in B'$ and every $c \in C$ that is not on the boundary (in the obvious sense), we have that $b + c \in C$, so C is $(B', 1 - \varepsilon)$-closed.

Now let us look at the example where A is the set of all sequences with at most $n/3$ 1s. This time let C be the set of all sequences that are 0 after the first $2n/3$ coordinates and have at most $n/3 + \varepsilon^{-1}$ 1s, and let $B' = \{e_1, \ldots, e_{2n/3}\}$. Then for any $x \in C$, the probability that $x - b_1 - \cdots - b_l \in A$ is at least $(\frac{1}{3})^l$, where $l = \varepsilon^{-1}$. Moreover, C is $(1 - \varepsilon)$-closed, again because adding an element of B' to a non-boundary element of C gives an element of C.

7.1.2 Our main result

Let us now see why a positive answer to Question 7.1.9 would imply Conjecture 7.1.4. The deduction will be easy once we have established the following theorem, which is the main result of this chapter. Similarly to Chapter 6, here $\mathcal{G}$ denotes $\mathbb{F}_2^{n_1} \otimes \cdots \otimes \mathbb{F}_2^{n_d}$ and $\mathcal{B}$ denotes the multiset $\{u_1 \otimes \cdots \otimes u_d : u_i \in \mathbb{F}_2^{n_i} \text{ for all } i\}$. Note that the notion of (B, η)-closedness can be generalized in an obvious way to multisets.

Theorem 7.1.10. *For any $\theta > 0$, there exists $\varepsilon = \varepsilon(d, \theta) > 0$ with the following property. Let $\delta > 0$. Then there exists a positive integer $k = k(d, \delta)$ with the following property. For any $\mathcal{B}' \subset \mathcal{B}$ with $|\mathcal{B}'| \geq \delta |\mathcal{B}|$ and any $\mathcal{A} \subset \mathcal{G}$ which is $(\mathcal{B}', 1 - \varepsilon)$-closed, there exists a k-simple set $\mathcal{D} \subset \mathcal{G}$ such that $|\mathcal{D} \cap \mathcal{A}| \geq (1 - \theta)|\mathcal{D}|$.*

We remark that in the case where $\mathcal{B}' = \mathcal{B}$ the proof is easy, and $\mathcal{D}$ can be chosen to be the whole of $\mathcal{G}$.

It is convenient to state the following corollary separately, which follows from Theorem 7.1.10 by taking $\theta = 1/2$.

Corollary 7.1.11. *There exists $\varepsilon = \varepsilon(d) > 0$ such that for any $\delta > 0$, there exists a positive integer $k = k(d, \delta)$ with the following property. For any $\mathcal{B}' \subset \mathcal{B}$ with $|\mathcal{B}'| \geq \delta|\mathcal{B}|$ and any $\mathcal{A} \subset \mathcal{G}$ which is $(\mathcal{B}', 1-\varepsilon)$-closed, there exists a k-simple set $\mathcal{D} \subset \mathcal{G}$ which has $|\mathcal{D} \cap \mathcal{A}| \geq \frac{1}{2}|\mathcal{D}|$.*

Let us see why Conjecture 7.1.4 follows from Corollary 7.1.11 and a positive answer to Question 7.1.9 in the case of the group $\mathcal{G}$ and the subset $B \subset \mathcal{G}$ of rank-1 tensors. Let $\eta > 0$. Pick $\varepsilon = \varepsilon(d)$ so that the conclusion of Corollary 7.1.11 holds. If the answer to Question 7.1.9 is positive for $\mathcal{G}$ and B, then we can choose $c > 0, \delta > 0$, and a positive integer l such that the conclusion of the question is true. Now let $\mathcal{A} \subset \mathcal{G}$ be η-closed. This is saying that $\mathcal{A}$ is (B, η)-closed. By the conclusion of Question 7.1.9, there exist a set $B' \subset B$ with $|B'| \geq \delta|B|$, and a non-empty subset $\mathcal{C} \subset \mathcal{G}$ such that $\mathcal{C}$ is $(B', 1-\varepsilon)$-closed and $\mathcal{C} \subset \{x \in \mathcal{G} : \mathbb{1}_{\mathcal{A}} * \mu_B * \cdots * \mu_B(x) \geq c\}$, where the number of μ_Bs in the convolution is l. Define $\mathcal{B}'$ to be the multiset that consists of the set B' together with the multiset of all $u_1 \otimes \cdots \otimes u_d$ with $u_i \in \mathbb{F}_2^{n_i}$ for each i and with at least one u_i equal to 0. Note that $|\mathcal{B}'| \geq \delta|\mathcal{B}|$ and $\mathcal{C}$ is $(\mathcal{B}', 1-\varepsilon)$-closed. By Corollary 7.1.11, there exists a k-simple set $\mathcal{D} \subset \mathcal{G}$, for some $k = k(d, \delta)$, which has $|\mathcal{D} \cap \mathcal{C}| \geq \frac{1}{2}|\mathcal{D}|$. Now pick $x \in \mathcal{D}$ and $b_1, \dots, b_l \in B$ uniformly and independently at random. The probability that $x - b_1 - \cdots - b_l \in \mathcal{A}$ is at least $c/2$. Therefore, there exists some $y \in \mathcal{G}$ such that when $x \in \mathcal{D}$ is randomly chosen, the probability that $x - y \in \mathcal{A}$ is at least $c/2$. That is, $|(\mathcal{D} - y) \cap \mathcal{A}| \geq \frac{1}{2}c|\mathcal{D}| = \frac{1}{2}c|\mathcal{D} - y|$. But $\mathcal{D} - y$ is a k-simple set, which finishes the proof of Conjecture 7.1.4.

Another simple corollary of Theorem 7.1.10 is the following result, which is Conjecture 7.1.4 in the case where $\mathcal{A}$ is a subspace.

Corollary 7.1.12. *For any $\eta > 0$ and any positive integer d, there exists some $k = k(d, \eta)$ with the following property. If $V \subset \mathcal{G}$ is a subspace which is η-closed, then V contains a k-simple subspace. That is, $V \supset \bigcap_{I \subset [d], I \neq \emptyset} (H_I \otimes \mathbb{F}_2^{I^c})$ for some collection of subspaces $H_I \subset \mathbb{F}_2^I$ of codimension at most k.*

Proof. Since V is a vector space, the condition that V is η-closed says that $u_1 \otimes \cdots \otimes u_d \in V$ for at least a proportion of η of all rank-1 tensors $u_1 \otimes \cdots \otimes u_d$. Thus, there exists some $\mathcal{B}' \subset \mathcal{B}$ with $|\mathcal{B}'| \geq \eta|\mathcal{B}|$ such that V is $(\mathcal{B}', 1)$-closed. Taking θ sufficiently close to 0 in Theorem 7.1.10, it follows that $V \supset \mathcal{D}$ for a k-simple set $\mathcal{D}$, where k depends only on d and η. Then $\mathcal{D}$ is a translate of $\bigcap_{I \subset [d], I \neq \emptyset} (H_I \otimes \mathbb{F}_2^{I^c})$ for some $H_I \subset \mathbb{F}_2^I$ of codimension at most k. Since V is a vector space, it follows that $V \supset \bigcap_{I \subset [d], I \neq \emptyset} (H_I \otimes \mathbb{F}_2^{I^c})$. □

In the next section, we shall prove Theorem 7.1.10. In the last section, we show that the answer to Question 7.1.9 is negative.

7.2 The proof of Theorem 7.1.10

In this section we shall use the notation introduced in Subsection 6.2.1. That is, $\mathcal{G} = \mathbb{F}_2^{n_1} \otimes \cdots \otimes \mathbb{F}_2^{n_d}$ is viewed as the set of d-dimensional $(n_1, \ldots, n_d)$-arrays over $\mathbb{F}_2$ which in turn is viewed as $\mathbb{F}_2^{n_1 n_2 \ldots n_d}$, equipped with the entry-wise dot product.

In order to prove Theorem 7.1.10, we recall Lemma 6.2.2 from Chapter 6. Qualitatively (and taking $\mathbb{F} = \mathbb{F}_2$), that lemma states the following.

Lemma 7.2.1. *For all $\delta > 0$ and $d \in \mathbb{N}$ there exist $f_1(d), f_3(d) > 0$ and $f_2(d, \delta) \in \mathbb{N}$ with the following property. Let $\mathcal{B}' \subset \mathcal{B}$ be a multiset such that $|\mathcal{B}'| \geq \delta|\mathcal{B}|$. Then there exists a multiset Q whose elements are chosen from $f_1(d)\mathcal{B}'$ (but with arbitrary multiplicity) with the following property. The set of arrays $r \in \mathcal{G}$ with $r.q = 0$ for at least $(1 - f_3(d))|Q|$ choices $q \in Q$ is contained in $\sum_{I \subset [d], I \neq \emptyset} V_I \otimes \mathbb{F}_2^{I^c}$ for a collection of subspaces $V_I \subset \mathbb{F}_2^I$ that each have dimension at most $f_2(d, \delta)$.*

In order to deduce Theorem 7.1.10 from this lemma, we shall use Fourier analysis. Recall that if A is a subset of $\mathcal{G}$ of density α, then by Parseval's identity we have $\alpha = \sum_r |\widehat{\mathbb{1}_A}(r)|^2$. Also, if B is a multiset in $\mathcal{G}$, then by Parseval's identity and the convolution law, $\langle \mathbb{1}_A * \mu_B, \mathbb{1}_A \rangle = \sum_r |\widehat{\mathbb{1}_A}(r)|^2 \widehat{\mu}_B(r)$ (for a multiset B, we define $\mu_B(x) = \frac{|G|}{|B|} B(x)$ where $B(x)$ is the multiplicity of x in B). Thus, the condition that A is (B, η)-closed can be rewritten as the inequality

$$\sum_r |\widehat{\mathbb{1}_A}(r)|^2 \widehat{\mu}_B(r) \geq \eta \sum_r |\widehat{\mathbb{1}_A}(r)|^2.$$

Another fact we shall use later is that if W is a subspace of $\mathcal{G}$, then $\widehat{\mu_W}(r)$ equals $\mathbb{E}_{w \in W}(-1)^{r.w}$, which is 1 if r belongs to the orthogonal complement of W and 0 otherwise.

Lemma 7.2.2. *Let G be an Abelian group, let $A \subset G$ be a finite subset, let $\eta_1, \eta_2 > 0$, and let $b_1, b_2 \in G$. Suppose that A is $(\{b_1\}, 1 - \eta_1)$-closed and $(\{b_2\}, 1 - \eta_2)$-closed in G. Then A is $(\{b_1 + b_2\}, 1 - \eta_1 - \eta_2)$-closed in G.*

Proof. Let $A_{\text{bad}} = \{a \in A : a + b_2 \notin A\}$. Then $|A_{\text{bad}}| \leq \eta_2|A|$, by hypothesis. So when $a \in A$ is chosen randomly, we have that

$$\mathbb{P}[a + b_1 + b_2 \notin A] \leq \mathbb{P}[a + b_1 \notin A] + \mathbb{P}[a + b_1 \in A_{\text{bad}}] \leq \eta_1 + \eta_2.$$

The result follows. □

We are now in a position to deduce Theorem 7.1.10 from Lemma 7.2.1. In the proof, whenever a new function g_i appears, we mean that there exists a function g_i with the claimed property.

Proof of Theorem 7.1.10. Let $\theta, \delta > 0$, $\mathcal{B}' \subset \mathcal{B}$ with $|\mathcal{B}'| \geq \delta|\mathcal{B}|$, and suppose that $\mathcal{A} \subset \mathcal{G}$ is $(\mathcal{B}', 1-\varepsilon)$-closed, where ε is to be specified. Let $\mathcal{B}'' = \{b \in \mathcal{B}' : \mathcal{A} \text{ is } (\{b\}, 1 - 2\varepsilon)\text{-closed}\}$. Clearly, $|\mathcal{B}''| \geq \frac{1}{2}|\mathcal{B}'|$. Using Lemma 7.2.1, we can find a multiset Q with elements chosen from $g_1(d)\mathcal{B}''$ such that the set of arrays $r \in \mathcal{G}$ with $r.q = 0$ for at least $(1 - g_2(d))|Q|$ choices $q \in Q$ (where $g_2(d) > 0$) is contained in $\sum_{I \subset [d], I \neq \emptyset} V_I \otimes \mathbb{F}_2^{I^c}$ for subspaces $V_I \subset \mathbb{F}_2^I$ of dimension at most $g_3(d, \delta)$. Then, by Lemma 7.2.2, $\mathcal{A}$ is $(\{b\}, 1 - 2g_1(d)\varepsilon)$-closed for every $b \in Q$, since each such b is an element of $g_1(d)\mathcal{B}''$. In particular, $\mathcal{A}$ is $(Q, 1 - 2g_1(d)\varepsilon)$-closed, and therefore

$$\sum_r |\widehat{\mathbb{1}_{\mathcal{A}}}(r)|^2 \widehat{\mu_Q}(r) \geq (1 - 2g_1(d)\varepsilon) \sum_r |\widehat{\mathbb{1}_{\mathcal{A}}}(r)|^2.$$

By Markov's inequality, it follows that

$$\sum_{r : \widehat{\mu_Q}(r) < 1 - 2g_2(d)} |\widehat{\mathbb{1}_{\mathcal{A}}}(r)|^2 \leq \frac{g_1(d)\varepsilon}{g_2(d)} \sum_r |\widehat{\mathbb{1}_{\mathcal{A}}}(r)|^2.$$

Choosing $\varepsilon = \varepsilon(d, \theta) > 0$ to be at most $\theta g_2(d)/g_1(d)$, we therefore have

$$\sum_{r : \widehat{\mu_Q}(r) \geq 1 - 2g_2(d)} |\widehat{\mathbb{1}_{\mathcal{A}}}(r)|^2 \geq (1 - \theta) \sum_r |\widehat{\mathbb{1}_{\mathcal{A}}}(r)|^2.$$

Now if $\widehat{\mu_Q}(r) \geq 1 - 2g_2(d)$ then $r.q = 0$ for at least $(1 - g_2(d))|Q|$ choices $q \in Q$. Thus, we have

$$\sum_{r \in T} |\widehat{\mathbb{1}_{\mathcal{A}}}(r)|^2 \geq (1 - \theta) \sum_{r \in \mathcal{G}} |\widehat{\mathbb{1}_{\mathcal{A}}}(r)|^2 \tag{7.1}$$

where $T = \sum_{I \subset [d], I \neq \emptyset} V_I \otimes \mathbb{F}_2^{I^c}$. Let $R = T^\perp = \bigcap_{I \subset [d], I \neq \emptyset} V_I^\perp \otimes \mathbb{F}_2^{I^c}$. Because $\widehat{\mu_R}$ is the characteristic function of T, (7.1) is equivalent to the inequality

$$\sum_{r \in \mathcal{G}} |\widehat{\mathbb{1}_{\mathcal{A}}}(r)|^2 \widehat{\mu_R}(r) \geq (1 - \theta) \sum_{r \in \mathcal{G}} |\widehat{\mathbb{1}_{\mathcal{A}}}(r)|^2,$$

which in physical space is the inequality

$$\langle \mathbb{1}_{\mathcal{A}} * \mathbb{1}_{\mathcal{A}}, \mu_R \rangle \geq (1 - \theta)\|\mathbb{1}_{\mathcal{A}}\|_2^2 = (1 - \theta)\alpha,$$

where α is the density of $\mathcal{A}$. Equivalently,

$$\langle \mu_{\mathcal{A}} * \mu_R, \mathbb{1}_{\mathcal{A}} \rangle \geq 1 - \theta,$$

which tells us that if a random element of $\mathcal{A}$ is added to a random element of R, then the sum belongs to $\mathcal{A}$ with probability at least $1 - \theta$. The number of triples $(a_1, a_2, r) \in \mathcal{A} \times \mathcal{A} \times R$ with $a_1 + a_2 = r$ is therefore at least $(1 - \theta)|\mathcal{A}||R|$, and therefore, by averaging, there exists $a \in \mathcal{A}$ such that $|(R - a) \cap \mathcal{A}| \geq (1 - \theta)|R| = (1 - \theta)|R - a|$. But $R - a$ is $g_3(d, \delta)$-simple, so we can take $k = g_3(d, \delta)$ and $\mathcal{D} = R - a$. $\square$

7.3 The counterexample to Question 7.1.9

We shall now present an example that gives a negative answer to Question 7.1.9. The example is easy to define, but it takes a little work to prove that it has the properties we require. In what follows, let $G = \mathbb{F}_2^n$. For a vector $v \in G$ write $|v|$ for the number of entries equal to 1 in v. Then our set A will be $\{v \in \mathbb{F}_2^n : |v| \leq n/2 - 10^{20}n^{3/4}\}$, and our set B will be $\{v \in \mathbb{F}_2^n : |v| = n^{1/2}\}$.

Note first that A is η-closed with respect to B where $\eta > 0$ is some absolute constant. Indeed, by the central limit theorem, when n sufficiently large, the probability that a random element $u \in A$ has $|u| \leq n/2 - 10^{20}n^{3/4} - n^{1/4}$ is at least some absolute constant η_1, and conditional on this, the probability that $|u + v| \in A$ for a random element $v \in B$ is at least some other absolute constant η_2, so we may take $\eta = \eta_1\eta_2$. What we shall prove is that for this η, with $\varepsilon = 0.99$, say, there do not exist c, δ and l with the properties described in Question 7.1.9. In fact, we shall prove the slightly stronger statement that for any $\delta > 0$ and positive integer l, if n is sufficiently large then there do not exist $C \subset A + lB$ and $B' \subset B$ with $|B'| \geq \delta|B|$ such that C is $(B', 0.99)$-closed. Since for sufficiently large n, we have $A + lB \subset A' = \{v \in \mathbb{F}_2^n : |v| \leq n/2 - 10^{15}n^{3/4}\}$, it suffices to prove the same statement but with $A + lB$ replaced by A'. From now on, we always assume that n is sufficiently large.

The proof relies on two lemmas and a definition.

Lemma 7.3.1. *If $B' \subset B$ has $|B'| \geq \delta|B|$, then $\widehat{\mu_{B'}}(u) \geq 0.98$ for at most $\exp(n^{2/3})$ vectors $u \in \mathbb{F}_2^n$.*

Definition 7.3.2. Given $B' \subset B$, we say $u \in A'$ is *B'-compatible* if the number of $w \in B'$ with $|u + w| \leq |u|$ is at least $|B'|/3$.

Note that if we fix some $u \in A'$ and take a random $w \in B$, then the probability that $|u+w| \leq |u|$ is much less than 1/3. Indeed, writing X for the expected number of indices i for which $w_i = 1$ and $u_i = 0$, and Y for the expected number of indices i for which $w_i = 1$ and $u_i = 1$, we have $\mathbb{E}[X - Y] \geq 2 \cdot 10^{15}n^{1/4}$, while the standard deviation of $X - Y$ is around $n^{1/4}$. So $X \leq Y$ holds with quite small probability.

Thus, for any large $B' \subset B$, intuitively we expect only a small proportion of elements of A' to be B'-compatible. The next lemma makes this precise.

Lemma 7.3.3. *Let $B' \subset B$ have $|B'| \geq \delta|B|$. Then the number of those $u \in A'$ which are B'-compatible is at most $\exp(-n^{3/4})|A'|$.*

Let us see why these two lemmas are sufficient. Suppose that $C \subset A'$ is $(B', 0.99)$-closed for some $B' \subset B$ with $|B'| \geq \delta|B|$. Let $w \in B'$ be chosen at random. Then the expected number of $u \in C$ such that $u + w \in C$ is at least $0.99|C|$, so by considering all such pairs $\{u, u + w\}$ and noting that $(u + w) + w = u$, we see that there are on average at least $\frac{0.99}{2}|C|$ choices of $u \in C$ such that $|u + w| \leq |u|$. Therefore, if $u \in C$ is chosen at random, the average number of $w \in B'$ such that $|u+w| \leq |u|$ is at least $\frac{0.99}{2}|B'|$. It follows that for at least $|C|/10$ elements of C the number of such w is at least $|B'|/3$, so at least $|C|/10$ elements of A' are B'-compatible. Lemma 7.3.3 then implies that C has density at most $10\exp(-n^{3/4})$ in G. Let us write γ for this density.

On the other hand, since C is $(B', 0.99)$-closed, we have the inequality

$$\sum_{u\in G} \widehat{\mu_{B'}}(u)|\hat{C}(u)|^2 \geq 0.99 \sum_{u\in G} |\hat{C}(u)|^2,$$

which implies that

$$\sum_{u\in G:\widehat{\mu_{B'}}(u)\geq 0.98} \widehat{\mu_{B'}}(u)|\hat{C}(u)|^2 \geq 0.01 \sum_{u\in G} |\hat{C}(u)|^2.$$

Using Lemma 7.3.1, together with the observations that $\widehat{\mu_{B'}}(u) \leq 1$ and $|\hat{C}(u)| \leq \gamma$ for every $u \in G$ and that $\sum_{u\in G} |\hat{C}(u)|^2 = \gamma$, we deduce that $\exp(n^{2/3})\gamma^2 \geq 0.01\gamma$, so $\gamma \geq 0.01\exp(-n^{2/3})$. For sufficiently large n, this contradicts the upper bound for γ that we obtained a few lines above.

It remains to prove the two lemmas. The next two results are preparation for the proof of Lemma 7.3.1.

Lemma 7.3.4. *Let V be a subspace of $\mathbb{F}_2^n$ of dimension d such that every non-zero $v \in V$ has $|v| \geq n^{8/15}$. Then V has a basis $\{v_1, \ldots, v_d\}$ such that for every i, the set $I_i = \{k \leq n : v_i(k) = 1, v_j(k) = 0$ for all $j \neq i\}$ has size at least $n^{8/15}/2^{d-1}$. Here and below, the kth entry of a vector v is denoted by $v(k)$.*

Proof. We use induction on d. The case $d = 1$ easily follows from the assumption on V. Let V' have dimension $d+1$ and suppose that for a d-dimensional subspace $V \subset V'$, $v_1, \ldots, v_d$ and $I_1, \ldots, I_d$ have been chosen satisfying the requirements. Choose some $v \in V' \setminus V$. Replacing v by $v - v_1$ if necessary, we may assume that $v(k) = 0$ for at least $|I_1|/2$ choices $k \in I_1$. Similarly, we may assume that $v(k) = 0$ for at least $|I_i|/2$ choices $k \in I_i$ for every $i \leq d$. Thus, there exist subsets $J_1, \ldots, J_d$ of $\{1, \ldots, n\}$ of size at least $n^{8/15}/2^d$ each such that for every $i \leq d$ and every $k \in J_i$ we have $v_i(k) = 1$ but $v_j(k) = 0$ for all j with $j \neq i, j \leq d$, and $v(k) = 0$. Let $J = \{k \leq n : v(k) = 1\}$. By the assumption on V', we have $|J| \geq n^{8/15}$. Now it is easy to see that we can define v_1' to be v_1 or $v_1 - v$ and achieve that $v_1'(k) = 0$ for at least $|J|/2$ choices of $k \in J$. Similarly, we can define $v_2', \ldots, v_d'$ such that each v_i' is v_i or $v_i - v$ and $v_1'(k) = \cdots = v_d'(k) = 0$ for at least $|J|/2^d$ choices $k \in J$. Then for any $i, j \leq d$, we have $v_i'(k) = v_i(k)$ for every $k \in J_j$, and it follows that for any $i \leq d$ and $k \in J_i$, we have $v_i'(k) = 1$ but $v_j'(k) = 0$ for all $j \neq i$, and $v(k) = 0$. Thus, the set $\{v_1', \ldots, v_d', v\}$ is suitable so the lemma is proved. $\square$

Corollary 7.3.5. *Let t be a positive integer not depending on n and let V be a subspace of $\mathbb{F}_2^n$ of dimension t such that every non-zero $v \in V$ has $|v| \geq n^{8/15}$. Then the density of those $w \in B$ with $w \cdot v = 0$ for all $v \in V$ is less than $(1.9)^{-(t-1)}$.*

Proof. We shall be slightly sketchy about the some of the details when they are very standard. As always, we assume that n is sufficiently large. Let $v_1, \ldots, v_t$ be a basis given by Lemma 7.3.4 with $d = t$. Let w be a random vector in B, let $i < t$, and let us consider the probability that $w.v_i = 0$ given that $w.v_j = 0$ for every $j < i$.

The expected number of non-zero coordinates of w in the union of the two intervals I_i and I_{i+1} is at least $n^{1/30}/2^{t-1}$, which tends to infinity, and the probability that it is at least half this number tends to 1 (very rapidly). If we condition further on this number, and if it is indeed at least $n^{1/30}/2^t$, then the probability that the number of non-zero coordinates of w in I_i is even is almost exactly 1/2. Therefore, the probability that $w.v_i = 0$ given that $w.v_j = 0$ for every $j < i$ is less than $1/(1.9)$.

Since this is true for every $i \leq t-1$, we obtain the result. $\square$

Proof of Lemma 7.3.1. Suppose that the result is not true. Let r be a positive integer to be specified later and pick $R = \{u_1, \ldots, u_r\}$ such that $\widehat{\mu_{B'}}(u_i) \geq 0.98$ for $i = 1, 2, \ldots, r$. Then for each i, we have $u_i \cdot w = 0$ for at least 99% of all $w \in B'$. Therefore there is a subset $B'' \subset B'$ with $|B''| \geq |B'|/2$ such that each $w \in B''$ has $u_i \cdot w = 0$ for at least 98% of the u_i. The number of subsets of R of size $\frac{49}{50}r$ is at most $\binom{r}{49r/50} = \binom{r}{r/50} \leq (50e)^{r/50} \leq (1.8)^{49r/50}$. Let $t = 49r/50$. Then there exists a subset T of R of size t such that the number of $w \in B$ with $w \cdot u = 0$ for all $u \in T$ is at least $\frac{|B''|}{(1.8)^t} \geq \frac{\delta}{2\cdot(1.8)^t}|B|$. Choose the smallest positive integer t with $\frac{\delta}{2\cdot(1.8)^t} \geq (1.9)^{-(t-1)}$ (and with $r = 50t/49$ an integer). Then the density of those $w \in B$ with $w \cdot u = 0$ for all $u \in T$ is at least $(1.9)^{-(t-1)}$.

Now let Q be the set of all $u \in \mathbb{F}_2^n$ with $\widehat{\mu_{B'}}(u) \geq 0.98$ and assume that $|Q| \geq \exp(n^{2/3})$. Let t and r be as above and choose $u_1, \ldots, u_r \in Q$ such that for every j, the (Hamming) distance of u_j from $\mathrm{span}(u_1, \ldots, u_{j-1})$ is at least $n^{8/15}$. (This is possible because the number of $u \in \mathbb{F}_2^n$ with Hamming distance at most $n^{8/15}$ from an r-dimensional vector space is at most $2^r \exp(O(n^{8/15} \log n)) < \exp(n^{2/3})$.) Applying Corollary 7.3.5 to $V = \mathrm{span}(T)$, where T is a subset of $\{u_1, \ldots, u_r\}$ of size t, we find that the density of those $w \in B$ with $w \cdot u = 0$ for all $u \in T$ is less than $(1.9)^{-(t-1)}$, which is a contradiction. $\square$

Proof of Lemma 7.3.3. In this proof, unless specified otherwise, we will view $\mathbb{F}_2^n$ as a subset of $\mathbb{R}^n$ and accordingly, the dot product is defined as $u \cdot w = \sum_i u(i)w(i)$ where the summation is in $\mathbb{R}$. Then $|u + w| \leq |u|$ is equivalent to $u \cdot w \geq |w|/2$. Hence u is B'-compatible if $u \cdot w \geq |w|/2$ for at least $|B'|/3$ vectors $w \in B'$.

Let t be a fixed positive integer, not depending on n, to be specified later. For a multiset $T = \{u_1, \ldots, u_t\} \subset A'$ write $s_T = \sum_{i=1}^t u_i - \frac{t}{2}q$ where q is the vector in $\mathbb{F}_2^n$ consisting of ones. Let $a_k = s_T(k)$ and $\sigma_T^2 = \sum_{k=1}^n a_k^2$. We say that T is *bad* if $\sigma_T^2 \geq 1000tn$.

Claim 1. If T is not bad, then the number of $w \in B$ with $u_i \cdot w \geq |w|/2$ for all i is at most $\frac{|B|}{100^t}$.

Proof of Claim 1. If $u_i \cdot w \geq |w|/2$ for all i, then $s_T \cdot w \geq 0$. Note that $s_T \cdot w = \sum_{k \leq n} a_k w(k)$. We shall view w as a random variable, chosen uniformly of all elements of B. What we need to prove is that $\mathbb{P}[\sum_{k\le n} a_k w(k) \ge 0] \le \frac{1}{100^t}$.

Let $m = n^{1/2}$ and let $w_1, \dots, w_m$ be standard basis vectors of $\mathbb{F}_2^n$, chosen independently and uniformly at random. Note that the expected number of $i \ne j$ such that $w_i = w_j$ is at most 1, so almost surely this number is at most $\log n$. In particular, almost surely we have $n^{1/2} - 2\log n \le |w_1 + \cdots + w_m| \le n^{1/2}$. Choose uniformly randomly an element $w \in B$ with minimal Hamming distance from $w_1 + \cdots + w_m \in \mathbb{F}_2^n$. This algorithm defines a uniformly random element of $w \in B$ such that almost surely we have $\sum_{k\le n} |\sum_{i\le m} w_i(k) - w(k)| \le \sum_{k\le n} |\sum_{i\le m} w_i(k) - (\sum_{i\le m} w_i)(k)| + \sum_{k\le n} |(\sum_{i\le m} w_i)(k) - w(k)| \le 2\log n + 2\log n = 4\log n$, where all the summations are taken in $\mathbb{R}$, except $\sum_{i\le m} w_i$, which is taken in $\mathbb{F}_2^n$.

At this point, we apply the following version of Chernoff's inequality, which appears as Theorem 3.4 in [19].

Let X_i $(1 \le i \le m)$ be independent random variables satisfying $X_i \le \mathbb{E}[X_i] + M$, for $1 \le i \le m$. We consider the sum $X = \sum_i X_i$ with expectation $\mathbb{E}[X] = \sum_i \mathbb{E}[X_i]$ and variance $\mathrm{Var}(X) = \sum_i \mathrm{Var}(X_i)$. Then, we have $\mathbb{P}(X \ge \mathbb{E}[X] + \lambda) \le \exp(-\frac{\lambda^2}{2(\mathrm{Var}(X) + M\lambda/3)})$.

We now take $X_i = \sum_{k\le n} a_k w_i(k)$ for $1 \le i \le m$. Since $|a_k| \le t$, the conditions of the theorem hold with $M = 2t$. As $u_i \in A'$ for all i, we have $\sum_{k\le n} a_k \le t(n/2 - 10^{15} n^{3/4}) - tn/2 = -10^{15} t n^{3/4}$. Then $\mathbb{E}[X] = m \frac{\sum_{k\le n} a_k}{n} \le -\frac{1}{2} 10^{15} t n^{1/4}$, and by the assumption that T is not bad, $\mathrm{Var}(X) \le m \frac{\sum_{k\le n} a_k^2}{n} \le 1000 t n^{1/2}$. Thus, taking $\lambda = 10^{14} t n^{1/4}$ in the above theorem it follows that

$$\begin{aligned}\mathbb{P}\left[\sum_{i\le m}\sum_{k\le n} a_k w_i(k) \ge -10^{14} t n^{1/4}\right] &\le \exp\left(-\frac{10^{28} t^2 n^{1/2}}{2(1000 t n^{1/2} + 2t \cdot 10^{14} t n^{1/4}/3)}\right) \\ &\le \frac{1}{2\cdot 100^t}.\end{aligned}$$

But

$$\begin{aligned}\sum_{k\le n} a_k w(k) &= \sum_{i\le m, k\le n} a_k w_i(k) + \sum_{k\le n} a_k (w(k) - \sum_{i\le m} w_i(k)) \\ &\le \sum_{i\le m, k\le n} a_k w_i(k) + t \sum_{k\le n} |(w(k) - \sum_{i\le m} w_i(k))|,\end{aligned}$$

and $\sum_{k\le n} |(w(k) - \sum_{i\le m} w_i(k))| \le 4\log n$ almost surely, it follows that $\mathbb{P}[\sum_{k\le n} a_k w(k) \ge 0] \le \frac{1}{100^t}$ and Claim 1 is proved.

Claim 2. If $u_1, u_2, \dots, u_t$ are independently and uniformly randomly chosen elements of A' then the probability that $T = \{u_1, \dots, u_t\}$ is bad is $o(\exp(-n^{7/8}))$.

Proof of Claim 2. Recall that T is bad if and only if $\sum_{k\le n} (\sum_{i\le t} u_i(k) - t/2)^2 \ge 1000tn$. $u_1, \dots, u_t$ are randomly chosen from A' but with probability $1 - o(\exp(n^{-7/8}))$ all of them have $|u_i| \ge n/2 - n^{99/100}$ so we may assume that $u_1, \dots, u_t$ are randomly chosen from the set $A'' = \{v \in \mathbb{F}_2^n : n/2 - n^{99/100} \le |v| \le n/2 - 10^{15} n^{3/4}\}$. It is not hard to see that we can write $u_i = x_i + y_i$ where x_i and y_i are random variables taking values in $\mathbb{F}_2^n$ and having the property that $x_i(k)$ are independent Bernoulli with parameter $1/2$ and $|y_i| \le 2n^{99/100}$

with probability $1 - o(\exp(-n^{7/8}))$. Then it suffices to prove that

$$\mathbb{P}\left[\sum_{k\leq n}(\sum_{i\leq t} x_i(k) - t/2)^2 \geq 500tn\right] = o(\exp(-n^{7/8}))$$

Let $X_i = (\sum_{i\leq t} x_i(k) - t/2)^2$ for $1 \leq k \leq n$. Then the X_i are iid random variables with $\mathbb{E}[X_i] = t/4$ and $\mathrm{Var}(X_i) = O(1)$. Thus, by Theorem 3.4 from [19] (which is the theorem stated above), taking $\lambda = 100tn$ and $M = t^2$, it follows that

$$\begin{aligned}\mathbb{P}\left[\sum_{k\leq n}(\sum_{i\leq t} x_i(k) - t/2)^2 \geq 500tn\right] &\leq \exp\left(-\frac{(100tn)^2}{2(nO(1) + t^2 \cdot 100tn/3)}\right)\\ &= o(\exp(-n^{7/8})),\end{aligned}$$

finishing the proof of Claim 2.

We are now in a position to complete the proof of the lemma. Let t be the smallest positive integer with $\frac{1}{100^t} < \frac{\delta}{100(6e)^t}$. Let the density of B'-compatible elements in A' be α. Pick $v_1, \ldots, v_{6t}$ independently and uniformly randomly from A'. Then with probability α^{6t}, every v_i is B'-compatible. If that is the case, then for every i, there are at least $|B'|/3$ vectors $w \in B'$ with $v_i \cdot w \geq |w|/2$. It follows that there is some $B'' \subset B'$ with $|B''| \geq |B'|/100$ such that for every $w \in B''$ we have $v_i \cdot w \geq |w|/2$ for at least t choices of i. The number of t-sets in $\{v_1, \ldots, v_{6t}\}$ is at most $(6e)^t$ so there must exist a t-set $T = \{u_1, \ldots, u_t\} \subset \{v_1, \ldots, v_{6t}\}$ (multisets are allowed) such that the number of $w \in B$ with $u_i \cdot w \geq |w|/2$ for each i is at least $|B''|/(6e)^t \geq \frac{\delta|B|}{100(6e)^t} > \frac{|B|}{100^t}$. By Claim 1, it follows that T is bad. Thus, the probability that $T = \{u_1, \ldots, u_t\}$ is bad when $u_1, \ldots, u_t$ are independently and uniformly randomly chosen from A', is at least $\frac{\alpha^{6t}}{\binom{6t}{t}}$. Hence, by Claim 2, we have $\frac{\alpha^{6t}}{\binom{6t}{t}} = o(\exp(-n^{7/8}))$, and we get $\alpha = o(\exp(-n^{3/4}))$. $\square$

Chapter 8

The two-step local density of graphs

8.1 Introduction

Kopylov [87] asked and answered the following question. Say that a graph G has the (p, q)-property if every set of p vertices contains a subset of q vertices which induces a complete graph. What is the minimal number of edges in a graph on n vertices having the (p, q)-property?

For $k \leq \frac{p-1}{q-1}$, consider the following graph G_k on n vertices. Let G_k have $m = p - 1 - k(q - 1)$ isolated vertices and let the remaining $n - m$ vertices induce a graph which is a union of k disjoint cliques of size as equal as possible. It is easy to see that G_k has the (p, q)-property. Indeed, let X be a set of p vertices. Then at least $k(q - 1) + 1$ of those vertices lie in one of the k disjoint cliques, so at least one of the cliques contains at least q vertices of X. In [87] it is shown that any extremal graph is of the form G_k for some k.

In this chapter, we consider a generalization of the above problem where instead of looking for a complete subgraph on q vertices inside every induced subgraph on p vertices, we look for a subgraph with q vertices and at least e edges. Let us choose parameters α, β, γ such that $p = \alpha n$, $q = \beta n$ and $e = \gamma\binom{\beta n}{2}$. The parameters α, β, γ can depend on n. We shall be interested in graphs satisfying the following property.

Definition 8.1.1. Let G be a graph on n vertices. Say that G has the (α, β, γ)-*property* if any subset $A \subset V(G)$ of size at least αn contains a further subset $B \subset A$ of size at least βn such that the edge density of $G[B]$ is at least γ.

This generalization makes the problem more complicated, because there is more flexibility to create extremal examples. To see this, consider the case $\alpha = \beta$. An easy averaging argument implies that a graph with the (α, α, γ)-property has density at least γ. This bound is essentially attained by a random graph with edge-probability γ. (If one wants it to be attained exactly, one can generalize the problem and allow weighted graphs with weights in $[0, 1]$. Then the graph with constant edge-weight γ obviously has the (α, α, γ)-property.) If $\gamma = 1/k$ for some integer $k > 1$, then an alternative example (for convenience, with loops) is given by taking k disjoint cliques with vertex sets $V_1, \dots, V_k$ of size n/k: if X is any set of vertices, then the number of ordered pairs $(x, y) \in X^2$ such that xy is an edge of G is $\sum_i |X \cap V_i|^2$, which is minimized when the sets $X \cap V_i$ have equal size n/k, so the density is at least $1/k = \gamma$. And that is not all. For instance, we can also take a convex combination of these two examples. That is, we can choose $\lambda \in (0,1)$ and set the edge weight of a pair (x, y) to be $\lambda + (1-\lambda)\gamma$ if $(x, y) \in V_i^2$ for some i, and $(1-\lambda)\gamma$ otherwise. And that is just the start: we can take convex combinations of more examples, with completely unrelated vertex partitions for the different examples.

The extra flexibility when $\alpha = \beta$ leads directly to extra flexibility for smaller β. For example, if α/β is an integer r, then we can partition the vertices into r sets $X_1, \dots, X_r$ of equal size, and then every set of vertices of density α intersects some X_i in a subset of density at least β. If we also arrange that every subset of X_i of density β (and therefore relative density α in X_i) induces a subgraph of density at least γ, then we have an example. And that can be achieved in different ways for different X_i, using the range of examples discussed in the previous paragraph.

One might still hope that every example is obtained by partitioning the vertices into a few sets, placing appropriate (β, β, γ) examples in each set, and using the pigeonhole principle to argue that every set of density α intersects one of the cells of the partition in a set of density at least β. However, a simple example shows that this does not work. Let $\alpha = 1/2$ and $\beta = 1/4 + \eta$ for some very small η. Then the best example of the above kind that has the (α, β, γ)-property is a block of size $(3/4 + \eta)n$ with edge density γ, whereas we could instead take two blocks of size $n/2$ with edge densities slightly greater than γ each to get a sparser example. When γ is close to 1, this might not be possible as the edge density in the blocks cannot be greater than 1. In this case, we have to take a few edges between the two blocks of size $n/2$ as well, further complicating the picture.

In the light of these examples, it seems difficult to characterize the extremal examples for all α, β, γ. However, our first result gives a lower bound which is asymptotically sharp for various values of the parameters, in particular when α/β is an integer.

Theorem 8.1.2. *Let $0 < \beta \le \alpha \le 1/2$ and $0 < \gamma \le 1$ be functions of n such that $\alpha = \omega(\frac{1}{n})$ and $\beta\gamma = \omega(\frac{(\log n)^{10}}{n})$. If G is a graph on n vertices which has the (α, β, γ)-property, then G has edge density at least $(1 - o(1))\frac{\beta\gamma}{\alpha}$.*

We will in fact prove a version of Theorem 8.1.2 for weighted graphs, from which Theorem 8.1.2 follows immediately. A weighted graph in this chapter has a non-negative, symmetric weight $w(u, v)$ assigned to each pair $\{u, v\}$ of vertices. Then, if G is a weighted graph, $e(G)$ is defined to be $\frac{1}{2}\sum_{u,v \in V(G)} w(u, v)$. Similarly to the simple graph case, we say that G has the (α, β, γ)-property if for every $A \subset V(G)$ of size at least αn, there exists a $B \subset A$ of size at least βn with $e(G[B]) \ge \gamma\binom{|B|}{2}$. Our generalization of Theorem 8.1.2 can now be stated as follows.

Theorem 8.1.3. *Let $0 < \beta \le \alpha \le 1/2$, $0 < \gamma \le 1$ and $M \le n$ be functions of n such* *that $\alpha = \omega(\frac{1}{n})$ and $\beta\gamma = \omega\big(\frac{M(\log n)^{10}}{n}\big)$. If G is a weighted graph on n vertices with edge weights at most M and G has the (α, β, γ)-property, then $e(G) \geq (1 - o(1))\frac{\beta\gamma}{\alpha}\binom{n}{2}$.*

The condition $\alpha \leq 1/2$ in these results is necessary. Indeed, take a set of $(1-\alpha+\beta)n$ vertices and place edges between them with probability roughly γ. Now any set of size αn intersects this set in at least βn vertices, and those βn vertices induce a graph with edge density roughly γ. The total density in this graph is roughly $(1-\alpha+\beta)^2\gamma$. So there is a graph having the (α, β, γ)-property with density roughly $(1-\alpha+\beta)^2\gamma$. Note that the inequality $(1-\alpha+\beta)^2 < \frac{\beta}{\alpha}$ is equivalent to the inequality $(\beta-\alpha)(\beta-(\alpha+1/\alpha-2)) < 0$. If $\alpha > 1/2$, then for $\alpha + 1/\alpha - 2 < \beta < \alpha$, this inequality is satisfied, so the conclusion of Theorem 8.1.2 fails.

In many cases, Theorem 8.1.2 is tight. To see this, consider the following family of graphs. Let $p \leq q$ be positive integers. If $\frac{p}{q} \geq \frac{\beta}{\alpha}$ and $p\gamma \leq 1$, then we can construct graphs G_n on n vertices which have the (α, β, γ)-property and which have average degree roughly $\frac{p}{q}\gamma n$. Indeed, let the vertex set of G_n be $S_1 \cup S_2 \cup \cdots \cup S_q$ where the sets $S_1, \dots, S_q$ are pairwise disjoint and have size roughly $\frac{n}{q}$. Define $G_n[S_i]$ to be an Erdős-Rényi random graph with density $p\gamma$, independently for all $1 \leq i \leq q$. Let all other pairs be non-edges. Almost surely, G_n has average degree at most roughly $p\gamma\frac{n}{q}$. Moreover, it is not hard to see that if $\beta\gamma = \omega(\frac{\log n}{n})$, then by the Chernoff bound, for every $Q \subset V(G_n)$ of size at least βn with $\sum_{1\leq i\leq q} |Q \cap S_i|^2 = \Omega(\frac{(\beta n)^2}{p})$, we have that $\sum_{1\leq i\leq q} e(G_n[Q \cap S_i])$ is at least roughly $\sum_{1\leq i\leq q} p\gamma\binom{Q\cap S_i}{2}$. If this holds, then G_n has the (α, β, γ)-property (approximately). Indeed, let $A \subset V(G_n)$ have $|A| \geq \alpha n$. Since $\frac{p}{q} \geq \frac{\beta}{\alpha}$, there exist $1 \leq i_1 < i_2 < \cdots < i_p \leq q$ such that $|A \cap (S_{i_1} \cup \cdots \cup S_{i_p})| \geq \beta n$. Take $B = A \cap (S_{i_1} \cup \cdots \cup S_{i_p})$. Then $|B| \geq \beta n$ and $\sum_{1\leq j\leq p} |B \cap S_{i_j}|^2 \geq \frac{(\beta n)^2}{p}$. Thus, $e(G_n[B]) = \sum_{1\leq j\leq p} e(G_n[B \cap S_{i_j}])$ is at least roughly $p\gamma\frac{|B|^2}{2p} = \gamma\frac{|B|^2}{2}$, so G_n indeed has the (α, β, γ)-property approximately.

Given this construction, it is easy to see that Theorem 8.1.2 is tight when $\gamma = o(1)$. This is because in this case we can take $p = \lfloor 1/\gamma \rfloor$ and choose a maximal positive integer q with $\frac{p}{q} \geq \frac{\beta}{\alpha}$. Since $p = \omega(1)$, we have $\frac{p}{q} = (1 + o(1))\frac{\beta}{\alpha}$.

The same construction shows that for every γ, Theorem 8.1.2 is tight when α/β is an integer. Moreover, if we place edges of weight $p\gamma$ instead of taking edges with probability $p\gamma$ in the construction above, we get a weighted graph that shows that Theorem 8.1.3 is tight for every α, β, γ. However, when β is close to α and γ is not too small, Theorem 8.1.2 is not tight for simple graphs. This is confirmed by the following theorem, which is our second main result.

Theorem 8.1.4. *Let $0 < \beta \leq \alpha \leq 1/2$ and $0 < \gamma \leq 1$ be functions of n such that $\alpha - \beta \leq \frac{\beta^3\gamma}{1000}$ and $\beta^3\gamma = \omega(\frac{\log n}{n^{1/2}})$. If G is a (simple) graph on n vertices which has the (α, β, γ)-property, then G has edge density at least $(1 - o(1))(1 - \alpha + \beta)^2\gamma$.*

Unlike in Theorems 8.1.2 and 8.1.3, the condition $\alpha \leq 1/2$ here is not necessary. Indeed, we can extend our result to $\alpha > 1/2$, but the proof of that is somewhat tedious, so we do not include it.

Note that $(1-\alpha+\beta)^2 - \frac{\beta}{\alpha} = (\alpha-\beta)(\frac{1}{\alpha} - 2 + (\alpha-\beta))$, so Theorem 8.1.4 improves Theorem 8.1.2 when $\alpha - \beta = \Omega(1)$. Moreover, Theorem 8.1.4 is tight, by the construction described immediately after Theorem 8.1.3.

We also consider the bipartite version of the above problem.

Definition 8.1.5. Let G be a bipartite graph with vertex sets X and Y of size n each. Say that G has the *bipartite* (α, β, γ)*-property* if for any $A \subset X$ and $B \subset Y$ with $|A|, |B| \geq \alpha n$, there exist $C \subset A$ and $D \subset B$ with $|C|, |D| \geq \beta n$ such that the bipartite edge density of $G[C, D]$ is at least γ.

We are interested in the minimal edge density of a bipartite graph with the bipartite (α, β, γ)-property.

Note that the bipartite case is somewhat more difficult than the graph case since already when $\gamma = 1$, when we are looking for a complete $K_{\beta n, \beta n}$, it is not clear what the extremal construction is. Indeed, the natural generalization of Kopylov's construction, namely a disjoint union of $K_{\theta n, \theta n}$s with $\theta \approx \beta/\alpha$, clearly does not work. Instead we take the following probabilistic construction. Let $\theta = \beta/\alpha$, let $m = c\theta^{-1}\log(1/\alpha)$, let $A_1, \dots, A_m$ be independent random subsets of X of size roughly θn and let $B_1, \dots, B_m$ be independent random subsets of Y of size roughly θn. The graph G_i is then defined by picking each edge between A_i and B_i independently at random with probability roughly γ. Finally, set $G = \cup_{i \leq m} G_i$. As we shall prove later, almost surely G has density at most $C\frac{\beta}{\alpha}\gamma\log(1/\alpha)$ and it has the (α, β, γ)-property. We obtain the following result.

Proposition 8.1.6. *There exists an absolute constant C with the following property. For any $0 < \beta < \alpha \leq 1/2$ and $0 < \gamma \leq 1$ depending on $n \in \mathbb{N}$ and satisfying $\beta\gamma = \omega(\frac{1}{n})$, for n sufficiently large there is a bipartite graph G_n on $n + n$ vertices with the bipartite (α, β, γ)-property and edge density at most $C\frac{\beta\gamma}{\alpha}\log(1/\alpha)$.*

We see that compared with Theorem 8.1.2, we have an extra $\log(1/\alpha)$ factor. We prove that this is necessary. More precisely, we prove the following theorem.

Theorem 8.1.7. *There exists an absolute constant $c > 0$ the following property. Let $0 < \beta < \alpha \leq 1/2$ and $0 < \gamma \leq 1$ be parameters depending on $n \in \mathbb{N}$ such that $\frac{\beta}{\alpha}\log_2(1/\alpha) \leq \frac{1}{10}$, $\alpha\beta\gamma = \omega(\frac{(\log n)^3}{n})$ and $\beta^2\gamma = \omega(\frac{\log n}{n})$. Then, for n sufficiently large, if G is a bipartite graph on $n + n$ vertices having the (α, β, γ)-property, then it has edge density at least $c\frac{\beta\gamma}{\alpha}\log(1/\alpha)$.*

Observe that a condition of the form $\frac{\beta}{\alpha}\log(1/\alpha) = O(1)$ is necessary for the theorem to hold. Indeed, we can easily find a bipartite graph with density roughly γ having the (α, β, γ)-property, so if $\frac{\beta}{\alpha}\log(1/\alpha) = \omega(1)$, then $c\frac{\beta\gamma}{\alpha}\log(1/\alpha)$ is not a lower bound.

We also prove a structural result in both the bipartite and the non-bipartite case (in fact, we prove an analogous result about r-partite graphs as well). Here we state the non-bipartite version.

Theorem 8.1.8. *Let α, β, γ be constants independent of n. Let G be a graph on n vertices having the (α, β, γ)-property. Then there exists a set $D \subset V(G)$ of size at least $\frac{\beta}{\alpha}n$ such that $G[D]$ has edge density at least $(1 - o(1))\gamma$.*

The rest of this chapter is organized as follows. In the next section, we prove lower bounds on the number of long paths in a weighted graph of given density. We then use these results in Sections 8.3 and 8.4 to prove Theorems 8.1.3 and 8.1.4, respectively. In Section 8.5, we prove Proposition 8.1.6 and Theorem 8.1.7. In Section 8.6 we prove Theorem 8.1.8. In the last section we give some concluding remarks and open problems.

8.2 The number of long paths in a weighted graph of given density

Recall from the introduction that a weighted graph has a nonnegative weight $w(u, v)$ assigned to each pair $\{u, v\}$ of vertices. The degree of the vertex u is $d_u = \sum_{v \in V(G)} w(u, v)$ while the average degree is $\bar{d} = \frac{1}{n} \sum_{u \in V(G)} d_u$.

Definition 8.2.1. Let G be a weighted graph (with edge weights w). For a positive integer ℓ and a vertex $v \in V(G)$, write

$$g_\ell(v, G) = \sum_{v_1, \ldots, v_\ell \in V(G)} w(v, v_1) w(v_1, v_2) \ldots w(v_{\ell-1}, v_\ell).$$

Let

$$h_\ell(G) = \sum_{v \in V(G)} g_\ell(v) = \sum_{v_0, \ldots, v_\ell} w(v_0, v_1) \ldots w(v_{\ell-1}, v_\ell).$$

For convenience we also set $g_0(v, G) = 1$ and $h_0(G) = n$.

Note that $h_\ell(G)$ can be viewed as the number of walks of length ℓ in G, whereas $g_\ell(v, G)$ is the number of walks of length ℓ starting at v.

We shall use the following fact relating the number of walks of various lengths.

Lemma 8.2.2 (Erdős–Godsil–Simonovits [39])**.** *Let $\ell \le k$ be positive integers and assume that k is even. Then*

$$\left(\frac{h_k(G)}{n}\right)^{1/k} \ge \left(\frac{h_\ell(G)}{n}\right)^{1/\ell}.$$

We will also need a classical result of Mulholland and Smith [99] and Blakley and Roy [10] which gives a lower bound for the number of walks of given length in a graph of given density.

Theorem 8.2.3 (Mulholland–Smith, Blakley–Roy)**.** *Let G be a weighted graph on n vertices with average degree $\bar{d}$. Then*

$$h_k(G) = \sum_{v_0, \ldots, v_k \in V(G)} w(v_0, v_1) \ldots w(v_{k-1}, v_k) \ge n\bar{d}^k.$$

We need a lower bound for the number of *paths*, rather than walks. The following lemma is the main result of this section.

Lemma 8.2.4. *Let G be a weighted graph on n vertices with average degree $\bar{d}$ and edge weights $|w(u,v)| \le M \le n$ such that $\bar{d} = \omega(Mk^4(\log n)^2)$. Let $k = \omega(1)$ be an even positive integer. Then*

$$\sum_{v_0,\dots,v_k \text{ distinct in } V(G)} w(v_0,v_1)\dots w(v_{k-1},v_k) \ge n(\bar{d} - o(\bar{d}))^k.$$

Proof. We distinguish between two cases. First, let us assume that there exists a set $S \subset V(G)$ of size at most $\frac{n}{\log n}$ such that the total weight of edges in G with at least one endpoint in S is $\Omega(n\bar{d})$. Write t for this total weight. Define sequences of sets $X_0 \supset X_1 \supset \dots$ and $Y_0 \supset Y_1 \supset \dots$ as follows. We take $X_0 = S$ and $Y_0 = V(G)$. Having defined X_i and Y_i, if there exists some $u \in X_i$ such that $\sum_{v\in Y_i} w(u,v) < \frac{t\log n}{4n}$, then set $X_{i+1} = X_i \setminus \{u\}$ and $Y_{i+1} = Y_i$. If such vertex does not exist, but there exists some $u \in Y_i$ such that $\sum_{v\in X_i} w(u,v) < \frac{t}{4n}$, then set $Y_{i+1} = Y_i \setminus \{u\}$ and $X_{i+1} = X_i$. In both cases, we say that u is discarded. Note that the process eventually stops with sets X_j and Y_j. Let $t_i = \sum_{u\in X_i, v\in Y_i} w(u,v)$. Note that $t_0 \ge t$. If in the $(i+1)$th step a vertex from X_i is discarded, then $t_{i+1} \ge t_i - \frac{t\log n}{4n}$, while if a vertex from Y_i is discarded, then $t_{i+1} \ge t_i - \frac{t}{4n}$. Since $|X_0| \le \frac{n}{\log n}$ and $|Y_0| \le n$, it follows that $t_j \ge t_0 - \frac{n}{\log n}\frac{t\log n}{4n} - n\frac{t}{4n} = t - \frac{t}{4} - \frac{t}{4} = \frac{t}{2}$. Moreover, for every $u \in X_j$ we have $\sum_{v\in Y_j} w(u,v) \ge \frac{t\log n}{4n}$, and for every $u \in Y_j$ we have $\sum_{v\in X_j} w(u,v) \ge \frac{t}{4n}$. Note that $\frac{t}{4n} = \Omega(\bar{d})$, so by assumption (for n sufficiently large) $\frac{t}{4n} \ge 2Mk$, hence it is easy to see that

$$\begin{aligned}
\sum_{\substack{v_0,v_2,\dots,v_k\in Y_j\\ v_1,v_3,\dots,v_{k-1}\in X_j\\ \text{all distinct}}} w(v_0,v_1)w(v_1,v_2)\dots w(v_{k-1},v_k) &\ge \sum_{v_0\in Y_j, v_1\in X_j} w(v_0,v_1)\left(\frac{t\log n}{8n}\right)^{k/2}\left(\frac{t}{8n}\right)^{k/2-1} \\
&= t_j\left(\frac{t\log n}{8n}\right)^{k/2}\left(\frac{t}{8n}\right)^{k/2-1} \\
&\ge \frac{t}{2}\left(\frac{t\log n}{8n}\right)^{k/2}\left(\frac{t}{8n}\right)^{k/2-1} \\
&\ge n\bar{d}^k,
\end{aligned}$$

provided that n is sufficiently large, where in the last inequality we used that $t = \Omega(n\bar{d})$.

We can therefore assume that for every set $S \subset V(G)$ of size at most $\frac{n}{\log n}$ the total weight in $G[V(G)\setminus S]$ is $\frac{n\bar{d}}{2}(1-o(1))$. Define a sequence $Z_0 \supset Z_1 \supset \dots$ of subsets of $V(G)$ and corresponding induced subgraphs $G_i = G[Z_i]$ as follows. We take $Z_0 = V(G)$. Let $\lambda = \frac{8k^2(\log n)^2}{n}$. If there exists some $1 \le \ell \le k$ and some $v_i \in Z_i$ with $g_\ell(v_i, G_i) > \lambda h_\ell(G_i)$, then choose such a v_i arbitrarily and set $Z_{i+1} = Z_i \setminus \{v_i\}$. In this case we have $h_\ell(G_{i+1}) < (1-\lambda)h_\ell(G_i)$.

Observe that for every $1 \le \ell \le k$, we have $h_\ell(G) \le n(Mn)^\ell \le n^{2k+1}$. On the other hand, $(1-\lambda)^{\frac{n}{2k\log n}} \le e^{-\lambda\frac{n}{2k\log n}} = e^{-4k\log n} = n^{-4k}$, so $h_\ell(G)(1-\lambda)^{\frac{n}{2k\log n}} < 1$. Thus, throughout the process of defining $Z_1, Z_2, \dots$, for any fixed $1 \le \ell \le k$, it happens at most $\lceil \frac{n}{2k\log n} \rceil \le \frac{n}{k\log n}$ times that the removed vertex v_i has $g_\ell(v_i, G_i) > \lambda h_\ell(G_i)$. In particular, the process must stop after at most $k \cdot \frac{n}{k\log n} = \frac{n}{\log n}$ steps. Hence, the final set Z_j satisfies $|Z_j| \ge n - \frac{n}{\log n}$. By our earlier discussion, it follows that the total weight in G_j is $\frac{n\bar{d}}{2}(1-o(1))$. Moreover, since the process terminates at Z_j, for every $1 \le \ell \le k$ and every $v \in Z_j$, we have $g_\ell(v, G_j) \le \lambda h_\ell(G_j)$.

We claim that

$$\sum_{v_0,\dots,v_k \in Z_j \text{ not all distinct}} w(v_0,v_1)\dots w(v_{k-1},v_k) = o(h_k(G_j)). \tag{8.1}$$

Once this is proved, we have

$$\begin{aligned}\sum_{v_0,\dots,v_k \in Z_j \text{ distinct}} w(v_0,v_1)\dots w(v_{k-1},v_k) &\ge (1-o(1))h_k(G_j)\\ &\ge (1-o(1))|Z_j|(\bar{d}(1-o(1)))^k\\ &\ge n(\bar{d}(1-o(1)))^k,\end{aligned}$$

where the second inequality follows from Theorem 8.2.3.

Therefore it remains to prove equation (8.1). For any $0 \le a < b \le k$,

$$\begin{aligned}&\sum_{v_0,\dots,v_k \in Z_j \text{ with } v_a=v_b} w(v_0,v_1)\dots w(v_{k-1},v_k)\\
&\le \sum_{v_0,\dots,v_k \in Z_j \text{ with } v_a=v_b} w(v_0,v_1)\dots w(v_{b-2},v_{b-1})\cdot M\cdot w(v_b,v_{b+1})\dots w(v_{k-1},v_k)\\
&= \sum_{v_0,\dots,v_b \in Z_j \text{ with } v_a=v_b} w(v_0,v_1)\dots w(v_{b-2},v_{b-1})\cdot M\cdot g_{k-b}(v_b,G_j)\\
&\le \sum_{v_0,\dots,v_{b-1} \in Z_j} w(v_0,v_1)\dots w(v_{b-2},v_{b-1})\cdot M\cdot \lambda h_{k-b}(G_j)\\
&= h_{b-1}(G_j)\cdot M\cdot\lambda h_{k-b}(G_j)\\
&\le n\left(\frac{h_k(G_j)}{n}\right)^{\frac{b-1}{k}}\cdot M\cdot\lambda n\left(\frac{h_k(G_j)}{n}\right)^{\frac{k-b}{k}} \qquad \text{(by Lemma 8.2.2)}\\
&= M\lambda n\left(\frac{h_k(G_j)}{n}\right)^{-\frac{1}{k}} h_k(G_j)\\
&\le M\lambda n\left(\frac{h_1(G_j)}{n}\right)^{-1} h_k(G_j) \qquad \text{(by Lemma 8.2.2)}\\
&= M\lambda n(\bar{d}(1-o(1))^{-1}h_k(G_j)\\
&= \frac{8k^2(\log n)^2M}{\bar{d}}(1+o(1))h_k(G_j),\end{aligned}$$

where the $o(1)$ term does not depend on a and b. Summing over all $0 \le a < b \le k$, we get

$$\sum_{\substack{v_0,\ldots,v_k\in Z_j\\ \text{not all distinct}}} w(v_0,v_1)\ldots w(v_{k-1},v_k) \le \binom{k+1}{2}\frac{8k^2(\log n)^2M}{\bar{d}}(1+o(1))h_k(G_j).$$

By $\bar{d}=\omega(Mk^4(\log n)^2)$, equation (8.1) follows. $\square$

The next result will not be needed in our paper, but we state it as it is a cleaner version of the previous lemma.

Corollary 8.2.5. *Let k be a positive integer and let G be a simple graph on n vertices with average degree $\bar{d}$ such that $\bar{d}=\omega(k^4(\log n)^2)$. Then the number of (directed) paths of length k in G is at least $n(\bar{d}-o(\bar{d}))^k$.*

Proof. For $k=O(1)$, this follows from Theorem 5 in [39]. In the $k=\omega(1)$ case, when k is even, this is immediate from Lemma 8.2.4. Moreover, in the case where $k=\omega(1)$ and k is odd, the corollary follows from Lemma 8.2.4 with k replaced by $k-1$, applied to a subgraph G' of G with at least $\frac{e(G)}{2}$ edges, average degree at least $\bar{d}$ and minimum degree at least $\frac{\bar{d}}{4}$. We leave the details to the interested reader. $\square$

8.3 Weighted graphs

In this section we prove Theorem 8.1.3. It is easy to see that it suffices to prove the following result.

Theorem 8.3.1. *Let $0<\beta\le\alpha\le 1/2$, $0<\gamma\le 1$ and $M\le n$ be functions of n such that $\alpha=\omega(\frac{1}{n})$ and $\beta\gamma=\omega\big(\frac{M(\log n)^{10}}{n}\big)$. Let $0<\epsilon<1$ be a constant. Suppose that G is a weighted graph on n vertices with edge weights at most M such that $e(G)\le(1-\epsilon)\frac{\beta\gamma}{2\alpha}n^2$. Then, if n is sufficiently large, there exists a set $A\subset V(G)$ of size at least αn such that for every $B\subset A$ of size at least βn, we have $e(G[B])\le(1-\frac{\epsilon}{4})\frac{\gamma|B|^2}{2}$.*

For the rest of this section, fix some $0<\epsilon<1$ constant, let n be sufficiently large and let G be a weighted graph on n vertices with edge weights at most M and $e(G)\le(1-\epsilon)\frac{\beta\gamma}{2\alpha}n^2$, where α,β,γ,M satisfy the conditions in the theorem.

Definition 8.3.2. Let $D=(1-\frac{\epsilon}{2})\beta\gamma n$. For vertices $v\in V(G)$, define

$$p_v=\begin{cases}1 & \text{if } d_v<D\\ \frac{D}{d_v} & \text{if } d_v\ge D.\end{cases}$$

Our strategy is to prove that if we define A to be a random subset of $V(G)$ where the vertex v is in A with probability p_v, then with positive probability A satisfies the conclusion of Theorem 8.3.1. The first step is to verify that with high probability A is large enough.

Lemma 8.3.3.
$$\sum_{v\in V(G)} p_v \geq \Big(1+\frac{\epsilon}{32}\Big)\alpha n.$$

Proof. Let $s = |\{v \in V(G) : d_v < D\}|$. Then

$$\begin{aligned}
\sum_{v\in V(G)} p_v &= s + \sum_{v:d_v\geq D} \frac{D}{d_v} \\
&\geq s + (n-s)\frac{D}{2e(G)/(n-s)} \\
&= s + (n-s)^2 \frac{(1-\frac{\epsilon}{2})\beta\gamma n}{2e(G)} \\
&\geq s + (n-s)^2 \frac{\alpha}{n}\cdot \frac{1-\frac{\epsilon}{2}}{1-\epsilon} \\
&\geq s + (n-s)^2 \frac{\alpha}{n}\Big(1+\frac{\epsilon}{2}\Big) \\
&= s + (n-s)^2\frac{\alpha}{n} + \frac{\epsilon\alpha}{2}\frac{(n-s)^2}{n}.
\end{aligned}$$

If $s \geq (1+\frac{\epsilon}{32})\alpha n$, then we are done. Otherwise, by $\alpha \leq 1/2$, we have $s \leq \frac{1+\epsilon/32}{2} n < \frac{3}{4}n$. Thus,
$$\frac{\epsilon\alpha}{2}\frac{(n-s)^2}{n} \geq \frac{\epsilon}{32}\alpha n.$$
Moreover,
$$s + (n-s)^2\frac{\alpha}{n} = \alpha n + s\Big(\frac{\alpha}{n}s + (1-2\alpha)\Big) \geq \alpha n,$$
which completes the proof of the lemma. □

This has the following simple corollary.

Corollary 8.3.4. *Let A be a random subset of $V(G)$ where each $v \in V(G)$ belongs to A with probability p_v, independently of the other vertices. Then almost surely $|A| \geq \alpha n$.*

Proof. By Lemma 8.3.3, $\mathbb{E}[|A|] \geq (1+\frac{\epsilon}{32})\alpha n$. By the Chernoff bound (see, for example, [50]),

$$\mathbb{P}\Big(|A| \leq \Big(1-\frac{\epsilon}{64}\Big)\mathbb{E}[|A|]\Big) \leq \exp\Big(-\frac{\epsilon^2\mathbb{E}[|A|]}{2\cdot 64^2}\Big) \leq \exp\Big(-\frac{(1+\frac{\epsilon}{32})\epsilon^2}{2\cdot 64^2}\alpha n\Big) = o(1)$$

since $\alpha = \omega(\frac{1}{n})$. Since $(1-\frac{\epsilon}{64})(1+\frac{\epsilon}{32}) \geq 1$, the corollary follows. □

To prove that for every $B \subset A$ of size at least βn, $G[B]$ is not too dense, we argue as follows. First, we prove that the expected number of paths in $G[A]$ of length roughly $(\log n)^2$ is small, so in particular it is small with probability at least $1/2$. However, if we have some $B \subset A$ of size at least βn such that $G[B]$ is fairly dense, then by our results from Section 8.2, $G[B]$, and consequently, $G[A]$, must contain many paths of length $(\log n)^2$, which is a contradiction.

Lemma 8.3.5. *Let A be a random subset of $V(G)$ where each $v \in V(G)$ belongs to A with probability p_v, independently of the other vertices. Then for any positive integer k,*

$$\mathbb{E}\Bigg[\sum_{v_0,\ldots,v_k \in A \text{ distinct}} w(v_0,v_1)w(v_1,v_2)\ldots w(v_{k-1},v_k)\Bigg] \le nD^k.$$

Proof. Let V' consist of the set of all vertices in $V(G)$ which have non-zero degree. Now

$$\begin{aligned}
&\mathbb{E}\Bigg[\sum_{v_0,\ldots,v_k \in A \text{ distinct}} w(v_0,v_1)w(v_1,v_2)\ldots w(v_{k-1},v_k)\Bigg]\\
&= \sum_{v_0,\ldots,v_k \in V' \text{ distinct}} w(v_0,v_1)w(v_1,v_2)\ldots w(v_{k-1},v_k)\mathbb{P}(v_0 \in A,\ldots,v_k \in A)\\
&= \sum_{v_0,\ldots,v_k \in V' \text{ distinct}} w(v_0,v_1)w(v_1,v_2)\ldots w(v_{k-1},v_k)p_{v_0}\ldots p_{v_k}\\
&\le \sum_{v_0,\ldots,v_k \in V' \text{ distinct}} w(v_0,v_1)w(v_1,v_2)\ldots w(v_{k-1},v_k)p_{v_0}\ldots p_{v_{k-1}}\\
&\le \sum_{v_0,\ldots,v_k \in V'} w(v_0,v_1)w(v_1,v_2)\ldots w(v_{k-1},v_k)\frac{D}{d_{v_0}}\cdot\cdots\cdot\frac{D}{d_{v_{k-1}}}\\
&= D^k \sum_{v_0 \in V'}\sum_{v_1,\ldots,v_k \in V'} \frac{w(v_0,v_1)}{d_{v_0}}\frac{w(v_1,v_2)}{d_{v_1}}\cdots\frac{w(v_{k-1},v_k)}{d_{v_{k-1}}}
\end{aligned}$$

However, for any $v_0 \in V'$,

$$\begin{aligned}
&\sum_{v_1,\ldots,v_k \in V'} \frac{w(v_0,v_1)}{d_{v_0}}\frac{w(v_1,v_2)}{d_{v_1}}\cdots\frac{w(v_{k-1},v_k)}{d_{v_{k-1}}}\\
&= \sum_{v_1,\ldots,v_{k-1} \in V'} \frac{w(v_0,v_1)}{d_{v_0}}\frac{w(v_1,v_2)}{d_{v_1}}\cdots\frac{w(v_{k-2},v_{k-1})}{d_{v_{k-2}}}\sum_{v_k \in V'}\frac{w(v_{k-1},v_k)}{d_{v_{k-1}}}\\
&= \sum_{v_1,\ldots,v_{k-1} \in V'} \frac{w(v_0,v_1)}{d_{v_0}}\frac{w(v_1,v_2)}{d_{v_1}}\cdots\frac{w(v_{k-2},v_{k-1})}{d_{v_{k-2}}}\\
&= 1
\end{aligned}$$

by induction on k. This completes the proof of the lemma. $\square$

Lemma 8.3.6. *Let A be a random subset of $V(G)$ where each $v \in V(G)$ belongs to A with probability p_v, independently of the other vertices. Then, with probability at least $1/2$, for every set $B \subset A$ of size at least βn, the average degree in $G[B]$ is at most $(1+\frac{\epsilon}{4})D$.*

Proof. Let $k = \lceil (\log n)^2 \rceil$ or $k = \lceil (\log n)^2 \rceil + 1$ chosen so that k is even. By Lemma 8.3.5, with probability at least $1/2$, we have

$$\sum_{v_0,\ldots,v_k \in A \text{ distinct}} w(v_0,v_1)w(v_1,v_2)\ldots w(v_{k-1},v_k) \le 2nD^k. \tag{8.2}$$

We claim that if this holds, then for every $B \subset A$ of size at least βn, the average degree of $G[B]$ is at most $(1+\frac{\epsilon}{4})D$. Indeed, suppose that there is some $B \subset A$ contradicting this claim. Then, using the condition $\beta\gamma = \omega\big(\frac{M(\log n)^{10}}{n}\big)$ in Theorem 8.3.1, $G[B]$ has average degree at least $(1+\frac{\epsilon}{4})D \geq D = (1-\frac{\epsilon}{2})\beta\gamma n = \omega(Mk^4(\log n)^2)$. Thus, by Lemma 8.2.4 applied to $G[B]$, we have

$$\sum_{v_0,\ldots,v_k \in B \text{ distinct}} w(v_0,v_1)w(v_1,v_2)\ldots w(v_{k-1},v_k) \geq |B|\big((1-o(1))(1+\frac{\epsilon}{4})D\big)^k$$
$$\geq \beta n\big((1+\frac{\epsilon}{4}-o(1))D\big)^k$$

Comparing this with equation (8.2), we obtain

$$2nD^k \geq \beta n\big((1+\frac{\epsilon}{4}-o(1))D\big)^k,$$

so

$$\frac{2}{\beta} \geq \big(1+\frac{\epsilon}{4}-o(1)\big)^k.$$

Since $k \geq (\log n)^2$ and $\beta \geq 1/n$, this is a contradiction. □

It is now easy to complete the proof of Theorem 8.3.1.

Proof of Theorem 8.3.1. Let A be a random subset of $V(G)$ where each $v \in V(G)$ belongs to A with probability p_v, independently of the other vertices.

By Corollary 8.3.4, almost surely $|A| \geq \alpha n$.

Moreover, by Lemma 8.3.6, with probability at least 1/2, every set $B \subset A$ of size at least βn has average degree at most $(1+\frac{\epsilon}{4})D \leq (1-\frac{\epsilon}{4})\beta\gamma n$. It follows that every such B has $e(G[B]) \leq (1-\frac{\epsilon}{4})\frac{\gamma|B|^2}{2}$.

So A is a suitable set with positive probability. □

8.4 Simple graphs

In this section we prove Theorem 8.1.4. It is easy to see that it suffices to prove the following result.

Theorem 8.4.1. *Let $0 < \beta \leq \alpha \leq 1/2$ and $0 < \gamma \leq 1$ be functions of n such that $\alpha - \beta \leq \frac{\beta^3\gamma}{1000}$ and $\beta^3\gamma = \omega(\frac{\log n}{n^{1/2}})$. Suppose that G is a simple graph on n vertices with $e(G) \leq (1-\epsilon)\frac{(1-\alpha+\beta)^2\gamma}{2}n^2$. Then there exists a set $A \subset V(G)$ of size at least αn such that for every $B \subset A$ of size at least βn, the number of edges in $G[B]$ is at most $(1-\frac{\epsilon}{4})\frac{\gamma|B|^2}{2}$.*

For the rest of this section, fix some $0 < \epsilon < 1/10$ constant, let n be sufficiently large and let G be a simple graph on n vertices with $e(G) \leq (1-\epsilon)\frac{(1-\alpha+\beta)^2\gamma}{2}n^2$, where α, β, γ satisfy the conditions described in the theorem. In what follows we shall ignore integrality issues as they do not make a genuine difference in the proofs.

Before turning to the proof of Theorem 8.4.1, let us outline the strategy, which is similar to that in the previous section, with one extra twist. Again, we will choose a suitable set A randomly. First we define probabilities p_v in a very similar fashion as before. If we defined p_v by the same formula as in the previous section, then they would not quite add up to αn because in this section the graph G has slightly more edges, hence higher average degree. Our key observation is the following. By the Chernoff bound, almost surely, for every $v \in V(G)$, the degree of v to A will be concentrated tightly around $\sum_{u \sim v} p_u$. Hence, intuitively it makes sense for us to keep those vertices v with probability 1 for which this sum is small, since those vertices will almost surely have low degree in $G[A]$. Accordingly, we define a new probability p'_v for each vertex v, which is the actual probability that we will keep the vertex with.

Definition 8.4.2. Recall from Definition 8.3.2 that $D = (1 - \frac{\epsilon}{2})\beta\gamma n$ and

$$p_v = \begin{cases} 1 & \text{if } d_v < D \\ \frac{D}{d_v} & \text{if } d_v \geq D. \end{cases}$$

Let $f(v) = \sum_{u \sim v} p_u$. Let T be the set of those $(\alpha - \beta)n$ vertices with the smallest value of $f(v)$. Let R be the subset of vertices $v \in T$ which have $d_v \geq \frac{\beta\gamma n}{2}$. Let $r = |R|$.

Define

$$p'_v = \begin{cases} 1 & \text{if } v \in R \text{ or } d_v < D' \\ \frac{D'}{d_v} & \text{otherwise} \end{cases}$$

where $D' = (1 - \frac{\epsilon}{2})(\beta\gamma n - \frac{\beta^2\gamma}{2(1-\alpha+\beta)^2} r)$.

We need to check that p'_v is not too far from p_v since we used $\sum_{u \sim v} p_u$ to estimate the degree of v to A.

Lemma 8.4.3. *If $v \notin R$, then $|p_v - p'_v| \leq \frac{\beta}{(1-\alpha+\beta)^2}\frac{r}{n}$.*

Proof. If $d_v < D'$, then $p_v = p'_v = 1$. If $D' \leq d_v < D$, then $p_v = 1$ and $p'_v = \frac{D'}{d_v}$. Then $|p_v - p'_v| = \frac{|d_v - D'|}{d_v} \leq \frac{D-D'}{D'}$. Note that $\alpha - \beta \leq \frac{\beta^3\gamma}{1000} \leq \frac{1}{1000}$, so

$$\begin{aligned} D' &= \left(1 - \frac{\epsilon}{2}\right)\left(\beta\gamma n - \frac{\beta^2\gamma}{2(1-\alpha+\beta)^2} r\right) \\ &\geq \left(1 - \frac{\epsilon}{2}\right)\left(\beta\gamma n - \frac{\beta^2\gamma}{2(1-\alpha+\beta)^2}(\alpha-\beta)n\right) \geq \frac{\beta\gamma n}{2}. \end{aligned}$$

Thus, $\frac{D-D'}{D'} \leq \frac{(1-\frac{\epsilon}{2})\frac{\beta^2\gamma}{2(1-\alpha+\beta)^2} r}{\frac{\beta\gamma n}{2}} \leq \frac{\beta}{(1-\alpha+\beta)^2}\frac{r}{n}$. Finally, if $d_v \geq D$, then $p_v = \frac{D}{d_v}$ and $p_v = \frac{D'}{d_v}$, so $|p_v - p'_v| = \frac{D-D'}{d_v} \leq \frac{D-D'}{D} \leq \frac{D-D'}{D'}$, which is at most $\frac{\beta}{(1-\alpha+\beta)^2}\frac{r}{n}$ as before. $\square$

Similarly to the previous section, we need to check that the expected size of A is large enough.

Lemma 8.4.4.
$$\sum_{v\in V(G)} p'_v \geq \Big(1+\frac{\epsilon}{64}\Big)\alpha n.$$

Proof. Let S be the set of those $v \in V(G)$ which have $p'_v = 1$. We claim that $S \supset T$. Indeed, if $v \in T$, then either $d_v \geq \frac{\beta\gamma n}{2}$ and $v \in R$, or $d_v < \frac{\beta\gamma n}{2} \leq D'$. Let $|S| = s$. Thus, $s \geq |T| = (\alpha - \beta)n$. Then

$$\begin{aligned}
\sum_{v\in V(G)} p'_v &= s + \sum_{v:p'_v<1} \frac{D'}{d_v} \geq s + (n-s)\frac{D'}{(\sum_{v:p'_v<1} d_v)/(n-s)} = s + \frac{(n-s)^2 D'}{\sum_{v:p'_v<1} d_v}\\
&\geq s + \frac{(n-s)^2 D'}{2e(G) - r\frac{\beta\gamma n}{2}} \geq s + \frac{(n-s)^2 D'}{(1-\epsilon)(1-\alpha+\beta)^2\gamma n^2 - r\frac{\beta\gamma n}{2}}\\
&\geq s + \frac{(n-s)^2 D'}{(1-\epsilon)((1-\alpha+\beta)^2\gamma n^2 - r\frac{\beta\gamma n}{2})}\\
&= s + \frac{1-\epsilon/2}{1-\epsilon}(n-s)^2\frac{\beta}{(1-\alpha+\beta)^2 n}\\
&\geq s + (1+\epsilon/2)(n-s)^2\frac{\beta}{(1-\alpha+\beta)^2 n}\\
&= s + (n-s)^2\frac{\beta}{(1-\alpha+\beta)^2 n} + \frac{\epsilon}{2}(n-s)^2\frac{\beta}{(1-\alpha+\beta)^2 n}.
\end{aligned}\tag{8.3}$$

Now note that $s + (n-s)^2\frac{\beta}{(1-\alpha+\beta)^2 n}$ is increasing in s in the range $s \geq 0$. Indeed, the coefficient of s in this expression is $1 - \frac{2\beta}{(1-\alpha+\beta)^2}$, which is non-negative as $\alpha \leq 1/2$.

Thus, since we know that $s = |S| \geq (\alpha-\beta)n$, we have

$$s + (n-s)^2\frac{\beta}{(1-\alpha+\beta)^2 n} \geq (\alpha-\beta)n + \beta n = \alpha n.$$

If $s \geq (1+\frac{\epsilon}{64})\alpha n$, then clearly (8.3) is at least $(1+\frac{\epsilon}{64})\alpha n$. Otherwise, by $\alpha \leq 1/2$, we have $s \leq \frac{1+\epsilon/64}{2}n < \frac{3}{4}n$. Then

$$\frac{\epsilon}{2}(n-s)^2\frac{\beta}{(1-\alpha+\beta)^2 n} \geq \frac{\epsilon}{32}\beta n \geq \frac{\epsilon}{64}\alpha n.$$

This completes the proof. □

Corollary 8.4.5. *Let A be a random subset of $V(G)$ where each $v \in V(G)$ belongs to A with probability p'_v, independently of the other vertices. Then almost surely $|A| \geq \alpha n$.*

Define G' to be the graph obtained by discarding all edges of G with at least one endpoint in R. Let d'_v be the degree of the vertex $v \in V(G') = V(G)$ in G'. For distinct vertices $u, v \in V(G)$, write $u \sim_{G'} v$ if uv is an edge in G'. The next lemma shows that the expected number of paths in G' is small. We will use this to conclude that G' does not have large dense subgraphs.

Lemma 8.4.6. *Let A be a random subset of $V(G)$ where each $v \in V(G)$ belongs to A with probability p'_v, independently of the other vertices. Then for any positive integer k,*

$$\mathbb{E}\left[\sum_{v_0,\ldots,v_k \in A \text{ distinct}} \mathbb{1}\{v_0 \sim_{G'} v_1 \sim_{G'} \cdots \sim_{G'} v_k\}\right] \le n(D')^k.$$

Proof. Note that for every $v \in V(G')$, $p'_v \le \frac{D'}{d'_v}$. So the proof of Lemma 8.3.5 (in the special case where the graph is a simple graph) works here. $\square$

Lemma 8.4.7. *Let A be a random subset of $V(G)$ where each $v \in V(G)$ belongs to A with probability p'_v, independently of the other vertices. Then, with probability at least $1/2$, for every set $B \subset A$ of size at least βn, the average degree in $G'[B]$ is at most $(1+\frac{\epsilon}{4})D'$.*

The proof of this is nearly identical to the proof of Lemma 8.3.6, so it is omitted.

Proof of Theorem 8.4.1. Let A be a random subset of $V(G)$ where each $v \in V(G)$ belongs to A with probability p'_v, independently of the other vertices. By Corollary 8.4.5 and Lemma 8.4.7, with positive probability $|A| \ge \alpha n$ and for every $B' \subset A$ of size at least βn we have

$$e(G'[B']) \le \frac{|B'| \cdot (1+\frac{\epsilon}{4})D'}{2} \le \left(1-\frac{\epsilon}{4}\right)\left(1-\frac{\beta}{2(1-\alpha+\beta)^2}\frac{r}{n}\right)\gamma\frac{|B'|^2}{2}. \tag{8.4}$$

Assume that there exists some $B \subset A$ with $|B| \ge \beta n$ and $e(G[B]) > (1-\frac{\epsilon}{4})\gamma\frac{|B|^2}{2}$. By passing to a suitable subset, we may assume that $|B| = \beta n$. Recall the sets $R, T \subset V(G)$ from Definition 8.4.2. Note that $|R| = r \le (\alpha-\beta)n$. Let $U = B \cap R$. Choose $U' \subset (A \setminus B) \setminus T$ arbitrarily with $|U'| = |U|$. Note that this is possible since $|(A \setminus B) \setminus T| = |A| - |B| - |T| + |B \cap T| \ge \alpha n - \beta n - (\alpha-\beta)n + |B \cap T| \ge |B \cap R| = |U|$. Let $B' = (B \setminus U) \cup U'$.

Claim. Almost surely we have $e(G[B]) - e(G[B']) \le \frac{\beta^3\gamma}{8(1-\alpha+\beta)^2}rn$.

Given this claim, since $B' \cap R = \emptyset$ we have

$$\begin{aligned} e(G'[B']) = e(G[B']) &> \left(1-\frac{\epsilon}{4}\right)\gamma\frac{(\beta n)^2}{2} - \frac{\beta^3\gamma}{8(1-\alpha+\beta)^2}rn \\ &\ge \left(1-\frac{\epsilon}{4}\right)\left(1-\frac{\beta}{2(1-\alpha+\beta)^2}\frac{r}{n}\right)\gamma\frac{(\beta n)^2}{2}, \end{aligned}$$

which contradicts equation (8.4).

Thus, we are left to prove the claim.

Proof of Claim. Note that

$$e(G[B]) - e(G[B']) \le e(G[U]) + \sum_{v \in U} d_v(B \cap B') - \sum_{v \in U'} d_v(B \cap B'), \tag{8.5}$$

where $d_v(S)$ denotes the number of neighbours of v in the set S. Since $|U| \le |R| = r$, we have

$$e(G[U]) \le \binom{r}{2}. \tag{8.6}$$

Also, $|B \cap B'| \ge \beta n - r$, so $|A \setminus (B \cap B')| \le (\alpha - \beta)n + r \le 2(\alpha - \beta)n$. Hence, for any $v \in V(G)$ we have $d_v(B \cap B') \ge d_v(A) - 2(\alpha - \beta)n$, so

$$\sum_{v \in U'} d_v(B \cap B') \ge \sum_{v \in U'} d_v(A) - 2r(\alpha - \beta)n. \tag{8.7}$$

Note that by Chernoff's bound (e.g. Corollary 5 in [50]),

$$\mathbb{P}\Big(|d_v(A) - \mathbb{E}[d_v(A)]| \ge (\log n)(\mathbb{E}[d_v(A)])^{1/2}\Big) \le 2\exp(-(\log n)^2/3) = o(1/n).$$

Thus, almost surely we have

$$|d_v(A) - \mathbb{E}[d_v(A)]| \le (\log n)(\mathbb{E}[d_v(A)])^{1/2} \le n^{1/2}\log n \tag{8.8}$$

for every $v \in V(G)$. Note that $\mathbb{E}[d_v(A)] = \sum_{u \sim_G v} p'_u$. By Lemma 8.4.3, we have

$$|\sum_{u \sim_G v} p'_u - \sum_{u \sim_G v} p_u| \le n \cdot \frac{\beta}{(1 - \alpha + \beta)^2}\frac{r}{n} + r = \left(\frac{\beta}{(1 - \alpha + \beta)^2} + 1\right) r. \tag{8.9}$$

However, by the definition of T, for every $v \in T$ and $v' \notin T$ we have $\sum_{u \sim_G v} p_u \le \sum_{u \sim_G v'} p_u$, so the same holds for every $v \in U$ and every $v' \in U'$. Thus, by equations (8.8) and (8.9), for every $v \in U$ and every $v' \in U'$, $d_v(A) \le d_{v'}(A) + 2(\frac{\beta}{(1-\alpha+\beta)^2} + 1)r + 2n^{1/2}\log n$, and therefore

$$\sum_{v \in U} d_v(A) \le \sum_{v \in U'} d_v(A) + 2\left(\frac{\beta}{(1 - \alpha + \beta)^2} + 1\right) r^2 + 2rn^{1/2}\log n. \tag{8.10}$$

Now by equations (8.5), (8.6), (8.7) and (8.10), we get

$$\begin{aligned} e(G[B]) - e(G[B']) &\le \binom{r}{2} + 2r(\alpha - \beta)n + 2\left(\frac{\beta}{(1 - \alpha + \beta)^2} + 1\right) r^2 + 2rn^{1/2}\log n \\ &\le r(\alpha - \beta)n + 2r(\alpha - \beta)n + 2\left(\frac{\beta}{(1 - \alpha + \beta)^2} + 1\right) r(\alpha - \beta)n + 2rn^{1/2}\log n. \end{aligned}$$

Note that by the condition $\alpha - \beta \le \frac{\beta^3\gamma}{1000}$ in the theorem, the first three terms on the right hand side are at most $\frac{\beta^3\gamma rn}{30}$ each, while by the condition $\beta^3\gamma = \omega(\frac{\log n}{n^{1/2}})$, the fourth term is $o(\beta^3\gamma rn)$, which completes the proof of the claim. $\square$

8.5 The bipartite case

Proof of Proposition 8.1.6. First, note that the condition $\beta\gamma = \omega(\frac{1}{n})$ implies $\beta n = \omega(1)$ and therefore also $\alpha n = \omega(1)$ and $\frac{\beta n}{\alpha} = \omega(1)$.

Let X and Y be sets of size n; these will be the parts of G_n. Let $m = \lceil K\frac{\alpha}{\beta}\log(1/\alpha)\rceil$ for some large absolute constant K. For each $1 \le i \le m$, define subsets $A_i \subset X$ and $B_i \subset Y$ of size $\lfloor\frac{2\beta}{\alpha}n\rfloor$ uniformly at random (and independently). The graph H_i is then defined by picking each edge between A_i and B_i independently at random with probability $\min(2\gamma, 1)$. Finally, set $G = G_n = \cup_{i\le m}H_i$. Notice that almost surely the bipartite edge density of G is at most $m(\frac{2\beta}{\alpha})^2 4\gamma \le 20K\frac{\beta\gamma}{\alpha}\log(1/\alpha)$.

Fix some $A \subset X$ and some $B \subset Y$ of size $\lceil\alpha n\rceil$ each. Note that $\mathbb{P}[|A \cap A_1| \le \beta n] \le \exp(-c\beta n)$ for some absolute constant $c > 0$. Similarly, $\mathbb{P}[|B \cap B_1| \le \beta n] \le \exp(-c\beta n)$. Now condition on the event that $|A \cap A_1|, |B \cap B_1| \ge \beta n$. The probability that the bipartite graph $H_1[A\cap A_1, B\cap B_1]$ has edge density less than γ is at most $\exp(-c'\gamma(\beta n)^2)$ for some absolute constant $c' > 0$. Thus, the probability that there do not exist $U \subset A$ and $V \subset B$ with $|U|, |V| \ge \beta n$ such that the edge density of $H_1[U, V]$ is at least γ is at most $2\exp(-c\beta n) + \exp(-c'\gamma(\beta n)^2)$. Since $\beta\gamma n = \omega(1)$, for n sufficiently large it is at most $3\exp(-c\beta n)$. Hence, the probability that the pair (A, B) witnesses that G does not have the bipartite (α, β, γ)-property is at most $(3\exp(-c\beta n))^m \le \exp(-\frac{c}{2}\beta nm) \le \exp(-10\alpha\log(1/\alpha)n)$ if K is sufficiently large. But the number of choices for $A \subset X$ and $B \subset Y$ of size $\lceil\alpha n\rceil$ each is $\binom{n}{\lceil\alpha n\rceil}^2 \le (e/\alpha)^{2\lceil\alpha n\rceil} = \exp(\log(e/\alpha)2\lceil\alpha n\rceil) = \exp((1 + \log(1/\alpha))2\lceil\alpha n\rceil) \le \exp(6\log(1/\alpha)\lceil\alpha n\rceil)$. It follows by the union bound that almost surely G has the bipartite (α, β, γ)-property. $\square$

We now turn to the proof of Theorem 8.1.7.

Lemma 8.5.1. *Let $0 < \beta < \alpha \le 1$ and $0 < \gamma \le 1$ be parameters depending on $n \in \mathbb{N}$ such that $\alpha\beta\gamma n = \omega((\log n)^3)$ and $\beta^2\gamma n = \omega(\log n)$. Let n be sufficiently large and let G be a bipartite graph with vertex sets X and Y such that $|X| = n$ and $|Y| \le n$. Suppose that for all $A \subset X$ with $|A| \ge \alpha n$ there exist $B \subset A$ and $C \subset Y$ such that $|B| \ge \beta n$, $|C| \ge m \ge \beta n$ and $G[B, C]$ has edge density at least γ. Then there exist $D \subset X$ and $E \subset Y$ with $|D| \ge \frac{\beta}{\alpha}n$ and $|E| \ge m$ such that $G[D, E]$ has edge density at least $\gamma/400$.*

We remark that we will only need the weaker condition that for a positive proportion of all $A \subset X$ with $|A| \ge \alpha n$ there exist $B \subset A$ and $C \subset Y$ with the above properties.

The proof of the lemma is somewhat technical, so let us give a brief sketch first. For simplicity, we assume that α, β and γ are constants not depending on n. For a set $S \subset Y$ and a positive integer $j \le \log n$, write $N_j(S)$ for the set of vertices $x \in X$ which have roughly $2^{-j}|S|$ neighbours in S. This way, for each $S \subset Y$ we get a dyadic partition of X. Let us consider all $S \subset Y$ of size roughly $\log n$. The number of such sets is at most roughly $\exp((\log n)^2)$. Thus, if A is a uniformly random subset of X of size αn, then by the union bound and Chernoff bound, almost surely $|N_j(S) \cap A| \approx \alpha|N_j(S)|$ for every S of size roughly $\log n$ and every $j \le \log n$. (This fails if $N_j(S)$ is very small, but let us ignore this issue here.) Thus, we can take some $A \subset X$ for which all these approximations hold. By assumption, there exist $B \subset A$ and $E \subset Y$ with $|B| \geq \beta n$ and $|E| \geq m$ such that $G[B, E]$ has density at least γ. This means that $\sum_{j \leq \log n} |N_j(E) \cap B| \cdot 2^{-j}|E| \approx e(G[B, E]) \geq \gamma|B||E|$. Let S be a uniformly random subset of E of size $\log n$. If $x \in N_j(E)$, then almost surely $x \in N_j(S)$ (note that this is not quite true; it can happen with non-negligible probability that $x \in N_{j-1}(S)$ or $x \in N_{j+1}(S)$, but this is easily dealt with in the proof below). Thus, with a bit of oversimplification $N_j(S)$ and $N_j(E)$ are roughly the same sets. Since $N_j(S) \cap B \subset N_j(S) \cap A$ and $|N_j(S) \cap A| \approx \alpha|N_j(S)|$, we may choose $D_j \subset N_j(S)$ such that $|D_j| \approx \frac{|N_j(S) \cap B|}{\alpha}$. Set $D = \cup_{j \leq \log n} D_j$. Then $|D| = \sum_{j \leq \log n} |D_j| \approx \frac{|B|}{\alpha} \geq \frac{\beta}{\alpha} n$. Since $D_j = N_j(S) \cap D$, we have (roughly) $D_j = N_j(E) \cap D$. Thus, $e(G[D, E]) \approx \sum_{j \leq \log n} |N_j(E) \cap D| \cdot 2^{-j}|E| \approx \sum_{j \leq \log n} |D_j| \cdot 2^{-j}|E| \approx \sum_{j \leq \log n} \frac{|N_j(S) \cap B|}{\alpha} \cdot 2^{-j}|E| \approx \frac{1}{\alpha} \sum_{j \leq \log n} |N_j(E) \cap B| \cdot 2^{-j}|E| \approx \frac{1}{\alpha} e(G[B, E]) \geq \frac{1}{\alpha}\gamma|B||E| \approx \gamma|D||E|$, so D and E are suitable.

The proof below makes this sketch precise without any significant new ideas. Note that we only get density $\gamma/400$ between D and E. However, by refining the partition $N_j(S)$ so that $N_j(S) = \{x \in X : \lambda^j|S| < |\Gamma_G(x) \cap S| \leq \lambda^{j-1}|S|\}$ for some $\lambda = 1 - o(1)$, we could obtain density $(1 - o(1))\gamma$ in $G[D, E]$, at least when α, β, γ are constants. This will be done (in the more general setting of r-partite graphs) in the next section. Nevertheless, we include this proof here as it is easier to read than the more general one later.

Proof of Lemma 8.5.1. For a set $S \subset Y$ and an integer $1 \leq j \leq \lceil \log_2 n \rceil + 1$, define $N_j(S)$ to be $\{x \in X : 2^{-j}|S| < |\Gamma_G(x) \cap S| \leq 2^{-(j-1)}|S|\}$. For $1 \leq j \leq \lceil \log_2 n \rceil + 1$, let $s_j = \lceil K \cdot 2^j \log n \rceil$ for a sufficiently large absolute constant K, and let $t_j = \lfloor 2^j \frac{\gamma \beta n}{100 \log_2 n} \rfloor$.

Let A be a uniformly random subset of X of size $\lceil \alpha n \rceil$. If a set $R \subset X$ has size at least t_j, then by the Chernoff bound

$$\mathbb{P}[|R \cap A| \geq 2\alpha|R|] \leq \exp(-c\alpha|R|) \leq \exp(-c\alpha t_j) \tag{8.11}$$

for some absolute constant $c > 0$. On the other hand, the number of sets of size s_j in Y is at most $\binom{n}{s_j} \leq n^{s_j} = \exp(s_j \log n)$. Let F_j be the event that there exists some $S_j \subset Y$ of size s_j such that $|N_{j-1}(S_j) \cup N_j(S_j) \cup N_{j+1}(S_j)| \geq t_j$ and $|N_{j-1}(S_j) \cup N_j(S_j) \cup N_{j+1}(S_j) \cap A| \geq 2\alpha|N_{j-1}(S_j) \cup N_j(S_j) \cup N_{j+1}(S_j)|$. From the assumption $\alpha\beta\gamma n = \omega((\log n)^3)$ it follows that $c\alpha t_j = \omega(s_j \log n)$, therefore by (8.11), for n sufficiently large we have $\mathbb{P}(F_j) < \frac{1}{10 \log_2 n}$ say. Thus, with positive probability none of the events F_j $(1 \leq j \leq \lceil \log_2 n \rceil + 1)$ occurs.

Pick some set $A \subset X$ of size $\lceil \alpha n \rceil$ for which this is the case. By assumption, there are $B \subset A$ and $E \subset Y$ such that $|B| = \lceil \beta n \rceil$, $|E| \geq m$ and $G[B, E]$ has edge density at least γ. Then

$$\sum_{j=1}^{\lceil \log_2 n \rceil + 1} |N_j(E) \cap B| \cdot 2^{-j+1}|E| \geq e(G[B, E]) \geq \gamma|B||E| \geq \beta\gamma n|E|.$$

After dividing both sides by $|E|$ and since $|N_j(E) \cap B| \le n$ for every j, we get

$$\sum_{j=1}^{\lceil -\log_2(\beta\gamma/4) \rceil} |N_j(E) \cap B| \cdot 2^{-j+1} \ge \beta\gamma n/2.$$

Thus, there is some $\epsilon \in \{0, 1, 2, 3, 4\}$ such that

$$\sum_{\substack{1 \le j \le \lceil -\log_2(\beta\gamma/4) \rceil \\ j \equiv \epsilon \mod 5}} |N_j(E) \cap B| \cdot 2^{-j+1} \ge \beta\gamma n/10.$$

Suppose that $|N_j(E) \cap B| < t_j$. Then by the definition of t_j, we have $|N_j(E) \cap B| \cdot 2^{-j+1} < \frac{\beta\gamma n}{50 \log_2 n}$. Thus

$$\sum_{\substack{1 \le j \le \lceil -\log_2(\beta\gamma/4) \rceil \\ j \equiv \epsilon \mod 5 \\ |N_j(E) \cap B| \ge t_j}} |N_j(E) \cap B| \cdot 2^{-j+1} \ge \beta\gamma n/20. \tag{8.12}$$

Claim. If j satisfies $1 \le j \le \lceil -\log_2(\beta\gamma/4) \rceil$ and $|N_j(E) \cap B| \ge t_j$, then we have $|N_{j-2}(E) \cup N_{j-1}(E) \cup N_j(E) \cup N_{j+1}(E) \cup N_{j+2}(E)| \ge \frac{1}{2\alpha} |N_j(E) \cap B|$.

Proof of Claim. First notice that since $j \le \lceil -\log_2(\beta\gamma/4) \rceil$, we have $s_j \le \lceil 2K \frac{4}{\beta\gamma} \log n \rceil$. Therefore the condition $\beta^2\gamma n = \omega(\log n)$ implies that for n sufficiently large we have $s_j \le \beta n \le m \le |E|$. Pick a random subset $S \subset E$ of size s_j. Let $x \in N_j(E)$. Then, by the Chernoff bound, $\mathbb{P}[x \notin N_{j-1}(S) \cup N_j(S) \cup N_{j+1}(S)] \le \exp(-c2^{-j} s_j)$ for some absolute constant $c > 0$. This is at most $1/n^2$ for K sufficiently large. On the other hand, again by the Chernoff bound, if $|i - j| \ge 3$ and $x \in N_i(E)$, then $\mathbb{P}[x \in N_{j-1}(S) \cup N_j(S) \cup N_{j+1}(S)] \le \exp(-c' 2^{-j} s_j) \le 1/n^2$. Thus, almost surely we have that

$$N_j(E) \subset N_{j-1}(S) \cup N_j(S) \cup N_{j+1}(S) \subset N_{j-2}(E) \cup N_{j-1}(E) \cup N_j(E) \cup N_{j+1}(E) \cup N_{j+2}(E).$$

In particular, $|N_{j-1}(S) \cup N_j(S) \cup N_{j+1}(S)| \ge |N_j(E)| \ge |N_j(E) \cap B| \ge t_j$, therefore $|N_{j-1}(S) \cup N_j(S) \cup N_{j+1}(S) \cap A| \le 2\alpha |N_{j-1}(S) \cup N_j(S) \cup N_{j+1}(S)|$ from the definition of A. Thus,

$$\begin{aligned}
&|N_{j-2}(E) \cup N_{j-1}(E) \cup N_j(E) \cup N_{j+1}(E) \cup N_{j+2}(E)| \\
&\ge |N_{j-1}(S) \cup N_j(S) \cup N_{j+1}(S)| \\
&\ge \frac{1}{2\alpha} |N_{j-1}(S) \cup N_j(S) \cup N_{j+1}(S) \cap A| \\
&\ge \frac{1}{2\alpha} |N_j(E) \cap A| \\
&\ge \frac{1}{2\alpha} |N_j(E) \cap B|.
\end{aligned}$$

This completes the proof of the claim.

By the claim, if $1 \le j \le \lceil -\log_2(\beta\gamma/4) \rceil$ and $|N_j(E) \cap B| \ge t_j$, then we may choose sets $D_j \subset N_{j-2}(E) \cup \cdots \cup N_{j+2}(E)$ with $|D_j| = \lceil \frac{1}{2\alpha} |N_j(E) \cap B| \rceil$. Note that these are pairwise disjoint for all $j \equiv \epsilon \mod 5$. By (8.12), we have

$$\sum_{\substack{1 \le j \le \lceil -\log_2(\beta\gamma/4) \rceil \\ j \equiv \epsilon \mod 5 \\ |N_j(E) \cap B| \ge t_j}} |D_j| \cdot 2^{-j+1} \ge \frac{\beta\gamma n}{40\alpha}.$$

Since any $x \in D_j$ has at least $2^{-j-2}|E|$ neighbours in E, we get

$$e\big(G\big[\bigcup_j D_j, E\big]\big) \ge \sum_j |D_j| \cdot 2^{-j-2}|E| \ge \frac{\beta\gamma n|E|}{320\alpha}.$$

On the other hand, $\sum_j |D_j| \le \frac{1}{2\alpha} \sum_j |N_j(E) \cap B| + \lceil \log_2 n \rceil + 1 \le \frac{|B|}{2\alpha} + \lceil \log_2 n \rceil + 1 \le \frac{\beta n}{\alpha}$ for n sufficiently large. So we may choose arbitrary $D \supset \cup_j D_j$ with $|D| = \lceil \frac{\beta n}{\alpha} \rceil$; then $G[D, E]$ has edge density at least $\frac{\gamma}{400}$. □

Corollary 8.5.2. *Let α, β, γ satisfy the conditions described in Lemma 8.5.1 and let n be sufficiently large. Let G be a bipartite graph with vertex sets U, V of size n having the bipartite (α, β, γ)-property. Then there exist $S \subset U$ and $T \subset V$ of size at least $\frac{\beta}{\alpha} n$ such that $G[S, T]$ has density at least $\gamma/400^2$.*

Proof. Fix $R \subset V$ with $|R| \ge \alpha n$. By the bipartite (α, β, γ)-property of G, Lemma 8.5.1 applies with $X = U, Y = R$ and $m = \beta n$. Thus, we get that for every $R \subset V$ with $|R| \ge \alpha n$ there are $A \subset R$ and $B \subset X$ with $|A| \ge \beta n$ and $|B| \ge \frac{\beta}{\alpha} n$ such that the edge density of $G[A, B]$ is at least $\gamma/400$. But now we can apply Lemma 8.5.1 with $X = V, Y = U, m = \frac{\beta}{\alpha} n$ and $\gamma/400$ in place of γ to find S and T with the required properties. □

Corollary 8.5.3. *Let α, β, γ satisfy the conditions described in Lemma 8.5.1 and assume in addition that $\beta/\alpha \le 1/10$. Let n be sufficiently large and let G be a bipartite graph with vertex sets X, Y of size n having the bipartite (α, β, γ)-property. Then there exist pairwise disjoint sets $S_1, \ldots, S_t \subset X$ and $T_1, \ldots, T_t \subset Y$ such that each S_i and T_j has size $\lceil \frac{\beta}{2\alpha} n \rceil$, $G[S_i, T_i]$ has edge density at least $\gamma/400^2$ for each i, t is even and $t \ge \frac{\alpha}{2\beta}$.*

Proof. Suppose we have chosen the first $k \le \frac{\alpha}{2\beta} + 1$ sets S_i and T_i. Let $X' = X \setminus \cup_{i \le k} S_i$ and $Y' = Y \setminus \cup_{i \le k} T_i$. Then $|X'| \ge n - (\frac{\alpha}{2\beta} + 1)\lceil \frac{\beta}{2\alpha} n \rceil \ge n/2$ and similarly $|Y'| \ge n/2$. Let $G' = G[X', Y']$. Then G' has the bipartite $(\alpha', \beta', \gamma)$-property for $\alpha' = \frac{|X|}{|X'|}\alpha$ and $\beta' = \frac{|X|}{|X'|}\beta$. Thus, by Corollary 8.5.2, there exist $S_{k+1} \subset X'$ and $T_{k+1} \subset Y'$ of size at least $\frac{\beta'}{\alpha'}|X'| \ge \frac{\beta}{2\alpha} n$ such that $G[S_{k+1}, T_{k+1}]$ has edge density at least $\gamma/400^2$. We can now replace both sets with suitable subsets of size $\lceil \frac{\beta}{2\alpha} n \rceil$ such that the density condition is still satisfied. □

We are in a position to complete the proof of Theorem 8.1.7.

Proof of Theorem 8.1.7. The proof goes by induction on $\log_2(1/\alpha)$. Let us first assume that $1/4 < \alpha \leq 1/2$. It is enough to prove that there is an absolute constant $c' > 0$ such that the edge density of G is at least $c'\beta\gamma$. Let X and Y be the parts of G. It is not hard to see that we may find pairwise disjoint sets $S_1, \ldots, S_k \subset X$ and $T_1, \ldots, T_k \subset Y$ of size $\lceil \beta n \rceil$ each such that $k\lceil \beta n \rceil \geq n/2$ and $G[S_i, T_i]$ has edge density at least γ for each i. This implies that $e(G) \geq k(\lceil \beta n \rceil)^2\gamma \geq \frac{\beta\gamma}{2}n^2$.

Suppose now that $\alpha \leq 1/4$. Choose sets S_i and T_j as provided by Corollary 8.5.3. Write $t = 2r$. Choose a uniformly random subset $I \subset [2r]$ of size r. Assume, for convenience, that n is even. Let $X'_1 = \bigcup_{i\in I} S_i$ and $Y'_1 = \bigcup_{i\in[2r]\setminus I} T_i$. Moreover, let X'_2 be a uniformly random subset of $X \setminus \bigcup_i S_i$ of size $\frac{1}{2}|X \setminus \bigcup_i S_i|$ and let Y'_2 be a uniformly random subset of $Y \setminus \bigcup_i T_i$ of size $\frac{1}{2}|Y \setminus \bigcup_i T_i|$. Let $X' = X'_1 \cup X'_2$ and let $Y' = Y'_1 \cup Y'_2$. Consider the graph $G' = G[X', Y']$. Note that $|X'| = |X|/2$ and $|Y'| = |Y|/2$, therefore G' is a bipartite graph on $n/2 + n/2$ vertices having the $(2\alpha, 2\beta, \gamma)$-property. By the induction hypothesis, we have

$$e(G') \geq c\frac{\beta\gamma}{\alpha}(\log_2(1/\alpha) - 1)\frac{n^2}{4} \tag{8.13}$$

We shall now estimate $\mathbb{E}[e(G')]$. Any edge in G which has an endpoint outside $\bigcup_i S_i \cup \bigcup_j T_j$ is present in G' with probability exactly $1/4$. Any edge in $G[S_i, T_i]$ for some i is present in G' with probability 0. Finally, any edge in $G[S_i, T_j]$ for some $i \neq j$ is present in G' with probability $\frac{r}{2(2r-1)} = 1/4 + \frac{1}{4(2r-1)}$. Since $\sum_i e(G[S_i, T_i]) \geq 2r \cdot \frac{\gamma}{400^2}(\frac{\beta}{2\alpha}n)^2$, it follows that

$$\mathbb{E}[e(G')] \leq \left(1/4 + \frac{1}{4(2r-1)}\right)e(G) - \frac{1}{4}2r\frac{\gamma}{400^2}\left(\frac{\beta}{2\alpha}n\right)^2.$$

Thus, by (8.13),

$$c\frac{\beta\gamma}{\alpha}(\log_2(1/\alpha) - 1)\frac{n^2}{4} \leq \left(1/4 + \frac{1}{4(2r-1)}\right)e(G) - \frac{1}{4}2r\frac{\gamma}{400^2}\left(\frac{\beta}{2\alpha}n\right)^2.$$

Recall that our aim is to prove that $e(G) \geq c\frac{\beta\gamma}{\alpha}\log_2(1/\alpha)n^2$. Suppose for contradiction that $e(G) < c\frac{\beta\gamma}{\alpha}\log_2(1/\alpha)n^2$. Then we get

$$c\frac{\beta\gamma}{\alpha}(\log_2(1/\alpha) - 1)\frac{n^2}{4} < \left(1/4 + \frac{1}{4(2r-1)}\right)c\frac{\beta\gamma}{\alpha}\log_2(1/\alpha)n^2 - \frac{1}{4}2r\frac{\gamma}{400^2}\left(\frac{\beta}{2\alpha}n\right)^2,$$

so

$$-c\frac{\beta\gamma}{\alpha}\frac{n^2}{4} < \frac{1}{4(2r-1)}c\frac{\beta\gamma}{\alpha}\log_2(1/\alpha)n^2 - \frac{1}{4}2r\frac{\gamma}{400^2}\left(\frac{\beta}{2\alpha}n\right)^2.$$

Dividing through by $\frac{\beta\gamma}{4\alpha}n^2$, we get

$$-c < \frac{1}{(2r-1)}c\log_2(1/\alpha) - 2r\frac{1}{400^2}\frac{\beta}{4\alpha},$$

so

$$\frac{\beta}{2\cdot 400^2\alpha}r < \left(\frac{\log_2(1/\alpha)}{2r-1}+1\right)c. \tag{8.14}$$

Since $r = t/2 \geq \frac{\alpha}{4\beta}$, the left hand side is at least $\frac{1}{8\cdot 400^2}$. On the other hand, the right hand side is at most $(\frac{1}{r}\log_2(1/\alpha)+1)c \leq (\frac{4\beta}{\alpha}\log_2(1/\alpha)+1)c \leq (4/10+1)c$. Thus, if c is a sufficiently small absolute constant (eg. $c = \frac{1}{16\cdot 400^2}$ would do), then (8.14) does not hold. This completes the proof of the theorem. □

8.6 Structural results

In this section we prove Theorem 8.1.8. We first need to generalize Lemma 8.5.1 from the previous section to r-partite graphs for all r. We also need to be more careful and obtain edge density $(1-o(1))\gamma$ between our sets, rather than $\gamma/400$ as in Lemma 8.5.1. These make the notation in the proof rather involved, but the main ideas are essentially the same.

Lemma 8.6.1. *Let r be a positive integer (independent of n) and let $\alpha, \beta, \gamma = \Omega(1)$ be functions of n. Let G be an r-partite graph with parts $X, Y_1, \ldots, Y_{r-1}$ such that $|X| = n$ and $|Y_1|, \ldots, |Y_{r-1}| \leq n$. Assume that for at least half of all $A \subset X$ of size $\lceil \alpha n \rceil$ there exist $B \subset A$ and $C_1 \subset Y_1$, $C_2 \subset Y_2$, $\ldots$, $C_{r-1} \subset Y_{r-1}$ such that $|B| \geq \beta n$, $|C_i| \geq m_i \geq \beta n$ for every i, and between any two of $B, C_1, \ldots, C_{r-1}$, the bipartite subgraph induced by G has density at least γ. Then there exist $D \subset X$ and $E_i \subset Y_i$ for every i such that $|D| \geq \frac{\beta}{\alpha}n$, $|C_i| \geq m_i$ for every i, and between any two of $D, E_1, \ldots, E_{r-1}$, the bipartite subgraph induced by G has density at least $(1-o(1))\gamma$.*

Proof. Let $\lambda = \exp(-\frac{1}{(\log n)^2})$ and let $k = \lfloor (\log n)^2 \log\log n \rfloor$. For a set $S \subset Y_1 \cup \cdots \cup Y_{r-1}$ and an integer $1 \leq j < k$, let $N_j(S) = \{x \in X : \lambda^j |S| < |\Gamma_G(x) \cap S| \leq \lambda^{j-1}|S|\}$. Moreover, let $N_k(S) = \{x \in X : |\Gamma_G(x) \cap S| \leq \lambda^{k-1}|S|\}$. Given an $(r-1)$-tuple $\vec{j} = (j_1, \ldots, j_{r-1})$ of integers and $\vec{S} = (S_1, \ldots, S_{r-1})$ such that $1 \leq j_i \leq k$ and $S_i \subset Y_i$ for every i, write $N_{\vec{j}}(\vec{S}) = \bigcap_{1\leq i\leq r-1} N_{j_i}(S_i)$. For $\vec{S} = (S_1, \ldots, S_{r-1})$ and sets $J_1, \ldots, J_{r-1} \subset \{1, 2, \ldots, k\}$, define

$$N'_{(J_1,\ldots,J_{r-1})}(\vec{S}) = \bigcup_{j_1\in J_1,\ldots,j_{r-1}\in J_{r-1}} N_{\vec{j}}(\vec{S}).$$

Moreover, for a vector $\vec{v}$ and nonnegative integers a, b, we write

$$[\vec{v}-a, \vec{v}+b] = ([v_1-a, v_1+b]_k, \ldots, [v_{r-1}-a, v_{r-1}+b]_k),$$

where $[c,d]_k$ denotes $[c,d] \cap \{1, 2, \ldots, k\}$.

Let $s = \lceil (\log n)^7 \rceil$, $t = \lfloor n^{2/3} \rfloor$ and $\ell = \lceil \log n \rceil$.

Let A be a uniformly random subset of X of size $\lceil \alpha n \rceil$. If $T \subset X$ has size at least t, then by the Chernoff bound

$$\mathbb{P}\big[|T \cap A| \geq \alpha|T| + \frac{\alpha|T|}{n^{1/20}}\big] \leq \exp\left(- c\frac{\alpha|T|}{n^{1/10}}\right) \leq \exp(-n^{1/2}) \tag{8.15}$$

for n sufficiently large. On the other hand, the number of $\vec{S} = (S_1, \dots, S_{r-1})$ with $S_i \subset Y_i$ and $|S_1| = \cdots = |S_{r-1}| = s$ is at most $\binom{n}{s}^r \leq n^{rs} = \exp(rs \log n) \leq \exp((\log n)^9)$ for n sufficiently large. Let F be the event that there exist $1 \leq v_1, \dots, v_{r-1} \leq k$ and some $S_1 \subset Y_1, \dots, S_{r-1} \subset Y_{r-1}$ of size s such that $|N'_{[\vec{v}-1, \vec{v}+\ell-4]}(\vec{S})| \geq t$ and $|N'_{[\vec{v}-1,\vec{v}+\ell-4]}(\vec{S}) \cap A| \geq \alpha(1 + \frac{1}{n^{1/20}})|N'_{[\vec{v}-1,\vec{v}+\ell-4]}(\vec{S})|$. Then, by the union bound and (8.15), we have $\mathbb{P}(F) = o(1)$.

Thus, we may choose some $A \subset X$ of size $\lceil \alpha n \rceil$ such that F does not hold but there exist $B \subset A$ and $E_1 \subset Y_1$, $E_2 \subset Y_2$, $\dots$, $E_{r-1} \subset Y_{r-1}$ with the property that $|B| \geq \beta n$, $|E_i| \geq m_i$ for every i, and between any two of $B, E_1, \dots, E_{r-1}$, the bipartite subgraph induced by G has density at least γ. Write $\vec{E} = (E_1, \dots, E_{r-1})$. Then for every $1 \leq i \leq r-1$,

$$\sum_{\vec{j}} |N_{\vec{j}}(\vec{E}) \cap B| \cdot \lambda^{j_i - 1}|E_i| \geq e(G[B, E_i]) \geq \gamma|B||E_i|, \tag{8.16}$$

where the summation is over all $\vec{j} = (j_1, \dots, j_{r-1})$ with $1 \leq j_1, \dots, j_{r-1} \leq k$.

Recall that $\ell = \lceil \log n \rceil$. For $1 \leq I \leq r-1$, choose integers $1 \leq u_I \leq \ell$ independently and uniformly at random. Clearly, for any $1 \leq i, I \leq r-1$,

$$\mathbb{E}\Big[\sum_{\substack{\vec{j}: \\ j_I \in \{u_I-4, u_I-3, u_I-2, u_I-1\} \mod \ell}} |N_{\vec{j}}(\vec{E}) \cap B| \cdot \lambda^{j_i-1} \Big] = \frac{4}{\ell} \sum_{\vec{j}} |N_{\vec{j}}(\vec{E}) \cap B| \cdot \lambda^{j_i-1},$$

and so

$$\mathbb{P}\Big[\sum_{\substack{\vec{j}: \\ j_I \in \{u_I-4, u_I-3, u_I-2, u_I-1\} \mod \ell}} |N_{\vec{j}}(\vec{E}) \cap B| \cdot \lambda^{j_i-1} > \frac{4r^2}{\ell} \sum_{\vec{j}} |N_{\vec{j}}(\vec{E}) \cap B| \cdot \lambda^{j_i-1} \Big] < \frac{1}{r^2}.$$

Thus, there exist $1 \leq u_1, \dots, u_{r-1} \leq \ell$ such that for every $1 \leq i, I \leq r-1$,

$$\sum_{\substack{\vec{j}: \\ j_I \in \{u_I-4, u_I-3, u_I-2, u_I-1\} \mod \ell}} |N_{\vec{j}}(\vec{E}) \cap B| \cdot \lambda^{j_i-1} \leq \frac{4r^2}{\ell} \sum_{\vec{j}} |N_{\vec{j}}(\vec{E}) \cap B| \cdot \lambda^{j_i-1},$$

and so for every $1 \leq i \leq r-1$,

$$\begin{aligned} \sum_{\substack{\vec{j}: \\ j_I \notin \{u_I-4, u_I-3, u_I-2, u_I-1\} \mod \ell \\ \text{for every } I}} |N_{\vec{j}}(\vec{E}) \cap B| \cdot \lambda^{j_i-1} &\geq \Big(1 - \frac{4r^3}{\ell}\Big) \sum_{\vec{j}} |N_{\vec{j}}(\vec{E}) \cap B| \cdot \lambda^{j_i-1} \\ &\geq \Big(1 - \frac{4r^3}{\ell}\Big)\gamma|B| \qquad \text{by (8.16).} \end{aligned}$$

Using the compact notation introduced at the beginning of the proof, we get that for every $1 \leq i \leq r-1$,

$$\sum_{\substack{\vec{v}: \\ \vec{v} \equiv \vec{u} \mod \ell}} |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \cdot \lambda^{v_i - 1} \geq \Big(1 - \frac{4r^3}{\ell}\Big)\gamma|B|.$$

However,

$$\sum_{\substack{\vec{v}: \\ |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| < t}} |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \cdot \lambda^{v_i - 1} \leq k^{r-1}t = o(\gamma|B|),$$

so

$$\sum_{\substack{\vec{v}: \\ \vec{v} \equiv \vec{u} \mod \ell \\ |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \geq t}} |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \cdot \lambda^{v_i - 1} \geq (1 - o(1))\gamma|B|. \tag{8.17}$$

Claim. For every $\vec{v}$, if $|N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \geq t$, then

$$|N'_{[\vec{v}-2,\vec{v}+\ell-3]}(\vec{E})| \geq \frac{1}{\alpha(1+n^{-1/20})} \cdot |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B|.$$

Proof of Claim. For every $1 \leq i \leq r-1$, pick random subsets $S_i \subset E_i$ of size s and write $\vec{S} = (S_1, \ldots, S_{r-1})$. Let $x \in N_j(E_i)$ for some $1 \leq i \leq r-1$ and some $1 \leq j < k$. If $x \notin N_{j-1}(S_i) \cup N_j(S_i) \cup N_{j+1}(S_i)$, then the random variable $|\{y \in S_i : xy \text{ is an edge}\}|$ deviates by at least $(1-\lambda)\mu$ from its mean μ. Hence, by the Chernoff bound, $\mathbb{P}[x \notin N_{j-1}(S_i) \cup N_j(S_i) \cup N_{j+1}(S_i)] \leq \exp(-c(1-\lambda)^2\mu)$ for some absolute constant $c > 0$. Since $\mu \geq \lambda^j s \geq \lambda^j(\log n)^7$, $\lambda^j \geq \lambda^k \geq \exp(-\log\log n) = \frac{1}{\log n}$ and $1 - \lambda \geq \frac{1}{2(\log n)^2}$, we have $\exp(-c(1-\lambda)^2\mu) \leq \frac{1}{n^2}$. It is not hard to see that also in the case $x \in N_k(E_i)$, we have $\mathbb{P}[x \notin N_{k-1}(S_i) \cup N_k(S_i)] \leq \frac{1}{n^2}$. Thus, almost surely, for every $1 \leq i \leq r-1$ and every $1 \leq j \leq k$, $N_j(E_i) \subset N_{j-1}(S_i) \cup N_j(S_i) \cup N_{j+1}(S_i)$. Then we also have that for every $1 \leq i \leq r-1$ and every $1 \leq j \leq k$,

$$\begin{aligned} N_j(E_i) &\subset N_{j-1}(S_i) \cup N_j(S_i) \cup N_{j+1}(S_i) \\ &\subset N_{j-2}(E_i) \cup N_{j-1}(E_i) \cup N_j(E_i) \cup N_{j+1}(E_i) \cup N_{j+2}(E_i). \end{aligned}$$

Hence,

$$N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \subset N'_{[\vec{v}-1,\vec{v}+\ell-4]}(\vec{S}) \subset N'_{[\vec{v}-2,\vec{v}+\ell-3]}(\vec{E}).$$

In particular, $|N'_{[\vec{v}-1,\vec{v}+\ell-4]}(\vec{S})| \geq |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E})| \geq |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \geq t$, therefore

$$|N'_{[\vec{v}-1,\vec{v}+\ell-4]}(\vec{S}) \cap A| \leq \alpha\Big(1 + \frac{1}{n^{1/20}}\Big)|N'_{[\vec{v}-1,\vec{v}+\ell-4]}(\vec{S})|$$

from the definition of A. Thus,

$$\begin{aligned}|N'_{[\vec{v}-2,\vec{v}+\ell-3]}(\vec{E})| &\geq |N'_{[\vec{v}-1,\vec{v}+\ell-4]}(\vec{S})| \\ &\geq \frac{1}{\alpha(1+n^{-1/20})}|N'_{[\vec{v}-1,\vec{v}+\ell-4]}(\vec{S}) \cap A| \\ &\geq \frac{1}{\alpha(1+n^{-1/20})}|N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap A| \\ &\geq \frac{1}{\alpha(1+n^{-1/20})}|N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B|.\end{aligned}$$

This completes the proof of the claim.

By the claim, if $|N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \geq t$, then we may choose sets $D_{\vec{v}} \subset N'_{[\vec{v}-2,\vec{v}+\ell-3]}(\vec{E})$ with $|D_{\vec{v}}| = \lceil \frac{1}{\alpha(1+n^{-1/20})}|N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \rceil$. Note that these are pairwise disjoint for all $\vec{v} \equiv \vec{u} \mod \ell$. By (8.17), for every $1 \leq i \leq r-1$, we have

$$\sum_{\substack{\vec{v}: \\ \vec{v} \equiv \vec{u} \mod \ell \\ |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \geq t}} |D_{\vec{v}}| \cdot \lambda^{v_i - 1} \geq (1 - o(1))\frac{\gamma}{\alpha}|B|.$$

If $v_i > k - \ell$, then $\lambda^{v_i-1} \leq \lambda^{\frac{1}{2}(\log n)^2 \log\log n} = \frac{1}{\sqrt{\log n}}$, so

$$\sum_{\substack{\vec{v}: \\ \vec{v} \equiv \vec{u} \mod \ell \\ |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \geq t \\ v_i > k-\ell}} |D_{\vec{v}}| \cdot \lambda^{v_i - 1} \leq \frac{1}{\sqrt{\log n}} \sum_{\substack{\vec{v}: \\ \vec{v} \equiv \vec{u} \mod \ell \\ |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \geq t}} |D_{\vec{v}}| \leq \frac{1}{\sqrt{\log n}}|X| = \frac{n}{\sqrt{\log n}},$$

therefore

$$\sum_{\substack{\vec{v}: \\ \vec{v} \equiv \vec{u} \mod \ell \\ |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \geq t \\ v_i \leq k-\ell}} |D_{\vec{v}}| \cdot \lambda^{v_i - 1} \geq (1 - o(1))\frac{\gamma}{\alpha}|B|.$$

If $v_i \leq k - \ell$, then $v_i + \ell - 3 < k$. But $D_{\vec{v}} \subset N'_{[\vec{v}-2,\vec{v}+\ell-3]}(\vec{E}) \subset \bigcup_{v_i-2 \leq j \leq v_i+\ell-3} N_j(E_i)$, so any $x \in D_{\vec{v}}$ has at least $\lambda^{v_i+\ell-2}|E_i|$ neighbours in E_i. Thus,

$$e\big(G\big[\bigcup_{\vec{v}} D_{\vec{v}}, E_i\big]\big) \geq \sum_{\vec{v}} |D_{\vec{v}}| \cdot \lambda^{v_i+\ell-2}|E_i| \geq (1-o(1))\frac{\gamma}{\alpha}|B||E_i|,$$

where the union and the summation are over all $\vec{v}$ with $\vec{v} \equiv \vec{u} \mod \ell$, $|N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \geq t$ and $v_i \leq k - \ell$. On the other hand,

$$\sum_{\substack{\vec{v}: \\ \vec{v} \equiv \vec{u} \mod \ell \\ |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \geq t}} |D_{\vec{v}}| \leq \sum_{\substack{\vec{v}: \\ \vec{v} \equiv \vec{u} \mod \ell \\ |N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \geq t}} \left\lceil \frac{1}{\alpha(1+n^{-1/20})}|N'_{[\vec{v},\vec{v}+\ell-5]}(\vec{E}) \cap B| \right\rceil \leq \frac{|B|}{\alpha}.$$

Now if $|\bigcup_{\vec{v}} D_{\vec{v}}| \geq \frac{\beta}{\alpha}n$, then choose $D = \bigcup_{\vec{v}} D_{\vec{v}}$, otherwise choose arbitrary $D \supset \bigcup_{\vec{v}} D_{\vec{v}}$

with $|D| = \lceil \frac{\beta}{\alpha} n \rceil$; then for every $1 \le i \le r-1$, $G[D, E_i]$ has edge density at least $(1 - o(1))\gamma$. □

The next result is the generalization of Corollary 8.5.2 from the previous section.

Corollary 8.6.2. *Let r be a positive integer (independent of n) and let $\alpha, \beta, \gamma = \Omega(1)$ be functions of n. Let G be an r-partite graph with parts $X_1, \ldots, X_r$ of size n each. Suppose that for a proportion $1 - o(1)$ of all choices $A_1 \subset X_1, \ldots, A_r \subset X_r$ with $|A_i| = \lceil \alpha n \rceil$, there exist $B_i \subset A_i$ of size at least βn each such that $G[B_i, B_j]$ has edge density at least γ for every $i \neq j$. Then there exist sets $S_i \subset X_i$ of size at least $\frac{\beta}{\alpha} n$ each such that $G[S_i, S_j]$ has edge density at least $(1 - o(1))\gamma$ for every $i \neq j$.*

Proof. We prove by induction on k that for every $0 \le k \le r$, for a proportion $1 - o(1)$ of all $A_{k+1} \subset X_{k+1}, \ldots, A_r \subset X_r$ of size $\lceil \alpha n \rceil$ there exist $U_1 \subset X_1, \ldots, U_k \subset X_k$ and $V_{k+1} \subset A_{k+1}, \ldots, V_r \subset A_r$ such that $|U_i| \ge \frac{\beta}{\alpha} n$ for every i, $|V_j| \ge \beta n$ for every j, and the edge density between any two of $U_1, \ldots, U_k, V_{k+1}, \ldots, V_r$ is at least $(1 - o(1))\gamma$. In particular, when $k = r$, this proves the corollary.

The case $k = 0$ is guaranteed by the conditions of the corollary. Let $k \ge 1$ and assume that we have already proved the statement for $k-1$. Assume that $A_{k+1} \subset X_{k+1}, \ldots, A_r \subset X_r$ are sets of size $\lceil \alpha n \rceil$ such that for at least half of all $A_k \subset X_k$ there exist $U_1 \subset X_1, \ldots, U_{k-1} \subset X_{k-1}$ and $V_k \subset A_k, \ldots, V_r \subset A_r$ such that $|U_i| \ge \frac{\beta}{\alpha} n$ for every i, $|V_j| \ge \beta n$ for every j, and the edge density between any two of $U_1, \ldots, U_{k-1}, V_k, \ldots, V_r$ is at least $(1 - o(1))\gamma$. Note that this holds for a proportion $1 - o(1)$ of all $A_{k+1}, \ldots, A_r$. But for any such $A_{k+1}, \ldots, A_r$, we can apply Lemma 8.6.1 to find $D \subset X_k$, $E_1 \subset X_1, \ldots, E_{k-1} \subset X_{k-1}$ and $E_{k+1} \subset A_{k+1}, \ldots, E_r \subset A_r$ such that $|D| \ge \frac{\beta}{\alpha} n$, $|E_1|, \ldots, |E_{k-1}| \ge \frac{\beta}{\alpha} n$, $|E_{k+1}|, \ldots, |E_r| \ge \beta n$, and the edge density between any two of $E_1, \ldots, E_{k-1}, D, E_{k+1}, \ldots, E_r$ is at least $(1 - o(1))\gamma$. This completes the inductive step. □

Proof of Theorem 8.1.8. It suffices to prove that for every fixed positive integer r, if n is sufficiently large, then there exists a set $T \subset V(G)$ of size at least $\frac{\beta}{\alpha} n$ such that $G[T]$ has edge density at least $(1 - \frac{2}{r})\gamma$.

For simplicity, assume that r divides n. Partition the vertex set of G into sets $X_1, \ldots, X_r$ of equal size uniformly at random.

Claim. Almost surely, for a proportion $1 - o(1)$ of all $A_1 \subset X_1, \ldots, A_r \subset X_r$ of size $\lceil \alpha \frac{n}{r} \rceil$ each, there exist $B_i \subset A_i$ of size at least $\beta \frac{n}{r}$ such that each $G[B_i, B_j]$ $(i \neq j)$ has edge density at least $(1 - o(1))\gamma$.

Proof of Claim. Let $A \subset V(G)$ have size $r \lceil \alpha \frac{n}{r} \rceil$. Since G has the (α, β, γ)-property, there exists some $B \subset A$ of size at least βn such that $G[B]$ has edge density at least γ. Conditional on the event that $|A \cap X_i| = \lceil \alpha \frac{n}{r} \rceil$ for every i, almost surely we have that $(1 - o(1)) \frac{\beta n}{r} \le |B \cap X_i|$ for every i and that the edge density of $G[B \cap X_i, B \cap X_j]$ is at least $(1 - o(1))\gamma$ for every $i \neq j$. But if these hold, then there exist $B_i \subset A \cap X_i$ of size at least $\beta \frac{n}{r}$ such that each $G[B_i, B_j]$ has edge density at least $(1 - o(1))\gamma$.

So for every $A \subset V(G)$ of size $r\lceil \alpha \frac{n}{r} \rceil$, conditional on the event that $|A \cap X_i| = \lceil \alpha \frac{n}{r} \rceil$ for every i, almost surely there exist $B_i \subset A \cap X_i$ of size at least $\beta \frac{n}{r}$ such that $G[B_i, B_j]$ has edge density at least $(1 - o(1))\gamma$. Hence, on average, the proportion of $A_1 \subset X_1, \ldots, A_r \subset X_r$ of size $\lceil \alpha \frac{n}{r} \rceil$ for which there exist suitable $B_i \subset A_i$ is $1 - o(1)$, which completes the proof of the claim.

Using the claim, there exists a partition of $V(G)$ to sets $X_1, \ldots, X_r$ of size $\frac{n}{r}$ such that for a proportion $1 - o(1)$ of all $A_1 \subset X_1, \ldots, A_r \subset X_r$ of size $\lceil \alpha \frac{n}{r} \rceil$, there exist $B_i \subset A_i$ of size at least $\beta \frac{n}{r}$ such that each $G[B_i, B_j]$ $(i \neq j)$ has edge density at least $(1 - o(1))\gamma$. Let G' be the r-partite graph $G[X_1, \ldots, X_r]$. By Corollary 8.6.2 applied to G', there exist sets $S_i \subset X_i$ of size at least $\frac{\beta}{\alpha} \frac{n}{r}$ each such that $G'[S_i, S_j]$ has edge density at least $(1 - o(1))\gamma$ for every $i \neq j$. Choose, for each i, a uniformly random subset $T_i \subset S_i$ of size $\lceil \frac{\beta}{\alpha} \frac{n}{r} \rceil$. Let $T = T_1 \cup \cdots \cup T_r$. Then $\frac{\beta}{\alpha} n \leq |T| \leq \frac{\beta}{\alpha} n + r$ and the expected number of edges in T is at least $\binom{r}{2} (\frac{\beta}{\alpha} \frac{n}{r})^2 (1 - o(1))\gamma = (1 - \frac{1}{r} - o(1))\gamma \frac{(\beta n/\alpha)^2}{2}$. Thus, for sufficiently large n, with positive probability $G[T]$ has edge density at least $(1 - \frac{2}{r})\gamma$. □

8.7 Concluding remarks

In this section we focus exclusively on the case where α, β, γ do not depend on n.

Definition 8.7.1. Let $f(\alpha, \beta, \gamma, n)$ be the minimum edge density of an n-vertex graph with the (α, β, γ)-property. Define $g(\alpha, \beta, \gamma) = \lim_{n \to \infty} f(\alpha, \beta, \gamma, n)$.

It is not hard to see, using Szemerédi's regularity lemma, that the limit exists.

When $\gamma = 1$, the problem is completely resolved by Kopylov's result [87], so let us assume that $\gamma < 1$. It is not hard to see that in this range g is continuous. Let us summarise what our main results say about g. Theorem 8.1.2 gives the following.

Theorem 8.7.2. *Let $\alpha \leq 1/2$. Then*

$$g(\alpha, \beta, \gamma) \geq \frac{\beta}{\alpha} \gamma.$$

The next result follows from the second construction defined after Theorem 8.1.3.

Proposition 8.7.3. *Let p, q be positive integers with $\frac{p}{q} \geq \frac{\beta}{\alpha}$ and $p\gamma \leq 1$. Then*

$$g(\alpha, \beta, \gamma) \leq \frac{p}{q} \gamma.$$

In particular, if $\alpha \leq 1/2$, $\frac{p}{q} = \frac{\beta}{\alpha}$ and $p\gamma \leq 1$, then

$$g(\alpha, \beta, \gamma) = \frac{\beta}{\alpha} \gamma.$$

Theorem 8.1.4 and the first construction after Theorem 8.1.3 yield the following.

Theorem 8.7.4. *If $\alpha \leq 1/2$ and $\alpha - \beta \leq \frac{\beta^3\gamma}{1000}$, then*

$$g(\alpha, \beta, \gamma) = (1 - \alpha + \beta)^2\gamma.$$

We remark that this holds for $\alpha > 1/2$ as well.

The next result is an easy corollary of Theorem 8.1.8.

Corollary 8.7.5. *Let $1 \leq \lambda \leq 1/\alpha$. Then any graph G on n vertices having the (α, β, γ)-property also has the $(\lambda\alpha, \lambda\beta, (1 - o(1))\gamma)$-property.*

Proof. Let us ignore ceilings and floor signs as they are not significant. Let $A \subset V(G)$ have size $\lambda\alpha n$. Note that $G[A]$ has the $(\frac{1}{\lambda}, \frac{\beta}{\lambda\alpha}, \gamma)$-property. Thus, by Theorem 8.1.8, there exists a set $D \subset A$ of size at least $\frac{\beta/(\lambda\alpha)}{1/\lambda}\lambda\alpha n = \lambda\beta n$ such that $G[D]$ has density at least $(1 - o(1))\gamma$. □

Corollary 8.7.6. *For any $1 \leq \lambda \leq 1/\alpha$,*

$$g(\alpha, \beta, \gamma) \geq g(\lambda\alpha, \lambda\beta, \gamma).$$

It would be interesting to understand the function g even better. Analogously to Theorem 8.7.4, we think that when $\beta < \frac{\alpha}{2}$, but $\frac{\alpha}{2} - \beta$ is small compared to β and γ, then the extremal construction is given by the disjoint union of two blocks of size $(\frac{1}{2} - \frac{\alpha}{2} + \beta)n$ with internal edge density roughly γ. Accordingly, we make the following (somewhat imprecise) conjecture.

Conjecture 8.7.7. *Let α, β, γ be constants. Assume that $\beta < \frac{\alpha}{2}$, but $\frac{\alpha}{2} - \beta$ is small compared to β and γ. Then*

$$g(\alpha, \beta, \gamma) = 2\Big(\frac{1}{2} - \frac{\alpha}{2} + \beta\Big)^2\gamma.$$

Chapter 9

The maximum number of induced C_5's in a planar graph

9.1 Introduction

The problem of maximizing the number of induced copies of a fixed graph H in a graph on n vertices has attracted a lot of attention recently, see, for example, [43,62,101]. Morrison and Scott determined the maximum possible number of induced cycles, without restriction on length, that can be contained in a graph on n vertices [97]. The maximal number of induced complete bipartite graphs and induced complete r-partite subgraphs have also been studied [12, 14, 16]. The problem of determining the maximum number of induced C_5's has been elusive for a long time and was finally solved by Balogh, Hu, Lidický and Pfender [6].

In this chapter we determine asymptotically the maximum possible number of induced C_5's in planar graphs on n vertices. Before we state our main result, let us mention some known results about the number of (not necessarily induced) subgraphs in planar graphs. Let the maximum number of (not necessarily induced) copies of the graph H in an n-vertex planar graph be denoted by $f(n, H)$. Győri et al. [58] proved that $f(n, C_5) = 2n^2-10n+12$ for $n \geq 8$ (and they also determined the value of $f(n, C_5)$ for $n \leq 7$). Hakimi and Schmeichel [59] showed that $f(n, C_4) = \frac{1}{2}(n^2 + 3n - 22)$ for $n \geq 4$ and classified the extremal graphs attaining this bound (a small correction to their result was given in [2]). It can be observed that if we take a planar graph on n vertices given by $K_{2,n-2}$ (see Figure 9.1 (b)), it contains exactly $\frac{1}{2}(n^2 - 5n + 6)$ induced 4-cycles. It follows that the maximum number of induced 4-cycles in a planar graph with n vertices is $\frac{1}{2}n^2 + O(n)$.

Very recently, Huynh, Joret and Wood [63] determined the order of magnitude of $f(n, H)$ for every graph H.

In this chapter, we give a tight asymptotic bound on the number of induced 5-cycles in a planar graph with given number of vertices.

Theorem 9.1.1. *Let G be a planar graph on n vertices. Then G contains at most $\frac{n^2}{3} + O(n)$ induced C_5's.*

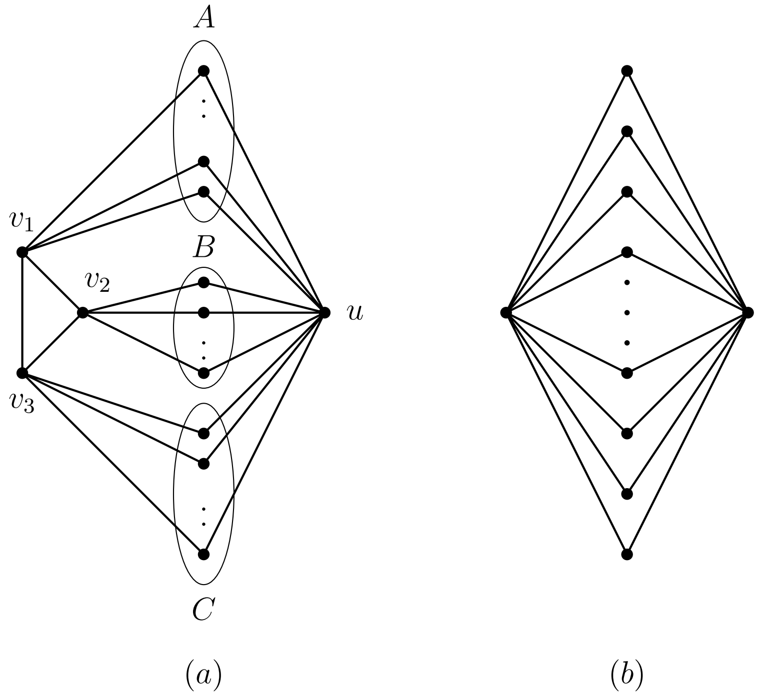

Figure 9.1: Planar graphs containing asymptotically maximum number of induced 5-cycles and 4-cycles, respectively

Let $3|(n-4)$ and let A, B and C be pairwise disjoint sets with $|A| = |B| = |C| = \frac{n-4}{3}$. We define an n-vertex planar graph G as follows. The vertex set of G is the union of A, B and C together with four other vertices, say v_1, v_2, v_3 and u. We define the edges of G as $E(G) = \{v_1v_2, v_2v_3, v_3v_1\} \cup \{v_1a, au : \forall a \in A\} \cup \{v_2b, bu : \forall b \in B\} \cup \{v_3c, cu : \forall c \in C\}$ (see Figure 9.1 (a)). It can be checked that G contains exactly $3 \cdot (\frac{n-4}{3})^2 = \frac{(n-4)^2}{3}$ induced C_5's. Thus, this construction shows that the bound we have in Theorem 9.1.1 is asymptotically best possible.

Our strategy to prove Theorem 9.1.1 is the following. We show that if n is sufficiently large, then there exists a vertex which is contained in at most $2n/3$ induced C_5's, unless the graph has a specific structure (see Lemma 9.4.1), in which case we argue directly that the graph contains at most $(\frac{2}{9} + o(1))n^2$ induced C_5's (see Lemma 9.4.2). This, combined with induction on n, implies Theorem 9.1.1.

9.2 A preliminary lemma

We start with a basic lemma, which we are going to use throughout the chapter.

Lemma 9.2.1. *Let G be a planar graph, let $v \in V(G)$, and let u and w be distinct neighbours of v. Let $X_0 = N(u) \setminus (N(w) \cup \{w\})$ and let $Y_0 = N(w) \setminus (N(u) \cup \{u\})$. Let X be the subset of X_0 consisting of those vertices that have at least one neighbour in Y_0, and let Y be the subset of Y_0 consisting of those vertices that have at least one neighbour in X_0. Then the number of induced C_5's in G containing u, v and w is at most $|X| + |Y| - 1$.*

Proof. Clearly any such C_5 contains precisely one vertex from each of X and Y. Hence, the number of such induced C_5's is at most the number of edges between X and Y. However, the induced bipartite subgraph of G with parts X and Y is acyclic. Indeed, suppose that there is a cycle $x_1y_1x_2y_2\dots x_ky_kx_1$ with $x_i \in X$ and for all i and $y_j \in Y$ for all j. The subgraph of G with vertices $u, v, w, x_1, y_1, \dots, x_k, y_k$ and edges $uv, vw, ux_1, ux_2, wy_1, wy_2, x_1y_1, y_1x_2, x_2y_2, y_2x_3, \dots, y_kx_1$ is a subdivision of $K_{3,3}$ with the parts being $\{u, y_1, y_2\}$ and $\{w, x_1, x_2\}$. Indeed, the only edge of this $K_{3,3}$ which is potentially not present in G is x_1y_2, but we have a path $y_2x_3y_3\dots x_ky_kx_1$ in G. Hence, G is not planar, which is a contradiction. Thus, the induced bipartite subgraph of G with parts X and Y is a forest, therefore it has at most $|X| + |Y| - 1$ edges. □

9.3 Finding an empty $K_{2,7}$

In this section we prove that if G does not contain an empty $K_{2,7}$, then there is even a vertex which is contained in at most $11n/20$ induced C_5's. Here an empty $K_{2,7}$ in a drawing of G means distinct vertices u and w, and $z_1, \dots, z_7 \in N(u) \cap N(w)$ in natural order such that the bounded region with boundary consisting of uz_1, z_1w, wz_7 and z_7u contains no vertex other than $z_2, \dots, z_6$.

Lemma 9.3.1. *Let n be sufficiently large and let G be a plane graph on n vertices. If G does not contain an empty (not necessary induced) $K_{2,7}$, then there is a vertex in G which is contained in at most $11n/20$ induced C_5's.*

To prove this, we need some preliminaries.

Lemma 9.3.2. *Let n be sufficiently large and let G be a planar graph on n vertices. If G does not contain a (not necessary induced) $K_{2,\frac{n}{10^6}}$, then there is a vertex in G which is contained in at most $n/2$ induced C_5's.*

Proof. Suppose otherwise. Let v be a vertex of degree at most 5 in G. Then v has distinct non-adjacent neighbours u and w such that the number of induced C_5's containing u, v and w is at least $n/20$. Define X and Y as in Lemma 9.2.1. By the same lemma, we have $|X| + |Y| \geq n/20$. Let G' be the induced bipartite subgraph of G with parts X and Y. By assumption, there is no vertex of degree at least $n/10^6$ in G'. Then since G' has at least $\frac{|X|+|Y|}{2} \geq n/40$ edges, there must exist a set of at least 10^4 independent edges in G'.

Let they be $x_1y_1, x_2y_2, \dots, x_{10^4}y_{10^4}$ such that $x_1, x_2, \dots, x_{10^4} \in X$, the edges $ux_1, ux_2, \dots, ux_{10^4}$ are in anti-clockwise order, and the bounded region with boundary consisting of edges ux_1, $x_1y_1, y_1w, wy_{10^4}, y_{10^4}x_{10^4}, x_{10^4}u$ contains all x_i and y_i. For $1 \leq i \leq 10^4 - 1$, let R_i be the bounded region with boundary consisting of $ux_i, x_iy_i, y_iw, wy_{i+1}, y_{i+1}x_{i+1}, x_{i+1}u$. Choose $11 \leq i \leq 10^4 - 12$ such that the number of vertices in $R_{i-10} \cup R_{i-9} \cup \dots \cup R_{i+11}$ is at most $n/300$. Let $R = R_i \cup R_{i+1}$.

Let S be the set of vertices of G in the interior of R which do not belong to $N(u) \cap N(w)$. Note that $x_{i+1} \in S$, so $S \neq \emptyset$. Now the graph $G'' = G[S]$ is planar, so there exists some $z \in S$ which has degree at most 5 in G''. But it is joined to at most 2 elements of $N(u) \cap N(w)$, so it has at most 7 neighbours in the interior of R. Hence (together with u, x_i, y_i, w, y_{i+2} and x_{i+2}), z has at most 13 neighbours.

By assumption, z is contained in at least $n/2$ induced C_5's. It is easy to see that any such C_5 is either contained entirely in $R_{i-10} \cup R_{i-9} \cup \cdots \cup R_{i+11}$ or it contains both u and w. In the former case, it can only use a set of at most $n/300$ vertices, and since z has degree at most 13, by Lemma 9.2.1 there are at most $\binom{13}{2} \cdot n/300 < n/3$ such induced C_5's. So there are at least $n/6$ induced C_5's containing z, u and w. Recall that u and w are non-adjacent and $z \notin N(u) \cap N(w)$. If $z \in N(u)$, then all these induced C_5's are of the form $uzswt$ for some $s \in N(z)$ and $t \in N(u) \cap N(w)$, while if $z \in N(w)$, then all these induced C_5's are of the form $uszwt$ for some $s \in N(z)$ and $t \in N(u) \cap N(w)$. In either case, since $|N(z)| \leq 13$, it follows that $|N(u) \cap N(w)| \geq \frac{n}{6 \cdot 13} > \frac{n}{10^6}$. This contradicts the condition in the lemma. □

Lemma 9.3.3. *Let n be sufficiently large and let G be a plane graph on n vertices. Let u and w be distinct vertices, and let $v_1, v_2, \ldots, v_6$ be some of their common neighbours, in natural order. Assume that the number of vertices in the interior of the bounded region with boundary consisting of uv_3, v_3w, wv_4 and v_4u is at least one but at most $n^{1/5}$ and that there is no common neighbour of u and w in the same region. Then G has a vertex which is contained in at most $11n/20$ induced C_5's.*

Proof. Suppose otherwise. Let R be the bounded region with boundary consisting of uv_3, v_3w, wv_4 and v_4u. Let x be an arbitrary vertex inside R. By assumption, $x \notin N(u) \cap N(w)$. Since there are at most $n^{1/5} + 4$ vertices in R (including its boundary), the number of induced C_5's containing x which lie entirely in R (possibly touching the boundary) is at most $(n^{1/5}+4)^4 \leq n/20$. Thus, since x is contained in at least $11n/20$ induced C_5's, there exist at least $n/2$ induced C_5's containing x which contain vertices outside R.

Take such an induced C_5 and call it C. We claim that C must contain both u and w, but does not contain v_3 and v_4. Indeed, if we go through the vertices of C one by one in natural order, starting with x, then there will be a vertex from the set $\{u, v_3, w, v_4\}$ right before the walk first leaves R, and then one in the same set when the walk first returns to R. Call these two vertices y and z, respectively. Since C contains the vertex x, which is in the interior of R, it follows that y and z are not neighbours in C, so they are also not neighbours in G. Thus, either $\{y, z\} = \{u, w\}$ or $\{y, z\} = \{v_3, v_4\}$. In the latter case, again since C is induced and contains x, C contains neither u nor w. So there exists a path of length at most 3 in C, and therefore also in G, from v_3 to v_4 outside of R which avoids both u and w. This is clearly not possible because of the vertices v_1, v_2, v_5 and v_6.

Thus, C indeed contains both u and w, and it is easy to see that it does not contain v_3 and v_4. Since $x \notin N(u) \cap N(w)$, it follows that either $x \in N(u)$ and $C = uxqwr$ for some $q \in N(x) \cap N(w) \setminus \{v_3, v_4\}$ and $r \in N(u) \cap N(w)$, or $x \in N(w)$ and $C = uqxwr$ for some $q \in N(x) \cap N(u) \setminus \{v_3, v_4\}$ and $r \in N(u) \cap N(w)$. In particular, it follows that $N(u)$ and $N(w)$ both have vertices in the interior of R.

Let X be the set of vertices of $N(u)$ in the interior of R and let Y be the set of vertices of $N(w)$ in the interior of R. Similarly as in the proof of Lemma 9.2.1, the induced bipartite subgraph of G with parts X and Y is acyclic. Thus, there is a vertex in that graph of degree at most one. Without loss of generality we may assume that some $x \in X$ has at most one neighbour in Y. Then, by the previous paragraph, there are at most $|N(u) \cap N(w)|$ induced C_5's containing x as well as vertices outside R. Thus, by the first paragraph, $|N(u) \cap N(w)| \geq n/2$.

By a simple averaging, it follows that there exist distinct $t_1, t_2, \ldots, t_7 \in N(u) \cap N(w)$ (in natural order) such that the region S bounded by ut_1, t_1w, wt_7, t_7u contains at most 100 vertices. Now any induced C_5 which contains t_4 and has vertices outside S must contain u and w. Such an induced C_5 cannot contain any vertices from $N(u) \cap N(w)$ other than t_4, so by Lemma 9.2.1, there are at most $n/2$ such induced C_5's. The number of induced C_5's containing t_4 but no vertices outside S is at most 100^5, so t_4 satisfies the conclusion of the lemma. □

Corollary 9.3.4. *Let n be sufficiently large and let G be a plane graph on n vertices with the property that G contains a (not necessarily induced) subgraph $K_{2,7\cdot\lceil n^{4/5}\rceil}$. Then in this $K_{2,7\cdot\lceil n^{4/5}\rceil}$ there is an empty $K_{2,7}$ or there is a vertex in G which is contained in at most $11n/20$ induced C_5's.*

Proof. Assume that there is no vertex in G which is contained in at most $n/2$ induced C_5's. Choose distinct u and w in G with $|N(u) \cap N(w)| \geq 7 \cdot \lceil n^{4/5}\rceil$. Let $v_1, v_2, \ldots, v_{7\cdot\lceil n^{4/5}\rceil} \in N(u) \cap N(w)$ in natural order. For each $1 \leq i \leq 7 \cdot \lceil n^{4/5}\rceil - 1$, let R_i be the bounded region with boundary consisting of the edges uv_i, v_iw, wv_{i+1} and $v_{i+1}u$. By Lemma 9.3.3, each R_i with $3 \leq i \leq 7 \cdot \lceil n^{4/5}\rceil - 3$ contains either zero or at least $n^{1/5}$ vertices in its interior. Hence, the number of non-empty R_i's is at most $n^{4/5} + 4$. Thus, there exists some $1 \leq i \leq 7 \cdot \lceil n^{4/5}\rceil - 6$ for which $u, w, v_i, v_{i+1}, \ldots, v_{i+6}$ define an empty $K_{2,7}$. □

Now Lemma 9.3.1 follows from Lemma 9.3.2 and Corollary 9.3.4.

9.4 Structure of the exceptional graphs

Lemma 9.4.1. *Let n be sufficiently large and let G be a planar graph on n vertices. Suppose there does not exist a vertex in G which is contained in at most $2n/3$ induced C_5's. Then there exist distinct non-adjacent vertices u and w with the following properties.*

1. $|N(u) \cap N(w)| \geq n/3 - n^{6/7}$.

2. *There exist sets $X \subset N(u) \setminus N(w)$ and $Y \subset N(w) \setminus N(u)$ such that $|X|+|Y| \geq 2n/3$ and every $x \in X$ is adjacent to at least 1 but at most $n^{5/6}$ elements of Y and every $y \in Y$ is adjacent to at least 1 but at most $n^{5/6}$ elements of X.*

Proof. Suppose otherwise. Take an arbitrary drawing of G. By Lemma 9.3.1, there exists an empty $K_{2,7}$ in G. Let u and w be the two vertices in the part of size 2 in $K_{2,7}$, and let v be the centre vertex in the part of size 7. Define X and Y as in the statement of Lemma 9.2.1. Since every induced C_5 containing v also contains u and w, and by assumption v is contained in more than $2n/3$ induced C_5's, it follows by Lemma 9.2.1 that $|X| + |Y| > 2n/3 + 1$. Moreover, since there exists an induced C_5 containing u, v and w, it follows that u and w are non-adjacent.

Let G' be the induced bipartite subgraph of G with parts X and Y.

Suppose first that G' has maximum degree at least $n^{5/6}$. By symmetry, we may assume that some $y \in Y$ has degree at least $n^{5/6}$ in G'. Then $|N(y) \cap X| \geq n^{5/6}$. For large enough n, together with u and y, these vertices form a $K_{2,7\lceil n^{4/5}\rceil}$. Thus, by Corollary 9.3.4, there are vertices $x_1, \ldots, x_7 \in N(y) \cap X$ such that together with u and y they form an empty (not necessarily induced) $K_{2,7}$. By assumption, x_4 is contained in at least $2n/3$ induced C_5's. However, note that any such induced C_5 also contains u and y. Let Z be the set of all vertices in $X \cup Y \setminus \{y, x_4\}$ which are contained in an induced C_5 containing x_4. Order the elements of Y as $y_1, y_2, \ldots, y_k$ such that the edges $wv, wy_1, \ldots, wy_k$ are in clockwise order.

Then y_iy_j is an edge only if $j = i + 1$. Indeed, for any $1 \leq \ell \leq k$ there exists a path from v to y_ℓ (through u and some $x \in X$) which avoids $\{w\} \cup Y \setminus \{y_\ell\}$. But if y_iy_j is an edge for some $j > i + 1$, then the triangle wy_iy_j separates y_{i+1} from v.

Now $y = y_i$ for some i.

Claim 1. If $k = 1$, then $Z = \emptyset$. Suppose that $k \geq 2$. If $i = 1$, then $Z \subset (N(y_2) \cap X) \cup \{y_2\} \setminus N(y_1)$. If $i = k$, then $Z \subset (N(y_{k-1}) \cap X) \cup \{y_{k-1}\} \setminus N(y_k)$. Otherwise $Z \subset ((N(y_{i-1}) \cup N(y_{i+1})) \cap X) \cup \{y_{i-1}, y_{i+1}\} \setminus N(y_i)$.

Proof. If $k = 1$, then $Y = \{y\}$ and $X \subset N(y) \cap N(u)$, so the first assertion is straightforward.

Suppose that $k \geq 2$. Let $z \in Z$. First assume that $z \in Y$. Then z is not a neighbour of u, so it must be a neighbour of $y = y_i$. Thus, $z = y_{i-1}$ or $z = y_{i+1}$.

Now assume that $z \in X$. Observe that since y, x_4, u and z are contained in an *induced* C_5, we have $z \notin N(y)$, and the fifth vertex in the C_5 is some $q \in N(y) \cap N(z)$.

Let us first assume that $2 \leq i \leq k-1$. Let r_1 be an arbitrary element in $N(y_{i-1}) \cap X$ and let r_2 be an arbitrary element in $N(y_{i+1}) \cap X$. Note that the edges wy_{i-1}, $y_{i-1}r_1$, r_1u, ur_2, r_2y_{i+1}, $y_{i+1}w$ divide the plane into two regions; let R be the one which contains y_i. Then either z is also in R (possibly on the boundary), or q is on the boundary of R. But ux_4yqz is an induced C_5, so $q \notin N(u)$. Thus, $q \notin X$ so $q \neq r_1$ and $q \neq r_2$. Also, $z \in X$, so $z \notin N(w)$, hence $q \neq w$. Moreover, q is distinct from u. Thus, if q is on the

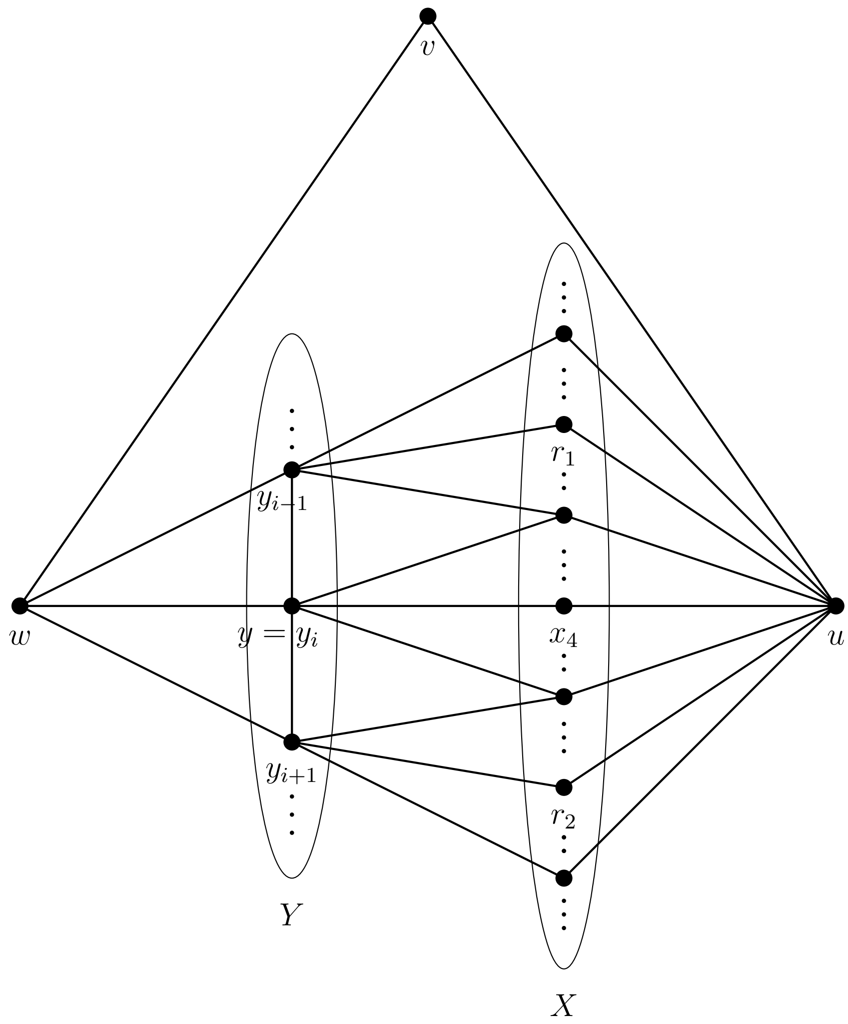

Figure 9.2: Proof of Claim 1

boundary of R, then $q = y_{i-1}$ or $q = y_{i+1}$. In either case $z \in N(y_{i-1}) \cup N(y_{i+1})$. If q is not on the boundary of R, then z is in R (possibly on the boundary). Also, $z \in X$, so z has a neighbour in Y. But $z \notin N(y_i)$, so $z \in N(y_{i-1}) \cup N(y_{i+1})$, as claimed.

Assume now that $i = 1$. Let r be an arbitrary element in $N(y_2) \cap X$. The edges wv, vu, ur, ry_2 and y_2w divide the plane into two regions; let R be the one containing y_1. Then either z is also in R (possibly on the boundary), or q is on the boundary of R. But $q \notin N(u)$ so $q \neq r$ and $q \neq v$. Also, $z \in X$, so $z \notin N(w)$, hence $q \neq w$. Moreover, q is distinct from u. Thus, if q is on the boundary of R, then $q = y_2$. Hence, $z \in N(y_2)$. If q is not on the boundary of R, then z is in R (possibly on the boundary). Also, $z \in X$, so z has a neighbour in Y. But $z \notin N(y_1)$, so $z \in N(y_2)$.

The case $i = k$ is very similar, so the claim is proved.

Since x_4 is contained in at least $2n/3$ induced C_5's, and any such C_5 contains u and y as well, it follows by Lemma 9.2.1 and Claim 1 that $n - |X \cup Y| + |((N(y_{i-1}) \cup N(y_{i+1})) \cap X) \cup \{y_{i-1}, y_{i+1}\} \setminus N(y_i)| \geq 2n/3 + 1$. Since $|X \cup Y| \geq 2n/3 - 1$, by symmetry we may assume that $|N(y_{i-1}) \cap X| \geq 7 \cdot \lceil n^{4/5} \rceil$. Then, by Corollary 9.3.4, there must exist vertices $x'_1, x'_2, \ldots, x'_7 \in N(y_{i-1}) \cap X$ which together with u and y_{i-1} form an empty (not necessarily induced) $K_{2,7}$. Let Z' be the set of all vertices in $X \cup Y \setminus \{y_{i-1}, x'_4\}$ which are contained in an induced C_5 containing x'_4. Then, by the same argument as in Claim 1, it follows that $Z' \subset ((N(y_{i-2}) \cup N(y_i)) \cap X) \cup \{y_{i-2}, y_i\} \setminus N(y_{i-1})$, and that $n - |X \cup Y| + |Z'| \geq 2n/3 + 1$. Thus, $|Z| + |Z'| \geq 2|X \cup Y| - 2n/3 + 2$. However, Z and Z' are disjoint, so $|Z| + |Z'| \leq |X \cup Y|$. Thus, $|X \cup Y| \leq 2n/3 - 2$, which is a contradiction.

So G' has maximum degree less than $n^{5/6}$.

Let $x \in X$ be an arbitrary vertex. We give an estimate for the number of induced C_5's containing x. We first count those C_5's which contain both u and w as vertices. Let us call these *type 1* C_5's. Since w is non-adjacent to both x and u, the number of type 1 C_5's containing x is at most $d_{G'}(x) \cdot t$, where $d_{G'}(x)$ is the degree of x in G' and $t = |N(u) \cap N(w)|$.

Call those induced C_5's which do not contain both u and w *type 2*. To bound the number of such C_5's, we will use the following claim.

Claim 2. For every $q \in V(G)$, the number of vertices $z \in X \cup Y$ for which there exists a path of length at most 3 between q and z avoiding both u and w is at most $100n^{5/6}$.

Proof. Take a maximal matching between X and Y. Let the edges in this matching be $x_{i_1}y_{i_1}, \ldots, x_{i_s}y_{i_s}$ such that $x_{i_j} \in X, y_{i_j} \in Y$ and the edges $wy_{i_1}, \ldots, wy_{i_s}$ are in clockwise order. For each $1 \leq j \leq s-1$, let R_j be the bounded region with boundary consisting of the edges $ux_{i_j}, x_{i_j}y_{i_j}, y_{i_j}w, wy_{i_{j+1}}, y_{i_{j+1}}x_{i_{j+1}}, x_{i_{j+1}}u$, and let R_0 be the unbounded region with boundary consisting of the edges $ux_{i_1}, x_{i_1}y_{i_1}, y_{i_1}w, wy_{i_s}, y_{i_s}x_{i_s}, x_{i_s}u$. Let $0 \leq j \leq s-1$. By the maximality of our matching, any element of $X \cup Y$ in the interior of R_j is a neighbour in G' of some vertex in $X \cup Y$ on the boundary of R_j. Since there are 4 vertices in $X \cup Y$ on the boundary of R_j, and G' has maximum degree less than $n^{5/6}$, there are at most $4n^{5/6}$ elements of $X \cup Y$ in the interior of R_j.

Let $q \in V(G) \setminus \{u, w\}$. Then q is in R_j (possibly on the boundary) for some $0 \le j \le s-1$. If there exists some $z \in X \cup Y$ for which there is a path of length at most 3 from q to z avoiding both u and w, then z is in $R_{j-4} \cup R_{j-3} \cup \ldots R_{j+4}$ (with the subscripts considered modulo s). But there are at most $9 \cdot 4n^{5/6}$ such vertices z, which finishes the proof of the claim.

Recall that G' is acyclic, so the number of edges in G' is at most $|X| + |Y| - 1$. Thus, if ℓ is the number of vertices of degree at least 3 in G', then $3\ell \le 2(|X| + |Y|)$, so the number of vertices of degree at most 2 in G' is $|X| + |Y| - \ell \ge \frac{|X|+|Y|}{3} \ge \frac{2n}{9}$.

The number of edges in G is at most $3n$, so the number of vertices in G of degree at least 60 is at most $n/10$.

Moreover, it follows from Claim 2 that the number of vertices $z \in X \cup Y$ for which there exist at least $1000n^{5/6}$ vertices $q \in V(G)$ with a path of length at most 3 between z and q and avoiding both u and w is at most $n/10$.

Thus, there exists a vertex $z \in X \cup Y$ which has degree at most 2 in G', degree at most 60 in G and for which the number of $q \in V(G)$ with a path of length at most 3 between z and q avoiding u and w is at most $1000n^{5/6}$.

Suppose that $q \in V(G)$ is distinct from z, u and w, and that there exists a type 2 induced C_5 containing both z and q. Then there exists a path of length at most 3 from q to z which contains neither u nor w. But there are at most $1000n^{5/6}$ such vertices $q \in V(G)$, so by Lemma 9.2.1, the number of type 2 induced C_5's containing z is at most $\binom{60}{2} \cdot (1000n^{5/6} + 2)$. Moreover, the number of type 1 induced C_5's containing z is at most $2t$, where $t = |N(u) \cap N(w)|$. Since the total number of induced C_5's containing z is at least $2n/3$, it follows that $|N(u) \cap N(w)| \ge n/3 - n^{6/7}$. This completes the proof of the lemma. $\square$

The next result completes the proof of Theorem 9.1.1.

Lemma 9.4.2. *Suppose that G is a planar graph in which there are distinct non-adjacent vertices u and w satisfying properties 1 and 2 from Lemma 9.4.1 and that there is no vertex which is contained in at most $11n/20$ induced C_5's. Then the number of induced C_5's in G is at most $(\frac{2}{9} + o(1))n^2$.*

Proof. In this proof we use the notation defined in the statement of Lemma 9.4.1.

Take a drawing of G. Let $N(u) \cap N(w) = \{v_1, \ldots, v_t\}$ such that $uv_1, uv_2, \ldots, uv_t$ are in anticlockwise order and the bounded region with boundary consisting of uv_1, v_1w, wv_t, v_tu contains all the v_i's. For $1 \le i \le t-1$, let R_i be the bounded region with boundary consisting of $uv_i, v_iw, wv_{i+1}, v_{i+1}u$. Suppose that there are at least $7 \cdot \lceil n^{4/5} \rceil$ values of i for which the interior of R_i contains a vertex of G. Then we can easily find a $K_{2,7\cdot\lceil n^{4/5}\rceil}$ in G in which no $K_{2,7}$ is empty, so by Corollary 9.3.4 there is a vertex in G that is contained in at most $11n/20$ induced C_5's, which is a contradiction. Thus, for all but $o(n)$ choices $6 \le i \le t-6$ the regions $R_{i-5}, R_{i-4}, \ldots, R_{i+5}$ contain no vertex in their interior. But for all such i, by property 1 we have that v_i is contained in at most $2n/3 + o(n)$ induced C_5's.

Let us remove the vertices v_i for these values of i from G and note that with this we remove at least $n/3 - o(1)$ vertices but at most $(\frac{2}{9} + o(1))n^2$ induced C_5's (since, by property 2, we have $|N(u) \cap N(w)| \leq n/3$). It suffices to show that in the remaining graph G' there are at most $o(n^2)$ induced C_5's. Let $S = V(G') \setminus (X \cup Y \cup \{u, w\})$. Note that $|S| = o(n)$.

Now we remove the vertices in S one by one in careful order, such that in each step we remove $O(n)$ induced C_5's. Note that any $v \in V(G)$ is joined to at most 6 vertices from $X \cup Y \cup \{u, w\}$. Thus, since G' is planar, we may remove the vertices of S one by one in a way that in each step the removed vertex has at most 11 neighbours in the current graph. This way, by Lemma 9.2.1, we remove at most $\binom{11}{2} \cdot n$ induced C_5's in each step. Thus, while removing the vertices in S, we remove at most $o(n^2)$ induced C_5's.

It remains to prove that in $G'' = G[X \cup Y \cup \{u, w\}]$ there are $o(n^2)$ induced C_5's. To show this, we prove that we may remove the vertices in $X \cup Y$ one by one such that in each step we remove $o(n)$ induced C_5's. Clearly, in each step we can remove a vertex $q \in X \cup Y$ which has degree at most 6 in the current graph. We claim that q is then contained in at most $o(n)$ induced C_5's. Let Z be the set of vertices $z \in X \cup Y$ for which there is a path of length at most 3 from q to z which avoids both u and w. Similarly as in Claim 2 in the previous lemma, it follows by property 2 that we have $|Z| = o(n)$. Since $N(u) \cap N(w) \cap (X \cup Y) = \emptyset$, there is no induced C_5 with vertices from $X \cup Y \cup \{u, w\}$ which contains both u and w, so any induced C_5 which contains q must consist of vertices from the set $Z \cup \{u, w\}$. Thus, as q has degree at most 6, by Lemma 9.2.1 there are at most $o(n)$ induced C_5's containing q. $\square$

Bibliography

[1] M. Ajtai, J. Komlós, and E. Szemerédi. A note on Ramsey numbers. *Journal of Combinatorial Theory, Series A*, 29(3):354–360, 1980.

[2] A. Alameddine. On the number of cycles of length 4 in a maximal planar graph. *Journal of Graph Theory*, 4:417–422, 1980.

[3] N. Alon. Explicit Ramsey graphs and orthonormal labelings. *The Electronic Journal of Combinatorics*, pages R12–R12, 1994.

[4] N. Alon, M. Krivelevich, and B. Sudakov. Turán numbers of bipartite graphs and related Ramsey-type questions. *Combinatorics, Probability and Computing*, 12:477–494, 2003.

[5] N. Alon, L. Rónyai, and T. Szabó. Norm-graphs: variations and applications. *Journal of Combinatorial Theory, Series B*, 76(2):280–290, 1999.

[6] J. Balogh, P. Hu, B. Lidický, and F. Pfender. Maximum density of induced 5-cycle is achieved by an iterated blow-up of 5-cycle. *European Journal of Combinatorics*, 52:47–58, 2016.

[7] B. Barak, P. K. Kothari, and D. Steurer. Small-set expansion in shortcode graph and the 2-to-2 conjecture. *arXiv preprint*, 2018.

[8] C. T. Benson. Minimal regular graphs of girths eight and twelve. *Canadian Journal of Mathematics*, 18:1091–1094, 1966.

[9] A. Bhowmick and S. Lovett. Bias vs structure of polynomials in large fields, and applications in effective algebraic geometry and coding theory. *arXiv preprint arXiv:1506.02047*, 2015.

[10] G. R. Blakley and P. Roy. A Hölder type inequality for symmetric matrices with nonnegative entries. In *Proc. Amer. Math. Soc*, volume 16, page 29, 1965.

[11] T. Bohman and P. Keevash. The early evolution of the H-free process. *Inventiones mathematicae*, 181(2):291–336, 2010.

[12] B. Bollobás, Y. Egawa, A. Harris, and G. Jin. The maximal number of induced r-partite subgraphs. *Graphs and Combinatorics*, 11:1–19, 1995.

[13] B. Bollobás and H. Hind. Graphs without large triangle free subgraphs. *Discrete Mathematics*, 87(2):119–131, 1991.

[14] B. Bollobás, C. Nara, and S.-i. Tachibana. The maximal number of induced complete bipartite graphs. *Discrete Mathematics*, 62:271–275, 1986.

[15] J. A. Bondy and M. Simonovits. Cycles of even length in graphs. *Journal of Combinatorial Theory, Series B*, 16(2):97–105, 1974.

[16] J. Brown and A. Sidorenko. The inducibility of complete bipartite graphs. *Journal of Graph Theory*, 18:629–645, 1994.

[17] B. Bukh. Random algebraic construction of extremal graphs. *Bulletin of the London Mathematical Society*, 47(6):939–945, 2015.

[18] B. Bukh and D. Conlon. Rational exponents in extremal graph theory. *J. Eur. Math. Soc.*, 20:1747–1757, 2018.

[19] F. Chung and L. Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Mathematics*, 3(1):79–127, 2006.

[20] D. Conlon. A sequence of triangle-free pseudorandom graphs. *Combinatorics, Probability and Computing*, 26(2):195–200, 2017.

[21] D. Conlon. Graphs with few paths of prescribed length between any two vertices. *Bull. London Math. Soc.*, to appear.

[22] D. Conlon, J. Fox, and B. Sudakov. Recent developments in graph Ramsey theory. *Surveys in combinatorics*, 424:49–118, 2015.

[23] D. Conlon, O. Janzer, and J. Lee. More on the extremal number of subdivisions. *Combinatorica*, to appear.

[24] D. Conlon and J. Lee. On the extremal number of subdivisions. *Int. Math. Res. Not.*, to appear.

[25] D. Conlon and M. Tyomkyn. Repeated patterns in proper colourings. *arXiv preprint arXiv:2002.00921*, 2020.

[26] S. Das, C. Lee, and B. Sudakov. Rainbow Turán problem for even cycles. *European Journal of Combinatorics*, 34(5):905–915, 2013.

[27] A. Dudek, T. Retter, and V. Rödl. On generalized Ramsey numbers of Erdős and Rogers. *Journal of Combinatorial Theory, Series B*, 109:213–227, 2014.

[28] A. Dudek and V. Rödl. On K_s-free subgraphs in K_{s+k}-free graphs and vertex Folkman numbers. *Combinatorica*, 31(1):39, 2011.

[29] A. Dudek and V. Rödl. On the function of Erdős and Rogers. In *Ramsey theory*, pages 63–76. Springer, 2011.

[30] P. Erdős. Some unsolved problems in graph theory and combinatorial analysis, 1971.

[31] P. Erdős and M. Simonovits. Lower bound for Turán number for bipartite non-degenerate graphs. `http://www.math.ucsd.edu/~erdosproblems/erdos/newproblems/TuranNondegenerate.html`.

[32] P. Erdős and M. Simonovits. A limit theorem in graph theory. *Studia Sci. Math. Hungar.*, 1:51–57, 1966.

[33] P. Erdős and A. H. Stone. On the structure of linear graphs. *Bull. Amer. Math. Soc.*, 52:1087–1091, 1946.

[34] P. Erdős. Some recent results on extremal problems in graph theory. results. *Theory of Graphs (Internat. Sympos., Rome, 1966)*, pages 117–123, 1967.

[35] P. Erdős. On the combinatorial problems which I would most like to see solved. *Combinatorica*, 1(1):25–42, 1981.

[36] P. Erdős. Some of my recent problems in combinatorial number theory, geometry and combinatorics. *Graph theory, combinatorics, and algorithms*, 1,2:335–349, 1995.

[37] P. Erdős and C. Rogers. The construction of certain graphs. *Canad. J. Math*, 14:702–707, 1962.

[38] P. Erdős and M. Simonovits. Some extremal problems in graph theory. In *Combinatorial theory and its applications, I*, (Proc. Colloq., Balatonfüred, 1969), pages 377–390. North-Holland, Amsterdam, 1970.

[39] P. Erdős and M. Simonovits. Compactness results in extremal graph theory. *Combinatorica*, 2(3):275–288, 1982.

[40] P. Erdős and M. Simonovits. Supersaturated graphs and hypergraphs. *Combinatorica*, 3(2):181–192, 1983.

[41] P. Erdős and G. Szekeres. A combinatorial problem in geometry. *Compositio mathematica*, 2:463–470, 1935.

[42] B. Ergemlidze, E. Győri, and A. Methuku. On the rainbow Turán number of paths. *The Electronic Journal of Combinatorics*, 26(1):P1.17, 2019.

[43] C. Even-Zohar and N. Linial. A note on the inducibility of 4-vertex graphs. *Graphs and Combinatorics*, 31:1367–1380, 2015.

[44] R. J. Faudree and M. Simonovits. On a class of degenerate extremal graph problems. *Combinatorica*, 3(1):83–93, 1983.

[45] S. Fish, C. Pohoata, and A. Sheffer. Local properties via color energy graphs and forbidden configurations. *SIAM Journal on Discrete Mathematics*, 34(1):177–187, 2020.

[46] Z. Füredi. On a Turán type problem of Erdős. *Combinatorica*, 11(1):75–79, 1991.

[47] Z. Füredi and M. Simonovits. The history of degenerate (bipartite) extremal graph problems. In *Erdös centennial*, volume 25 of *Bolyai Soc. Math. Stud.*, pages 169–264. János Bolyai Math. Soc., Budapest, 2013.

[48] Z. Füredi and D. B. West. Ramsey theory and bandwidth of graphs. *Graphs and Combinatorics*, 17(3):463–471, 2001.

[49] D. Ghosh, E. Győri, O. Janzer, A. Paulos, N. Salia, and O. Zamora. The maximum number of induced C_5's in a planar graph. *arXiv preprint arXiv:2004.01162*, 2020.

[50] M. Goemans. Chernoff bounds, and some applications. `http://math.mit.edu/~goemans/18310S15/chernoff-notes.pdf`.

[51] W. T. Gowers and O. Janzer. Subsets of Cayley graphs that induce many edges. *Theory of Computing*, 15(20):1-29, 2019.

[52] W. T. Gowers and O. Janzer. Improved bounds for the Erdős-Rogers function. *Advances in Combinatorics*, 2020:3, 27pp.

[53] W. T. Gowers and O. Janzer. The two-step local density of graphs. Work in progress.

[54] W. T. Gowers and J. Wolf. Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$. *Geometric and Functional Analysis*, 21(1):36–69, 2011.

[55] B. Green. The asymmetric Balog-Szemerédi-Gowers theorem. `http://people.maths.ox.ac.uk/greenbj/papers/asym-BSG.pdf`.

[56] B. Green and T. Tao. The distribution of polynomials over finite fields, with applications to the Gowers norm. *Contributions to Discrete Mathematics*, 4(2):1–36, 2009.

[57] A. Grzesik, O. Janzer, and Z. L. Nagy. The Turán number of blow-ups of trees. *Journal of Combinatorial Theory, Series B*, to appear.

[58] E. Győri, A. Paulos, N. Salia, C. Tompkins, and O. Zamora. The maximum number of pentagons in a planar graph. arXiv:1909.13532, 2019.

[59] S. Hakimi and E. Schmeichel. On the number of cycles of length k in a maximal planar graph. *Journal of Graph Theory*, 3:69–86, 1979.

[60] E. Haramaty and A. Shpilka. On the structure of cubic and quartic polynomials. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 331–340. ACM, 2010.

[61] H. Hatami, P. Hatami, and S. Lovett. Higher-order Fourier analysis and applications, 2016. `https://cseweb.ucsd.edu/~slovett/files/survey-higher_order_fourier.pdf`.

[62] J. Hirst. The inducibility of graphs on four vertices. *Journal of Graph Theory*, 75:231–243, 2014.

[63] T. Huynh, G. Joret, and D. R. Wood. Subgraph densities in a surface. arXiv:2003.13777, 2020.

[64] O. Janzer. Low analytic rank implies low partition rank for tensors. *arXiv preprint arXiv:1809.10931*, 2018.

[65] O. Janzer. Improved bounds for the extremal number of subdivisions. *The Electronic Journal of Combinatorics*, 26(3):P3.3, 2019.

[66] O. Janzer. The extremal number of the subdivisions of the complete bipartite graph. *SIAM Journal on Discrete Mathematics*, 34(1):241–250, 2020.

[67] O. Janzer. Rainbow Turán number of even cycles, repeated patterns and blow-ups of cycles. *arXiv preprint arXiv:2006.01062*, 2020.

[68] O. Janzer. Polynomial bound for the partition rank vs the analytic rank of tensors. *Discrete Analysis*, 2020:7, 17pp.

[69] O. Janzer. The extremal number of longer subdivisions. *Bulletin of the London Mathematical Society*, to appear.

[70] O. Janzer, A. Methuku, and Z. L. Nagy. On the Turán number of the blow-up of the hexagon. *arXiv preprint arXiv:2006.05897*, 2020.

[71] T. Jiang. Compact topological minors in graphs. *J. Graph Theory*, 67:139–152, 2011.

[72] T. Jiang, J. Ma, and L. Yepremyan. On Turán exponents of bipartite graphs. *arXiv preprint arXiv:1806.02838*, 2018.

[73] T. Jiang and A. Newman. Small dense subgraphs of a graph. *SIAM Journal on Discrete Mathematics*, 31(1):124–142, 2017.

[74] T. Jiang and Y. Qiu. Many Turán exponents via subdivisions. *arXiv preprint arXiv:1908.02385*, 2019.

[75] T. Jiang and Y. Qiu. Turán numbers of bipartite subdivisions. *SIAM Journal on Discrete Mathematics*, 34(1):556–570, 2020.

[76] T. Jiang and R. Seiver. Turán numbers of subdivided graphs. *SIAM J. Discrete Math.*, 26:1238–1255, 2012.

[77] D. Johnston, C. Palmer, and A. Sarkar. Rainbow Turán problems for paths and forests of stars. *The Electronic Journal of Combinatorics*, 24(1):P1.34, 2017.

[78] D. Johnston and P. Rombach. Lower bounds for rainbow Turán numbers of paths and other trees. *arXiv preprint arXiv:1901.03308*, 2019.

[79] D. Y. Kang, J. Kim, and H. Liu. On the rational Turán exponent conjecture. *arXiv preprint arXiv:1811.06916*, 2018.

[80] T. Kaufman and S. Lovett. Worst case to average case reductions for polynomials. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science(FOCS)*, pages 166–175, 2008.

[81] D. Kazhdan and T. Ziegler. Approximate cohomology. *Selecta Mathematica*, 24(1):499–509, 2018.

[82] P. Keevash, D. Mubayi, B. Sudakov, and J. Verstraëte. Rainbow Turán problems. *Combinatorics, Probability and Computing*, 16(1):109–126, 2007.

[83] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, pages 767–775. ACM, 2002.

[84] S. Khot, D. Minzer, and M. Safra. Pseudorandom sets in Grassmann graph have near-perfect expansion. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 25, 2018.

[85] Y. Kohayakawa, B. Nagle, V. Rödl, and M. Schacht. Weak hypergraph regularity and linear hypergraphs. *J. Combin. Theory Ser. B*, 100(2):151–160, 2010.

[86] J. Kollár, L. Rónyai, and T. Szabó. Norm-graphs and bipartite Turán numbers. *Combinatorica*, 16:399–406, 1996.

[87] G. N. Kopylov. A generalization of Turán's theorem. *Mathematical notes of the Academy of Sciences of the USSR*, 26(4):786–791, 1979.

[88] A. Kostochka and L. Pyber. Small topological complete subgraphs of "dense" graphs. *Combinatorica*, 8:83–86, 1988.

[89] T. Kővári, V. Sós, and P. Turán. On a problem of K. Zarankiewicz. *Colloquium Math.*, 3:50–57, 1954.

[90] M. Krivelevich. K_s-free graphs without large K_r-free subgraphs. *Combinatorics, Probability and Computing*, 3(3):349–354, 1994.

[91] M. Krivelevich. Bounding Ramsey numbers through large deviation inequalities. *Random Structures & Algorithms*, 7(2):145–155, 1995.

[92] A. Lampert. Bias implies low rank for quartic polynomials. *arXiv preprint arXiv:1902.10632*, 2019.

[93] S. Lovett. The analytic rank of tensors is subadditive, and its applications. *Discrete Analysis*, 2019.

[94] W. Mader. Homomorphieeigenschaften und mittlere kantendichte von graphen. *Mathematische Annalen*, 174(4):265–268, 1967.

[95] C. McDiarmid. *On the method of bounded differences*, pages 148–188. London Mathematical Society Lecture Note Series. Cambridge University Press, 1989.

[96] L. Milićević. Polynomial bound for partition rank in terms of analytic rank. *Geometric and Functional Analysis*, pages 1–28, 2019.

[97] N. Morrison and A. Scott. Maximising the number of induced cycles in a graph. *Journal of Combinatorial Theory, Series B*, 126:24–61, 2017.

[98] D. Mubayi and J. Verstraëte. A note on pseudorandom Ramsey graphs. *arXiv preprint arXiv:1909.01461*, 2019.

[99] H. Mulholland and C. A. Smith. An inequality arising in genetical theory. *The American Mathematical Monthly*, 66(8):673–683, 1959.

[100] E. Naslund. The partition rank of a tensor and k-right corners in $\mathbb{F}_q^n$. *Journal of Combinatorial Theory, Series A*, 174:105190, 2020.

[101] A. Razborov. On the minimal density of triangles in graphs. *Combinatorics, Probability and Computing*, 17:603–618, 2008.

[102] T. Sanders. On the Bogolyubov–Ruzsa lemma. *Analysis & PDE*, 5(3):627–655, 2012.

[103] R. Singleton. On minimal graphs of maximum even girth. *Journal of Combinatorial Theory*, 1(3):306–332, 1966.

[104] J. Spencer. Asymptotic lower bounds for Ramsey functions. *Discrete Mathematics*, 20:69–76, 1977.

[105] B. Sudakov. Large K_r-free subgraphs in K_s-free graphs and some other Ramsey-type problems. *Random Structures & Algorithms*, 26(3):253–265, 2005.

[106] B. Sudakov. A new lower bound for a Ramsey-type problem. *Combinatorica*, 25(4):487–498, 2005.

[107] B. Sudakov and I. Tomon. Turán number of bipartite graphs with no $K_{t,t}$. *Proceedings of the American Mathematical Society*, 2020.

[108] T. Tao and V. H. Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.

[109] G. Wolfovitz. K_4-free graphs without large induced triangle-free subgraphs. *Combinatorica*, 33(5):623–631, Oct 2013.

[110] Z. Xu, T. Zhang, Y. Jing, and G. Ge. Color isomorphic even cycles and a related Ramsey problem. *arXiv preprint arXiv:2004.01932*, 2020.