Sparse regularity and relative Szemerédi theorems

by

Yufei Zhao

S.B., S.B., Massachusetts Institute of Technology (2010)
M.A.St., University of Cambridge (2011)

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mathematics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

Signature redacted

Author ....
Department of Mathematics
March 31, 2015

Signature redacted

Certified by ....
Jacob Fox
Professor
Thesis Supervisor

Signature redacted

Accepted by ....
Peter Shor
Co-Chair, Department Committee on Graduate Students

Sparse regularity and relative Szemerédi theorems

by

Yufei Zhao

Submitted to the Department of Mathematics
on March 31, 2015, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mathematics

Abstract

We extend various fundamental combinatorial theorems and techniques from the dense setting to the sparse setting.

First, we consider Szemerédi's regularity lemma, a fundamental tool in extremal combinatorics. The regularity method, in its original form, is effective only for dense graphs. It has been a long standing problem to extend the regularity method to sparse graphs. We solve this problem by proving a so-called "counting lemma," thereby allowing us to apply the regularity method to relatively dense subgraphs of sparse pseudorandom graphs.

Next, by extending these ideas to hypergraphs, we obtain a simplification and extension of the key technical ingredient in the proof of the celebrated Green–Tao theorem, which states that there are arbitrarily long arithmetic progressions in the primes. The key step, known as a relative Szemerédi theorem, says that any positive proportion subset of a pseudorandom set of integers contains long arithmetic progressions. We give a simple proof of a strengthening of the relative Szemerédi theorem, showing that a much weaker pseudorandomness condition is sufficient.

Finally, we give a short simple proof of a multidimensional Szemerédi theorem in the primes, which states that any positive proportion subset of $\mathcal{P}^d$ (where $\mathcal{P}$ denotes the primes) contains constellations of any given shape. This has been conjectured by Tao and recently proved by Cook, Magyar, and Titichetrakun and independently by Tao and Ziegler.

Thesis Supervisor: Jacob Fox
Title: Professor

Acknowledgments

I am deeply indebted to my advisor Jacob Fox. I thank him for all his help and guidance throughout my time in graduate school. Under his mentorship, I was able to grow as a research mathematician. He has always been extremely generous with his time, resources, and ideas. I could not have imagined a more fruitful student-advisor relationship.

I thank Henry Cohn and Eyal Lubetzky for letting me intern with them and learn from them during my summers at Microsoft Research, which has been a really nice and productive summer getaway for me. I am grateful to Microsoft Research for their hospitality, as well as for generously funding my graduate studies during my final two years at MIT.

A special thanks to David Conlon for being a wonderful collaborator from whom I have learned much from.

I would also like to thank Richard Stanley, Joe Gallian, and Michel Goemans for their guidance early on and for encouraging me to pursue research during my undergraduate years. Without their influence, I might not have started my PhD in the first place.

The MIT Mathematics Department has provided a very productive work environment for me over the years. I also thank the Akamai Presidential Fellowship for funding my first year when I was still trying to find my way as a graduate student.

My time at MIT has been made much more enjoyable thanks to my friends and classmates. Being with them allowed me to grow as a person. I will miss our conversations and activities, which I have always found to be delightful and encouraging. I have also enjoyed being a part of the vibrant community at Sidney Pacific during my entire time as a graduate student at MIT.

Last but certainly not the least, none of this would have been possible without the love and support of my parents, and I could not thank them enough for all they have done.

Contents

1 Introduction **9**

2 Sparse reguarity **13**

2.1 Background and results 15

2.2 Counting strategy 32

2.3 Counting in Γ 38

2.4 Counting in G 47

2.5 Inheriting regularity 66

2.6 One-sided counting 74

2.7 Counting cycles 79

2.8 Applications 92

2.9 Concluding remarks 101

3 A relative Szemerédi theorem **117**

3.1 Definitions and results 120

3.2 The relative Szemerédi theorem 129

3.3 The relative hypergraph removal lemma 133

3.4 The weak regularity lemma 134

3.5 The counting lemma 139

3.6 Concluding remarks 147

4 Arithmetic transference **155**

4.1 Dense model theorem 157

4.2 Counting lemma 160
4.3 Proof of the relative Szemerédi theorem 161

5 Multidimensional Szemerédi theorem in the primes **163**

Bibliography **171**

Chapter 1

Introduction

This thesis contains a collection of related results on the theme of extending classical combinatorial theorems from the dense setting to the sparse setting. We develop methods for *transferring* combinatorial theorems about dense sets to the sparse setting, so that these theorems can be applied as a black-box to relatively dense subsets of sparse pseudorandom sets. We consider applications both in graph theory (primarily Turán, Ramsey, and removal-type results) as well as in additive combinatorics (Szemerédi's theorem).

Chapter 2 is based on the paper

[31] D. Conlon, J. Fox, and Y. Zhao, *Extremal results in sparse pseudorandom graphs*, Adv. Math. **256** (2014), 206–290.

It extends the graph regularity method from the dense setting to the sparse setting.

Szemerédi's regularity lemma [121] is a fundamental tool in extremal combinatorics. However, the original version is effective only for dense graphs. In the 1990s, Kohayakawa [75] and Rödl (unpublished) proved an analogue of Szemerédi's regularity lemma for sparse graphs as part of a general program toward extending extremal results to sparse graphs. Many of the key applications of Szemerédi's regularity lemma use an associated counting lemma. In order to prove extensions of these results which also apply to sparse graphs, it remained a well-known open problem to prove a counting lemma in sparse graphs.

The main advance of Chapter 2 lies in a new counting lemma, which complements the sparse regularity lemma of Kohayakawa and Rödl, allowing us to count small graphs in regular subgraphs of a sufficiently pseudorandom graph. We use this to prove sparse extensions of several well-known combinatorial theorems, including the removal lemmas for graphs and groups, the Erdős-Stone-Simonovits theorem and Ramsey's theorem. These results extend and improve upon a substantial body of previous work.

Chapter 3, the most significant part of this thesis, is based on the paper

[30] D. Conlon, J. Fox, and Y. Zhao, *A relative Szemerédi theorem.*, Geom. Funct. Anal., to appear.

It simplifies and strengthens the key technical result in the proof of the Green–Tao theorem [69] by extending some ideas used in Chapter 2 to hypergraphs.

The celebrated Green–Tao theorem states that there are arbitrarily long arithmetic progressions in the primes. A key input to the proof of the Green–Tao theorem is Szemerédi's theorem [120], which says that every subset of integers with positive upper density contains arbitrarily long arithmetic progressions. Szemerédi's theorem is a deep and important result and it has had a huge impact on the subsequent development of combinatorics and, in particular, was responsible for the introduction of the regularity method.

The primes have zero density in the integers, so the Green-Tao theorem does not follow directly from Szemerédi's theorem. Instead, the key idea of Green and Tao is to embed the primes as a relatively dense subset of "almost primes," which is a sparse pseudorandom set of integers. They established a *transference principle*, allowing them to apply Szemerédi's theorem to the sparse setting.

The proof of the Green-Tao theorem has two key steps. The first step, which Green and Tao refer to as the "main new ingredient" of their proof, is to establish a *relative Szemerédi theorem.* Informally speaking, such a result says that if S is a (sparse) set of integers satisfying certain pseudorandomness conditions and A is a subset of S with positive relative density, then A contains long arithmetic progressions. The second step is to construct an appropriate superset of the primes and verify that it has the desired pseudorandomness properties.

In the work of Green and Tao, the pseudorandomness conditions on the ground set are known as the *linear forms condition* and the *correlation condition.* Roughly speaking, both conditions say that, in terms of the number of solutions to certain linear systems of equations, the set behaves like a random set of the same density. Of the two conditions, the correlation condition is more technical and seems less relevant to arithmetic progressions. A natural question is whether a relative Szemerédi theorem holds under weaker pseudorandomness hypotheses. In Chapter 3, we answer this question in the affirmative, showing that a weak linear forms condition is sufficient for the relative Szemerédi theorem to hold, thereby completely removing the need for a correlation condition.

Our strengthened version can be applied to give the first relative Szemerédi theorem for k-term arithmetic progressions in pseudorandom subsets of $\mathbb{Z}_N$ of density N^{-c_k}.

The key component in our proof is an extension of the regularity method to sparse pseudorandom hypergraphs, which is interesting in its own right. From this we derive a relative extension of the hypergraph removal lemma. This is a strengthening of an earlier theorem used by Tao [123] in his proof that the Gaussian primes contain arbitrarily shaped constellations and, by standard arguments, allows us to deduce the relative Szemerédi theorem.

Although the techniques in Chapter 3 were inspired by the work that had been completed earlier and presented in Chapter 2, these two chapters are logically independent of each other. The approach presented in Chapter 3 is more streamlined in comparison.

Chapter 4 is based on

[134] Y. Zhao, *An arithmetic transference proof of a relative Szemerédi theorem*, Math. Proc. Cambridge Philos. Soc. **156** (2014), 255–261.

This chapter contains an alternate roadmap for proving the relative Szemerédi theorem in Chapter 3. Instead of applying the hypergraph removal lemma, we show that one can in fact more directly apply Szemerédi's theorem as a black box. This approach provides a somewhat more direct route to establishing the relative Szemerédi theorem, and it gives better quantitative bounds.

Chapter 5 is based on

[50] J. Fox and Y. Zhao, *A short proof of the multidimensional Szemerédi theorem in the primes*, Amer. J. Math., to appear.

It gives a short proof of the following theorem: every subset of $\mathcal{P}^d$ (where $\mathcal{P}$ denotes the set of primes) of positive relative density contains constellations of any given shape. This had been conjectured by Tao [123], and it was very recently proved by Cook, Magyar, and Titichetrakun [35] and independently by Tao and Ziegler [125]. Here we give a simple proof using the Green–Tao theorem on linear equations in primes and the Furstenberg–Katznelson multidimensional Szemerédi theorem.

Chapter 2

Sparse reguarity

Szemerédi's regularity lemma is one of the most powerful tools in extremal combinatorics. Roughly speaking, it says that the vertex set of every graph can be partitioned into a bounded number of parts so that the induced bipartite graph between almost all pairs of parts is pseudorandom. Many important results in graph theory, such as the graph removal lemma and the Erdős-Stone-Simonovits theorem on Turán numbers, have straightforward proofs using the regularity lemma.

Crucial to most applications of the regularity lemma is the use of a counting lemma. A counting lemma, roughly speaking, is a result that says that the number of embeddings of a fixed graph H into a pseudorandom graph G can be estimated by pretending that G were a genuine random graph. The combined application of the regularity lemma and a counting lemma is known as the regularity method, and has important applications in graph theory, combinatorial geometry, additive combinatorics and theoretical computer science. For surveys on the regularity method and its applications, see [78, 83, 105].

One of the limitations of Szemerédi's regularity lemma is that it is only meaningful for dense graphs. While an analogue of the regularity lemma for sparse graphs has been proven by Kohayakawa [75] and by Rödl (see also [57, 112]), the problem of proving an associated counting lemma for sparse graphs has turned out to be much more difficult. In random graphs, proving such an embedding lemma is a famous problem, known as the KŁR conjecture [76], which has only been resolved very recently

[10, 33, 110].

Establishing an analogous result for pseudorandom graphs has been a central problem in this area. Certain partial results are known in this case [80, 82], but it has remained an open problem to prove a counting lemma for embedding a general fixed subgraph. We resolve this difficulty, proving a counting lemma for embedding any fixed small graph into subgraphs of sparse pseudorandom graphs.

As applications, we prove sparse extensions of several well-known combinatorial theorems, including the removal lemmas for graphs and groups, the Erdős-Stone-Simonovits theorem, and Ramsey's theorem. Proving such sparse analogues for classical combinatorial results has been an important trend in modern combinatorics research. For example, a sparse analogue of Szemerédi's theorem was an integral part of Green and Tao's proof [69] that the primes contain arbitrarily long arithmetic progressions.

Organization. We will begin in the next section by giving an overview of the background and our results. In Section 2.2, we give a high level overview of the proof of our counting lemmas. In Section 2.3, we prove some useful statements about counting in the pseudorandom graph Γ. Then, in Section 2.4, we prove the sparse counting lemma, Theorem 2.1.12. The short proof of Proposition 2.1.13 and some related propositions about inheritance are given in Section 2.5. The proof of the one-sided counting lemma, which uses inheritance, is then given in Section 2.6. In Section 2.7, we take a closer look at one-sided counting in cycles. The sparse counting lemma has a large number of applications extending many classical results to the sparse setting. In Section 2.8, we discuss a number of them in detail, including sparse extensions of the Erdős-Stone-Simonovits theorem, Ramsey's theorem, the graph removal lemma and the removal lemma for groups. In Section 2.9 we briefly discuss a number of other applications, such as relative quasirandomness, induced Ramsey numbers, algorithmic applications and multiplicity results.

2.1 Background and results

2.1.1 Pseudorandom graphs

The binomial random graph $G_{n,p}$ is formed by taking an empty graph on n vertices and choosing to add each edge independently with probability p. These graphs tend to be very well-behaved. For example, it is not hard to show that with high probability all large vertex subsets X, Y have density approximately p between them. Motivated by the question of determining when a graph behaves in a random-like manner, Thomason [127, 128] began the first systematic study of this property. Using a slight variant of Thomason's notion, we say that a graph on vertex set V is *(p, β)-jumbled* if, for all vertex subsets $X, Y \subseteq V$,

$$|e(X,Y) - p|X||Y|| \leq \beta\sqrt{|X||Y|}.$$

The random graph $G_{n,p}$ is, with high probability, (p, β)-jumbled with $\beta = O(\sqrt{pn})$. It is not hard to show [41, 43] that this is optimal and that a graph on n vertices with $p \leq 1/2$ cannot be (p, β)-jumbled with $\beta = o(\sqrt{pn})$. Nevertheless, there are many explicit examples which are optimally jumbled in that $\beta = O(\sqrt{pn})$. The Paley graph with vertex set $\mathbb{Z}_p$, where $p \equiv 1 (\mathrm{mod}\ 4)$ is prime, and edge set given by connecting x and y if their difference is a quadratic residue is such a graph with $p = \frac{1}{2}$ and $\beta = O(\sqrt{n})$. Many more examples are given in the excellent survey [86].

A fundamental result of Chung, Graham and Wilson [23] states that for graphs of density p, where p is a fixed positive constant, the property of being $(p, o(n))$-jumbled is equivalent to a number of other properties that one would typically expect in a random graph. The following theorem details some of these many equivalences.

Theorem. *For any fixed $0 < p < 1$ and any sequence of graphs $(\Gamma_n)_{n \in \mathbb{N}}$ with $|V(\Gamma_n)| = n$ the following properties are equivalent.*

P_1: *Γ_n is $(p, o(n))$-jumbled, that is, for all subsets $X, Y \subseteq V(\Gamma_n)$, $e(X,Y) = p|X||Y| + o(n^2)$;*

P_2: *$e(\Gamma_n) \geq p\binom{n}{2} + o(n^2)$, $\lambda_1(\Gamma_n) = pn + o(n)$ and $|\lambda_2(\Gamma_n)| = o(n)$, where $\lambda_i(\Gamma_n)$ is the ith largest eigenvalue, in absolute value, of the adjacency matrix of Γ_n;*

P_3: *for all graphs H, the number of labeled induced copies of H in Γ_n is $p^{e(H)}(1-p)^{\binom{\ell}{2}-e(H)}n^\ell + o(n^\ell)$, where $\ell = V(H)$;*

P_4: *$e(\Gamma_n) \geq p\binom{n}{2} + o(n^2)$ and the number of labeled cycles of length 4 in Γ_n is at most $p^4n^4 + o(n^4)$.*

Any graph sequence which satisfies any (and hence all) of these properties is said to be *p-quasirandom.* The most surprising aspect of this theorem, already hinted at in Thomason's work, is that if the number of cycles of length 4 is as one would expect in a binomial random graph then this is enough to imply that the edges are very well-spread. This theorem has been quite influential. It has led to the study of quasirandomness in other structures such as hypergraphs [20, 62], groups [64], tournaments, permutations and sequences (see [22] and it references), and progress on problems in different areas (see, e.g., [26, 63, 64]). It is also closely related to Szemerédi's regularity lemma and its recent hypergraph generalization [63, 96, 106, 124] and all proofs of Szemerédi's theorem [120] on long arithmetic progressions in dense subsets of the integers use some notion of quasirandomness.

For sparser graphs, the equivalences between the natural generalizations of these properties are not so clear cut (see [21, 77, 82] for discussions). In this case, it is natural to generalize the jumbledness condition for dense graphs by considering graphs which are $(p, o(pn))$-jumbled. Otherwise, we would not even have control over the density in the whole set. However, it is no longer the case that being $(p, o(pn))$-jumbled implies that the number of copies of any subgraph H agrees approximately with the expected count. For $H = K_{3,3}$ and $p = n^{-1/3}$, it is easy to see this by taking the random graph $G_{n,p}$ and changing three vertices u, v and w so that they are each connected to everything else. This does not affect the property of being $(p, o(pn))$-jumbled but it does affect the $K_{3,3}$ count, since as well as the roughly $p^9n^6 = n^3$ copies of $K_{3,3}$ that one expects in a random graph, one gets a further $\Omega(n^3)$ copies of $K_{3,3}$ containing all of u, v and w.

However, for any given graph H one can find a function $\beta_H := \beta_H(p, n)$ such that if Γ is a (p, β_H)-jumbled graph on n vertices then Γ contains a copy of H. Our chief concern in this paper will be to determine jumbledness conditions which are sufficient to imply other properties. In particular, we will be concerned with determining conditions under which certain combinatorial theorems continue to hold within jumbled graphs.

One particularly well-known class of (p, β)-jumbled graphs is the collection of (n, d, λ)*-graphs*. These are graphs on n vertices which are d-regular and such that all eigenvalues of the adjacency matrix, save the largest, are smaller in absolute value than λ. The famous expander mixing lemma tells us that these graphs are (p, β)-jumbled with $p = d/n$ and $\beta = \lambda$. Bilu and Linial [14] proved a converse of this fact, showing that every (p, β)-jumbled d-regular graph is an (n, d, λ)-graph with $\lambda = O(\beta \log(d/\beta))$. This shows that the jumbledness parameter β and the second largest in absolute value eigenvalue λ of a regular graph are within a logarithmic factor of each other.

Pseudorandom graphs have many surprising properties and applications and have recently attracted a lot of attention both in combinatorics and theoretical computer science (see, e.g., [86]). Here we will focus on their behavior with respect to extremal properties. We discuss these properties in the next section.

2.1.2 Extremal results in pseudorandom graphs

In this paper, we study the extent to which several well-known combinatorial statements continue to hold relative to pseudorandom graphs or, rather, (p, β)-jumbled graphs and (n, d, λ)-graphs.

One of the most important applications of the regularity method is the graph removal lemma [3, 108]. In the following statement and throughout the paper, $v(H)$ and $e(H)$ will denote the number of vertices and edges in the graph H, respectively. The graph removal lemma states that for every fixed graph H and every $\epsilon > 0$ there exists $\delta > 0$ such that if G contains at most $\delta n^{v(H)}$ copies of H then G may be made H-free by removing at most ϵn^2 edges. This innocent looking result, which follows easily from Szemerédi's regularity lemma and the graph counting lemma, has surprising applications in diverse areas, amongst others a simple proof of Roth's theorem on 3-term arithmetic progressions in dense subsets of the integers. It is also much more difficult to prove than one might expect, the best known bound [46] on δ^{-1} being a tower function of height on the order of $\log \epsilon^{-1}$.

An analogue of this result for random graphs (and hypergraphs) was proven in [32]. For pseudorandom graphs, the following analogue of the triangle removal lemma was recently proven by Kohayakawa, Rödl, Schacht and Skokan [80].

Theorem. *For every $\epsilon > 0$, there exist $\delta > 0$ and $c > 0$ such that if $\beta \leq cp^3 n$ then any (p, β)-jumbled graph Γ on n vertices has the following property. Any subgraph of Γ containing at most $\delta p^3 n^3$ triangles may be made triangle-free by removing at most $\epsilon p n^2$ edges.*

Here we extend this result to all H. The *degeneracy* $d(H)$ of a graph H is the smallest nonnegative integer d for which there exists an ordering of the vertices of H such that each vertex has at most d neighbors which appear earlier in the ordering. Equivalently, it may be defined as $d(H) = \max\{\delta(H') : H' \subseteq H\}$, where $\delta(H)$ is the minimum degree of H. Throughout the paper, we will also use the standard notation $\Delta(H)$ for the maximum degree of H.

The parameter we will use in our theorems, which we refer to as the 2-*degeneracy* $d_2(H)$, is related to both of these natural parameters. Given an ordering $v_1, \ldots, v_m$ of the vertices of H and $i \leq j$, let $N_{i-1}(j)$ be the number of neighbors v_h of v_j with $h \leq i-1$. We then define $d_2(H)$ to be the minimum d for which there is an ordering of the edges as $v_1, \ldots, v_m$ such that for any edge $v_i v_j$ with $i < j$ the sum $N_{i-1}(i) + N_{i-1}(j) \leq 2d$. Note that $d_2(H)$ may be a half-integer. For comparison with degeneracy, note that $\frac{d(H)}{2} \leq d_2(H) \leq d(H) - \frac{1}{2}$ and both sides can be sharp.

Theorem 2.1.1. *For every graph H and every $\epsilon > 0$, there exist $\delta > 0$ and $c > 0$ such that if $\beta \leq cp^{d_2(H)+3} n$ then any (p, β)-jumbled graph Γ on n vertices has the following property. Any subgraph of Γ containing at most $\delta p^{e(H)} n^{v(H)}$ copies of H may be made H-free by removing at most $\epsilon p n^2$ edges.*

We remark that for many graphs H, the constant 3 in the exponent of this theorem may be improved, and this applies equally to all of the theorems stated below. While we will not dwell on this comment, we will call attention to it on occasion throughout the paper, pointing out where the above result may be improved. Note that the above theorem generalizes the graph removal lemma by considering the case $\Gamma = K_n$, which is (p, β)-jumbled with $p = 1$ and $\beta = 1$. For the same reason, the other results we establish extend the original versions.

Green [68] developed an arithmetic regularity lemma and used it to deduce an arithmetic removal lemma in abelian groups which extends Roth's theorem. Green's proof of the arithmetic regularity lemma relies on Fourier analytic techniques. Král, Serra and Vena [84] found a new proof of the arithmetic removal lemma using the removal lemma for directed cycles which extends to all groups. They proved that for each $\epsilon > 0$ and integer $m \geq 3$ there is $\delta > 0$ such that if G is a group of order n and $A_1, \ldots, A_m$ are subsets of G such that there are at most δn^{m-1} solutions to the equation $x_1 x_2 \cdots x_m = 1$ with $x_i \in A_i$ for all i, then it is possible to remove at most ϵn elements from each set A_i so as to obtain sets A_i' for which there are no solutions to $x_1 x_2 \cdots x_m = 1$ with $x_i \in A_i'$ for all i.

By improving the bound in Theorem 2.1.1 for cycles, we obtain the following sparse extension of the removal lemma for groups. The *Cayley graph* $G(S)$ of a subset S of a group G has vertex set G and (x, y) is an edge of G if $x^{-1}y \in S$. We say that a subset S of a group G is (p, β)-jumbled if the Cayley graph $G(S)$ is (p, β)-jumbled. When G is abelian, if $\left|\sum_{x \in S} \chi(x)\right| \leq \beta$ for all nontrivial characters $\chi \colon G \to \mathbb{C}$, then S is $(\frac{|S|}{|G|}, \beta)$-jumbled (see [80, Lemma 16]). Let $k_3 = 3$, $k_4 = 2$, $k_m = 1 + \frac{1}{m-3}$ if $m \geq 5$ is odd, and $k_m = 1 + \frac{1}{m-4}$ if $m \geq 6$ is even. Note that k_m tends to 1 as $m \to \infty$.

Theorem 2.1.2. *For each $\epsilon > 0$ and integer $m \geq 3$, there are $c, \delta > 0$ such that the following holds. Suppose $B_1, \ldots, B_m$ are subsets of a group G of order n such that each B_i is (p, β)-jumbled with $\beta \leq cp^{k_m} n$. If subsets $A_i \subseteq B_i$ for $i = 1, \ldots, m$ are such that there are at most $\delta |B_1| \cdots |B_m| / n$ solutions to the equation $x_1 x_2 \cdots x_m = 1$ with $x_i \in A_i$ for all i, then it is possible to remove at most $\epsilon |B_i|$ elements from each set A_i so as to obtain sets A_i' for which there are no solutions to $x_1 x_2 \cdots x_m = 1$ with*

$x_i \in A_i'$ *for all* i.

This result easily implies a Roth-type theorem in quite sparse pseudorandom subsets of a group. We say that a subset B of a group G is (ϵ, m)-*Roth* if for all integers $a_1, \ldots, a_m$ which satisfy $a_1 + \cdots + a_m = 0$ and $\gcd(a_i, |G|) = 1$ for $1 \le i \le m$, every subset $A \subseteq B$ which has no nontrivial solution to $x_1^{a_1} x_2^{a_2} \cdots x_m^{a_m} = 1$ has $|A| \le \epsilon |B|$.

Corollary 2.1.3. *For each* $\epsilon > 0$ *and integer* $m \ge 3$, *there is* $c > 0$ *such that the following holds. If* G *is a group of order* n *and* B *is a* (p, β)*-jumbled subset of* G *with* $\beta \le c p^{k_m} n$, *then* B *is* (ϵ, m)*-Roth.*

Note that Roth's theorem on 3-term arithmetic progressions in dense sets of integers, follows from the special case of this result with $B = G = \mathbb{Z}_n$, $m = 3$, and $a_1 = a_2 = 1$, $a_3 = -2$. The rather weak pseudorandomness condition in Corollary 2.1.3 shows that even quite sparse pseudorandom subsets of a group have the Roth property. For example, if B is optimally jumbled, in that $\beta = O(\sqrt{pn})$ and $p \ge C n^{-\frac{1}{2k_m - 1}}$, then B is (ϵ, m)-Roth. This somewhat resembles a key part of the proof of the Green-Tao theorem that the primes contain arbitrarily long arithmetic progressions, where they show that pseudorandom sets of integers have the Szemerédi property. As Corollary 2.1.3 applies to quite sparse pseudorandom subsets, it may lead to new applications in number theory.

Our methods are quite general and also imply similar results for other well-known combinatorial theorems. We say that a graph Γ is (H, ϵ)-*Turán* if any subgraph of Γ with at least

$$\left(1 - \frac{1}{\chi(H) - 1} + \epsilon\right) e(\Gamma)$$

edges contains a copy of H. Turán's theorem itself [132], or rather a generalization known as the Erdős-Stone-Simonovits theorem [44], says that K_n is (H, ϵ)-Turán provided that n is large enough.

To find other graphs which are (H, ϵ)-Turán, it is natural to try the random graph $G_{n,p}$. A recent result of Conlon and Gowers [32], also proved independently by Schacht [111], states that for every $t \ge 3$ and $\epsilon > 0$ there exists a constant C such that if $p \ge C n^{-2/(t+1)}$ the graph $G_{n,p}$ is with high probability (K_t, ϵ)-Turán. This confirms a conjecture of Haxell, Kohayakawa, Łuczak and Rödl [74, 76] and, up to the constant C, is best possible. Similar results also hold for more general graphs H and hypergraphs.

For pseudorandom graphs and, in particular, (n, d, λ)-graphs, Sudakov, Szabó and Vu [119] showed the following. A similar result, but in a slightly more general context, was proved by Chung [19].

Theorem. *For every $\epsilon > 0$ and every positive integer $t \geq 3$, there exists $c > 0$ such that if $\lambda \leq cd^{t-1}/n^{t-2}$ then any (n, d, λ)-graph is (K_t, ϵ)-Turán.*

For $t = 3$, an example of Alon [2] shows that this is best possible. His example gives something even stronger, a triangle-free (n, d, λ)-graph for which $\lambda \leq c\sqrt{d}$ and $d \geq n^{2/3}$. Therefore, no combinatorial statement about the existence of triangles as subgraphs can surpass the threshold $\lambda \leq cd^2/n$. It has also been conjectured [47, 86, 119] that $\lambda \leq cd^{t-1}/n^{t-2}$ is a natural boundary for finding K_t as a subgraph in a pseudorandom graph, but no examples of such graphs exist for $t \geq 4$. Finding such graphs remains an important open problem on pseudorandom graphs.

For triangle-free graphs H, Kohayakawa, Rödl, Schacht, Sissokho and Skokan [79] proved the following result which gives a jumbledness condition that implies that a graph is (H, ϵ)-Turán.

Theorem. *For any fixed triangle-free graph H and any $\epsilon > 0$, there exists $c > 0$ such that if $\beta \leq cp^{\nu(H)}n$ then any (p, β)-jumbled graph on n vertices is (H, ϵ)-Turán. Here $\nu(H) = \frac{1}{2}(d(H) + D(H) + 1)$, where $D(H) = \min\{2d(H), \Delta(H)\}$.*

More recently, the case where H is an odd cycle was studied by Aigner-Horev, Hàn and Schacht [1], who proved the following result, optimal up to the logarithmic factors [4].

Theorem. *For every odd integer $\ell \geq 3$ and any $\epsilon > 0$, there exists $c > 0$ such that if $\beta \log^{\ell-3} n \leq cp^{1+1/(\ell-2)}n$ then any (p, β)-jumbled graph on n vertices is (C_ℓ, ϵ)-Turán.*

In this paper, we prove that a similar result holds, but for general graphs H and, in most cases, with a better bound on β.

Theorem 2.1.4. *For every graph H and every $\epsilon > 0$, there exists $c > 0$ such that if $\beta \leq cp^{d_2(H)+3}n$ then any (p, β)-jumbled graph on n vertices is (H, ϵ)-Turán.*

We may also prove a structural version of this theorem, known as a stability result. In the dense case, this result, due to Erdős and Simonovits [115], states that if an H-free graph contains almost $\left(1 - \frac{1}{\chi(H)-1}\right)\binom{n}{2}$ edges, then it must be very close to being $(\chi(H) - 1)$-partite.

Theorem 2.1.5. *For every graph H and every $\epsilon > 0$, there exist $\delta > 0$ and $c > 0$ such that if $\beta \leq cp^{d_2(H)+3}n$ then any (p, β)-jumbled graph Γ on n vertices has the following property. Any H-free subgraph of Γ with at least $\left(1 - \frac{1}{\chi(H)-1} - \delta\right) p\binom{n}{2}$ edges may be made $(\chi(H) - 1)$-partite by removing at most ϵpn^2 edges.*

The final main result that we will prove concerns Ramsey's theorem [98]. This states that for any graph H and positive integer r, if n is sufficiently large, any r-coloring of the edges of K_n contains a monochromatic copy of H.

To consider the analogue of this result in sparse graphs, let us say that a graph Γ is (H, r)-*Ramsey* if, in any r-coloring of the edges of Γ, there is guaranteed to be a monochromatic copy of H. For $G_{n,p}$, a result of Rödl and Ruciński [104] determines the threshold down to which the random graph is (H, r)-Ramsey with high probability. For most graphs, including the complete graph K_t, this threshold is the same as for the Turán property. These results were only extended to hypergraphs comparatively recently, by Conlon and Gowers [32] and by Friedgut, Rödl and Schacht [52].

Very little seems to be known about the (H, r)-Ramsey property relative to pseudorandom graphs. In the triangle case, results of this kind are implicit in some recent papers [37, 91] on Folkman numbers, but no general theorem seems to be known. We prove the following.

Theorem 2.1.6. *For every graph H and every positive integer $r \geq 2$, there exists $c > 0$ such that if $\beta \leq cp^{d_2(H)+3}n$ then any (p, β)-jumbled graph on n vertices is (H, r)-Ramsey.*

One common element to all these results is the requirement that $\beta \leq cp^{d_2(H)+3}n$. It is not hard to see that this condition is almost sharp. Consider the binomial random graph on n vertices where each edge is chosen with probability $p = cn^{-2/d(H)}$, where $c < 1$. By the definition of degeneracy, there exists some subgraph H' of H such that $d(H)$ is the minimum degree of H'. Therefore, $e(H') \geq v(H')d(H)/2$ and the expected number of copies of H' is at most

$$p^{e(H')}n^{v(H')} \leq (p^{d(H)/2}n)^{v(H')} < 1.$$

We conclude that with positive probability $G_{n,p}$ does not contain a copy of H' or, consequently, of H. On the other hand, with high probability, it is (p, β)-jumbled with

$$\beta = O(\sqrt{pn}) = O(p^{(d(H)+2)/4}n).$$

Since $d_2(H)$ differs from $d(H)$ by at most a constant factor, we therefore see that, up to a multiplicative constant in the exponent of p, our results are best possible.

If $H = K_t$, it is sufficient, for the various combinatorial theorems above to hold, that the graph Γ be $(p, cp^t n)$-jumbled. For triangles, the example of Alon shows that there are $(p, cp^2 n)$-jumbled graphs which do not contain any triangles and, for $t \geq 4$, it is conjectured [47, 86, 119] that there are $(p, cp^{t-1}n)$-jumbled graphs which do not contain a copy of K_t. If true, this would imply that in the case of cliques all of our results are sharp up to an additive constant of one in the exponent. A further discussion relating to the optimal exponent of p for general graphs is in the concluding remarks.

2.1.3 Regularity and counting lemmas

One of the key tools in extremal graph theory is Szemerédi's regularity lemma [121]. Roughly speaking, this says that any graph may be partitioned into a collection of vertex subsets so that the bipartite graph between most pairs of vertex subsets is random-like. To be more precise, we introduce some notation. It will be to our advantage to be quite general from the outset.

A *weighted graph* on a set of vertices V is a symmetric function $G\colon V \times V \to [0, 1]$.

Here symmetric means that $G(x,y) = G(y,x)$. A weighted graph is bipartite (or multipartite) if it is supported on the edge set of a bipartite (or multipartite graph). A graph can be viewed as a weighted graph by taking G to be the characteristic function of the edges.

Note that here and throughout the remainder of the paper, we will use integral notation for summing over vertices in a graph. For example, if G is a bipartite graph with vertex sets X and Y, and f is any function $X \times Y \to \mathbb{R}$, then we write

$$\int_{\substack{x\in X\\ y\in Y}} f(x,y)\ dxdy := \frac{1}{|X|\,|Y|}\sum_{x\in X}\sum_{y\in Y} f(x,y).$$

The measure dx will always denote the uniform probability distribution on X. The advantage of the integral notation is that we do not need to keep track of the number of vertices in G. All our formulas are, in some sense, scale-free with respect to the order of G. Consequently, our results also have natural extensions to graph limits [89], although we do not explore this direction here.

Definition 2.1.7 (DISC). A weighted bipartite graph $G\colon X \times Y \to [0,1]$ is said to satisfy the *discrepancy* condition $\mathrm{DISC}(q,\epsilon)$ if

$$\left|\int_{\substack{x\in X\\ y\in Y}} (G(x,y) - q)u(x)v(y)\ dxdy\right| \le \epsilon \tag{2.1}$$

for all functions $u\colon X \to [0,1]$ and $v\colon Y \to [0,1]$. In any weighted graph G, if X and Y are subsets of vertices of G, we say that the pair $(X,Y)_G$ satisfies $\mathrm{DISC}(q,\epsilon)$ if the induced weighted graph on $X \times Y$ satisfies $\mathrm{DISC}(q,\epsilon)$.

The usual definition for discrepancy of an (unweighted) bipartite graph G is that for all $X' \subseteq X$, $Y' \subseteq Y$, we have $|e(X',Y') - q\,|X'|\,|Y'|| \le \epsilon\,|X|\,|Y|$. It is not hard to see that the two notions of discrepancy are equivalent (with the same ϵ).

A partition $V(G) = V_1 \cup \cdots \cup V_k$ is said to be *equitable* if all pieces in the partition are of comparable size, that is, if $||V_i| - |V_j|| \le 1$ for all i and j. Szemerédi's regularity lemma now says the following.

Theorem (Szemerédi's regularity lemma). *For every $\epsilon > 0$ and every positive integer m_0, there exists a positive integer M such that any weighted graph G has an equitable partition into k pieces with $m_0 \le k \le M$ such that all but at most ϵk^2 pairs of vertex subsets (V_i, V_j) satisfy* $\mathrm{DISC}(q_{ij}, \epsilon)$ *for some q_{ij}.*

On its own, the regularity lemma would be an interesting result. But what really makes it so powerful is the fact that the discrepancy condition allows us to count small subgraphs. In particular, we have the following result, known as a counting lemma.

Proposition 2.1.8 (Counting lemma in dense graphs). *Let G be a weighted m-partite graph with vertex subsets $X_1, X_2, \ldots, X_m$. Let H be a graph with vertex set $\{1, \ldots, m\}$ and with $e(H)$ edges. For each edge (i, j) in H, assume that the induced bipartite graph $G(X_i, X_j)$ satisfies* $\mathrm{DISC}(q_{ij}, \epsilon)$. *Define*

$$G(H) := \int_{x_1 \in X_1, \ldots, x_m \in X_m} \prod_{(i,j) \in E(H)} G(x_i, x_j)\ dx_1 \cdots dx_m$$

and

$$q(H) := \prod_{(i,j) \in E(H)} q_{ij}.$$

Then

$$|G(H) - q(H)| \le e(H)\epsilon.$$

The above result, for an unweighted graph, is usually stated in the following equivalent way: the number of embeddings of H into G, where the vertex $i \in V(H)$ lands in X_i, differs from $|X_1|\,|X_2| \cdots |X_m| \prod_{(i,j) \in E(H)} q_{ij}$ by at most $e(H)\epsilon\,|X_1|\,|X_2| \cdots |X_m|$. Our notation $G(H)$ can be viewed as the probability that a random embedding of vertices of H into their corresponding parts in G gives rise to a valid embedding as a subgraph.

Proposition 2.1.8 may be proven by telescoping (see, e.g., Theorem 2.7 in [16]). Consider, for example, the case where H is a triangle. Then

$$G(x_1, x_2)G(x_1, x_3)G(x_2, x_3) - q_{12}q_{13}q_{23}$$

may be rewritten as

$$(G(x_1,x_2)-q_{12})G(x_1,x_3)G(x_2,x_3)+q_{12}(G(x_1,x_3)-q_{13})G(x_2,x_3) \\ +q_{12}q_{13}(G(x_2,x_3)-q_{23}). \quad (2.2)$$

Applying the discrepancy condition (2.1), we see that, after integrating the above expression over all $x_1 \in X_1, x_2 \in X_2, x_3 \in X_3$, each term in (2.2) is at most ϵ in absolute value. The result follows for triangles. The general case follows similarly.

In order to prove extremal results in sparse graphs, we would like to transfer some of this machinery to the sparse setting. Because the number of copies of a subgraph in a sparse graph G is small, the error between the expected count and the actual count must also be small for a counting lemma to be meaningful. Another way to put this is that we aim to achieve a small multiplicative error in our count.

Since we require smaller errors when counting in sparse graphs, we need stronger discrepancy hypotheses. In the following definition, we should view p as the order of magnitude density of the graph, so that the error terms should be bounded in the same order of magnitude. In a dense graph, $p=1$. We assume that $q \le p$. It may be helpful to think of q/p as bounded below by some positive constant, although this is not strictly required.

Definition 2.1.9 (DISC). A weighted bipartite graph $G\colon X \times Y \to [0,1]$ is said to satisfy $\mathrm{DISC}(q,p,\epsilon)$ if

$$\left|\int_{\substack{x\in X\\ y\in Y}} (G(x,y)-q)u(x)v(y)\;dxdy\right| \le \epsilon p$$

for all functions $u\colon X \to [0,1]$ and $v\colon Y \to [0,1]$.

Unfortunately, discrepancy alone is not strong enough for counting in sparse graphs. Consider the following example. Let G be a tripartite graph with vertex sets X_1, X_2, X_3, such that $(X_1,X_2)_G$ and $(X_2,X_3)_G$ satisfy $\mathrm{DISC}(q,p,\frac{\epsilon}{2})$. Let X_2' be a subset of X_2 with size $\frac{\epsilon}{2}p\,|X_2|$. Let G' be modified from G by adding the complete bipartite graph between X_1 and X_2', as well as the complete bipartite graph between

X_2' and X_3. The resulting pairs $(X_1, X_2)_{G'}$ and $(X_2, X_3)_{G'}$ satisfy DISC(q, p, ϵ). Consider the number of paths in G and G' with one vertex from each of X_1, X_2, X_3 in turn. Given the densities, we expect there to be approximately $q^2 |X_1| |X_2| |X_3|$ such paths, and we would like the error to be $\delta p^2 |X_1| |X_2| |X_3|$ for some small δ that goes to zero as ϵ goes to zero. However, the number of paths in G' from X_1 to X_2' to X_3 is $\frac{\epsilon}{2} p |X_1| |X_2| |X_3|$, which is already too large when p is small.

For our counting lemma to work, G needs to be a relatively dense subgraph of a much more pseudorandom host graph Γ. In the dense case, Γ can be the complete graph. In the sparse world, we require Γ to satisfy the jumbledness condition. In practice, we will use the following equivalent definition. The equivalence follows by considering random subsets of X and Y, where x and y are chosen with probabilities $u(x)$ and $v(y)$, respectively.

Definition 2.1.10 (Jumbledness). Let $\Gamma = (X \cup Y, E_\Gamma)$ be a bipartite graph. We say that Γ is $(p, \gamma\sqrt{|X| |Y|})$*-jumbled* if

$$\left| \int_{\substack{x \in X \\ y \in Y}} (\Gamma(x,y) - p) u(x) v(y) \, dx dy \right| \leq \gamma \sqrt{\int_{x \in X} u(x) \, dx} \sqrt{\int_{y \in Y} v(y) \, dy} \tag{2.3}$$

for all functions $u \colon X \to [0,1]$ and $v \colon Y \to [0,1]$.

With the discrepancy condition defined as in Definition 2.1.9, we may now state a regularity lemma for sparse graphs. Such a lemma was originally proved independently by Kohayakawa [75] and by Rödl (see also [57, 112]). The following result, tailored specifically to jumbled graphs, follows easily from the main result in [75].

Theorem 2.1.11 (Regularity lemma in jumbled graphs). *For every $\epsilon > 0$ and every positive integer m_0, there exists $\eta > 0$ and a positive integer M such that if Γ is a $(p, \eta p n)$-jumbled graph on n vertices any weighted subgraph G of Γ has an equitable partition into k pieces with $m_0 \leq k \leq M$ such that all but at most ϵk^2 pairs of vertex subsets (V_i, V_j) satisfy* DISC(q_{ij}, p, ϵ) *for some q_{ij}.*

The main result of this paper is a counting lemma which complements this regularity lemma. Proving such an embedding lemma has remained an important open problem ever since Kohayakawa and Rödl first proved the sparse regularity lemma. Most of the work has focused on applying the sparse regularity lemma in the context of random graphs. The key conjecture in this case, known as the KŁR conjecture, concerns the probability threshold down to which a random graph is, with high probability, such that any regular subgraph contains a copy of a particular subgraph H. This conjecture has only been resolved very recently [10, 33, 110]. For pseudorandom graphs, it has been a wide open problem to prove a counting lemma which complements the sparse regularity lemma. The first progress on proving such a counting lemma was made recently in [80], where Kohayakawa, Rödl, Schacht and Skokan proved a counting lemma for triangles. Here, we prove a counting lemma which works for any graph H. Even for triangles, our counting lemma gives an improvement over the results in [80], since our results have polynomial-type dependence on the discrepancy parameters, whereas the results in [80] require exponential dependence since a weak regularity lemma was used as an intermediate step during their proof of the triangle counting lemma.

Our results are also related to the work of Green and Tao [69] on arithmetic progressions in the primes. What they really prove is the stronger result that Szemerédi's theorem on arithmetic progressions holds in subsets of the primes. In order to do this, they first show that the primes, or rather the almost primes, are a pseudorandom subset of the integers and then that Szemerédi's theorem continues to hold relative to such pseudorandom sets. In the language of their paper, our counting lemma is a generalized von Neumann theorem.

Here is the statement of our first counting lemma. Note that, given a graph H, the line graph $L(H)$ is the graph whose vertices are the edges of H and where two vertices are adjacent if their corresponding edges in H share an endpoint. Recall that $d(\cdot)$ is the degeneracy and $\Delta(\cdot)$ is the maximum degree.

Theorem 2.1.12. *Let H be a graph with vertex set $\{1, \ldots, m\}$ and with $e(H)$ edges. For every $\theta > 0$, there exist $c, \epsilon > 0$ of size at least polynomial in θ so that the following holds.*

Let $p > 0$ and let Γ be a graph with vertex subsets $X_1, \ldots, X_m$ and suppose that

Table 2.1: Sufficient conditions on k in the jumbledness hypothesis $(p, cp^k\sqrt{|X_i|\,|X_j|})$ for the counting lemmas of various graphs. Two-sided counting refers to results of the form $|G(H) - q(H)| \leq \theta p^{e(H)}$ while one-sided counting refers to result of the form $G(H) \geq q(H) - \theta p^{e(H)}$.

H	Two-sided counting $k \geq$	One-sided counting $k \geq$	
K_t	t	t	$t \geq 3$
C_ℓ	2	2	$\ell = 4$
	2	$1 + \frac{1}{2\lfloor(\ell-3)/2\rfloor}$	$\ell \geq 5$
$K_{2,t}$	$\frac{t+2}{2}$	$\frac{5}{2}$	$t \geq 3$
$K_{s,t}$	$\frac{s+t+1}{2}$	$\frac{s+3}{2}$	$3 \leq s \leq t$
Tree	$\frac{\Delta(H)+1}{2}$	No jumbledness needed	See Prop. 2.4.9 and 2.7.7
$K_{1,2,2}$	4	4	

the bipartite graph $(X_i, X_j)_\Gamma$ is $(p, cp^k\sqrt{|X_i|\,|X_j|})$-jumbled for every $i < j$, where $k \geq \min\left\{\frac{\Delta(L(H))+4}{2}, \frac{d(L(H))+6}{2}\right\}$. Let G be a subgraph of Γ, with the vertex i of H assigned to the vertex subset X_i of G. For each edge ij in H, assume that $(X_i, X_j)_G$ satisfies $\mathrm{DISC}(q_{ij}, p, \epsilon)$. *Define*

$$G(H) := \int_{x_1 \in X_1, \ldots, x_m \in X_m} \prod_{(i,j)\in E(H)} G(x_i, x_j)\ dx_1 \cdots dx_m$$

and

$$q(H) := \prod_{(i,j)\in E(H)} q_{ij}.$$

Then

$$|G(H) - q(H)| \leq \theta p^{e(H)}.$$

For some graphs H, our methods allow us to achieve slightly better values of k in Theorem 2.1.12. However, the value given in the theorem is the cleanest general statement. See Table 2.1 for some example of hypotheses on k for various graphs H. To see that the value of k is never far from best possible, we first note that $\Delta(H) - 1 \leq d(L(H)) \leq \Delta(H) + d(H) - 2$.

Let H have maximum degree Δ. By considering the random graph $G_{n,p}$ with

$p = n^{-1/\Delta}$, we can find a $(p, cp^{\Delta/2}n)$-jumbled graph Γ containing approximately $p^{e(H)}n^{v(H)}$ labeled copies of H. We modify Γ to form Γ' by fixing one vertex v and connecting it to everything else. It is easy to check that the resulting graph Γ' is $(p, c'p^{\Delta/2}n)$-jumbled. However, the number of copies of H disagrees with the expected count, since there are approximately $p^{e(H)}n^{v(H)}$ labeled copies from the original graph Γ and a further approximately $p^{e(H)-\Delta}n^{v(H)-1} = p^{e(H)}n^{v(H)}$ labeled copies containing v. We conclude that for $k < \Delta/2$ we cannot hope to have such a counting lemma and, therefore, the value of k in Theorem 2.1.12 is close to optimal.

Since we are dealing with sparse graphs, the discrepancy condition 2.1.9 appears, at first sight, to be rather weak. Suppose, for instance, that we have a sparse graph satisfying DISC(q, p, ϵ) between each pair of sets from V_1, V_2 and V_3 and we wish to embed a triangle between the three sets. Then, a typical vertex v in V_1 will have neighborhoods of size roughly $q|V_2|$ and $q|V_3|$ in V_2 and V_3, respectively. But now the condition DISC(q, p, ϵ) tells us nothing about the density of edges between the two neighborhoods. They are simply too small.

To get around this, Gerke, Kohayakawa, Rödl and Steger [55] showed that if (X, Y) is a pair satisfying DISC(q, p, ϵ) then, with overwhelmingly high probability, a small randomly chosen pair of subsets $X' \subseteq X$ and $Y' \subseteq Y$ will satisfy DISC(q, p, ϵ'), where ϵ' tends to zero with ϵ. We say that the pair inherits regularity. This may be applied effectively to prove embedding lemmas in random graphs (see, for example, [56, 81]). For pseudorandom graphs, the beginnings of such an approach may be found in [80].

Our approach in this paper works in the opposite direction. Rather than using the inheritance property to aid us in proving counting lemmas, we first show how one may prove the counting lemma and then use it to prove a strong form of inheritance in jumbled graphs. For example, we have the following theorem.

Proposition 2.1.13. *For any $\alpha > 0$, $\xi > 0$ and $\epsilon' > 0$, there exists $c > 0$ and $\epsilon > 0$ of size at least polynomial in α, ξ, ϵ' such that the following holds.*

Let $p \in (0, 1]$ and $q_{XY}, q_{XZ}, q_{YZ} \in [\alpha p, p]$. Let Γ be a tripartite graph with vertex sets X, Y and Z and G be a subgraph of Γ. Suppose that

- $(X,Y)_\Gamma$ *is* $(p, cp^4\sqrt{|X|\,|Y|})$*-jumbled and* $(X,Y)_G$ *satisfies* $\mathrm{DISC}(q_{XY}, p, \epsilon)$*; and*
- $(X,Z)_\Gamma$ *is* $(p, cp^2\sqrt{|X|\,|Z|})$*-jumbled and* $(X,Z)_G$ *satisfies* $\mathrm{DISC}(q_{XZ}, p, \epsilon)$*; and*
- $(Y,Z)_\Gamma$ *is* $(p, cp^3\sqrt{|Y|\,|Z|})$*-jumbled and* $(Y,Z)_G$ *satisfies* $\mathrm{DISC}(q_{YZ}, p, \epsilon)$.

Then at least $(1-\xi)\,|Z|$ *vertices* $z \in Z$ *have the property that* $|N_X(z)| \geq (1-\xi)q_{XZ}\,|X|$, $|N_Y(z)| \geq (1-\xi)q_{YZ}\,|Y|$*, and* $(N_X(z), N_Y(z))_G$ *satisfies* $\mathrm{DISC}(q_{XY}, p, \epsilon')$.

The question now arises as to why one would prove that the inheritance property holds if we already know its intended consequence. Surprisingly, there is another counting lemma, giving only a lower bound on $G(H)$, which is sufficient to establish the various extremal results but typically requires a much weaker jumbledness assumption. The proof of this statement relies on the inheritance property in a critical way. The notations $G(H)$ and $q(H)$ were defined in Theorem 2.1.12.

Theorem 2.1.14. *For every fixed graph* H *on vertex set* $\{1, 2, \ldots, m\}$ *and every* $\alpha, \theta > 0$*, there exist constants* $c > 0$ *and* $\epsilon > 0$ *such that the following holds.*

Let $p > 0$ *and let* Γ *be a graph with vertex subsets* $X_1, \ldots, X_m$ *and suppose that the bipartite graph* $(X_i, X_j)_\Gamma$ *is* $(p, cp^{d_2(H)+3}\sqrt{|X_i|\,|X_j|})$*-jumbled for every* $i < j$ *with* $ij \in E(H)$*. Let* G *be a subgraph of* Γ*, with the vertex* i *of* H *assigned to the vertex subset* X_i *of* G*. For each edge* ij *of* H*, assume that* $(X_i, X_j)_G$ *satisfies* $\mathrm{DISC}(q_{ij}, p, \epsilon)$*, where* $\alpha p \leq q_{ij} \leq p$*. Then*

$$G(H) \geq (1-\theta)q(H).$$

We refer to Theorem 2.1.14 as a one-sided counting lemma, as we get a lower bound for $G(H)$ but no upper bound. However, in order to prove the theorems of Section 2.1.2, we only need a lower bound. The proof of Theorem 2.1.14 is a sparse version of a classical embedding strategy in regular graphs (see, for example, [24, 49, 66]). Note that, as in the theorems of Section 2.1.2, the exponent $d_2(H) + 3$ can be improved for certain graphs H. We will say more about this later. Moreover, one cannot hope to do better than $\beta = O(p^{(d(H)+2)/4}n)$, so that the condition on β is sharp up to a multiplicative constant in the exponent of p. We suspect that the exponent may even be sharp up to an additive constant.

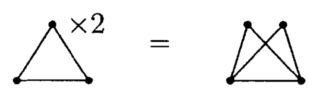
2.2 Counting strategy

In this section, we give a general overview of our approach to counting. There are two types of counting results: two-sided counting and one-sided counting. Two-sided counting refers to results of the form $|G(H) - q(H)| \leq \theta p^{e(H)}$ while one-sided counting refers to results of the form $G(H) \geq q(H) - \theta p^{e(H)}$. One-sided counting is always implied by two-sided counting, although sometimes we are able to obtain one-sided counting results under weaker hypotheses.

2.2.1 Two-sided counting

There are two main ingredients to the proof: *doubling* and *densification*. These two procedures reduce the problem of counting embeddings of H to the same problem for some other graphs H'.

If $a \in V(H)$, *the graph H with a doubled*, denoted $H_{a\times 2}$, is the graph created from $V(H)$ by adding a new vertex a' whose neighbors are precisely the neighbors of a. In the assignment of vertices of $H_{a\times 2}$ to vertex subsets of Γ, the new vertex a' is assigned to the same vertex subset of Γ. For example, the following figure shows a triangle with a vertex doubled.



A typical reduction using doubling is summarized in Figure 2-1. Each graph represents the claim that the number of embeddings of the graph drawn, where the straight edges must land in G and the wavy edges must land in Γ, is approximately what one would expect from multiplying together the appropriate edge densities between the vertex subsets of G and Γ.

The top arrow in Figure 2-1 is the doubling step. This allows us to reduce the problem of counting H to that of counting a number of other graphs, each of which may have some edges which embed into G and some which embed into Γ. For example, if we let H_{-a} be the graph that we get by omitting every edge which is connected to a particular vertex a, we are interested in the number of copies of H_{-a} in both G and

Γ. We are also interested in the original graph H, but now on the understanding that the edges incident with a embed into G while those that do not touch a embed into Γ. Finally, we are interested in the graph $H_{a\times 2}$ formed by doubling the vertex a, but again the edges which do not touch a or its copy a' only have to embed into Γ. This reduction, which is justified by an application of the Cauchy-Schwarz inequality, will be detailed in Section 2.4.1.

The bottom two arrows in Figure 2-1 are representative of another reduction, where we can reduce the problem of counting a particular graph, with edges that map to both G and Γ, into one where we only care about the edges that embed into G. We can make such a reduction because counting embeddings into Γ is much easier due to its jumbledness. We will discuss this reduction, amongst other properties of jumbled graphs Γ, in Section 2.3.

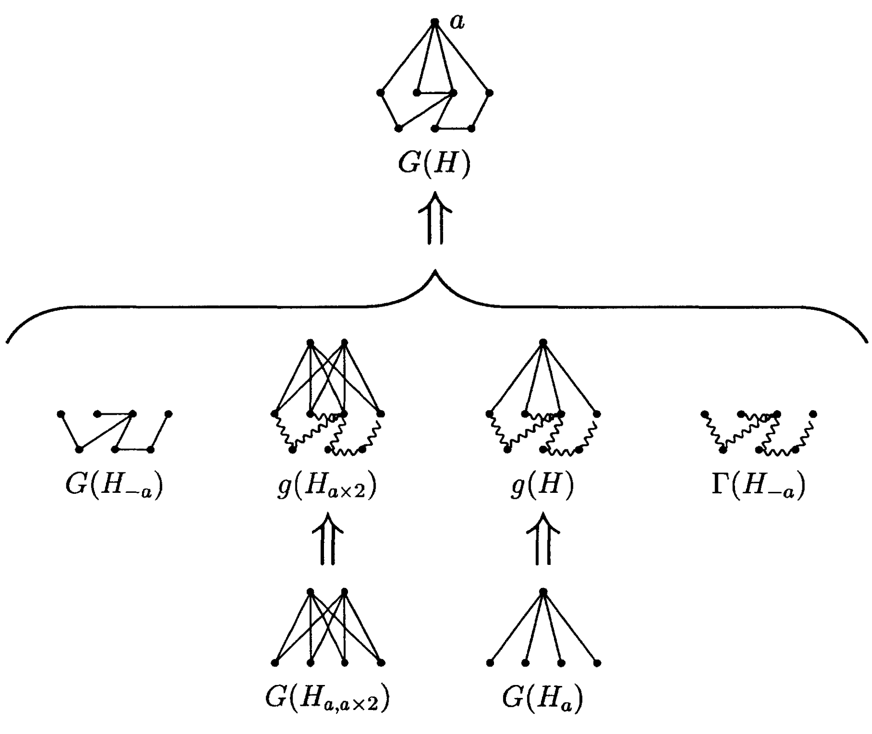


Figure 2-1: The doubling reduction. Each graph represents some counting lemma. The straight edges must embed into G while wavy edges must embed into the jumbled graph Γ. The labels have not yet been defined, but they appear in Lemma 2.4.3.

For triangles, a similar reduction is shown in Figure 2-2. In the end, we have changed the task of counting triangles to the task of counting the number of cycles of length 4. It would be natural now to apply doublng to the 4-cycle but, unfortunately, this process is circular. Instead, we introduce an alternative reduction process which we refer to as densification.

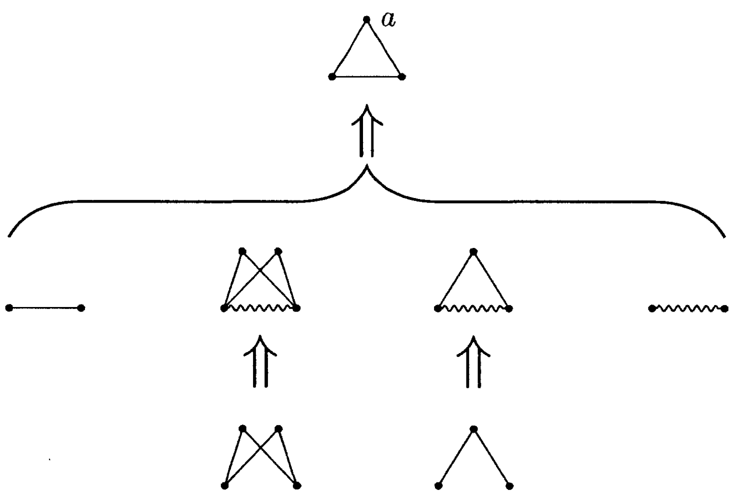
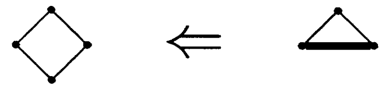


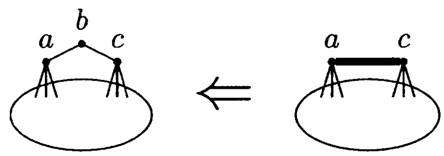
Figure 2-2: The doubling reduction for counting triangles.

In the above reduction from triangles to 4-cycles, two of the vertices of the 4-cycle are embedded into the same part X_i of G. We actually consider the more general setting where the vertices of the 4-cycle lie in different parts, X_1, X_2, X_3, X_4, of G.

Assume without loss of generality that there is no edge between X_1 and X_3 in G. Let us add a weighted graph between X_1 and X_3, where the weight on the edge x_1x_3 is proportional to the number of paths $x_1x_4x_3$ for $x_4 \in X_4$. Since $(X_1, X_4)_G$ and $(X_3, X_4)_G$ satisfy discrepancy, the number of paths will be on the order of $q_{14}q_{34}\,|X_4|$ for most pairs (x_1, x_3). After discarding a negligible set of pairs (x_1, x_3) that give too many paths, and then appropriately rescaling the weights of the other edges x_1x_3, we create a weighted bipartite graph between X_1 and X_3 that behaves like a *dense* weighted graph satisfying discrepancy. Furthermore, counting 4-cycles in X_1, X_2, X_3, X_4 is equivalent to counting triangles in X_1, X_2, X_3 due to the choice of weights. We call this process densification. It is illustrated below. In the figure, a thick edge signifies that the bipartite graph that it embeds into is dense.



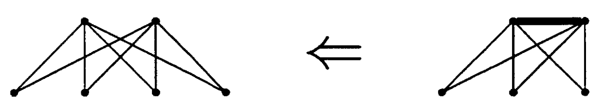
More generally, if b is a vertex of H of degree 2, with neighbors $\{a, c\}$, such that a and c are not adjacent, then densification allows us to transform H by removing the edges ab and bc and adding a dense edge ac, as illustrated below. For more on this process, we refer the reader to Section 2.4.2.



We needed to count 4-cycles in order to count triangles, so it seems at first as if our reduction from 4-cycles to triangles is circular. However, instead of counting triangles in a sparse graph, we now have a dense bipartite graph between one of the pairs of vertex subsets. Since it is easier to count in dense graphs than in sparse graphs, we have made progress. The next step is to do doubling again. This is shown in Figure 2-3. The bottommost arrow is another application of densification.

We have therefore reduced the problem of counting triangles in a sparse graph to that of counting triangles in a dense weighted graph, which we already know how to do. This completes the counting lemma for triangles.

In Figure 2-1, doubling reduces counting in a general H to counting H with one vertex deleted (which we handle by induction) as well as graphs of the form $K_{1,t}$ and $K_{2,t}$. Trees like $K_{1,t}$ are not too hard to count. It therefore remains to count $K_{2,t}$. As with counting C_4 (the case $t = 2$), we first perform a densification.



The graph on the right can be counted using doubling and induction, as shown in Figure 2-4. Note that the C_4 count is required as an input to this step. This then completes the proof of the counting lemma.

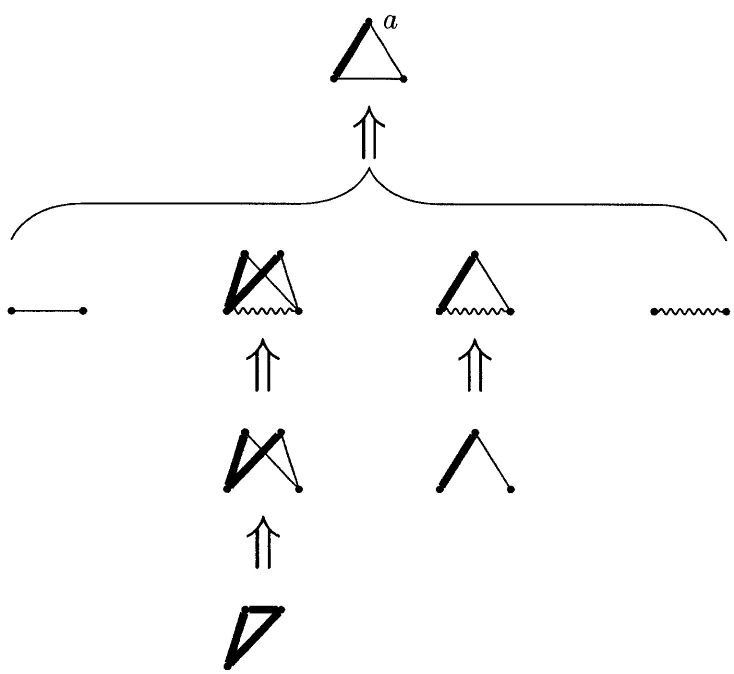


Figure 2-3: The doubling reduction for triangles with one dense edge.

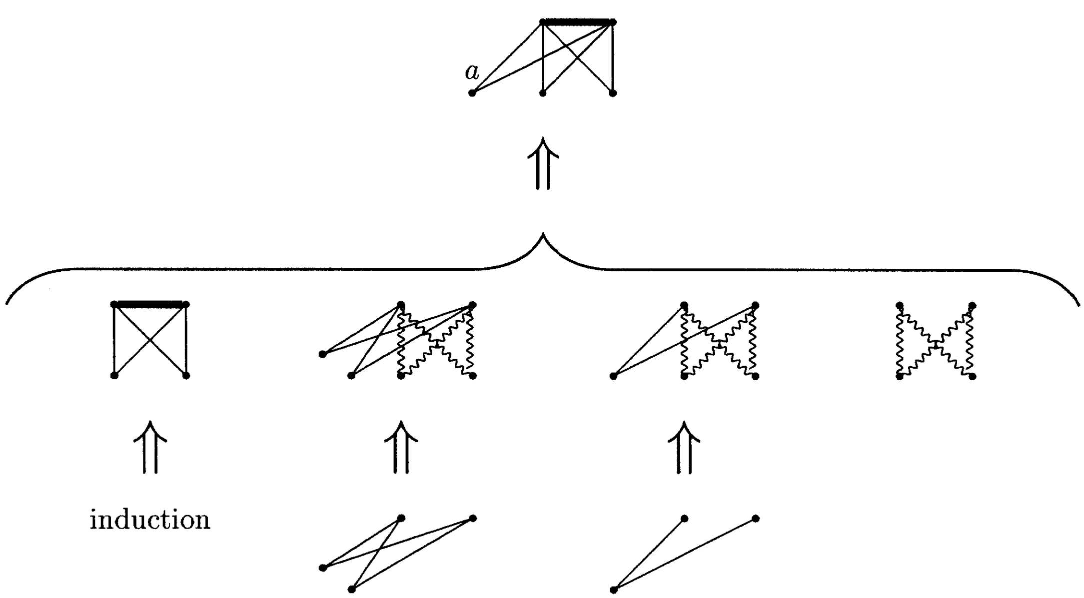


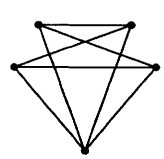
Figure 2-4: The doubling reduction for counting $K_{2,t}$.

2.2.2 One-sided counting

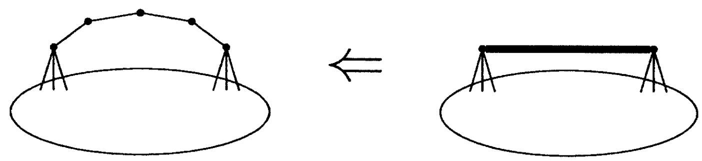
For one-sided counting, we embed the vertices of H into those of G one at a time. By making a choice for where a vertex a of H lands in G, we shrink the set of possible targets for each neighbor of a. These target sets shrink by a factor roughly corresponding to the edge densities of G, as most vertices of G have close to the expected number of neighbors due to discrepancy. This allows us to obtain a lower bound on the number of embeddings of H into G.

The above argument is missing one important ingredient. When we shrink the set of possible targets of vertices in H, we do not know if G restricted to these smaller vertex subsets still satisfies the discrepancy condition, which is needed for embedding later vertices. When G is dense, this is not an issue, since the restricted vertex subsets have size at least a constant factor of the original vertex subsets, and thus discrepancy is inherited. When G is sparse, the restricted vertex subsets can become much smaller than the original vertex subsets, so discrepancy is not automatically inherited.

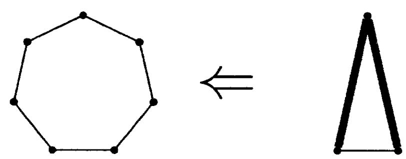
To address this issue, we observe that discrepancy between two vertex sets follows from some variant of the $K_{2,2}$ count (and the counting lemma shows that they are in fact equivalent). By our counting lemma, we also know that the graph below has roughly the expected count. This in turn implies that discrepancy is inherited in the neighborhoods of G since, roughly speaking, it implies that almost every vertex has roughly the expected number of 4-cycles in its neighborhood. The one-sided counting approach sketched above then carries through. For further details on inheritance of discrepancy, see Section 2.5. The proof of the one-sided counting lemma may be found in Section 2.6.



We also prove a one-sided counting lemma for large cycles using much weaker jumbledness hypotheses. The idea is to extend densification to more than two edges at a time. We will show how to transform a multiply subdivided edge into a single dense edge, as illustrated below.



Starting with a long cycle, we can perform two such densifications, as shown below. The resulting triangle is easy to count, since a typical embedding of the top vertex gives a linear-sized neighborhood. The full details may be found in Section 2.7.



2.3 Counting in Γ

In this section, we develop some tools for counting in Γ. Here is the setup for this section.

Setup 2.3.1. Let Γ be a graph with vertex subsets $X_1, \dots, X_m$. We have quantities $p, c \in (0,1]$ and $k \geq 1$. Let H be a graph with vertex set $\{1, \dots, m\}$, with vertex a assigned to X_a. For every edge ab in H, assume that one of the following two holds:

- $(X_a, X_b)_\Gamma$ is $(p, cp^k\sqrt{|X_a|\,|X_b|})$-jumbled, in which case we set $p_{ab} = p$ and say that ab is a *sparse edge*, or
- $(X_a, X_b)_\Gamma$ is a complete bipartite graph, in which case we set $p_{ab} = 1$ and say that ab is a *dense edge.*

Let H^{sp} denote the subgraph of H consisting of sparse edges.

2.3.1 Example: counting triangles in Γ

We start by showing, as an example, how to prove the counting lemma in Γ for triangles. Most of the ideas found in the rest of this section can already be found in this special case.

Proposition 2.3.2. *Assume Setup 2.3.1. Let H be a triangle with vertices $\{1,2,3\}$. Assume that $k \geq 2$. Then $|\Gamma(H) - p^3| \leq 5cp^3$.*

Proof. In the following integrals, we assume that x, y and z vary uniformly over X_1, X_2 and X_3, respectively. We have the telescoping sum

$$\begin{aligned}\Gamma(H) - p^3 &= \int_{x,y,z} (\Gamma(x,y) - p)\Gamma(x,z)\Gamma(y,z)\ dxdydz \\ &\quad + \int_{x,y,z} p(\Gamma(x,z) - p)\Gamma(y,z)\ dxdydz + \int_{x,y,z} p^2(\Gamma(y,z) - p)\ dxdydz. \quad (2.4)\end{aligned}$$

The third integral on the right-hand side of (2.4) is bounded in absolute value by cp^3 by the jumbledness of Γ. In particular, this implies that $\int_{y,z} \Gamma(y,z)\ dydz \leq (1+c)p$. Similarly we have $\int_{x,z} \Gamma(x,z)\ dxdz \leq (1+c)p$. Using the jumbledness condition (2.3) followed by the Cauchy-Schwarz inequality, the second integral above is bounded in absolute value by

$$\int_y cp^3 \sqrt{\int_z \Gamma(y,z)\ dz}\ dy \leq cp^3 \sqrt{\int_{y,z} \Gamma(y,z)\ dydz} \leq cp^3 \sqrt{(1+c)p}.$$

Finally, the first integral on the right-hand side of (2.4) is bounded in absolute value by, using (2.3) and the Cauchy-Schwarz inequality,

$$\begin{aligned}&\int_z cp^2 \sqrt{\int_x \Gamma(x,z)\ dx} \sqrt{\int_y \Gamma(y,z)\ dy}\ dz \\ &\qquad \leq cp^2 \sqrt{\int_{x,z} \Gamma(x,z)\ dxdz} \sqrt{\int_{y,z} \Gamma(y,z)\ dydz} \leq c(1+c)p^3. \quad (2.5)\end{aligned}$$

Therefore, (2.4) is bounded in absolute value by $5cp^3$. ■

Remark. (1) In the more general proof, the step corresponding to (2.5) will be slightly different but is similar in its application of the Cauchy-Schwarz inequality.
(2) The proof shows that we do not need the full strength of the jumbledness everywhere—we only need $(p, cp^{3/2}\sqrt{|X|\,|Z|})$-jumbledness for $(X,Z)_\Gamma$ and $(p, cp\sqrt{|Y|\,|Z|})$-jumbledness

for $(Y, Z)_\Gamma$. In Section 2.6, it will be useful to have a counting lemma with such non-balanced jumbledness assumptions in order to optimize our result. To keep things simple and clear, we will assume balanced jumbledness conditions here and remark later on the changes needed when we wish to have optimal non-balanced ones.

2.3.2 Notation

In the proof of the counting lemmas we frequently encounter expressions such as $G(x_1, x_2)G(x_1, x_3)G(x_2, x_3)$ and their integrals. We introduce some compact notation for such products and integrals. Note that if we are counting copies of H, we will usually assign each vertex a of H to some vertex subset X_a and we will only be interested in counting those embeddings where each vertex of H is mapped into the vertex subset assigned to it. If $U \subseteq V(H)$, a map $U \to V(G)$ or $U \to V(\Gamma)$ is called *compatible* if each vertex of U gets mapped into the vertex set assigned to it. We can usually assume without loss of generality that the vertex subsets X_a are disjoint for different vertices of H, as we can always create a new multipartite graph with disjoint vertex subsets X_a with the same H-embedding counts as the original graph.

If f is a symmetric function on pairs of vertices of G (actually we only care about its values on $X_a \times X_b$ for $ab \in E(H)$) and $\mathbf{x}\colon V(H) \to V(G)$ is any compatible map (we write $\mathbf{x}(a) = x_a$), then we define

$$f(H \mid \mathbf{x}) := \prod_{ab \in E(H)} f(x_a, x_b).$$

By taking the expectation as $\mathbf{x}$ varies uniformly over all compatible maps $V(H) \to V(G)$, we can define the value of a function on a graph.

$$f(H) := \mathbb{E}_{\mathbf{x}} \left[f(H \mid \mathbf{x}) \right] = \int_{\mathbf{x}} f(H \mid \mathbf{x}) \; d\mathbf{x}.$$

We shall always assume that the measure $d\mathbf{x}$ is the uniform probability measure on compatible maps.

Following the above notation, the quantities that arise most frequently in our work are

$$\Gamma(H), \quad p(H), \quad G(H), \quad q(H).$$

For unweighted graphs, we use G and Γ to denote the characteristic function of the edge set of the graph, so that $G(H)$ is the probability that a uniformly random compatible map $V(H) \to V(G)$ is a graph homomorphism from H to G. For weighted graphs, the value on the edges are the edge weights. For $p(H)$ and $q(H)$, we view $p(\cdot,\cdot)$ and $q(\cdot,\cdot)$ as constant functions on each $X_a \times X_b$, taking values p_{ab} and q_{ab}, so that $p(H) = \prod_{ab\in E(H)} p_{ab}$ and $q(H) = \prod_{ab\in E(H)} q_{ab}$. Since p_{ab} only takes value p or 1 depending on whether ab is a sparse edge or a dense edge we have $p(H) = p^{e(H^{\mathrm{sp}})}$, where recall that H^{sp} is the subgraph of H consisting of the sparse edges. For counting lemmas, we are interested in comparing $G(H)$ with $q(H)$.

It will be useful to have some notation for the conditional sum of a function f given that some vertices have been fixed. If $U \subseteq V(H)$ and $\mathbf{y}\colon U \to V(G)$ is any compatible map, then

$$f(H \mid \mathbf{y}) := \mathbb{E}_{\mathbf{x}}\left[f(H \mid \mathbf{x}) \mid \mathbf{x}|_U = \mathbf{y}\right] = \int_{\mathbf{z}} f(H \mid \mathbf{y}, \mathbf{z})\ d\mathbf{z},$$

where, in the integral, $\mathbf{z}$ varies uniformly over all compatible maps $V(H) \setminus U \to V(G)$, and the notation $\mathbf{y}, \mathbf{z}$ denotes the compatible map $V(H) \to V(G)$ built from combining $\mathbf{y}$ and $\mathbf{z}$. Note that when $U = \emptyset$, $f(H \mid \mathbf{y}) = f(H)$. When $U = V(H)$, the two definitions of $f(H \mid \mathbf{y})$ agree. When $U = \{a_1, \dots, a_t\}$, we sometimes write $\mathbf{y}$ as $a_1 \to y_1, \dots, a_t \to y_t$, so we can write $f(H \mid a_1 \to y_1, \dots, a_t \to y_t)$.

Since we work with approximations frequently, it will be convenient if we introduce some shorthand. If A, B, P are three quantities, we write

$$A \underset{c,\epsilon}{\overset{P}{\approx}} B$$

to mean that for every $\theta > 0$, we can find $c, \epsilon > 0$ of size at least polynomial in θ (i.e., $c, \epsilon = \Omega(\theta^r)$ as $\theta \to 0$ for some $r > 0$) so that $|A - B| \le \theta P$. Sometimes one of c or ϵ is omitted from the $\approx$ notation if θ does not depend on the parameter. Note that the dependencies do not depend on the parameters p and q, but may depend on the graphs to be embedded, e.g., H. For instance, the main counting lemma (Theorem 2.1.12) can be phrased as

$$G(H) \overset{p^{e(H)}}{\underset{c,\epsilon}{\approx}} q(H).$$

2.3.3 Counting graphs in Γ

We begin by giving a counting lemma in Γ, which is significantly easier than counting in G. We remark that a similar counting lemma for Γ an (n, d, λ) regular graph was proven by Alon (see [86, Thm. 4.10]).

Proposition 2.3.3. *Assume Setup 2.3.1. If $k \geq \frac{d(L(H^{\mathrm{sp}}))+2}{2}$, then*

$$|\Gamma(H) - p(H)| \leq \left((1+c)^{e(H^{\mathrm{sp}})} - 1\right) p(H).$$

The exact coefficient of $p(H)$ in the bound is not important. Any bound of the form $c^{\Omega(1)}p(H)$ suffices.

Dense edges play no role, so it suffices to consider the case when all edges of H are sparse. We prove Proposition 2.3.3 by iteratively applying the following inequality.

Lemma 2.3.4. *Let H be a graph with vertex set $\{1, \ldots, m\}$. Let Γ be a graph with vertex subsets $X_1, \ldots, X_m$. Let $ab \in E(H)$. Let H_{-ab} denote H with the edge ab removed. Let $H_{-a,-b}$ denote H with all edges incident to a or b removed. Assume that $\Gamma(X_a, X_b)$ is $(p, \gamma\sqrt{|X_a|\,|X_b|})$-jumbled. Let $f \colon V(\Gamma) \times V(\Gamma) \to [0,1]$ be any symmetric function. Then*

$$\left| \int_{\substack{x \in X_a \\ y \in X_b}} (\Gamma(x,y) - p) f\left(H_{-ab} \mid a \to x, b \to y\right)\, dxdy \right| \leq \gamma \sqrt{f(H_{-ab}) f(H_{-a,-b})}.$$

Proof. Let $H_{a,-ab}$ denote the edges of H_{-ab} incident to a, and let $H_{b,-ab}$ be the edges of H_{-ab} incident to b. Since $H_{-ab} = H_{-a,-b} \uplus H_{a,-ab} \uplus H_{b,-ab}$, as a disjoint union of edges, for any compatible map $\mathbf{x}\colon V(H) \to V(\Gamma)$ we have

$$f\left(H_{-ab} \mid \mathbf{x}\right) = f\left(H_{-a,-b} \mid \mathbf{x}\right) f\left(H_{a,-ab} \mid \mathbf{x}\right) f\left(H_{b,-ab} \mid \mathbf{x}\right). \tag{2.6}$$

In the following calculation, $\mathbf{z}$ varies uniformly over compatible maps $V(H)\setminus\{a,b\} \to V(\Gamma)$, x varies uniformly over X_a, and y varies uniformly over X_b. The first equality follows from (2.6) while the three inequalities follow, in order, from the triangle inequality, the jumbledness condition, and the Cauchy-Schwarz inequality.

$$\begin{aligned}
&\left|\int_{x,y} (\Gamma(x,y)-p) f\left(H_{-ab} \mid a \to x, b \to y\right)\, dxdy\right| \\
&= \left|\int_{\mathbf{z}} f\left(H_{-a,-b} \mid \mathbf{z}\right) \int_{x,y} (\Gamma(x,y)-p) f\left(H_{a,-ab} \mid a \to x, \mathbf{z}\right) f\left(H_{b,-ab} \mid b \to y, \mathbf{z}\right)\, dxdyd\mathbf{z}\right| \\
&\le \int_{\mathbf{z}} f\left(H_{-a,-b} \mid \mathbf{z}\right) \left|\int_{x,y} (\Gamma(x,y)-p) f\left(H_{a,-ab} \mid a \to x, \mathbf{z}\right) f\left(H_{b,-ab} \mid b \to y, \mathbf{z}\right)\, dxdy\right| d\mathbf{z} \\
&\le \int_{\mathbf{z}} f\left(H_{-a,-b} \mid \mathbf{z}\right) \gamma \sqrt{\int_x f\left(H_{a,-ab} \mid a \to x, \mathbf{z}\right)\, dx} \sqrt{\int_y f\left(H_{b,-ab} \mid b \to y, \mathbf{z}\right)\, dy} d\mathbf{z} \\
&= \gamma \int_{\mathbf{z}} f\left(H_{-a,-b} \mid \mathbf{z}\right) \sqrt{f\left(H_{a,-ab} \mid \mathbf{z}\right)} \sqrt{f\left(H_{b,-ab} \mid \mathbf{z}\right)} d\mathbf{z} \\
&\le \gamma \sqrt{\int_{\mathbf{z}} f\left(H_{-a,-b} \mid \mathbf{z}\right)\, d\mathbf{z}} \sqrt{\int_{\mathbf{z}} f\left(H_{-a,-b} \mid \mathbf{z}\right) f\left(H_{a,-ab} \mid \mathbf{z}\right) f\left(H_{b,-ab} \mid \mathbf{z}\right) d\mathbf{z}} \\
&= \gamma \sqrt{f(H_{-a,-b}) f(H_{-ab})}.
\end{aligned}$$

■

Proof of Proposition 2.3.3. As remarked after the statement of the proposition, it suffices to prove the result in the case when all edges of H are sparse. We induct on the number of edges of H. If H has no edges, then $\Gamma(H) = p(H) = 1$. So assume that H has at least one edge. Since $k \ge \frac{1}{2}(d(L(H)) + 2)$, we can find an edge ab of H such that $\deg_H(a) + \deg_H(b) \le d(L(H)) + 2 \le 2k$. Let H_{-ab} and $H_{-a,-b}$ be as in Lemma 2.3.4. Since $L(H)$ is $(2k-2)$-degenerate, the line graph of any subgraph of

H is also $(2k-2)$-degenerate. By the induction hypothesis, we have

$$|\Gamma(H_{-ab}) - p(H_{-ab})| \le ((1+c)^{e(H)-1} - 1)p(H_{-ab}) \qquad \text{and}$$
$$|\Gamma(H_{-a,-b}) - p(H_{-a,-b})| \le ((1+c)^{e(H)-1} - 1)p(H_{-a,-b}).$$

The following identity allows us to "split off" the edge ab. It may be helpful to compare this calculation with the telescoping identity in (2.4) (there we split off all the edges of the triangle, one edge at a time, whereas here we only split off one edge). We have

$$\begin{aligned}\Gamma(H) - p(H) &= \int_{\substack{x\in X_a\\ y\in X_b}} \Gamma(x,y)\Gamma\left(H_{-ab} \mid a\to x, b\to y\right)\, dxdy - p\cdot p(H_{-ab})\\ &= \int_{\substack{x\in X_a\\ y\in X_b}} (\Gamma(x,y)-p)\Gamma\left(H_{-ab} \mid a\to x, b\to y\right)\, dxdy\\ &\qquad + p\cdot\left(\Gamma(H_{-ab}) - p(H_{-ab})\right).\end{aligned}$$

The second term on the right is bounded in absolute value by $((1+c)^{e(H)-1}-1)p(H)$. For the first term, by Lemma 2.3.4 and the induction hypothesis, we have

$$\begin{aligned}&\left|\int_{x,y} (\Gamma(x,y)-p)\Gamma\left(H_{-ab} \mid a\to x, b\to y\right)\, dxdy\right|\\ &\le cp^k\sqrt{\Gamma(H_{-ab})\Gamma(H_{-a,-b})}\\ &\le cp^k(1+c)^{e(H)-1}\sqrt{p(H_{-ab})p(H_{-a,-b})}\\ &\le c(1+c)^{e(H)-1}p(H).\end{aligned} \tag{2.7}$$

The last inequality is where we used $2k \ge \deg_H(a) + \deg_H(b)$. Combining the two estimates gives the desired result. ■

2.3.4 Counting partial embeddings into Γ

As outlined in Section 2.2, we need to count embeddings of H where some edges are embedded into G (the straight edges in the figures) and some edges are embedded into Γ (the wavy edges). We prove counting estimates for these embeddings here. The main result of this section is summarized in the figure after Lemma 2.3.6. The proofs are almost identical to that of Proposition 2.3.3. We just need to be a little more careful with the exponents on the jumbledness parameter.

First we consider the case where exactly one edge needs to be embedded into Γ and the other edges are embedded into some subgraph of Γ. To state the result requires a little notation. Suppose that $H = H' \cup H''$ is an edge disjoint partition of the graph H into two subgraphs H' and H''. We define $d(L(H', H''))$ to be the smallest d such there is an ordering of the edges of H with the edges of H' occurring before the edges of H'' such that every edge e has at most d neighbors, that is, edges containing either of the endpoints of e, which appear earlier in the ordering.

Lemma 2.3.5. *Assume Setup 2.3.1. Let $ab \in E(H)$ and H_{-ab} be the graph H with edge ab removed. Assume $k \geq \frac{d(L(H^{\mathrm{sp}}_{-ab}, ab^{\mathrm{sp}}))+2}{2}$. Let G be any weighted subgraph of Γ (i.e., $0 \leq G \leq \Gamma$ as functions). Let g denote the function that agrees with Γ on $X_a \times X_b$ and with G everywhere else. Then*

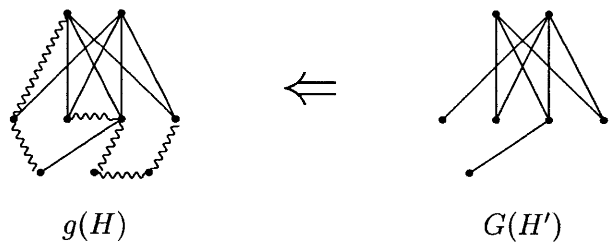
$$|g(H) - p_{ab}G(H_{-ab})| \leq c(1+c)^{e(H^{\mathrm{sp}})-1}p(H).$$

The lemma follows from essentially the same calculation as (2.7), except that we take ab as our first edge to remove (this is why there is a stronger requirement on k) and then use $G \leq \Gamma$.

Iterating the lemma, we obtain the following result where multiple edges need to be embedded into Γ. It can be proved by iterating Lemma 2.3.5 or mimicking the proof of Proposition 2.3.3. We include a figure illustrating the lemma.

Lemma 2.3.6. *Assume Setup 2.3.1. Let H' be a subgraph of H. Assume $k \geq \frac{d(L(H'^{\mathrm{sp}}, (H \setminus H')^{\mathrm{sp}}))+2}{2}$. Let G be a weighted subgraph of Γ. Let g be a function that agrees with Γ on $X_a \times X_b$ when $ab \in E(H \setminus H')$ and with G otherwise. Then*

$$|g(H) - p(H \setminus H')G(H')| \leq \left((1+c)^{e(H^{\mathrm{sp}})} - 1\right) p(H).$$



2.3.5 Exceptional sets

This section contains a couple of lemmas about Γ that we will need later on. The reader may choose to skip this section until the results are needed.

We begin with a standard estimate for the number of vertices in a jumbled graph whose degrees deviate from the expected value. The proof follows immediately from the definition of jumbledness.

Lemma 2.3.7. *Let Γ be a $(p, \gamma\sqrt{|X|\,|Y|})$-jumbled graph between vertex subsets X and Y. Let $v\colon Y \to [0,1]$ and let $\xi > 0$. If*

$$U \subseteq \left\{ x \in X \;\middle|\; \int_{y \in Y} \Gamma(x,y)v(y)\ dy \geq (1+\xi)p\mathbb{E}v \right\}$$

or

$$U \subseteq \left\{ x \in X \;\middle|\; \int_{y \in Y} \Gamma(x,y)v(y)\ dy \leq (1-\xi)p\mathbb{E}v \right\},$$

then

$$\frac{|U|}{|X|} \leq \frac{\gamma^2}{\xi^2 p^2 \mathbb{E}v}.$$

The next lemma says that restrictions of the count $\Gamma(H)$ to small sets of vertices or pairs of vertices yield small counts. This will be used in Section 2.4.2 to bound the contributions from exceptional sets.

Lemma 2.3.8. *Assume Setup 2.3.1 with $k \geq \frac{d(L(H^{\mathrm{sp}}))+2}{2}$. Let $\mathbf{x}\colon V(H) \to V(\Gamma)$ vary uniformly over compatible maps. Let $u\colon V(\Gamma) \to [0,1]$ be any function and write $u(\mathbf{x}) = \prod_{a \in V(H)} u(x_a)$. Let E' be a weighted graph with the same vertices as Γ whose edge set is supported on $X_a \times X_b$ for $ab \notin H^{\mathrm{sp}}$. Let H' be any graph with the same* *vertices as H. Then*

$$\int_{\mathbf{x}} \Gamma(H \mid \mathbf{x})\, u(\mathbf{x}) E'(H' \mid \mathbf{x})\ d\mathbf{x} \leq \left((1+c)^{e(H^{\mathrm{sp}})} - 1 + \int_{\mathbf{x}} u(\mathbf{x}) E'(H' \mid \mathbf{x})\ d\mathbf{x}\right) p(H).$$

Lemma 2.3.8 follows by showing that

$$\left|\int_{\mathbf{x}} \left(\Gamma(H \mid \mathbf{x}) - p(H)\right) u(\mathbf{x}) E'(H' \mid \mathbf{x})\ d\mathbf{x}\right| \leq \left((1+c)^{e(H^{\mathrm{sp}})} - 1\right) p(H).$$

The proof is similar to that of Proposition 2.3.3. In the step analogous to (2.7), after applying the jumbledness condition as our first inequality, we bound u and E' by 1 and then continue exactly the same way.

2.4 Counting in G

In this section we develop the counting lemma for subgraphs G of Γ, as outlined in Section 2.2. The two key ingredients are doubling and densification, which are discussed in Sections 2.4.1 and 2.4.2, respectively. Here is the common setup for this section.

Setup 2.4.1. Assume Setup 2.3.1. We have some quantity $\epsilon > 0$. Let G be a weighted subgraph of Γ. For every edge $ab \in E(H)$, assume that $(X_a, X_b)_G$ satisfies $\mathrm{DISC}(q_{ab}, p_{ab}, \epsilon)$, where $0 \leq q_{ab} \leq p_{ab}$.

Unlike in Section 2.3, we do not make an effort to keep track of the unimportant coefficients of $p(H)$ in the error bounds, as it would be cumbersome to do so. Instead, we use the $\approx$ notation introduced in Section 2.3.2.

The goal of this section is to prove the following counting lemma. This is slightly more general than Theorem 2.1.12 in that it allows H to have both sparse and dense edges.

Theorem 2.4.2. *Assume Setup 2.4.1 with* $k \geq \min\left\{\frac{\Delta(L(H^{\mathrm{sp}}))+4}{2}, \frac{d(L(H^{\mathrm{sp}}))+6}{2}\right\}$. *Then*

$$G(H) \overset{p(H)}{\underset{c,\epsilon}{\approx}} q(H).$$

The requirement on k stated in Theorem 2.4.2 is not necessarily best possible. The proof of the counting lemma will be by induction on the vertices of H, removing one vertex at a time. A better bound on k can sometimes be obtained by tracking the requirements on k at each step of the procedure, as explained in a tutorial in Section 2.4.5.

2.4.1 Doubling

Doubling is a technique used to reduce the problem of counting embeddings of H in G to the problem of counting embeddings of H with one vertex deleted.

If $a \in V(H)$, $H_{a\times 2}$ is the graph H with vertex a doubled. In the assignment of vertices of $H_{a\times 2}$ to vertex subsets of Γ, the new vertex a' is assigned to the same vertex subset as a. Let H_a be the subgraph of H consisting of edges with a as an endpoint, and let $H_{a,a\times 2}$ be H_a with a doubled. Let H_{-a} be the subgraph of H consisting of edges not having a as an endpoint. We refer to Figure 2-1 for an illustration.

Lemma 2.4.3. *Assume Setup 2.4.1. Fix a vertex $a \in V(G)$. Let g be a function that agrees with G on $X_i \times X_j$ whenever $a \in \{i, j\}$ and with Γ on $X_i \times X_j$ whenever $a \notin \{i, j\}$. Then*

$$\begin{aligned} |G(H) - q(H)| \le q(H_a)\,|G(H_{-a}) - q(H_{-a})| \\ + G(H_{-a})^{1/2}\left(g(H_{a\times 2}) - 2q(H_a)g(H) + q(H_a)^2\Gamma(H_{-a})\right)^{1/2}. \quad (2.8) \end{aligned}$$

Proof. Let $\mathbf{y}$ vary uniformly over compatible maps $V(H)\setminus\{a\} \to V(G)$ where $\mathbf{y}(b) \in X_b$ for each $b \in V(H)\setminus\{a\}$. We have

$$G(H) - q(H) = q(H_a)(G(H_{-a}) - q(H_{-a})) + \int_{\mathbf{y}} \left(G\left(H_a \mid \mathbf{y}\right) - q(H_a)\right) G\left(H_{-a} \mid \mathbf{y}\right)\, d\mathbf{y}.$$

It remains to bound the integral, which we can do using the Cauchy-Schwarz inequality.

$$\begin{aligned}
&\left(\int_{\mathbf{y}} (G(H_a \mid \mathbf{y}) - q(H_a))\, G(H_{-a} \mid \mathbf{y})\; d\mathbf{y}\right)^2 \\
&\le \left(\int_{\mathbf{y}} G(H_{-a} \mid \mathbf{y})\; d\mathbf{y}\right)\left(\int_{\mathbf{y}} (G(H_a \mid \mathbf{y}) - q(H_a))^2\, G(H_{-a} \mid \mathbf{y})\; d\mathbf{y}\right) \\
&= G(H_{-a}) \int_{\mathbf{y}} (G(H_a \mid \mathbf{y}) - q(H_a))^2\, G(H_{-a} \mid \mathbf{y})\; d\mathbf{y} \\
&\le G(H_{-a}) \int_{\mathbf{y}} (G(H_a \mid \mathbf{y}) - q(H_a))^2\, \Gamma(H_{-a} \mid \mathbf{y})\; d\mathbf{y} \\
&= G(H_{-a}) \left(g(H_{a\times 2}) - 2q(H_a)g(H) + q(H_a)^2\Gamma(H_{-a})\right).
\end{aligned}$$

■

Note that we did not need to assume the full strength of Setup 2.4.1 in the above lemma, since we did not use anything about the jumbledness of Γ or the discrepancy in G in its proof. But these assumptions are useful in what comes next.

Using Proposition 2.3.3 and Lemma 2.3.6, we know that under appropriate hypotheses, we have

$$\begin{aligned}
g(H_{a\times 2}) &\overset{p(H_{a\times 2})}{\underset{c}{\approx}} p(H_{-a})G(H_{a,a\times 2}), \\
g(H) &\overset{p(H)}{\underset{c}{\approx}} p(H_{-a})G(H_a) \\
\text{and}\quad \Gamma(H_{-a}) &\overset{p(H_{-a})}{\underset{c}{\approx}} p(H_{-a}).
\end{aligned} \tag{2.9}$$

If we can show that

$$\begin{aligned}
G(H_{a,a\times 2}) &\overset{p(H_{a,a\times 2})}{\underset{c,\epsilon}{\approx}} q(H_{a,a\times 2}) \\
\text{and}\quad G(H_a) &\overset{p(H_a)}{\underset{c,\epsilon}{\approx}} q(H_a),
\end{aligned}$$

then the rightmost term in (2.8) is $\approx_{c,\epsilon}^{p(H)} 0$, which would reduce the problem to showing that $G(H_{-a}) \approx_{c,\epsilon}^{p(H_{-a})} q(H_{-a})$. This reduction step is spelled out below. See Figure 2-5 for an illustration.

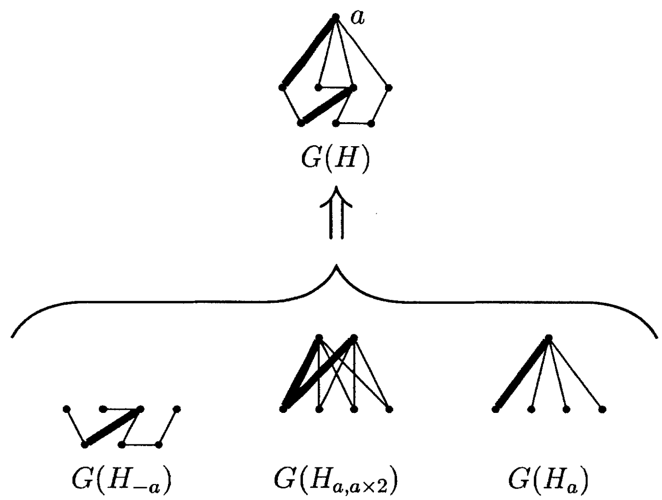


Figure 2-5: The doubling reduction; cf. Figure 2-1 and (2.9).

Lemma 2.4.4 (Doubling). *Assume Setup 2.4.1. Let $a \in V(H)$. Suppose that $k \geq \frac{d(L(H^{\mathrm{sp}}_{a,a\times 2}, H^{\mathrm{sp}}_{-a}))+2}{2}$. Suppose that*

$$G(H_{-a}) \overset{p(H_{-a})}{\underset{c,\epsilon}{\approx}} q(H_{-a}), \quad G(H_a) \overset{p(H_a)}{\underset{c,\epsilon}{\approx}} q(H_a) \text{ and } G(H_{a,a\times 2}) \overset{p(H_{a,a\times 2})}{\underset{c,\epsilon}{\approx}} q(H_{a,a\times 2}).$$

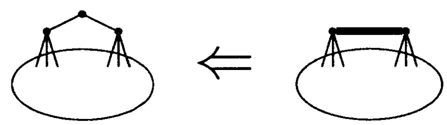
Then

$$G(H) \overset{p(H)}{\underset{c,\epsilon}{\approx}} q(H).$$

Remark. We do not always need the full strength of Setup 2.4.1 (although it is convenient to state it as such). For example, when H is a triangle with vertices $\{1, 2, 3\}$, H_{-1} is a single edge, so we do not need discrepancy on $(X_2, X_3)_G$ to obtain $G(H_{-a}) \approx^{p}_{c,\epsilon} q_{23}$. In particular, our approach gives the triangle counting lemma in the form stated in Kohayakawa et al. [80], where discrepancy is assumed for only two of the three pairs of vertex subsets of G.

2.4.2 Densification

Densification is the technique that allows us to transform a subdivided edge of H into a single dense edge, as summarized in the figure below. This section also contains a counting lemma for trees (Proposition 2.4.9).



We introduce the following notation for the density analogues of degree and codegree. If G (and similarly Γ) is a weighted graph with vertex subsets X, Y, Z, then for $x \in X$ and $z \in Z$, we write

$$\begin{aligned} G(x,Y) &= \int_{y\in Y} G(x,y)\ dy, \\ \text{and}\quad G(x,Y,z) &= \int_{y\in Y} G(x,y)G(y,z)\ dy. \end{aligned}$$

Now we state the goal of this section.

Lemma 2.4.5 (Densification). *Assume Setup 2.4.1 with* $k \geq \frac{d(L(H^{sp}))+2}{2}$. *Let* 1, 2, 3 *be vertices in* H *such that* 1 *and* 3 *are the only neighbors of* 2 *in* H, *and* $13 \notin E(H)$. *Let* H' *denote the graph obtained from* H *by deleting edges* 12 *and* 23 *and adding edge* 13. *Construct a weighted graph* G' *on the same set of vertices of* G *as follows.*

- *For each* $ab \in E(H')\backslash\{13\}$, *set* $G'(x_a, x_b) = G(x_a, x_b)$ *for all* $(x_a, x_b) \in X_a \times X_b$.
- *On* $X_1 \times X_3$, *define* G' *by*

$$G'(x_1, x_3) = \frac{1}{2p_{12}p_{23}} \min\{G(x_1, X_2, x_3), 2p_{12}p_{23}\}. \tag{2.10}$$

Also, set $q'_{ab} = q_{ab}$ *for all* $ab \in E(H') \setminus \{13\}$, *and set* $q'_{13} = \frac{q_{12}q_{23}}{2p_{12}p_{23}}$. *Then* $(X_1, X_3)_{G'}$ *satisfies* $\mathrm{DISC}(q'_{13}, 1, 2\epsilon + 18c)$ *and*

$$|G(H) - 2p_{12}p_{23}G'(H')| \leq ((1+c)^{e(H^{sp})} - 1 + 26c^2)p(H).$$

Note that $q(H) = 2p_{12}p_{23}q'(H')$. So we obtain the following reduction step as a corollary (here $p(H')$ is defined to be $p(H_{-2})$, where H_{-2} denotes H with edges 12 and 23 deleted).

Corollary 2.4.6. *Continuing with Lemma 2.4.5. If*

$$G'(H') \overset{p(H')}{\underset{c,\epsilon}{\approx}} q'(H')$$

then

$$G(H) \overset{p(H)}{\underset{c,\epsilon}{\approx}} q(H).$$

Note that if in place of (2.10), we had instead defined

$$G''(x_1, x_2) = (2p_{12}p_{23})^{-1} G(x_1, X_2, x_3),$$

then $G(H) = 2p_{12}p_{23}G''(H')$ (to see this, when evaluating the integral for $G(H)$, integrate out the variable x_2 first). Since $G'' \geq G'$ everywhere, we have

$$2p_{12}p_{23}G'(H') \leq G(H). \tag{2.11}$$

Lemma 2.4.5 claims that the gap in (2.11) is small. The reason for introducing the cutoff in (2.10) is so that $G' \leq 1$ on $X_1 \times X_3$, so that 13 becomes a dense edge for G'.

The proof of Lemma 2.4.5 consists of the following steps:

1. Show that the weighted graph on $X_1 \times X_3$ with weights $G(x_1, X_2, x_3)$ satisfies discrepancy.

2. Show that the capping of weights has negligible effect on discrepancy.

3. Show that the capping of weights has negligible effect on the H-count.

Steps 2 and 3 are done by bounding the contribution from pairs of vertices in $X_1 \times X_3$ which have too high co-degree with X_2 in Γ.

We shall focus on the more difficult case when both edges 12 and 23 are sparse. The case when at least one of the two edges is dense is analogous and much easier. Let us start with a warm-up by showing how to do step 1 for the latter dense case. We shall omit the rest of the details in this case.

Lemma 2.4.7. *Let $0 \le q_1 \le p_1 \le 1$, $0 \le q_2 \le 1$, $\epsilon > 0$. Let G be a weighted graph with vertex subsets X, Y, Z, such that $(X,Y)_G$ satisfies* $\mathrm{DISC}(q_1, p_1, \epsilon)$ *and $(Y,Z)_G$ satisfies* $\mathrm{DISC}(q_2, 1, \epsilon)$. *Then the graph G' on (X,Z) defined by $G'(x,z) = G(x,Y,z)$ satisfies* $\mathrm{DISC}(q_1q_2, p_1, 2\epsilon)$.

Proof. Let $u\colon X \to [0,1]$ and $w\colon Z \to [0,1]$ be arbitrary functions. In the following integrals, let x, y and z vary uniformly over X, Y and Z, respectively. We have

$$\begin{aligned}
&\int_{x,z} u(x)(G(x,Y,z) - q_1q_2)w(z)\ dxdz \\
&= \int_{x,y,z} u(x)(G(x,y)G(y,z) - q_1q_2)w(z)\ dxdydz \\
&= \int_{x,y,z} u(x)(G(x,y) - q_1)G(y,z)w(z)\ dxdydz \\
&\quad + q_1 \int_{x,y,z} u(x)(G(y,z) - q_2)w(z)\ dxdydz. \qquad (2.12)
\end{aligned}$$

Each of the two integrals in the last sum is bounded by ϵp_1 in absolute value by the discrepancy hypotheses. Therefore $(X,Z)_{G'}$ satisfies $\mathrm{DISC}(q_1q_2, p_1, 2\epsilon)$. ■

The next lemma is step 1 for the sparse case.

Lemma 2.4.8. *Let $c, p, \epsilon \in (0,1]$ and $q_1, q_2 \in [0,p]$. Let Γ be a graph with vertex subsets X, Y, Z and G a weighted subgraph of Γ. Suppose that*

- *$(X,Y)_\Gamma$ is $(p, cp^{3/2}\sqrt{|X|\,|Y|})$-jumbled and $(X,Y)_G$ satisfies* $\mathrm{DISC}(q_1, p, \epsilon)$; *and*
- *$(Y,Z)_\Gamma$ is $(p, cp^{3/2}\sqrt{|Y|\,|Z|})$-jumbled and $(Y,Z)_G$ satisfies* $\mathrm{DISC}(q_2, p, \epsilon)$.

Then the graph G' on (X,Z) defined by $G'(x,z) = G(x,Y,z)$ satisfies $\mathrm{DISC}(q_1q_2, p^2, 3\epsilon + 6c)$.

Remark. By unraveling the proof of Lemma 2.3.8, we see that the exponent of p in the jumbledness of $(X,Y)_\Gamma$ can be relaxed from $\frac{3}{2}$ to 1.

Proof. We begin the proof the same way as Lemma 2.4.7. In (2.12), the second term is bounded in absolute value by $q_1\epsilon p \le \epsilon p^2$. We need to do more work to bound the first integral.

Define $v\colon Y \to [0,1]$ by

$$v(y) = \int_z G(y,z)w(z)\ dz.$$

So the first integral in (2.12), the quantity we need to bound, equals

$$\int_{x,y} u(x)(G(x,y) - q_1)v(y)\ dxdy. \tag{2.13}$$

If we apply discrepancy immediately, we get a bound of ϵp, which is not small enough, as we need a bound on the order of $o(p^2)$. The key observation is that $v(y)$ is bounded above by $2p$ on most of Y. Indeed, let

$$Y' = \{y \in Y \mid \Gamma(y,Z) > 2p\}.$$

By Lemma 2.3.7 we have $|Y'| \leq c^2 p\,|Y|$. Since $v1_{Y\setminus Y'}$ is bounded above by $2p$, we can apply discrepancy on $(X,Y)_G$ with the functions u and $\frac{1}{2p}v1_{Y\setminus Y'}$ to obtain

$$\left|\int_{x,y} u(x)(G(x,y) - q_1)v(y)1_{Y\setminus Y'}\ dxdy\right| \leq 2\epsilon p^2.$$

In the following calculation, the first inequality follows from the triangle inequality and the second inequality follows from expanding $v(y)$ and using $G \leq \Gamma$ and $u, w \leq 1$.

$$\begin{aligned}
&\left|\int_{x,y} u(x)(G(x,y) - q_1)v(y)1_{Y'}(y)\ dxdy\right| \\
&\leq \int_{x,y} u(x)G(x,y)v(y)1_{Y'}(y)\ dxdy + q_1\int_{x,y} u(x)v(y)1_{Y'}(y)\ dxdy \\
&\leq \int_{x,y,z} \Gamma(x,y)1_{Y'}(y)\Gamma(y,z)\ dxdydz + q_1\int_{y,z} 1_{Y'}(y)\Gamma(y,z)\ dydz
\end{aligned} \tag{2.14}$$

Now we apply Lemma 2.3.8 (with the u in the lemma being $1_{Y'}$ on Y and 1 on X and Z, and the H' in the lemma being the empty graph so that $E'(H' \mid \mathbf{x}) = 1$ for all $\mathbf{x}$). We get

$$\int_{x,y,z} \Gamma(x,y)\Gamma(y,z)1_{Y'}(y)\ dxdydz \leq \left((1+c)^2 - 1 + \frac{|Y'|}{|Y|}\right) p^2 \leq 4cp^2.$$

and

$$\int_{y,z} \Gamma(y,z)1_{Y'}(y)\ dydz \leq \left((1+c) - 1 + \frac{|Y'|}{|Y|}\right) p \leq 2cp.$$

It follows that (2.14) is bounded by $4cp^2 + 2cpq_1 \leq 6cp^2$. Therefore, (2.13) is at most $(2\epsilon + 6c)p^2$ in absolute value. Recall that the second integral in (2.12) was bounded by ϵp^2. The result follows from combining these two estimates. ■

The technique used in Lemma 2.4.8 also allows us to count trees in G.

Proposition 2.4.9. *Assume Setup 2.4.1 with H a tree and $k \geq \frac{\Delta(H^{\mathrm{sp}})+1}{2}$. Then*

$$G(H) \overset{p(H)}{\underset{c,\epsilon}{\approx}} q(H).$$

In fact, it can be shown that the error has the form

$$|G(H) - q(H)| \leq M_H(c+\epsilon)p(H)$$

for some real number $M_H > 0$ depending on H.

To prove Proposition 2.4.9, we formulate a weighted version and induct on the number of edges. The weighted version is stated below.

Lemma 2.4.10. *Assume the same setup as in Proposition 2.4.9. Let $u\colon V(G) \to [0,1]$ be any function. Let $\mathbf{x}$ vary uniformly over all compatible maps $V(H) \to V(G)$. Write $u(\mathbf{x}) = \prod_{a\in V(H)} u(x_a)$. Then,*

$$\int_{\mathbf{x}} G\left(H \mid \mathbf{x}\right) u(\mathbf{x})\ d\mathbf{x} \overset{p(H)}{\underset{c,\epsilon}{\approx}} q(H) \int_{\mathbf{x}} u(\mathbf{x})\ d\mathbf{x}.$$

To prove Lemma 2.4.10, we remove one leaf of H at a time and use the technique in the proof of Lemma 2.4.8 to transfer the weight of the leaf to its unique neighboring vertex and use Lemma 2.3.8 to bound the contributions of the vertices with large degrees in Γ. We omit the details.

Continuing with the proof of densification, the following estimate is needed for steps 2 and 3.

Lemma 2.4.11. *Let Γ be a graph with vertex subsets X, Y, Z, such that $(X,Y)_\Gamma$ is $(p, cp\sqrt{|X|\,|Y|})$-jumbled and $(Y,Z)_\Gamma$ is $(p, cp^{3/2}\sqrt{|Y|\,|Z|})$-jumbled. Let*

$$E' = \left\{(x,z) \in X \times Z \;\middle|\; \Gamma(x,Y,z) > 2p^2\right\}.$$

Then $|E'| \le 26c^2\,|X|\,|Z|$.

Proof. Let

$$X' = \left\{x \in X \;\middle|\; |\Gamma(x,Y) - p| > \frac{p}{2}\right\}.$$

Then, by Lemma 2.3.7, $|X'| \le 8c^2\,|X|$. For every $x \in X$, let

$$Z'_x = \left\{z \in Z \;\middle|\; \Gamma(x,Y,z) > 2p^2\right\}.$$

For $x \in X \setminus X'$, we have, again by Lemma 2.3.7, that $|Z'_x| \le 18c^2\,|Z|$. The result follows by noting that $E' \subseteq (X' \times Z) \cup \{(x,z) \mid x \in X \setminus X', z \in Z'_x\}$. ■

The following lemma is step 2 in the program.

Lemma 2.4.12. *Let $c, \epsilon, p \in (0,1]$ and $q_1, q_2 \in [0,p]$. Let Γ be a graph with vertex subsets X, Y, Z, and G a weighted subgraph of Γ. Suppose that*

- *$(X,Y)_\Gamma$ is $(p, cp\sqrt{|X|\,|Y|})$-jumbled and $(X,Y)_G$ satisfies $\mathrm{DISC}(q_1, p, \epsilon)$; and*
- *$(Y,Z)_\Gamma$ is $(p, cp^{3/2}\sqrt{|Y|\,|Z|})$-jumbled and $(Y,Z)_G$ satisfies $\mathrm{DISC}(q_2, p, \epsilon)$.*

Then the graph G' on (X,Z) defined by $G'(x,z) = \min\{G(x,Y,z), 2p^2\}$ satisfies $\mathrm{DISC}(q_1q_2, p^2, 3\epsilon + 35c)$.

Proof. Let $u\colon X \to [0,1]$ and $w\colon Z \to [0,1]$ be any functions. In the following integrals, x, y and z vary uniformly over X, Y and Z, respectively. We have

$$\begin{aligned}
&\int_{x,z} (G'(x,z) - q_1q_2)u(x)w(z)\ dxdz \\
&= \int_{x,z} (G(x,Y,z) - q_1q_2)u(x)w(z)\ dxdz - \int_{x,z} (G(x,Y,z) - G'(x,z))u(x)w(z)\ dxdz.
\end{aligned}$$

The first integral on the right-hand side can be bounded in absolute value by $(3\epsilon+6c)p^2$ by Lemma 2.4.8. For the second integral, let $E' = \{(x,z) \in X \times Z \mid \Gamma(x,Y,z) > 2p^2\}$. We have

$$\begin{aligned}
0 &\le \int_{x,z} (G(x,Y,z) - G'(x,z))u(x)w(z)\ dxdz \\
&\le \int_{x,z} G(x,Y,z)1_{E'}(x,z)\ dxdz \\
&\le \int_{x,y,z} \Gamma(x,y)\Gamma(y,z)1_{E'}(x,z)\ dxdydz \\
&\le \left((1+c)^2 - 1 + \frac{|E'|}{|X|\,|Z|}\right)p^2 \\
&\le 29cp^2
\end{aligned}$$

by Lemmas 2.3.8 and 2.4.11. The result follows by combining the estimates. ∎

Finally we prove step 3 in the program, thereby completing the proof of densification.

Proof of Lemma 2.4.5. We prove the result when both edges 12 and 23 are sparse. When at least one of 12 and 23 is dense, the proof is analogous and easier.

Lemma 2.4.12 implies that $(X_1, X_3)_{G'}$ satisfies $\mathrm{DISC}(q_{13}, \frac{1}{2}, 3\epsilon + 35c)$, and hence it must also satisfy $\mathrm{DISC}(q_{13}, 1, 2\epsilon + 18c)$. Let

$$E' = \{(x_1, x_3) \mid \Gamma(x_1, X_2, x_3) > 2p_{12}p_{23}\}\,.$$

We have $|E'| \le 26c^2 |X_1| |X_3|$ by Lemma 2.4.11. Let

$$E'' = \{(x_1, x_3) \mid G(x_1, X_2, x_3) > 2p_{12}p_{23}\}.$$

So E'' is the set of all (x_1, x_3) for which we applied the cap in (2.10) when constructing G'. Since $G \le \Gamma$, we have $E'' \subseteq E'$, so $|E''| \le |E'| \le 26c^2 |X_1| |X_3|$.

Recall the observation (2.11) that $G(H) \ge 2p_{12}p_{23}G'(H')$. The gap in the inequality comes from the cutoffs (2.10) of weights on edges $(x_1, x_3) \in E''$. Let H_{-2} denote H with edges 12 and 23 deleted. Let $\mathbf{x}_{-2}\colon V(H) \setminus \{2\} \to V(\Gamma)$, $\mathbf{x}\colon V(H) \to V(\Gamma)$ vary uniformly over all compatible maps. We have

$$\begin{aligned}
0 &\le G(H) - 2p_{12}p_{23}G'(H') \\
&= \int_{\mathbf{x}_{-2}} G(H_{-2} \mid \mathbf{x}_{-2}) (G(x_1, X_2, x_3) - 2p_{12}p_{23}) 1_{E''}(x_1, x_3) \, d\mathbf{x}_{-2} \\
&\le \int_{\mathbf{x}_{-2}} \Gamma(H_{-2} \mid \mathbf{x}_{-2}) \Gamma(x_1, X_2, x_3) 1_{E''}(x_1, x_3) \, d\mathbf{x}_{-2} \\
&= \int_{\mathbf{x}} \Gamma(H \mid \mathbf{x}) \, 1_{E''}(x_1, x_3) \, d\mathbf{x} \\
&\le \left((1+c)^{e(H^{\mathrm{sp}})} - 1 + \frac{|E''|}{|X_1| \, |X_3|} \right) p(H) \\
&\le \left((1+c)^{e(H^{\mathrm{sp}})} - 1 + 26c^2 \right) p(H)
\end{aligned}$$

where the third inequality is by Lemma 2.3.8. ■

2.4.3 Counting C_4

With the tools of doubling and densification, we are now ready to count embeddings in G. We start by showing how to count C_4, as it is an important yet tricky step.

Proposition 2.4.13. *Assume Setup 2.4.1 with $H = C_4$ and $k \ge 2$. Then*

$$|G(C_4) - q(C_4)| \le 100(c + \epsilon)^{1/2} p(C_4).$$

The constant 100 is unimportant. It can be obtained by unraveling the calculations. We omit the details.

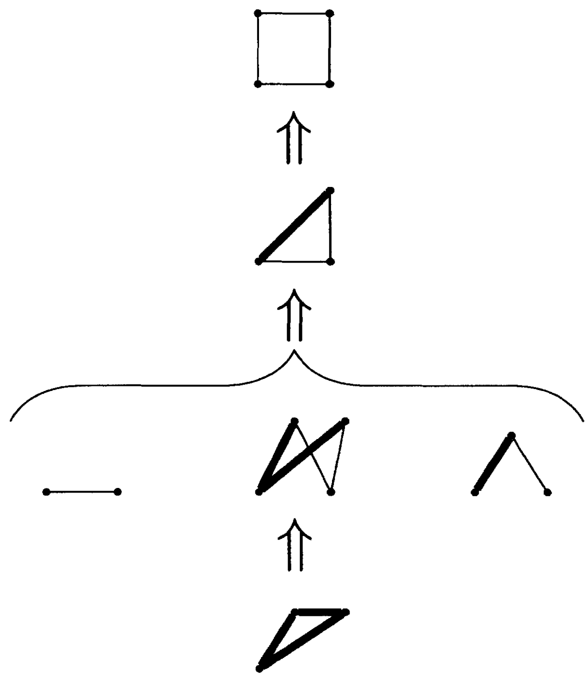


Figure 2-6: The proof that $G(C_4) \approx_{c,\epsilon}^{p^4} q(C_4)$. The vertical arrows correspond to densification, doubling the top vertex and densification, respectively.

Proposition 2.4.13 follows from repeated applications of doubling (Lemma 2.4.4) and densification (Corollary 2.4.6). The chain of implications is summarized in Figure 2-6 in the case when all four edges of C_4 are sparse (the other cases are easier). In the figure, each graph represents a claim of the form $G(H) \approx_{c,\epsilon}^{p(H)} q(H)$. The sparse and dense edges are distinguished by thickness. The claim for the dense triangle follows from the counting lemma for dense graphs (Proposition 2.1.8) and the claim for the rightmost graph follows from Lemma 2.4.7.

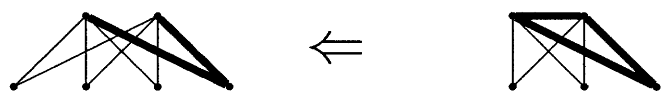
2.4.4 Finishing the proof of the counting lemma

Given a graph H, we can use the doubling lemma, Lemma 2.4.4, to reduce the problem of counting H in G to the problem of counting H_{-a} in G, where H_{-a} is H with some vertex a deleted, provided we can also count H_a and $H_{a,a\times 2}$. Suppose a has degree t in H and degree t' in H^{sp}. The graph H_a is isomorphic to some $K_{1,t}$. Since $K_{1,t}$ is a tree, we can count copies using Proposition 2.4.9, provided that the exponent of p in the jumbledness of Γ satisfies $k \geq \frac{t'+1}{2}$. The following lemma shows that we can count embeddings of $H_{a,a\times 2}$ as well.

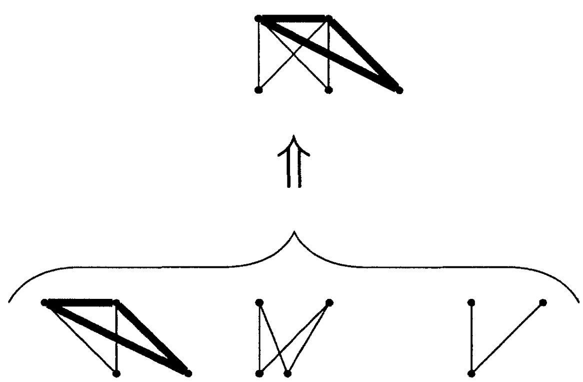
Lemma 2.4.14. *Assume Setup 2.4.1 where $H = K_{2,t}$ with vertices $\{a_1, a_2; b_1, \ldots, b_t\}$. Assume that the edges a_ib_j are sparse for $1 \leq j \leq t'$ and dense for $j > t'$. If $k \geq \frac{t'+2}{2}$, then*

$$G(H) \overset{p(H)}{\underset{c,\epsilon}{\approx}} q(H).$$

Proof. When $t' = 0$, all edges of H are dense, so the result follows from the dense counting lemma. So assume $t' \geq 1$. First we apply densification as follows:



When $t' = 1$, we get a dense graph so we are done. Otherwise, the result follows by induction using doubling as shown below, where we use Propositions 2.4.13 and 2.4.9 to count C_4 and $K_{1,2}$, respectively.



■

Once we can count H_a and $H_{a,a\times 2}$, we obtain the following reduction result via doubling.

Lemma 2.4.15. *Assume Setup 2.4.1. Let a be a vertex of H. If $k \geq \frac{d(L(H^{\mathrm{sp}}_{a,a\times 2}, H^{\mathrm{sp}}_{-a}))+2}{2}$, then*

$$G(H) \overset{p(H)}{\underset{c,\epsilon}{\approx}} q(H_a)G(H_{-a}).$$

The proof of the counting lemma follows once we keep track of the requirements on k.

Proof of Theorem 2.4.2. When H has no sparse edges, the result follows from the dense counting lemma (Proposition 2.1.8). Otherwise, using Lemma 2.4.15, it remains to show that if $k \geq \min\left\{\frac{\Delta(L(H^{\mathrm{sp}}))+4}{2}, \frac{d(L(H^{\mathrm{sp}}))+6}{2}\right\}$, then there exists some vertex a of H satisfying $k \geq \frac{d(L(H^{\mathrm{sp}}_{a,a\times 2}, H^{\mathrm{sp}}_{-a}))+2}{2}$. Actually, the hypothesis on k is strong enough that any a will do. Indeed, we have $\Delta(L(H^{\mathrm{sp}})) + 2 \geq \Delta(L(H^{\mathrm{sp}}_{a\times 2})) \geq d(L(H^{\mathrm{sp}}_{a,a\times 2}, H^{\mathrm{sp}}_{-a}))$ since doubling a increases the degree of every vertex by at most 1. We also have $d(L(H^{\mathrm{sp}})) \geq d(L(H^{\mathrm{sp}}_{a,a\times 2}, H^{\mathrm{sp}}_{-a})) - 4$ since every edge in H^{sp}_{-a} shares an endpoint with at most 4 edges in $H^{\mathrm{sp}}_{a,a\times 2}$. ■

2.4.5 Tutorial: determining jumbledness requirements

The jumbledness requirements stated in our counting lemmas are often not the best that come out of our proofs. We had to make a tradeoff between strength and simplicity while formulating the results. In this section, we give a short tutorial on finding the jumbledness requirements needed for our counting lemma to work for any particular graph H. These fine-tuned bounds can be extracted from a careful examination of our proofs, with no new ideas introduced in this section.

We work in a more general setting where we allow non-balanced jumbledness conditions between vertex subsets of Γ. This will arise naturallly in Section 2.6 when we prove a one-sided counting lemma.

Setup 2.4.16. Let Γ be a graph with vertex subsets $X_1, \dots, X_m$. Let $p, c \in (0, 1]$. Let H be a graph with vertex set $\{1, \dots, m\}$, with vertex a assigned to X_a. For every edge ab in H, one of the following two holds:

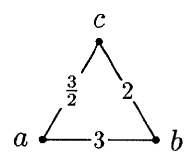
- $(X_a, X_b)_\Gamma$ is $(p, cp^{k_{ab}}\sqrt{|X_a|\,|X_b|})$-jumbled for some $k_{ab} \geq 1$, in which case we set $p_{ab} = p$ and say that ab is a *sparse edge*, or

- $(X_a, X_b)_\Gamma$ is a complete bipartite graph, in which case we set $p_{ab} = 1$ and say that ab is a *dense edge*.

Let H^{sp} denote the subgraph of H consisting of sparse edges.

Let $\epsilon > 0$. Let G be a weighted subgraph of Γ. For every edge $ab \in E(H)$, assume that $(X_a, X_b)_G$ satisfies $\mathrm{DISC}(q_{ab}, p_{ab}, \epsilon)$, where $0 \le q_{ab} \le p_{ab}$.

In the figures in this section, we label the edges by the lower bounds on k_{ab} that are sufficient for the two-sided counting lemma to hold. For instance, the figure below shows the jumbledness conditions that are sufficient for the triangle counting lemma[1], namely $k_{ab} \ge 3$, $k_{bc} \ge 2$, $k_{ac} \ge \frac{3}{2}$.



Although we are primarily interested in embeddings of H into G, we need to consider partial embeddings where some of the edges of H are allowed to embed into Γ. So we encounter three types of edges of H, summarized in Table 2.2. (Note that for dense edges ab, $(X_a, X_b)_\Gamma$ is a complete bipartite graph, so such embeddings are trivial and ab can be ignored.)

Table 2.2: Types of edges in H.

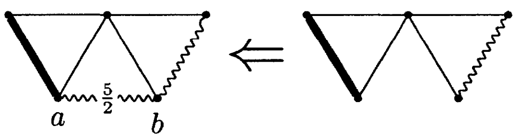
Figure	Name	Description
∿κ∿	Jumbled edge	An edge to be embedded in $(X_a, X_b)_\Gamma$ with $p_{ab} = p$ and $k_{ab} \ge \kappa$.
▬▬	Dense edge	An edge to be embedded in $(X_a, X_b)_G$ with $p_{ab} = 1$.
—κ—	Sparse edge	An edge to be embedded in $(X_a, X_b)_G$ with $p_{ab} = p$ and $k_{ab} \ge \kappa$.

Our counting lemma is proved through a number of reduction procedures. At each step, we transform H into one or more other graphs H'. At the end of the reduction procedure, we should arrive at a graph which only has dense edges. To determine the

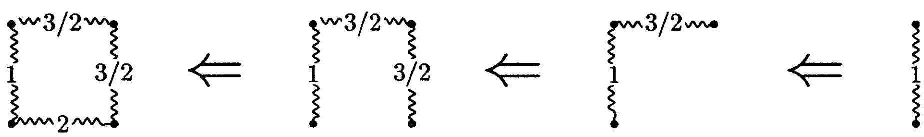
[1]As mentioned in the remark after Lemma 2.4.4, we do not actually need DISC on $(X_a, X_b)_G$, since edge density is enough. We do not dwell on this point in this section and instead focus on jumbledness requirements.

jumbledness conditions required to count some H, we perform these reduction steps and keep track of the requirements at each step. We explain how to do this for each reduction procedure.

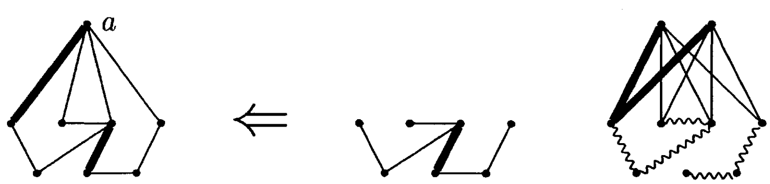
Removing a jumbled edge. To remove a jumbled edge ab from H, we need k_{ab} to be at least the average of the sparse degrees (i.e., counting both sparse and jumbled edges) at the endpoints of ab, i.e., $k_{ab} \geq \frac{1}{2}(\deg_{H^{\text{sp}}}(a) + \deg_{H^{\text{sp}}}(b))$. See Lemma 2.3.5. For example, $k_{ab} \geq \frac{5}{2}$ is sufficient to remove the edge ab in the graph below.



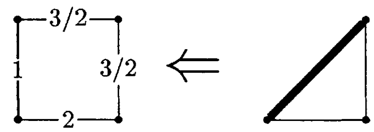
By removing jumbled edges one at a time, we can find conditions that are sufficient for counting embeddings into Γ (Proposition 2.3.3). The following figure shows how this is done for a 4-cycle.



Doubling The figure below illustrates doubling. If the jumbledness hypotheses are sufficient to count the two graphs on the right, then they are sufficient to count the original graph. The first graph is produced by deleting all edges with a as an endpoint, and the second graph is produced by doubling a and then, for all edges not adjacent to a, deleting the dense edges and converting sparse edges to jumbled ones.

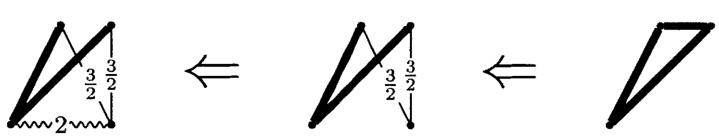


Densification To determine the jumbledness needed to perform densification, delete all dense edges, transform all sparse edges into jumbled edges, and use the earlier method to determine the jumbledness required to count embeddings into Γ. For example, the jumbledness on the left figure below shows the requirements on C_4 needed to perform the densification step. It may be the case that even stronger hypotheses are needed to count the new graph (although for this example this is not the case).



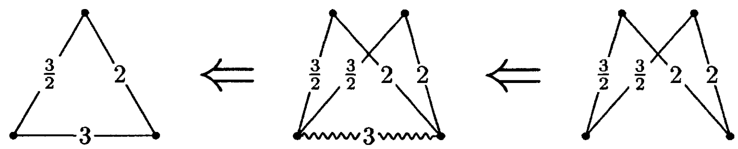
Trees To determine the jumbledness needed to count some tree H, delete all dense edges in H, transform all sparse edges into jumbled edges and use the earlier method, removing one leaf at a time to determine the jumbledness required to count embeddings into Γ (Proposition 2.4.9).

Example 2.4.17 (C_4). Let us check that the labeling of C_4 in the densification paragraph gives sufficient jumbledness to count C_4. It remains to check that the jumbledness hypotheses are sufficient to count the triangle with a single edge. We can double the top vertex so that it remains to check the first graph below (the other graph produced from doubling is a single edge, which is trivial to count). We can remove the jumbled edge, and then perform densification to get a dense triangle, which we know how to count.

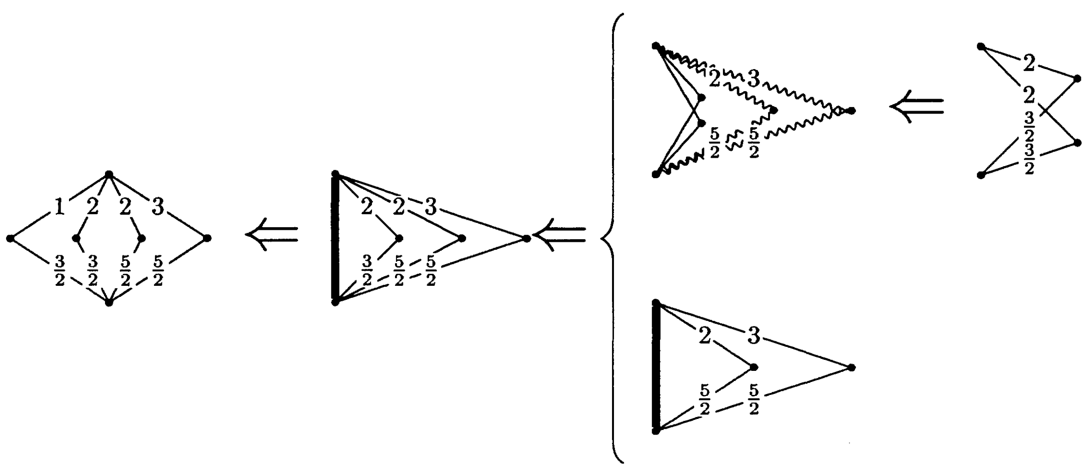


Example 2.4.18 (K_3). The following diagram illustrates the process of checking sufficient jumbledness hypotheses to count triangles (again, the first graph resulting from doubling is a single edge and is thus omitted from the figure). The sufficiency

for C_4 follows from the previous example.

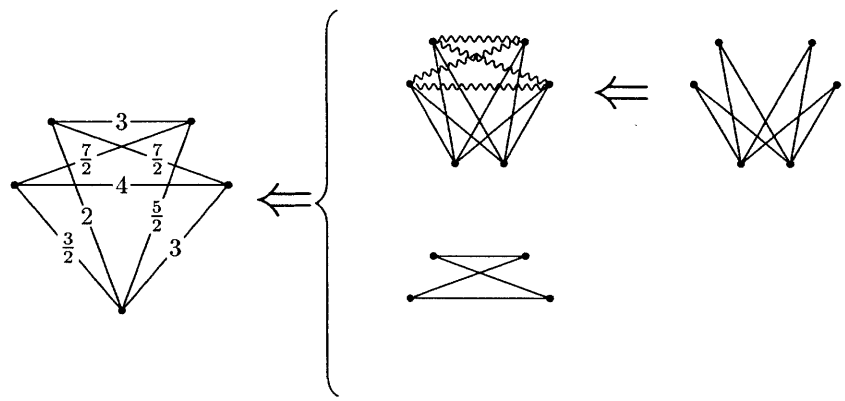


Example 2.4.19 ($K_{2,t}$)**.** The following diagram shows sufficient jumbledness to count $K_{2,4}$. The same pattern holds for $K_{2,t}$. The reduction procedure was given in the proof of Lemma 2.4.14. First we perform densification to the two leftmost edges, and then apply doubling to the remaining middle vertices in order from left to right.



Example 2.4.20 ($K_{1,2,2}$)**.** The following diagram shows sufficient jumbledness to count $K_{1,2,2}$. The edge labels for the graphs on the right are inherited from the graph on the left and are omitted from the figure to avoid clutter. This example will be

used in the next section on inheriting regularity.



2.5 Inheriting regularity

Regularity is inherited on large subsets, in the sense that if $(X, Y)_G$ satisfies $\mathrm{DISC}(q, 1, \epsilon)$, then for any $U \subseteq X$ and $V \subseteq Y$, the induced pair $(U, V)_G$ satisfies $\mathrm{DISC}(q, 1, \epsilon')$ with $\epsilon' = \frac{|X||Y|}{|U||V|}\epsilon$. This is a trivial consequence of the definition of discrepancy, and the change in ϵ comes from rescaling the measures dx and dy after restricting the uniform distribution to a subset. The loss in ϵ is a constant factor as long as $\frac{|U|}{|X|}$ and $\frac{|V|}{|Y|}$ are bounded from below. So if G is a dense tripartite graph with vertex subsets X, Y, Z, with each pair being dense and regular, then we expect that for most vertices $z \in Z$, its neighborhoods $N_X(z)$ and $N_Y(z)$ are large, and hence they induce regular pairs with only a constant factor loss in the discrepancy parameter ϵ.

The above argument does not hold in sparse pseudorandom graphs. It is still true that if $(X, Y)_G$ satisfies $\mathrm{DISC}(q, p, \epsilon)$ then for any $U \subseteq X$ and $V \subseteq Y$ the induced pair $(U, V)_G$ satisfies $\mathrm{DISC}(q, p, \epsilon')$ with $\epsilon' = \frac{|X||Y|}{|U||V|}\epsilon$. However, in the tripartite setup from the previous paragraph, we expect most $N_X(z)$ to have size on the order of $p\,|X|$. So the naive approach shows that most $z \in Z$ induce a bipartite graph satisfying $\mathrm{DISC}(q, p, \epsilon')$ where ϵ' is on the order of $\frac{\epsilon}{p^2}$. This is undesirable, as we do not want ϵ to depend on p.

It turns out that for most $z \in Z$, the bipartite graph induced by the neighborhoods satisfies $\mathrm{DISC}(q, p, \epsilon')$ for some ϵ' depending on ϵ but not p. In this section we prove this fact using the counting lemma developed earlier in the paper. We recall the statement from the introduction.

Proposition 2.1.13 *For any $\alpha > 0$, $\xi > 0$ and $\epsilon' > 0$, there exists $c > 0$ and $\epsilon > 0$ of size at least polynomial in α, ξ, ϵ' such that the following holds.*

Let $p \in (0, 1]$ and $q_{XY}, q_{XZ}, q_{YZ} \in [\alpha p, p]$. Let Γ be a tripartite graph with vertex subsets X, Y and Z and G be a subgraph of Γ. Suppose that

- *$(X, Y)_\Gamma$ is $(p, cp^4\sqrt{|X|\,|Y|})$-jumbled and $(X, Y)_G$ satisfies $\mathrm{DISC}(q_{XY}, p, \epsilon)$; and*
- *$(X, Z)_\Gamma$ is $(p, cp^2\sqrt{|X|\,|Z|})$-jumbled and $(X, Z)_G$ satisfies $\mathrm{DISC}(q_{XZ}, p, \epsilon)$; and*
- *$(Y, Z)_\Gamma$ is $(p, cp^3\sqrt{|Y|\,|Z|})$-jumbled and $(Y, Z)_G$ satisfies $\mathrm{DISC}(q_{YZ}, p, \epsilon)$.*

Then at least $(1-\xi)\,|Z|$ vertices $z \in Z$ have the property that $|N_X(z)| \geq (1-\xi)q_{XZ}\,|X|$, $|N_Y(z)| \geq (1-\xi)q_{YZ}\,|Y|$, and $(N_X(z), N_Y(z))_G$ satisfies $\mathrm{DISC}(q_{XY}, p, \epsilon')$.

The idea of the proof is to first show that a bound on the $K_{2,2}$ count implies DISC and then to use the $K_{1,2,2}$ count to bound the $K_{2,2}$ count between neighborhoods.

We also state a version where only one side gets smaller. While the previous proposition is sufficient for embedding cliques, this second version will be needed for embedding general graphs H.

Proposition 2.5.1. *For any $\alpha > 0$, $\xi > 0$ and $\epsilon' > 0$, there exists $c > 0$ and $\epsilon > 0$ of size at least polynomial in α, ξ, ϵ' such that the following holds.*

Let $p \in (0, 1]$ and $q_{XY}, q_{XZ} \in [\alpha p, p]$. Let Γ be a tripartite graph with vertex subsets X, Y and Z and G be a subgraph of Γ. Suppose that

- *$(X, Y)_\Gamma$ is $(p, cp^{5/2}\sqrt{|X|\,|Y|})$-jumbled and $(X, Y)_G$ satisfies $\mathrm{DISC}(q_{XY}, p, \epsilon)$; and*
- *$(X, Z)_\Gamma$ is $(p, cp^{3/2}\sqrt{|X|\,|Z|})$-jumbled and $(X, Z)_G$ satisfies $\mathrm{DISC}(q_{XZ}, p, \epsilon)$.*

Then at least $(1-\xi)\,|Z|$ vertices $z \in Z$ have the property that $|N_X(z)| \geq (1-\xi)q_{XZ}\,|X|$ and $(N_X(z), Y)_G$ satisfies $\mathrm{DISC}(q_{XY}, p, \epsilon')$.

2.5.1 C_4 implies DISC

From our counting lemma we already know that if G is a subgraph of a sufficiently jumbled graph with vertex subsets X and Y such that $(X,Y)_G$ satisfies DISC(q,p,ϵ), then the number of $K_{2,2}$ in G across X and Y is roughly $q^4\left|X\right|^2\left|Y\right|^2$. In this section, we show that the converse is true, that the $K_{2,2}$ count implies discrepancy, even without any jumbledness hypotheses.

In what follows, for any function $f\colon X\times Y\to\mathbb{R}$, we write

$$f(K_{s,t})=\int_{\substack{x_1,\dots,x_s\in X\\ y_1,\dots,y_t\in Y}}\prod_{i=1}^{s}\prod_{j=1}^{t}f(x_i,y_j)\;dx_1\cdots dx_s dy_1\cdots dy_t.$$

The following lemma shows that a bound on the "de-meaned" C_4-count implies discrepancy.

Lemma 2.5.2. *Let G be a bipartite graph between vertex sets X and Y. Let $0\le q\le p\le 1$ and $\epsilon>0$. Define $f\colon X\times Y\to\mathbb{R}$ by $f(x,y)=G(x,y)-q$. If $f(K_{2,2})\le\epsilon^4p^4$ then $(X,Y)_G$ satisfies* DISC(q,p,ϵ).

Proof. Let $u\colon X\to[0,1]$ and $v\colon Y\to[0,1]$ be any functions. Applying the Cauchy-Schwarz inequality twice, we have

$$\begin{aligned}
&\left(\int_{x\in X}\int_{y\in Y}f(x,y)u(x)v(y)\;dydx\right)^4\\
&\le\left(\int_{x\in X}\left(\int_{y\in Y}f(x,y)u(x)v(y)\;dy\right)^2\,dx\right)^2\\
&=\left(\int_{x\in X}u(x)^2\left(\int_{y\in Y}f(x,y)v(y)\;dy\right)^2\,dx\right)^2\\
&\le\left(\int_{x\in X}\left(\int_{y\in Y}f(x,y)v(y)\;dy\right)^2\,dx\right)^2\\
&=\left(\int_{x\in X}\int_{y,y'\in Y}f(x,y)f(x,y')v(y)v(y')\;dydy'dx\right)^2\\
&\le\int_{y,y'\in Y}\left(\int_{x\in X}f(x,y)f(x,y')v(y)v(y')\;dx\right)^2\,dydy'
\end{aligned}$$

$$
\begin{aligned}
&= \int_{y,y'\in Y} v(y)^2 v(y')^2 \left(\int_{x\in X} f(x,y)f(x,y')\,dx\right)^2 dydy' \\
&\le \int_{y,y'\in Y} \left(\int_{x\in X} f(x,y)f(x,y')\,dx\right)^2 dydy' \\
&= \int_{y,y'\in Y}\int_{x,x'\in X} f(x,y)f(x,y')f(x',y)f(x',y')\,dxdx'dydy' \\
&= f(K_{2,2}) \\
&\le \epsilon^4 p^4.
\end{aligned}
$$

Thus

$$\left|\int_{x\in X}\int_{y\in Y} (G(x,y)-q)u(x)v(y)\,dydx\right| \le \epsilon p.$$

Hence $(X,Y)_G$ satisfies $\mathrm{DISC}(q,p,\epsilon)$. ■

Lemma 2.5.3. *Let G be a bipartite graph between vertex sets X and Y. Let $0 \le q \le p \le 1$ and $\epsilon > 0$. Let $U \subseteq X$ and $V \subseteq Y$. Let $\mu = \frac{|U|}{|X|}$ and $\nu = \frac{|V|}{|Y|}$. Define $f\colon X \times Y \to \mathbb{R}$ by $f(x,y) = (G(x,y) - q)1_U(x)1_V(y)$. If $f(K_{2,2}) \le \epsilon^4 p^4 \mu^2 \nu^2$, then $(U,V)_G$ satisfies* $\mathrm{DISC}(q,p,\epsilon)$.

Proof. This lemma is equivalent to Lemma 2.5.2 after appropriate rescaling of the measures dx and dy. ■

The above lemmas are sufficient for proving inheritance of regularity, so that the reader may now skip to the next subsection. The rest of this subsection contains a proof that an upper bound on the actual C_4 count implies discrepancy, a result of independent interest which is discussed further in Section 2.9.2 on relative quasirandomness.

Proposition 2.5.4. *Let G be a bipartite graph between vertex sets X and Y. Let $0 \le q \le 1$ and $\epsilon > 0$. Suppose $G(K_{1,1}) \ge (1-\epsilon)q$ and $G(K_{2,2}) \le (1+\epsilon)^4 q^4$, then $(X,Y)_G$ satisfies* $\mathrm{DISC}(q,q,4\epsilon^{1/36})$.

The hypotheses in Proposition 2.5.4 actually imply two-sided bounds on $G(K_{1,1})$, $G(K_{1,2})$, $G(K_{2,1})$, and $G(K_{2,2})$, by the following lemma.

Lemma 2.5.5. *Let G be a bipartite graph between vertex sets X and Y and $f\colon X \times Y \to \mathbb{R}$ be any function. Then $f(K_{1,1})^4 \le f(K_{1,2})^2 \le f(K_{2,2})$.*

Proof. The result follows from two applications of the Cauchy-Schwarz inequality.

$$\begin{aligned}
f(K_{2,2}) &= \int_{y,y' \in Y} \int_{x,x' \in X} f(x,y)f(x,y')f(x',y)f(x',y')\ dxdx'dydy' \\
&= \int_{y,y' \in Y} \left(\int_{x \in X} f(x,y)f(x,y')\ dx \right)^2\ dydy' \\
&\ge \left(\int_{y,y' \in Y} \int_{x \in X} f(x,y)f(x,y')\ dx\ dydy' \right)^2 \\
&= f(K_{1,2})^2 \\
&= \left(\int_{x \in X} \left(\int_{y \in Y} f(x,y)\ dy \right)^2\ dx \right)^2 \\
&\ge \left(\int_{x \in X} \int_{y \in Y} f(x,y)\ dy\ dx \right)^4 \\
&= f(K_{1,1})^4.
\end{aligned}$$

■

A bound on $K_{1,2}$ is a second moment bound on the degree distribution, so we can bound the number of vertices of low degree using Chebyshev's inequality, as done in the next lemma. Recall the notation $G(x,S) = \int_{y \in Y} G(x,y)1_S(y)\ dy$ for $S \subseteq Y$ as the normalized degree.

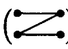
Lemma 2.5.6. *Let G be a bipartite graph between vertex sets X and Y. Let $0 \le q \le 1$ and $\epsilon > 0$. Suppose $G(K_{1,1}) \ge (1-\epsilon)q$ and $G(K_{1,2}) \le (1+\epsilon)^2 q^2$. Let $X' = \{x \in X \mid G(x,Y) < (1 - 2\epsilon^{1/3})q\}$. Then $|X'| \le 2\epsilon^{1/3}|X|$.*

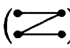
Proof. We have

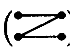
$$\frac{|X'|}{|X|}\left(2\epsilon^{1/3}q\right)^2 \leq \int_{x\in X} \left(G(x,Y)-q\right)^2\ dx$$
$$= G(K_{1,2}) - 2qG(K_{1,1}) + q^2$$
$$\leq (1+\epsilon)^2q^2 - 2(1-\epsilon)q^2 + q^2$$
$$\leq 5\epsilon q^2.$$

Thus $|X'| \leq \frac{5}{4}\epsilon^{1/3}|X|$. ■

We write

G () $= \int_{\substack{x,x'\in X\\ y,y'\in Y}} G(x,y)G(x',y)G(x',y')\ dx dx' dy dy'.$

The next lemma proves a lower bound on G () by discarding vertices of low degree.

Lemma 2.5.7. *Let G be a bipartite graph between vertex sets X and Y. Let $0 \leq q \leq 1$ and $\epsilon > 0$. Suppose $G(K_{1,1}) \geq (1-\epsilon)q$, $G(K_{1,2}) \leq (1+\epsilon)^2q^2$ and $G(K_{2,1}) \leq (1+\epsilon)^2q^2$. Then G () $\geq (1-14\epsilon^{1/9})q^3$.*

Proof. Let

$$X' = \left\{x \in X \mid G(x,Y) < (1-2\epsilon^{1/3})q\right\}.$$

Let G' denote the subgraph of G where we remove all edges with an endpoint in X'. Then $G'(K_{2,1}) \leq G(K_{2,1}) \leq (1+\epsilon)^2q^2$ and, by Lemma 2.5.6

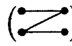
$$G'(K_{1,1}) \geq \frac{|X\setminus X'|}{|X|}(1-2\epsilon^{1/3})q \geq (1-2\epsilon^{1/3})^2q \geq (1-4\epsilon^{1/3})q.$$

Let

$$Y' = \left\{y \in Y \mid G(X\setminus X', y) < (1-4\epsilon^{1/9})q\right\}.$$

So $|Y'| \leq 4\epsilon^{1/9}$ by applying Lemma 2.5.6 again. Restricting to paths with vertices in

$X \setminus X', Y \setminus Y', X \setminus X', Y$, we find that

$G\Big($  $\Big) \geq \frac{|Y \setminus Y'|}{|Y|} \left(\min_{y \in Y \setminus Y'} G(X \setminus X', y) \right)^2 \left(\min_{x \in X \setminus X'} G(x, Y) \right)$

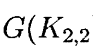
$$\begin{aligned} &\geq (1 - 4\epsilon^{1/9})^3 (1 - 2\epsilon^{1/3}) q^3 \\ &\geq (1 - 14\epsilon^{1/9}) q^3. \end{aligned}$$

■

Proof of Proposition 2.5.4. Using Lemma 2.5.5, we have

$$G(K_{1,2}) \leq (1+\epsilon)^2 q^2, \quad G(K_{2,1}) \leq (1+\epsilon)^2 q^2, \quad \text{and} \quad G(K_{1,1}) \leq (1+\epsilon) q.$$

Let $f(x,y) = G(x,y) - q$. Applying Lemma 2.5.7, we have

$f(K_{2,2}) = G(K_{2,2}) - 4qG\Big($  $\Big) + 2q^2 G(K_{1,1})^2 + 4q^2 G(K_{1,2}) - 4q^3 G(K_{1,1}) + q^4$

$$\begin{aligned} &\leq (1+\epsilon)^4 q^4 - 4(1 - 14\epsilon^{1/9}) q^4 + 2(1+\epsilon)^2 q^4 + 4(1+\epsilon)^2 q^4 - 4(1-\epsilon) q^4 + q^4 \\ &\leq 100\epsilon^{1/9} q^4. \end{aligned}$$

Thus, by Lemma 2.5.2, $(X,Y)_G$ satisfies DISC$(q, q, 4\epsilon^{1/36})$. ■

The arguments here can be modified to show that a bound on $K_{1,2}$ implies one-sided counting for trees. We state the generalization and omit the proof.

Proposition 2.5.8. *Let H be a tree on vertices $\{1, 2, \ldots, m\}$. For every $\theta > 0$ there exists $\epsilon > 0$ of size polynomial in θ so that the following holds.*

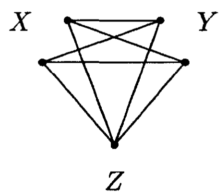
Let G be a weighted graph with vertex subsets $X_1, \ldots, X_m$. For every edge ab of H, assume there is some $q_{ab} \in [0,1]$ so that the bipartite graph $(X_a, X_b)_G$ satisfies $G(K_{1,1}) \geq (1-\epsilon) q_{ab}$, $G(K_{1,2}) \leq (1+\epsilon)^2 q_{ab}^2$ and $G(K_{2,1}) \leq (1+\epsilon)^2 q_{ab}^2$. Then $G(H) \geq (1-\theta) q(H)$.

2.5.2 $K_{1,2,2}$ implies inheritance

We now prove Propositions 2.1.13 and 2.5.1 using Lemma 2.5.3.

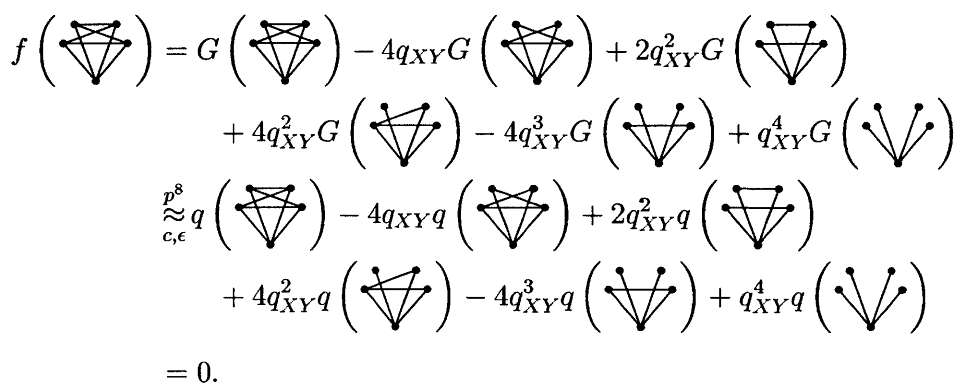
Proof of Proposition 2.1.13. First we show that only a small fraction of vertices in Z have very few neighbors in X and Y. Let Z_1 be the set of all vertices in Z with fewer than $(1-\xi)q_{XZ}|X|$ neighbors in X. Applying discrepancy to (X, Z_1) yields $\xi q_{XZ}|Z_1| \le \epsilon p|Z|$. If we assume that $\epsilon \le \frac{1}{3}\alpha\xi^2$, we have $|Z_1| \le \frac{\epsilon p}{\xi q_{XZ}}|Z| \le \frac{\xi}{3}|Z|$. Similarly let Z_2 be the set of all vertices in Z with fewer than $(1-\xi)q_{YZ}|Y|$ neighbors in Y, so that $|Z_2| \le \frac{\xi}{3}|Z|$ as well.

Define $f\colon V(G)\times V(G) \to \mathbb{R}$ to be a function which agrees with G on pairs (X,Z) and (Y,Z), and agrees with $G - q_{XY}$ on (X,Y). Let us assign each vertex of $K_{2,2,1}$ to one of $\{X,Y,Z\}$ as follows (two vertices are assigned to each of X and Y).

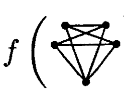


The stated jumbledness hypotheses suffice for counting $K_{1,2,2}$ and its subgraphs; we refer to the tutorial in Section 2.4.5 for an explanation.

By expanding all the $(G(x,y) - q_{XY})$ factors and using our counting lemma, we get

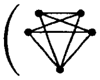


Therefore, by choosing ϵ and c to be sufficiently small (but polynomial in ξ, α, ϵ'), we can guarantee that

f  $\le \frac{1}{3}\xi(1-\xi)^4\epsilon'^4\alpha^4p^8.$

Let $K_{2,2}$ denote the subgraph of the above $K_{1,2,2}$ that gets mapped between X and Y.

For each $z \in Z$, let $f_z \colon X \times Y \to \mathbb{R}$ be defined by $(G(x,y) - q_{XY})1_{N_X(z)}(x)1_{N_Y(z)}(y)$. We have

$$f\left(\right.$$  $$\left. \right) = \int_{z \in Z} f_z(K_{2,2})\, dz.$$

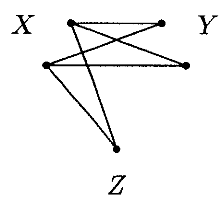
By Lemma 2.5.5, $f_z(K_{2,2}) \geq 0$ for all $z \in Z$. Let Z_3 be the set of vertices z in Z such that $f_z(K_{2,2}) > \epsilon'^4(1-\xi)^4\alpha^4p^8$. Then $|Z_3| \leq \frac{\xi}{3}|Z|$.

Let $Z' = Z \setminus (Z_1 \cup Z_2 \cup Z_3)$. So $|Z'| \geq (1-\xi)|Z|$. Furthermore, for any $z \in Z_1$,

$$f_z(K_{2,2}) \leq \epsilon'^4(1-\xi)^4\alpha^4p^8 \leq \epsilon'^4(1-\xi)^4p^4q_{XZ}^2q_{YZ}^2 \leq \epsilon'^4p^4\left(\frac{|N_X(z)|}{|X|}\right)^2\left(\frac{|N_Y(z)|}{|Y|}\right)^2.$$

It follows by Lemma 2.5.3 that $(N_X(z), N_Y(z))_G$ satisfies $\mathrm{DISC}(q_{XY}, p, \epsilon')$. ■

Proof of Proposition 2.5.1. The proof is essentially the same as the proof of Proposition 2.1.13 with the difference being that we now use the following graph. We omit the details.



■

2.6 One-sided counting

We are now in a position to prove Theorem 2.1.14, which we now recall.

Theorem 2.1.14 *For every fixed graph H on vertex set $\{1, 2, \ldots, m\}$ and every $\alpha, \theta > 0$, there exist constants $c > 0$ and $\epsilon > 0$ such that the following holds.*

Let Γ be a graph with vertex subsets $X_1, \ldots, X_m$ and suppose that the bipartite graph $(X_i, X_j)_\Gamma$ is $(p, cp^{d_2(H)+3}\sqrt{|X_i|\,|X_j|})$-jumbled for every $i < j$ with $ij \in E(H)$. Let G be a subgraph of Γ, with the vertex i of H assigned to the vertex subset X_i of G. For each edge ij of H, assume that $(X_i, X_j)_G$ satisfies $\mathrm{DISC}(q_{ij}, p, \epsilon)$, where $\alpha p \leq q_{ij} \leq p$. Then $G(H) \geq (1-\theta)q(H)$.

The idea is to embed vertices of H one at a time. At each step, the set of potential targets for each unembedded vertex shrinks, but we can choose our embedding so that it doesn't shrink too much and discrepancy is inherited.

Proof. Suppose that $v_1, v_2, \ldots, v_m$ is an ordering of the vertices of H which yields the 2-degeneracy $d_2(H)$ and that the vertex v_i is to be embedded in X_i. Let $L(j) = \{v_1, v_2, \ldots, v_j\}$. For $i > j$, let $N(i,j) = N(v_i) \cap L(j)$ be the set of neighbors v_h of v_i with $h \leq j$. Let $q(j) = \prod q_{ab}$, where the product is taken over all edges $v_a v_b$ of H with $1 \leq a < b \leq j$ and $q(i,j) = \prod_{v_h \in N(i,j)} q_{hi}$. Note that $q(j) = q(j, j-1)q(j-1)$.

We need to define several constants. To begin, we let $\theta_m = \theta$ and $\epsilon_m = 1$. Given θ_j and ϵ_j, we define $\xi_j = \frac{\theta_j}{6m^2}$ and $\theta_{j-1} = \frac{\theta_j}{2}$. We apply Propositions 2.1.13 and 2.5.1 with α, ξ_j and ϵ_j to find constants c_{j-1} and ϵ^*_{j-1} such that the conclusions of the two propositions hold. We let $\epsilon_{j-1} = \min(\epsilon^*_{j-1}, \frac{\alpha\theta_j^2}{72m})$, $c = \frac{1}{2}\alpha^{d_2(H)}c_0$ and $\epsilon = \epsilon_0$.

We will find many embeddings $f : V(H) \to V(G)$ by embedding the vertices of H one by one in increasing order. We will prove by induction on j that there are $(1-\theta_j)q(j)|X_1||X_2|\ldots|X_j|$ choices for $f(v_1), f(v_2), \ldots, f(v_j)$ such that the following conditions hold. Here, for each $i > j$, we let $T(i,j)$ be the set of vertices in X_i which are adjacent to $f(v_h)$ for every $v_h \in N(i,j)$. That is, it is the set of possible vertices into which, having embedded $v_1, v_2, \ldots, v_j$, we may embed v_i.

- For $1 \leq a < b \leq j$, $(f(v_a), f(v_b))$ is an edge of G if (v_a, v_b) is an edge of H;
- $|T(i,j)| \geq (1 - \frac{\theta_j}{6})q(i,j)|X_i|$ for every $i > j$;
- For each $i_1, i_2 > j$ with $v_{i_1}v_{i_2}$ an edge of H, the graph $(T(i_1,j), T(i_2,j))_G$ satisfies the discrepancy condition $\mathrm{DISC}(q_{ab}, p, \epsilon_j)$.

The base case $j = 0$ clearly holds by letting $T(i,0) = X_i$. We may therefore assume that there are $(1-\theta_{j-1})q(j-1)|X_1||X_2|\ldots|X_{j-1}|$ embeddings of $v_1, \ldots, v_{j-1}$ satisfying the conditions above. Let us fix such an embedding f. Our aim is to find a set $W(j) \subseteq T(j, j-1)$ with $|W(j)| \geq (1 - \frac{\theta_j}{2})q(j, j-1)|X_j|$ such that for every $w \in W(j)$ the following three conditions hold.

1. For each $i > j$ with $v_i \in N(v_j)$, there are at least $(1 - \frac{\theta_j}{6})q(i,j)|X_i|$ vertices in $T(i, j-1)$ which are adjacent to w;

2. For each $i_1, i_2 > j$ with $v_{i_1}v_{i_2}$, $v_{i_1}v_j$ and $v_{i_2}v_j$ edges of H, the induced subgraph of G between $N(w) \cap T(i_1, j)$ and $N(w) \cap T(i_2, j)$ satisfies the discrepancy condition $\text{DISC}(q_{ab}, p, \epsilon_j)$;

3. For each $i_1, i_2 > j$ with $v_{i_1}v_{i_2}$ and $v_{i_1}v_j$ edges of H and $v_{i_2}v_j$ not an edge of H, the induced subgraph of G between $N(w) \cap T(i_1, j)$ and $T(i_2, j-1)$ satisfies the discrepancy condition $\text{DISC}(q_{ab}, p, \epsilon_j)$.

Note that once we have found such a set, we may take $f(v_j) = w$ for any $w \in W(j)$. By using the induction hypothesis to count the number of embeddings of the first $j-1$ vertices, we see that there are at least

$$\left(1 - \frac{\theta_j}{2}\right) q(j, j-1)|X_j|(1 - \theta_{j-1})q(j-1)|X_1||X_2|\ldots|X_{j-1}| \\ \geq (1 - \theta_j)q(j)|X_1||X_2|\ldots|X_j|$$

ways of embedding $v_1, v_2, \ldots, v_j$ satisfying the necessary conditions. Here we used that $q(j) = q(j, j-1)q(j-1)$ and $\theta_{j-1} = \frac{\theta_j}{2}$. The induction therefore follows by letting $T(i, j) = N(w) \cap T(i, j-1)$ for all $i > j$ with $v_i \in N(v_j)$ and $T(i, j) = T(i, j-1)$ otherwise.

It remains to show that there is a large subset $W(j)$ of $T(j, j-1)$ satisfying the required conditions. For each $i > j$, let $A_i(j)$ be the set of vertices in $T(j, j-1)$ for which $|N(w) \cap T(i, j-1)| \leq (1 - \frac{\theta_j}{12})q_{ij}|T(i, j-1)|$. Then, since the graph between $T(i, j-1)$ and $T(j, j-1)$ satisfies $\text{DISC}(q_{ji}, p, \epsilon_{j-1})$, we have that $\epsilon_{j-1}p|T(j, j-1)| \geq \frac{\theta_j}{12}q_{ij}|A_i(j)|$. Hence, since $q_{ij} \geq \alpha p$,

$$|A_i(j)| \leq \frac{12\epsilon_{j-1}}{\alpha\theta_j}|T(j, j-1)|.$$

Note that for any $w \in T(j, j-1) \setminus A_i(j)$,

$$|N(w) \cap T(i, j-1)| \geq \left(1 - \frac{\theta_j}{12}\right) q_{ij} |T(i, j-1)| \geq \left(1 - \frac{\theta_j}{6}\right) q(i, j) |X_i|.$$

For each $i_1, i_2 > j$ with $v_{i_1}v_{i_2}$, $v_{i_1}v_j$ and $v_{i_2}v_j$ edges of H, let $B_{i_1,i_2}(j)$ be the set of vertices w in $T(j, j-1)$ for which the graph between $N(w) \cap T(i_1, j-1)$ and $N(w) \cap T(i_2, j-1)$ does not satisfy $\mathrm{DISC}(q_{i_1 i_2}, p, \epsilon_j)$. Note that

$$\begin{aligned}|T(i_1, j-1)||T(i_2, j-1)| &\geq \left(1 - \frac{\theta_{j-1}}{6}\right)^2 q(i_1, j-1) q(i_2, j-1) |X_{i_1}||X_{i_2}| \\ &\geq \frac{\alpha^{2d_2(H)-2}}{2} p^{2d_2(H)-2} |X_{i_1}||X_{i_2}|,\end{aligned}$$

where we get $2d_2(H) - 2$ because j is a neighbor of both i_1 and i_2 with $j < i_1, i_2$. Similarly, $|T(i_1, j-1)||T(j, j-1)|$ and $|T(i_2, j-1)||T(j, j-1)|$ are at least $\frac{\alpha^{2d_2(H)}}{2} p^{2d_2(H)} |X_{i_1}||X_j|$ and $\frac{\alpha^{2d_2(H)}}{2} p^{2d_2(H)} |X_{i_2}||X_j|$, respectively.

Since

$$\begin{aligned}cp^{d_2(H)+3}\sqrt{|X_{i_1}||X_{i_2}|} &\leq \frac{1}{2}\alpha^{d_2(H)} c_0 p^4 \sqrt{p^{2d_2(H)-2}|X_{i_1}||X_{i_2}|} \\ &\leq c_0 p^4 \sqrt{|T(i_1, j-1)||T(i_2, j-1)|},\end{aligned}$$

the induced subgraph of Γ between $T(i_1, j-1)$ and $T(i_2, j-1)$ is

$$(p, c_0 p^4 \sqrt{|T(i_1, j-1)||T(i_2, j-1)|})\text{-jumbled.}$$

Similarly, the induced subgraph of Γ between the sets $T(j, j-1)$ and $T(i_1, j-1)$ is

$$(p, c_0 p^3 \sqrt{|T(j, j-1)||T(i_1, j-1)|})\text{-jumbled}$$

and the induced subgraph between $T(j, j-1)$ and $T(i_2, j-1)$ is

$$(p, c_0 p^3 \sqrt{|T(j, j-1)||T(i_2, j-1)|})\text{-jumbled.}$$

By our choice of ϵ_{j-1}, we may therefore apply Proposition 2.1.13 to show that $|B_{i_1,i_2}(j)| \leq \xi_j |T(j, j-1)|$.

For each $i_1, i_2 > j$ with $v_{i_1}v_{i_2}$ and $v_{i_1}v_j$ edges of H and $v_{i_2}v_j$ not an edge of H, let $C_{i_1,i_2}(j)$ be the set of vertices w in $T(j, j-1)$ for which the graph between $N(w) \cap T(i_1, j-1)$ and $T(i_2, j-1)$ does not satisfy DISC$(q_{i_1 i_2}, p, \epsilon_j)$. As with $B_{i_1,i_2}(j)$, we may apply Proposition 2.5.1 to conclude that $|C_{i_1,i_2}(j)| \leq \xi_j |T(j, j-1)|$.

Counting over all possible bad events and using that $|T(j, j-1)| \geq (1 - \frac{\theta_{j-1}}{6}) q(j, j-1)|X_j|$, we see that the set $W(j)$ of good vertices has size at least $(1-\sigma)q(j, j-1)|X_j|$, where

$$\sigma \leq \frac{\theta_{j-1}}{6} + \frac{12m\epsilon_{j-1}}{\alpha\theta_j} + 2\binom{m}{2}\xi_j \leq \frac{\theta_j}{12} + \frac{\theta_j}{6} + \frac{\theta_j}{6} \leq \frac{\theta_j}{2},$$

as required. Here we used $\theta_j = \frac{\theta_{j-1}}{2}$, $\epsilon_{j-1} \leq \frac{\alpha\theta_j^2}{72m}$ and $\xi_j = \frac{\theta_j}{6m^2}$. This completes the proof. ■

Note that for the clique K_t, we have $d_2(K_t) + 3 = t + 1$. In this case, it is better to use the bound coming from two-sided counting, which gives the exponent t.

Another case of interest is when the graph H is triangle-free. Here it is sufficient to always apply the simpler inheritance theorem, Proposition 2.5.1, to maintain discrepancy. Then, since

$$p^{d_2(H)+2}\sqrt{|X_{i_1}||X_{i_2}|} = p^{\frac{5}{2}}\sqrt{p^{2d_2(H)-1}|X_{i_1}||X_{i_2}|},$$

we see that an exponent of $d_2(H) + 2$ is sufficient in this case. In particular, for $H = K_{s,t}$ with $s \leq t$, we get an exponent of $d_2(K_{s,t}) + 2 = \frac{s-1}{2} + 2 = \frac{s+3}{2}$, as quoted in Table 2.1.

It is also worth noting that a one-sided counting lemma for Γ holds under the slightly weaker assumption that $\beta \leq cp^{d_2(H)+1}n$. We omit the details since the proof is a simpler version of the previous one, without the necessity for tracking inheritance of discrepancy.

Proposition 2.6.1. *For every fixed graph H on vertex set $\{1, 2, \ldots, m\}$ and every $\theta > 0$ there exists a constant $c > 0$ such that the following holds.*

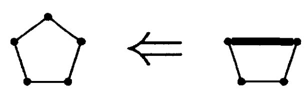
Let Γ be a graph with vertex subsets $X_1, \ldots, X_m$ where vertex i of H is assigned to the vertex subset X_i of Γ and suppose that the bipartite graph $(X_i, X_j)_\Gamma$ is $(p, cp^{d_2(H)+1}\sqrt{|X_i|\,|X_j|})$-jumbled for every $i < j$ with $ij \in E(H)$. Then $\Gamma(H) \geq (1-\theta)p(H)$.

2.7 Counting cycles

Using the tools of doubling and densification, we already know how to count all cycles. For cycles of length 4 or greater, (p, cp^2n)-jumbledness suffices.

Proposition 2.7.1. *Assume Setup 2.4.1 with $H = C_\ell$ and $k \geq 3$ if $\ell = 3$ or $k \geq 2$ if $\ell \geq 4$. Then $G(C_\ell) \approx_{c,\epsilon}^{p(C_\ell)} q(C_\ell)$.*

Proof. When $\ell = 4$, see Proposition 2.4.13. When $\ell = 3$, see Section 2.2 for the doubling procedure. For $\ell \geq 5$, we can perform densification to reduce the problem to counting $C_{\ell-1}$ with at least one dense edge, so we proceed by induction. ■



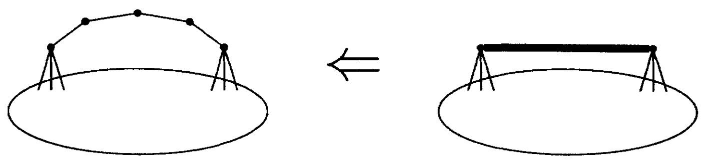
The goal of this section is to prove a one-sided counting lemma for cycles that requires much weaker jumbledness.

Proposition 2.7.2. *Assume Setup 2.4.1 with $H = C_\ell$, where $\ell \geq 5$, and all edges sparse. Let $k \geq 1+\frac{1}{\ell-3}$ if ℓ is odd and $1+\frac{1}{\ell-4}$ if ℓ is even. Then $G(C_\ell) \geq q(C_\ell) - \theta p(C_\ell)$ with $\theta \leq 100(\epsilon^{1/(2\ell)} + \ell c^{2/3})$.*

The strategy is via subdivision densification, as outlined in Section 2.2.

2.7.1 Subdivision densification

In Section 2.4.2 we showed how to reduce a counting problem by transforming a singly subdivided edge of H into a dense edge. In this section, we show how to transform a multiply subdivided edge of H into a dense edge, using much weaker hypotheses on jumbledness, at least for one-sided counting. The idea is that a long subdivision allows more room for mixing, and thus requires less jumbledness at each step.



We introduce a weaker variant of discrepancy for one-sided counting.

Definition 2.7.3. Let G be a graph with vertex subsets X and Y. We say that $(X,Y)_G$ satisfies $\mathrm{DISC}_{\geq}(q,p,\epsilon)$ if

$$\int_{\substack{x\in X\\ y\in Y}} (G(x,y)-q)u(x)v(y)\ dxdy \geq -\epsilon p \tag{2.15}$$

for all functions $u\colon X\to[0,1]$ and all $v\colon Y\to[0,1]$.

In a graph H, we say that $a_0a_1a_2\cdots a_m$ is a *subdivided edge* if the neighborhood of a_i in H is $\{a_{i-1},a_{i+1}\}$ for $1\leq i\leq m-1$. Say that it is sparse if every edge a_ia_{i+1}, $0\leq i\leq m-1$, is sparse.

For a graph Γ or G with vertex subsets $X_0,X_1,\ldots,X_m$, $x_0\in X_0$, $x_m\in X_m$ and $X_i'\subseteq X_i$, we write

$$G(x_0,X_1',X_2',\ldots,X_m') = \int_{\substack{x_1\in X_1\\ \cdots\\ x_m\in X_m}} G(x_0,x_1)1_{X_1'}(x_1)\cdots G(x_{m-1},x_m)1_{X_m'}(x_m)\ dx_1dx_2\cdots dx_m,$$

and

$$G(x_0,X_1',X_2',\ldots,x_m) = \int_{\substack{x_1\in X_1\\ \cdots\\ x_{m-1}\in X_{m-1}}} G(x_0,x_1)1_{X_1'}(x_1)\cdots G(x_{m-1},x_m)\ dx_1dx_2\cdots dx_{m-1}.$$

These quantities can be interpreted probabilistically. The first expression is the probability that a randomly chosen sequence of vertices with one endpoint fixed is a path in G with the vertices landing in the chosen subsets. For the second expression, both endpoints are fixed.

Lemma 2.7.4 (Subdivision densification). *Assume Setup 2.4.1. Let $\ell \geq 2$ and let $a_0 a_1 \cdots a_\ell$ be a sparse subdivided edge and assume that $a_0 a_\ell$ is not an edge of H. Assume $k \geq 1 + \frac{1}{2\ell - 2}$. Replace the induced bipartite graph (X_{a_0}, X_{a_ℓ}) by the weighted bipartite graph given by*

$$G(x_0, x_\ell) = \frac{1}{4p^\ell} \min \left\{ G(x_0, X_1, X_2, \ldots, X_{\ell-1}, x_\ell), 4p^\ell \right\}$$

for $(x_0, x_\ell) \in X_0 \times X_\ell$. Let H' be H with the path $a_0 a_1 \cdots a_\ell$ deleted and the edge $a_0 a_\ell$ added. Let $p_{a_0 a_\ell} = 1$ and $q_{a_0 a_\ell} = \frac{1}{4p^\ell} q_{a_0 a_1} q_{a_1 a_2} \cdots q_{a_{\ell-1} a_\ell}$. Then $G(H) \geq 4p^\ell G(H')$ and $(X_0, X_\ell)_G$ satisfies $\mathrm{DISC}_{\geq}(q_{a_0 a_\ell}, 1, 18(\epsilon^{1/(2\ell)} + \ell c^{2/3}))$.

Remark. If there is at least one dense edge in the subdivision, then using arguments similar to the ones in Section 2.4.2, modified for one-sided counting, we can show that $k \geq 1$ suffices for subdivision densification.

The idea of the proof is very similar to densification in Section 2.4.2. The claim $G(H) \geq 4p^\ell G(H')$ follows easily from the new edge weights. It remains to show that $(X_0, X_\ell)_G$ satisfies $\mathrm{DISC}_{\geq}$. So Lemma 2.7.4 follows from the next result.

Lemma 2.7.5. *Let $m \geq 2$, $c, \epsilon, p \in (0, 1]$, and $q_1, q_2, \ldots, q_m \in [0, p]$. Let Γ be any weighted graph with vertex subsets $X_0, X_1, \ldots, X_m$ and let G be a subgraph of Γ. Suppose that, for each $i = 1, \ldots, m$, $(X_{i-1}, X_i)_\Gamma$ is $(p, cp^{1+\frac{1}{2m-2}} \sqrt{|X_{i-1}| |X_i|})$-jumbled and $(X_{i-1}, X_i)_G$ satisfies $\mathrm{DISC}_{\geq}(q_i, p, \epsilon)$. Then the weighted graph G' on (X_0, X_m) defined by*

$$G'(x_0, x_m) = \min \left\{ G(x_0, X_1, X_2, \ldots, X_{m-1}, x_m), 4p^m \right\}$$

satisfies $\mathrm{DISC}_{\geq}(q_1 q_2 \cdots q_m, p^m, 72(\epsilon^{1/(2m)} + mc^{2/3}))$.

Here are the steps for the proof of Lemma 2.7.5.

1. Show that the graph on $X_0 \times X_m$ with weights $G(x_0, X_1, X_2, \ldots, X_{m-1}, x_m)$ satisfies $\mathrm{DISC}_{\geq}$.

2. Under the assumption that every vertex X_i has roughly the same number of neighbors in X_{i+1} for every i, show that capping of the edge weights has negligible effect on discrepancy.

3. Show that we can delete a small subset from each vertex subset X_i so that the assumption in step 2 is satisfied.

Step 2 is the most difficult. Since we are only proving lower bound discrepancy, it is okay to delete vertices in step 3. This is also the reason why this proof, without significant modification, cannot prove two-sided discrepancy (which may require stronger hypotheses), as we may have deleted too many edges in the process. Also, unlike the densification in Section 2.4.2, we do not have to worry about the effect of the edge weight capping on the overall H-count, as we are content with a lower bound.

The next two lemmas form step 1 of the program.

Lemma 2.7.6. *Let G be a weighted graph with vertex subsets X, Y, Z. Let $p_1, p_2, \epsilon \in (0, 1]$ and $q_1 \in [0, p_1]$, $q_2 \in [0, p_2]$. If $(X, Y)_G$ satisfies $\mathrm{DISC}_{\geq}(q_1, p_1, \epsilon)$ and $(Y, Z)_G$ satisfies $\mathrm{DISC}_{\geq}(q_2, p_2, \epsilon)$, then the induced weighted bipartite graph G' on (X, Z) whose weight is given by*

$$G'(x, z) = G(x, Y, z)$$

satisfies $\mathrm{DISC}_{\geq}(q_1 q_2, p_1 p_2, 6\sqrt{\epsilon})$.

Note that no jumbledness hypothesis is needed for the lemma.

Proof. Let $u \colon X \to [0, 1]$ and $w \colon Z \to [0, 1]$ be arbitrary functions. Let

$$Y' = \left\{ y \in Y \;\middle|\; \int_{x \in X} (G(x, y) - q_1) u(x) \; dx \leq -\sqrt{\epsilon} p_1 \right\}.$$

Then applying (2.15) to u and $1_{Y'}$ yields $|Y'| \leq \sqrt{\epsilon} \, |Y|$. Similarly, let

$$Y'' = \left\{ y \in Y \;\middle|\; \int_{z \in Z} (G(y, z) - q_2) w(z) \; dx \leq -\sqrt{\epsilon} p_2 \right\}.$$

Then $|Y''| \le \sqrt{\epsilon}\,|Y|$ as well. So

$$\begin{aligned}
&\int_{\substack{x\in X\\ z\in Z}} G'(x,z)u(x)w(z)\ dxdz \\
&= \int_{\substack{x\in X\\ y\in Y\\ z\in Z}} u(x)G(x,y)G(y,z)w(z)\ dxdydz \\
&\ge \int_{y\in Y\setminus(Y'\cup Y'')} \left(\int_{x\in X} G(x,y)u(x)\ dx\right)\left(\int_{z\in Z} G(y,z)w(z)\ dz\right)\ dy \\
&\ge \int_{y\in Y\setminus(Y'\cup Y'')} (q_1\mathbb{E}u - \sqrt{\epsilon}p_1)(q_2\mathbb{E}w - \sqrt{\epsilon}p_2)\ dy \\
&\ge (1-2\sqrt{\epsilon})(q_1\mathbb{E}u - \sqrt{\epsilon}p_1)(q_2\mathbb{E}w - \sqrt{\epsilon}p_2) \\
&\ge q_1q_2\mathbb{E}u\mathbb{E}w - 6\sqrt{\epsilon}p_1p_2.
\end{aligned}$$

■

The above proof can be extended to prove a one-sided counting lemma for trees without any jumbledness hypotheses. We omit the details.

Proposition 2.7.7. *Let H be a tree on vertices $\{1,2,\dots,m\}$. For every $\theta > 0$, there exists $\epsilon > 0$ of size at least polynomial in θ such that the following holds.*

Let G be a weighted graph with vertex subsets $X_1,\dots,X_m$. For each edge ab of H, suppose that $(X_a,X_b)_G$ satisfies $\mathrm{DISC}_{\ge}(q_{ab},p_{ab},\epsilon)$ for some $0 \le q_{ab} \le p_{ab} \le 1$. Then $G(H) \ge q(H) - \theta p(H)$.

By Lemma 2.7.6 and induction, we obtain the following lemma about counting paths in G.

Lemma 2.7.8. *Let G be a weighted graph with vertex subsets $X_0, X_1, \dots, X_m$. Let $0 < \epsilon < 1$. Suppose that for each $i = 1,2,\dots,m$, $(X_{i-1},X_i)_G$ satisfies $\mathrm{DISC}_{\ge}(q_i,p_i,\epsilon)$ for some numbers $0 \le q_i \le p_i \le 1$. Then the induced weighted bipartite graph G' on $X_0 \times X_m$ whose edge weights are given by*

$$G'(x_0,x_m) = G(x_0,X_1,X_2,\dots,X_{m-1},x_m)$$

satisfies $\mathrm{DISC}_{\ge}(q_1q_2\cdots q_m, p_1p_2\cdots p_m, 36\epsilon^{1/(2m)})$.

Proof. Applying Lemma 2.7.6, we see that the auxiliary weighted graphs on (X_0, X_2), (X_2, X_4), ... satisfy $\mathrm{DISC}_{\geq}(q_1q_2, p_1p_2, 36\epsilon^{1/2})$, etc. Next we apply Lemma 2.7.6 again, and we deduce that the auxiliary weighted graph on $(X_0, X_4), (X_4, X_8)$ must satisfy $\mathrm{DISC}_{\geq}(q_1q_2q_3q_4, p_1p_2p_3p_4, 36\epsilon^{1/4})$, etc. Continuing, we find that $(X_0, X_m)_{G'}$ satisfies $\mathrm{DISC}_{\geq}(q_1q_2\cdots q_m, p_1p_2\cdots p_m, \epsilon')$ with $\epsilon' = 36\epsilon^{2^{-(\log_2 m+1)}} = 36\epsilon^{1/(2m)}$. ∎

For step 2 of the proof, we need to assume some degree-regularity between the parts. We note that the order of X and Y is important in the following definition.

Definition 2.7.9. Let Γ be a weighted graph with vertex subsets X and Y. We say that $(X, Y)_\Gamma$ is *(p, ξ, η)-bounded* if $|\Gamma(x, Y) - p| \leq \xi p$ for all $x \in X$ and $\Gamma(x, y) \leq \eta$ for all $x \in X$ and $y \in Y$.

Here is the idea of the proof. Fix a vertex $x_0 \in X_0$, and consider its successive neighborhoods in $X_1, X_2, \ldots$. Let us keep track of the number of paths from x_0 to each endpoint. We expect the number of paths to be somewhat evenly distributed among vertices in the successive neighborhoods and, therefore, we do not expect many vertices in X_i to have disproportionately many paths to x_0. In particular, capping the weights of $\Gamma(x_0, X_1, \ldots, X_{m-1}, x_m)$ has a negligible effect.

Here is a back-of-the-envelope calculation. Suppose every pair $(X_i, X_{i+1})_\Gamma$ is $(p, \gamma\sqrt{|X_i|\,|X_{i+1}|})$-jumbled. First we remove a small fraction of vertices from each vertex subset X_i so that in the remaining graph Γ is bounded, i.e., every vertex has roughly the expected number of neighbors in the next vertex subset. Let $S \subseteq X_i$, and let $N(S)$ be its neighborhood in X_{i+1}. Then the number of edges $e(S, N(S))$ between S and $N(S)$ is roughly $p\,|S|\,|X_{i+1}|$ by the degree assumptions on X_i. On the other hand, by jumbledness, $e(S, N_{i+1}(S)) \leq \gamma\sqrt{|X_i|\,|X_{i+1}|\,|S|\,|N(S)|} + p\,|S|\,|N(S)|$. When S is small, the first term dominates, and by comparing the two estimates we get that $\frac{|N(S)|}{|X_{i+1}|}$ is at least roughly $p^2\gamma^{-2}\frac{|S|}{|X_i|}$. Now fix a vertex $x_0 \in X_0$. It has about $p\,|X_1|$ neighbors in X_1. At each step, the fraction of X_i occupied by the successive neighborhood of x_0 expands by a factor of about $p^2\gamma^{-2}$, until the successive neighborhood saturates some X_i. Note that for $\gamma = cp^{1+\frac{1}{2m-2}}$, we have $p(p^2\gamma^{-2})^{m-1} \gg 1$, so the successive neighborhood of x_0 in X_m is essentially all of X_m. So we can expect the resulting weighted graph to be dense.

We will use induction. We show that from a fixed $x_0 \in X_0$, if we can bound the number of paths to each vertex in X_i, then we can do so for X_{i+1} as well.

The next result is the key technical lemma. It is an induction step for the lemma that follows. One should think of X, Y and Z as X_0, X_i and X_{i+1}, respectively.

Lemma 2.7.10. *Let $p_1, p_2, \xi_1, \xi_2, \xi_3 \in (0, 1]$, and $\eta_1, \gamma_2 > 0$. Let Γ be a weighted graph with vertex subsets X, Y, Z. Assume that $(X, Y)_\Gamma$ is (p_1, ξ_1, η_1)-bounded and $(Y, Z)_\Gamma$ is $(p_2, \xi_2, 1)$-bounded and $(p_2, \gamma_2\sqrt{|Y|\,|Z|})$-jumbled. Let $\eta' = \max\left\{4\gamma_2^2 p_2^{-1}\xi_3^{-1}\eta_1, 4p_1p_2\right\}$ and $\xi' = \xi_1 + 2\xi_2 + 2\xi_3$. Then the weighted graph Γ' on (X, Z) given by*

$$\Gamma'(x, z) = \min\left\{\Gamma(x, Y, z), \eta'\right\}$$

is (p_1p_2, ξ', η')-bounded.

Proof. We have $\Gamma'(x, z) \leq \eta'$ for all $x \in X, z \in Z$. Also, by the boundedness assumptions, we have $\Gamma'(x, Z) \leq \Gamma(x, Y, Z) \leq (1 + \xi_1)(1 + \xi_2)p_1p_2 \leq (1 + \xi')p_1p_2$. It only remains to prove that $\Gamma'(x, Z) \geq (1 - \xi')p_1p_2$ for all $x \in X$.

Fix any $x \in X$. Let

$$Z'_x = \left\{z \in Z \mid \Gamma(x, Y, z) > \eta'\right\}.$$

Note that $\Gamma'(x, Z) \geq \Gamma(x, Y, Z) - \Gamma(x, Y, Z'_x)$, so we would like to find an upper bound for $\Gamma(x, Y, Z'_x)$.

Apply the jumbledness criterion (2.3) to $(Y, Z)_\Gamma$ with the functions $u(y) = \Gamma(x, y)\eta_1^{-1}$ and $v(z) = 1_{Z'_x}$. Note that $0 \leq u \leq 1$ due to boundedness. We have

$$\int_{\substack{y \in Y \\ z \in Z}} \Gamma(x, y)\eta_1^{-1}(\Gamma(y, z) - p_2)1_{Z'_x}(z)\; dydz$$
$$\leq \gamma_2\sqrt{\Gamma(x, Y)\eta_1^{-1}\frac{|Z'_x|}{|Z|}} \leq \gamma_2\sqrt{(1 + \xi_1)p_1\eta_1^{-1}\frac{|Z'_x|}{|Z|}}.$$

The integral equals $\eta_1^{-1}\left(\Gamma(x,Y,Z'_x) - p_2\Gamma(x,Y)\frac{|Z'_x|}{|Z|}\right)$, so we have

$$\Gamma(x,Y,Z'_x) - p_2\Gamma(x,Y)\frac{|Z'_x|}{|Z|} \leq \gamma_2\sqrt{(1+\xi_1)p_1\eta_1\frac{|Z'_x|}{|Z|}}. \tag{2.16}$$

On the other hand, we have

$$\Gamma(x,Y,Z'_x) - p_2\Gamma(x,Y)\frac{|Z'_x|}{|Z|} \geq \eta'\frac{|Z'_x|}{|Z|} - (1+\xi_1)p_1p_2\frac{|Z'_x|}{|Z|} \geq \frac{\eta'}{2}\frac{|Z'_x|}{|Z|}. \tag{2.17}$$

Combining (2.17) with (2.16), we get

$$\frac{|Z'_x|}{|Z|} \leq \frac{4\gamma_2^2(1+\xi_1)p_1\eta_1}{\eta'^2}. \tag{2.18}$$

Substituting (2.18) back into (2.16), we have

$$\begin{aligned}
\Gamma(x,Y,Z'_x) &\leq (1+\xi_1)p_1p_2\frac{|Z'_x|}{|Z|} + \gamma_2\sqrt{(1+\xi_1)p_1\eta_1\frac{|Z'_x|}{|Z|}} \\
&\leq \frac{4\gamma_2^2(1+\xi_1)^2p_1^2p_2\eta_1}{\eta'^2} + \frac{2\gamma_2^2(1+\xi_1)p_1\eta_1}{\eta'} \\
&\leq \frac{4\gamma_2^2(1+\xi_1)^2p_1^2p_2\eta_1}{(4\gamma_2^2p_2^{-1}\xi_3^{-1}\eta_1)(4p_1p_2)} + \frac{2\gamma_2^2(1+\xi_1)p_1\eta_1}{4\gamma_2^2p_2^{-1}\xi_3^{-1}\eta_1} \\
&= \frac{1}{4}(1+\xi_1)^2\xi_3p_1p_2 + \frac{1}{2}(1+\xi_1)\xi_3p_1p_2 \\
&\leq 2\xi_3p_1p_2.
\end{aligned}$$

Therefore,

$$\Gamma'(x,Z) \geq \Gamma(x,Y,Z) - \Gamma(x,Y,Z'_x) \geq (1-\xi_1)(1-\xi_2)p_1p_2 - 2\xi_3p_1p_2 \geq (1-\xi')p_1p_2.$$

This completes the proof that Γ' is (p_1p_2, ξ', η')-bounded. ■

By repeated applications of Lemma 2.7.10, we obtain the following lemma for embedding paths in Γ.

Lemma 2.7.11. *Let $0 < 4c^2 < \xi < \frac{1}{4m}$ and $0 < p \leq 1$. Let Γ be a graph with* *vertex subsets $X_0, X_1, \ldots, X_m$. Suppose that, for each $i = 1, \ldots, m$, $(X_{i-1}, X_i)_\Gamma$ is $(p, \xi, 1)$-bounded and $(p, cp^{1+\frac{1}{2m-2}}\sqrt{|X_{i-1}|\,|X_i|})$-jumbled. Then the weighted bipartite graph Γ' on (X_0, X_m) defined by*

$$\Gamma'(x_0, x_m) = \min\{\Gamma(x_0, X_1, X_2, \ldots, X_{m-1}, x_m), 4p^m\}$$

is $(p^m, 4m\xi, 4p^m)$-bounded.

Proof. Since $\Gamma'(x_0, X_m) \le \Gamma(x_0, X_1, X_2, \ldots, X_m) \le (1+\xi)^m p^m \le e^{m\xi} p^m \le (1 + 4m\xi)p^m$ for all $x_0 \in X_0$, it remains to show that $\Gamma'(x_0, X_m) \ge (1 - 4m\xi)p^m$ for all $x_0 \in X_0$.

For every $i = 1, \ldots, m$, define a weighted graph $\Gamma^{(i)}$ on vertex sets X_0, X_i, X_{i+1} (with $\Gamma^{(m)}$ only defined on X_0 and X_m) as follows. Set $(X_i, X_{i+1})_{\Gamma^{(i)}} = (X_i, X_{i+1})_\Gamma$ for each $1 \le i \le m-1$. Set $(X_0, X_1)_{\Gamma^{(1)}} = (X_0, X_1)_\Gamma$ and

$$\Gamma^{(i+1)}(x_0, x_{i+1}) = \min\left\{\Gamma^{(i)}(x_0, X_i, x_{i+1}), \eta_{i+1}\right\}$$

for each $1 \le i \le m-1$, where

$$\eta_i = \max\left\{(4c^2\xi^{-1})^{i-1} p^{(i-1)\left(1+\frac{1}{m-1}\right)}, 4p^i\right\}$$

for every i. So $\Gamma^{(i)}(x_0, x_i) \le \Gamma(x_0, X_1, \ldots, X_{i-1}, x_i)$ for every i and every $x_0 \in X_0, x_i \in X_i$. Let $\gamma = cp^{1+\frac{1}{2m-2}}$. Note that $\eta_{i+1} = \max\{4\gamma^2 p^{-1}\xi^{-1}\eta_i, 4p^{i+1}\}$ for every i. So it follows by Lemma 2.7.10 and induction that $(X_0, X_i)_{\Gamma^{(i)}}$ is $(p^i, 4i\xi, \eta_i)$-bounded for every i. Since $\eta_m = 4p^m$, $\Gamma'(x_0, X_m) \ge \Gamma^{(m)}(x_0, X_m) \ge (1 - 4m\xi)p^m$, as desired. ■

To complete step 2 of the proof, we show that the boundedness assumptions imply that the edge weight capping has negligible effect on discrepancy.

Lemma 2.7.12. *Let $0 < 4c^2 < \xi$ and $0 < p \le 1$. Let Γ be a graph with vertex subsets $X_0, X_1, \ldots, X_m$ and let G be a subgraph of Γ. Suppose that, for each $i = 1, \ldots, m$, $(X_{i-1}, X_i)_\Gamma$ is $(p, \xi, 1)$-bounded and $(p, cp^{1+\frac{1}{2m-2}}\sqrt{|X_{i-1}|\,|X_i|})$-jumbled and $(X_{i-1}, X_i)_G$*

satisfies $\mathrm{DISC}_{\geq}(q_i, p_i, \epsilon)$. *Then the weighted graph* G' *on* (X_0, X_m) *defined by*

$$G'(x_0, x_m) = \min\{G(x_0, X_1, X_2, \ldots, X_{m-1}, x_m), 4p^m\}$$

satisfies $\mathrm{DISC}_{\geq}(q_1 q_2 \cdots q_m, p^m, 36\epsilon^{1/(2m)} + 8m\xi)$.

Proof. We may assume that $\xi < \frac{1}{4m}$ since otherwise the claim is trivial as every graph satisfies $\mathrm{DISC}_{\geq}(q, p, \epsilon)$ when $\epsilon \geq 1$. Let Γ' be constructed as in Lemma 2.7.11. To simplify notation, let us write

$$\begin{aligned} G(x_0, x_m) &= G(x_0, X_1, \cdots, X_{m-1}, x_m) \\ \text{and} \quad \Gamma(x_0, x_m) &= \Gamma(x_0, X_1, \cdots, X_{m-1}, x_m) \end{aligned}$$

for $x_0 \in X_0$, $x_m \in X_m$. We have

$$\begin{aligned} G(x_0, x_m) - G'(x_0, x_m) &= \max\{0, G(x_0, x_m) - 4p^m\} \\ &\leq \max\{0, \Gamma(x_0, x_m) - 4p^m\} = \Gamma(x_0, x_m) - \Gamma'(x_0, x_m). \end{aligned}$$

Let $q = q_1 q_2 \cdots q_m$. For any functions $u\colon X \to [0,1]$ and $v\colon Y \to [0,1]$, we have

$$\begin{aligned} &\int_{\substack{x_0 \in X_0 \\ x_m \in X_m}} (G'(x_0, x_m) - q) u(x_0) v(x_m) \, dx_0 dx_m \\ &\geq \int_{\substack{x_0 \in X_0 \\ x_m \in X_m}} (G(x_0, x_m) - q) u(x_0) v(x_m) \, dx_0 dx_m \\ &\qquad - \int_{\substack{x_0 \in X_0 \\ x_m \in X_m}} (\Gamma(x_0, x_m) - \Gamma'(x_0, x_m)) u(x_0) v(x_m) \, dx_0 dx_m. \end{aligned}$$

The first term is at least $-36\epsilon^{1/(2m)} p^m$ by Lemma 2.7.8. For the second term, we use the boundedness of Γ and Γ' to get

$$\begin{aligned}
&\int_{\substack{x_0\in X_0\\ x_m\in X_m}} (\Gamma(x_0,x_m)-\Gamma'(x_0,x_m))u(x_0)v(x_m)\ dx_0dx_m\\
&\le \int_{\substack{x_0\in X_0\\ x_m\in X_m}} (\Gamma(x_0,x_m)-\Gamma'(x_0,x_m))\ dx_0dx_m\\
&\le (1+\xi)^m p^m-(1-4\xi m)p^m\\
&\le 8\xi m p^m.
\end{aligned}$$

It follows that G' satisfies $\mathrm{DISC}_{\geq}(q, p^m, 36\epsilon^{1/(2m)} + 8m\xi)$. ■

This completes step 2 of the program. Finally, we need to show that we have a large subgraph of Γ satisfying boundedness, so that we can apply Lemma 2.7.12 and then transfer the results back to the original graph.

Lemma 2.7.13. *Let $0 < \delta,\gamma,\xi,p < 1$ satisfy $2\gamma^2 \leq \delta\xi^2p^2$. Let Γ be a graph with vertex subsets $X_0, X_1, \ldots, X_m$ and suppose that, for each $i = 1,\ldots,m$, $(X_{i-1}, X_i)_\Gamma$ is $(p,(1-\delta)\gamma\sqrt{|X_{i-1}|\,|X_i|})$-jumbled. Then we can find $\widetilde{X}_i \subseteq X_i$ with $\left|\widetilde{X}_i\right| \geq (1-\delta)\,|X_i|$ for every i such that, for every $0 \leq i \leq m-1$, the induced bipartite graph $(\widetilde{X}_i, \widetilde{X}_{i+1})_\Gamma$ is $(p,\xi,1)$-bounded and $(p, \gamma\sqrt{\left|\widetilde{X}_i\right|\left|\widetilde{X}_{i+1}\right|})$-jumbled.*

Proof. The jumbledness condition follows directly from the size of $|X_i|$, so it suffices to make the bipartite graphs bounded. Let $\widetilde{X}_m = X_m$. For each $i = m-1, m-2, \ldots, 0$, in this order, set $\widetilde{X}_i$ to be the vertices in X_i with $(1\pm\xi)p\left|\widetilde{X}_{i+1}\right|$ neighbors in X_{i+1}. So $(\widetilde{X}_i, \widetilde{X}_{i+1})$ is $(p,\xi,1)$-bounded. Lemma 2.3.7 gives us $\left|X_i \setminus \widetilde{X}_i\right| \leq \frac{2\gamma^2}{\xi^2p^2}\,|X_i| \leq \delta\,|X_i|$. ■

Lemma 2.7.14. *Let $0 \leq q \leq p \leq 1$ and $\epsilon,\delta,\delta' > 0$. Let G be a weighted bipartite graph with vertex sets X and Y. Let $\widetilde{X} \subseteq X$ and $\widetilde{Y} \subseteq Y$ satisfy $\left|\widetilde{X}\right| \geq (1-\delta)\,|X|$ and $\left|\widetilde{Y}\right| \geq (1-\delta)\,|Y|$. Let $\widetilde{G}$ be a weighted bipartite graph on $(\widetilde{X},\widetilde{Y})$ such that $G(x,y) \geq (1-\delta')\widetilde{G}(x,y)$ for all $x \in \widetilde{X}, y \in \widetilde{Y}$. If $(\widetilde{X},\widetilde{Y})_{\widetilde{G}}$ satisfies $\mathrm{DISC}_{\geq}(q,p,\epsilon)$, then $(X,Y)_G$ satisfies $\mathrm{DISC}_{\geq}(q,p,\epsilon+2\delta+\delta')$.*

Proof. For this proof we use sums instead of integrals since the integrals corresponding to $(X,Y)_G$ and $(\widetilde{X},\widetilde{Y})_G$ have different normalizations and can be somewhat confusing.

Let $u \colon X \to [0,1]$ and $v \colon Y \to [0,1]$. We have

$$\begin{aligned}
&\sum_{x\in X}\sum_{y\in Y} G(x,y)u(x)v(y) \\
&\geq (1-\delta')\sum_{x\in\widetilde{X}}\sum_{y\in\widetilde{Y}} \widetilde{G}(x,y)u(x)v(y) \\
&\geq q(1-\delta')\left(\sum_{x\in\widetilde{X}} u(x)\right)\left(\sum_{y\in\widetilde{Y}} v(y)\right) - \epsilon p\left|\widetilde{X}\right|\left|\widetilde{Y}\right| \\
&\geq q(1-\delta')\left(\sum_{x\in X} u(x) - \delta\left|X\right|\right)\left(\sum_{y\in Y} v(y) - \delta\left|Y\right|\right) - \epsilon p\left|X\right|\left|Y\right| \\
&\geq qu(X)v(Y) - (\epsilon + 2\delta + \delta')p\left|X\right|\left|Y\right|.
\end{aligned}$$

■

Proof of Lemma 2.7.5. We apply Lemma 2.7.13 to find large subsets of vertices for which the induced subgraph of Γ is bounded and then apply Lemma 2.7.12 to show that G restricted to this subgraph satisfies DISC$_{\geq}$. Finally, we use Lemma 2.7.14 to pass the result back to the original graph.

Here are the details. Let $\xi = 8c^{2/3}$ and $\delta = \frac{1}{4}c^{2/3}$, so that the hypotheses of Lemma 2.7.13 are satisfied with $\gamma = \frac{c}{1-\delta}p^{1+\frac{1}{2m-2}}$. Therefore, we can find $\widetilde{X}_i \subseteq X_i$ with $\left|\widetilde{X}_i\right| \geq (1-\delta)X_i$ for each i so that $(\widetilde{X}_i, \widetilde{X}_{i+1})_\Gamma$ is $(p,\xi,1)$-bounded and $(p, \frac{c}{1-\delta}p^{1+\frac{1}{2m-2}}\sqrt{\left|\widetilde{X}_i\right|\left|\widetilde{X}_{i+1}\right|})$-jumbled for every $0 \leq i \leq m-1$. Let $\widetilde{G}$ denote the graph G restricted to $\widetilde{X}_0, \ldots, \widetilde{X}_m$. Note that the normalizations of G and $\widetilde{G}$ are different. For instance, for any $S \subseteq \widetilde{X}_1$ and any $x_0 \in \widetilde{X}_0$ and $x_2 \in \widetilde{X}_2$, we write

$$G(x_0, S, x_2) = \frac{1}{|X_1|}\sum_{x_1\in S} G(x_0,x_1)G(x_1,x_2)$$

while

$$\widetilde{G}(x_0, S, x_2) = \frac{1}{\left|\widetilde{X}_1\right|}\sum_{x_1\in S} G(x_0,x_1)G(x_1,x_2).$$

So $(\widetilde{X}_{i-1}, \widetilde{X}_i)_{\widetilde{G}}$ satisfies DISC$_{\geq}(q_i, p, \epsilon')$ with $\epsilon' \leq \frac{\epsilon}{(1-\delta)^2} \leq 2\epsilon$. Let $\widetilde{G}'$ denote the weighted bipartite graph on $(\widetilde{X}_0, \widetilde{X}_m)$ given by

$$\widetilde{G}'(x_0, x_m) = \min\left\{\widetilde{G}(x_0, \widetilde{X}_1, \ldots, \widetilde{X}_{m-1}, x_m), 4p^m\right\}.$$

Since $4(\frac{c}{1-\delta})^2 \le 8c^2 \le \xi$, we can apply Lemma 2.7.12 to $\widetilde{G}$ to find that $(\widetilde{X}_0, \widetilde{X}_m)_{\widetilde{G}'}$ satisfies $\mathrm{DISC}_{\ge}(q_1 \cdots q_m, p^m, 72\epsilon^{1/(2m)} + 8m\xi)$. To pass the result back to G', we note that

$$\begin{aligned}
G'(x_0, x_m) &= \min\left\{G(x_0, X_1, \ldots, X_{m-1}, x_m), 4p^m\right\} \\
&\ge \min\left\{G(x_0, \widetilde{X}_1, \ldots, \widetilde{X}_{m-1}, x_m), 4p^m\right\} \\
&= \min\left\{\frac{\left|\widetilde{X}_1\right| \cdots \left|\widetilde{X}_{m-1}\right|}{|X_1| \cdots |X_{m-1}|}\widetilde{G}(x_0, \widetilde{X}_1, \ldots, \widetilde{X}_{m-1}, x_m), 4p^m\right\} \\
&\ge (1-\delta)^{m-1}\widetilde{G}'(x_0, x_m) \\
&\ge (1-(m-1)\delta)\widetilde{G}'(x_0, x_m).
\end{aligned}$$

It follows by Lemma 2.7.14 that $(X_0, X_m)_{G'}$ satisfies $\mathrm{DISC}_{\ge}(q_1 \cdots q_m, p^m, \epsilon')$ with $\epsilon' \le 72\epsilon^{1/(2m)} + 8m\xi + 2\delta + (m-1)\delta \le 72(\epsilon^{1/(2m)} + mc^{2/3})$. ■

2.7.2 One-sided cycle counting

If we can perform densification to reduce H to a triangle with two dense edges, then we have a counting lemma for H, as shown by the following lemma. Note that we do not even need any jumbledness assumptions on the remaining sparse edge.

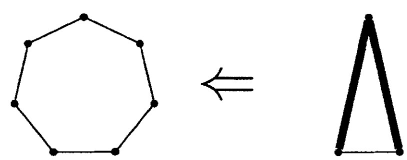
Lemma 2.7.15. *Let K_3 denote the triangle with vertex set $\{1, 2, 3\}$. Let G be a weighted graph with vertex subsets X_1, X_2, X_3 such that, for all $i \ne j$, $(X_i, X_j)_G$ satisfies $\mathrm{DISC}_{\ge}(q_{ij}, p_{ij}, \epsilon)$, where $p_{13} = p_{23} = 1$, $0 \le p_{12} \le 1$, and $0 \le q_{ij} \le p_{ij}$. Then $G(K_3) \ge q_{12}q_{13}q_{23} - 3\epsilon p_{12}$.*

Proof. We have

$$
\begin{aligned}
&G(K_3) - q_{12}q_{13}q_{23} \\
&\quad = \int_{x_1,x_2,x_3} (G(x_1,x_2) - q_{12})G(x_1,x_3)G(x_2,x_3)\ dx_1dx_2dx_3 \\
&\qquad + q_{12}\int_{x_1,x_2,x_3} (G(x_1,x_3) - q_{13})G(x_2,x_3)\ dx_1dx_2dx_3 \\
&\qquad\qquad + q_{12}q_{13}\int_{x_1,x_2,x_3} (G(x_2,x_3) - q_{13})\ dx_1dx_2dx_3.
\end{aligned}
$$

The first integral can be bounded below by $-\epsilon p_{12}$ and the latter two integrals by $-\epsilon q_{12}$. This gives the desired bound. ■

The one-sided counting lemma can be proved by performing subdivision densification as shown below.



Proof of Proposition 2.7.2. Let the vertices of C_ℓ be $\{1, 2, \ldots, \ell\}$ in that order. Apply subdivision densification (Lemma 2.7.4) to the subdivided edge $(1, 2, \ldots, \lceil \ell/2 \rceil)$, as well as to the subdivided edge $(\lceil \ell/2 \rceil, \lceil \ell/2 \rceil + 1, \ldots, \ell)$. Conclude with Lemma 2.7.15. ■

2.8 Applications

It is now relatively straightforward to prove our sparse pseudorandom analogues of Turán's theorem, Ramsey's theorem and the graph removal lemma. All of the proofs have essentially the same flavour. We begin by applying the sparse regularity lemma for jumbled graphs, Theorem 2.1.11. We then apply the dense version of the theorem we are considering to the reduced graph to find a copy of our graph H. The counting lemma then implies that our original sparse graph must also contain many copies of H.

In order to apply the counting lemma, we will always need to clean up our regular partition, removing all edges which are not contained in a dense regular pair. The following lemma is sufficient for our purposes.

Lemma 2.8.1. *For every $\epsilon, \alpha > 0$ and positive integer m, there exists $c > 0$ and a positive integer M such that if Γ is a (p, cpn)-jumbled graph on n vertices then any subgraph G of Γ is such that there is a subgraph G' of G with $e(G') \geq e(G) - 4\alpha e(\Gamma)$ and an equitable partition of the vertex set into k pieces $V_1, V_2, \ldots, V_k$ with $m \leq k \leq M$ such that the following conditions hold.*

1. *There are no edges of G' within V_i for any $1 \leq i \leq k$.*
2. *Every non-empty subgraph $(V_i, V_j)_{G'}$ has $d_{G'}(V_i, V_j) = q_{ij} \geq \alpha p$ and satisfies* $\mathrm{DISC}(q_{ij}, p, \epsilon)$.

Proof. Let $m_0 = \max(32\alpha^{-1}, m)$ and $\theta = \frac{\alpha}{32}$. An application of Theorem 2.1.11, the sparse regularity lemma for jumbled graphs, using $\min\{\theta, \epsilon\}$ as the parameter ϵ in the regularity lemma, tells us that there exists an $\eta > 0$ and a positive integer M such that if Γ is $(p, \eta pn)$-jumbled then there is an equitable partition of the vertices of G into k pieces with $m_0 \leq k \leq M$ such that all but θk^2 pairs of vertex subsets $(V_i, V_j)_G$ satisfy $\mathrm{DISC}(q_{ij}, p, \epsilon)$. Let $c = \min(\eta, \frac{1}{8M^2})$.

Since Γ is (p, β)-jumbled with $\beta \leq cpn$, $c \leq \frac{1}{8M^2}$ and $n \leq 2M|V_i|$ for all i, the number of edges between V_i and V_j satisfies

$$|e(V_i, V_j) - p|V_i||V_j|| \leq cpn^2 \leq \frac{1}{2}p|V_i||V_j|$$

and thus lies between $\frac{1}{2}p|V_i||V_j|$ and $\frac{3}{2}p|V_i||V_j|$. Note that this also holds for $i = j$, allowing for the fact that we will count all edges twice.

Therefore, if we remove all edges contained entirely within any V_i, we remove at most $2pk\left(\frac{2n}{k}\right)^2 = \frac{8pn^2}{k} \leq \frac{\alpha}{4}pn^2$ edges. Here we used that $|V_i| \leq \lceil \frac{n}{k} \rceil \leq \frac{2n}{k}$ for all i. If we remove all edges contained within pairs which do not satisfy the discrepancy condition, the number of edges we are removing is at most $2p\theta k^2 \left(\frac{2n}{k}\right)^2 = 8p\theta n^2 = \frac{\alpha}{4}pn^2$. Finally, if we remove all edges contained within pairs whose density is smaller than αp, we remove at most $\alpha p\binom{n}{2} \le \frac{\alpha}{2}pn^2$ edges. Overall, we have removed at most $\alpha pn^2 \le 4\alpha e(\Gamma)$ edges. We are left with a graph G' with $e(G') \ge e(G) - 4\alpha e(\Gamma)$ edges, as required. ■

2.8.1 Erdős-Stone-Simonovits theorem

We are now ready to prove the Erdős-Stone-Simonovits theorem in jumbled graphs. We first recall the statement. Recall that a graph Γ is (H,ϵ)-*Turán* if any subgraph of Γ with at least $\left(1 - \frac{1}{\chi(H)-1} + \epsilon\right) e(\Gamma)$ edges contains a copy of H.

Theorem 2.1.4 *For every graph H and every $\epsilon > 0$, there exists $c > 0$ such that if $\beta \le cp^{d_2(H)+3}n$ then any (p,β)-jumbled graph on n vertices is (H,ϵ)-Turán.*

Proof. Suppose that H has vertex set $\{1, 2, \ldots, m\}$, Γ is a (p,β)-jumbled graph on n vertices, where $\beta \le cp^{d_2(H)+3}n$, and G is a subgraph of Γ containing at least $\left(1 - \frac{1}{\chi(H)-1} + \epsilon\right) e(\Gamma)$ edges.

We will need to apply the one-sided counting lemma, Lemma 2.1.14, with $\alpha = \frac{\epsilon}{8}$ and θ. We get constants c_0 and $\epsilon_0 > 0$ such that if Γ is $(p, c_0p^{d_2(H)+3}\sqrt{|X_i|\,|X_j|})$-jumbled and G satisfies $\mathrm{DISC}(q_{ij}, p, \epsilon_0)$, where $\alpha p \le q_{ij} \le p$, between sets X_i and X_j for every $1 \le i < j \le m$ with $ij \in E(H)$, then $G(H) \ge (1-\theta)q(H)$.

Apply Lemma 2.8.1 with $\alpha = \epsilon/8$ and ϵ_0. This yields constants c_1 and M such that if Γ is (p, c_1pn)-jumbled then there is a subgraph G' of G with

$$e(G') \ge \left(1 - \frac{1}{\chi(H)-1} + \epsilon - 4\alpha\right) e(\Gamma) \ge \left(1 - \frac{1}{\chi(H)-1} + \frac{\epsilon}{2}\right) e(\Gamma),$$

where we used that $\alpha = \frac{\epsilon}{8}$. Moreover, there is an equitable partition of the vertex set into $k \le M$ pieces $V_1, \ldots, V_k$ such that every non-empty subgraph $(V_i, V_j)_{G'}$ has $d(V_i, V_j) = q_{ij} \ge \alpha p$ and satisfies $\mathrm{DISC}(q_{ij}, p, \epsilon_0)$.

We now consider the reduced graph R, considering each piece V_i of the partition as a vertex v_i and placing an edge between v_i and v_j if and only if the graph between V_i and V_j is non-empty. Since Γ is (p, cpn)-jumbled and $n \le 2M|V_i|$, the number of edges between any two pieces differs from $p|V_i||V_j|$ by at most $cpn^2 \le \frac{\epsilon}{20}p|V_i||V_j|$

provided that $c \leq \frac{\epsilon}{80M^2}$. Note, moreover, that $|V_i| \leq \lceil \frac{n}{k} \rceil \leq (1 + \frac{\epsilon}{20})\frac{n}{k}$ provided that $n \geq \frac{20M}{\epsilon}$. Therefore, the number of edges in the reduced graph R is at least

$$e(R) \geq \frac{e(G')}{(1+\frac{\epsilon}{20})p\lceil \frac{n}{k} \rceil^2} \geq \frac{(1 - \frac{1}{\chi(H)-1} + \frac{\epsilon}{2})e(\Gamma)}{(1+\frac{\epsilon}{20})^3 p(\frac{n}{k})^2} \geq \left(1 - \frac{1}{\chi(H)-1} + \frac{\epsilon}{4}\right)\binom{k}{2},$$

where the final step follows from $e(\Gamma) \geq (1 - \frac{\epsilon}{20})p\binom{n}{2}$.

Applying the Erdős-Stone-Simonovits theorem to the reduced graph implies that it contains a copy of H. But if this is the case then we have a collection of vertex subsets $X_1, \dots, X_m$ such that, for every edge $ij \in E(H)$, the induced subgraph $(X_i, X_j)_{G'}$ has $d(X_i, X_j) = q_{ij} \geq \alpha p$ and satisfies DISC(q_{ij}, p, ϵ_0). By the counting lemma, provided $c \leq \frac{c_0}{2M}$, we have $G(H) \geq G'(H) \geq (1-\theta)(\alpha p)^{e(H)}(2M)^{-v(H)}$. Therefore, for $c = \min(\frac{c_0}{2M}, c_1, \frac{\epsilon}{80M^2})$, we see that G contains a copy of H. ∎

The proof of the stability theorem, Theorem 2.1.5, is similar to the proof of Theorem 2.1.4, so we confine ourselves to a sketch. Suppose that Γ is a (p, β)-jumbled graph on n vertices, where $\beta \leq cp^{d_2(H)+3}n$, and G is a subgraph of Γ containing $\left(1 - \frac{1}{\chi(H)-1} - \delta\right) e(\Gamma)$ edges. An application of Lemma 2.8.1 as in the proof above allows us to show that there is a subgraph G' of G formed by removing at most $\frac{\delta}{4}pn^2$ edges and a regular partition of G' into k pieces such that the reduced graph has at least $\left(1 - \frac{1}{\chi(H)-1} - 2\delta\right)\binom{k}{2}$ edges. This graph can contain no copies of H - otherwise the original graph would have many copies of H as in the last paragraph above. From the dense version of the stability theorem [115] it follows that if δ is sufficiently small then we may make R into a $(\chi(H)-1)$-partite graph by removing at most $\frac{\epsilon}{16}k^2$ edges. We imitate this removal process in the graph G'. That is, if we remove edges between v_i and v_j in R then we remove all of the edges between V_i and V_j in G'. Since the number of edges between V_i and V_j is at most $2p|V_i||V_j|$, we will remove at most

$$\frac{\epsilon}{16}k^2 2p\left\lceil \frac{n}{k} \right\rceil^2 \leq \frac{\epsilon}{2}pn^2$$

edges in total from G'. Since we have already removed all edges which are contained within any V_i the resulting graph is clearly $(\chi(H)-1)$-partite. Moreover, the total number of edges removed is at most $\frac{\delta}{4}pn^2 + \frac{\epsilon}{2}pn^2 \leq \epsilon pn^2$, as required.

2.8.2 Ramsey's theorem

In order to prove that the Ramsey property also holds in sparse jumbled graphs, we need the following lemma which says that we may remove a small proportion of edges from any sufficiently large clique and still maintain the Ramsey property.

Lemma 2.8.2. *For any graph H and any positive integer $r \geq 2$, there exist $a, \eta > 0$ such that if n is sufficiently large and G is any subgraph of K_n of density at least $1 - \eta$, any r-coloring of the edges of G will contain at least $an^{v(H)}$ monochromatic copies of H.*

Proof. Suppose first that the edges of K_n have been r-colored. Ramsey's theorem together with a standard averaging argument tells us that for n sufficiently large there exists a_0 such that there are at least $a_0 n^{v(H)}$ monochromatic copies of H. Since G is formed from K_n by removing at most ηn^2 edges, this deletion process will force us to delete at most $\eta n^{v(H)}$ copies of H. Therefore, provided that $\eta \leq \frac{a_0}{2}$, the result follows with $a = \frac{a_0}{2}$. ■

We also need a slight variant of the sparse regularity lemma, Theorem 2.1.11, which allows us to take a regular partition which works for more than one graph.

Lemma 2.8.3. *For every $\epsilon > 0$ and integers $\ell, m_0 \geq 1$, there exist $\eta > 0$ and a positive integer M such that if Γ is a $(p, \eta pn)$-jumbled graph on n vertices and $G_1, G_2, \ldots G_\ell$ is a collection of weighted subgraphs of Γ then there is an equitable partition into $m_0 \leq k \leq M$ pieces such that for each G_i, $1 \leq i \leq \ell$, all but at most ϵk^2 pairs of vertex subsets $(V_a, V_b)_{G_i}$ satisfy $\mathrm{DISC}(q_{ab}^{(i)}, p, \epsilon)$ for some $q_{ab}^{(i)}$.*

There is also an appropriate analogue of Lemma 2.8.1 to go with this regularity lemma.

Lemma 2.8.4. *For every $\epsilon, \alpha > 0$ and positive integer m, there exist $c > 0$ and a positive integer M such that if Γ is a (p, cpn)-jumbled graph on n vertices then any* *collection of subgraphs $G_1, G_2, \ldots, G_\ell$ of Γ will be such that there are subgraphs G_i' of G_i with $e(G_i') \geq e(G_i) - 4\alpha e(\Gamma)$ and an equitable partition of the vertex set into k pieces $V_1, V_2, \ldots, V_k$ with $m \leq k \leq M$ such that the following conditions hold.*

1. *There are no edges of G_i' within V_a for any $1 \leq i \leq \ell$ and any $1 \leq a \leq k$.*
2. *Every subgraph $(V_a, V_b)_{G_i'}$ containing any edges from G_i' has $d_{G_i'}(V_a, V_b) = q_{ab}^{(i)} \geq \alpha p$ and satisfies $\mathrm{DISC}(q_{ab}^{(i)}, p, \epsilon)$.*

The proof of the sparse analogue of Ramsey's theorem now follows along the lines of the proof of Theorem 2.1.4 above.

Theorem 2.1.6 *For every graph H and every positive integer $r \geq 2$, there exists $c > 0$ such that if $\beta \leq cp^{d_2(H)+3}n$ then any (p, β)-jumbled graph on n vertices is (H, r)-Ramsey.*

Proof. Suppose that H has vertex set $\{1, 2, \ldots, m\}$, Γ is a (p, β)-jumbled graph on n vertices, where $\beta \leq cp^{d_2(H)+3}n$, and $G_1, G_2, \ldots, G_r$ are subgraphs of Γ where G_i is the subgraph whose edges have been colored in color i.

Let a, η be the constants given by Lemma 2.8.2. That is, for $n \geq n_0$, any subgraph of K_n of density at least $1 - \eta$ is such that any r-coloring of its edges contains at least $an^{v(H)}$ monochromatic copies of H. We will need to apply the one-sided counting lemma, Theorem 2.1.14, with $\alpha = \frac{\eta}{8r}$ and θ. We get constants c_0 and $\epsilon_0 > 0$ such that if Γ is $(p, c_0 p^{d_2(H)+3}\sqrt{|X_i|\,|X_j|})$-jumbled and G satisfies $\mathrm{DISC}(q_{ij}, p, \epsilon_0)$, where $\alpha p \leq q_{ij} \leq p$, between sets X_i and X_j for every $1 \leq i < j \leq m$ with $ij \in E(H)$, then $G(H) \geq (1 - \theta)q(H)$.

We apply Lemma 2.8.4 to the collection G_i with $\alpha = \frac{\eta}{8r}$, ϵ_0 and $m = n_0$. This yields $c_1 > 0$ and a positive integer M such that if Γ is $(p, c_1 pn)$-jumbled then there is a collection of graphs G_i' such that $e(G_i') \geq e(G_i) - 4\alpha e(\Gamma)$ and every subgraph $(V_a, V_b)_{G_i'}$ containing any edges from G_i' has $d_{G_i'}(V_a, V_b) = q_{ab}^{(i)} \geq \alpha p$ and satisfies $\mathrm{DISC}(q_{ab}^{(i)}, p, \epsilon_0)$. Adding over all r graphs, we will have removed at most $4r\alpha e(\Gamma) = \frac{\eta}{2}e(\Gamma)$ edges. Let G' be the union of the G_i'. This graph has density at least $1 - \frac{\eta}{2}$ in Γ.

We now consider the colored reduced (multi)graph R, considering each piece V_a of the partition as a vertex v_a and placing an edge of color i between v_a and v_b

if the graph between V_a and V_b contains an edge of color i. Since Γ is (p, cpn)-jumbled and $n \leq 2M|V_i|$, the number of edges between any two pieces differs from $p|V_i||V_j|$ by at most $cpn^2 \leq \frac{\eta}{20}p|V_i||V_j|$ provided that $c \leq \frac{\eta}{80M^2}$. Note, moreover, that $|V_i| \leq \lceil \frac{n}{k} \rceil \leq (1 + \frac{\eta}{20})\frac{n}{k}$ provided that $n \geq \frac{20M}{\eta}$. Therefore, the number of edges in the reduced graph R is at least

$$e(R) \geq \frac{e(G')}{(1 + \frac{\eta}{20})p\lceil \frac{n}{k} \rceil^2} \geq \frac{(1 - \frac{\eta}{2})e(\Gamma)}{(1 + \frac{\eta}{20})^3 p(\frac{n}{k})^2} \geq (1 - \eta)\binom{k}{2},$$

where the final step follows from $e(\Gamma) \geq (1 - \frac{\eta}{20})p\binom{n}{2}$.

We now apply Lemma 2.8.2 to the reduced graph. Since $k \geq m = n_0$, there exists a monochromatic copy of H in the reduced graph, in color i, say. But if this is the case then we have a collection of vertex subsets $X_1, \ldots, X_m$ such that, for every edge $ab \in E(H)$, the induced subgraph $(X_a, X_b)_{G'_i}$ has $d_{G'_i}(X_a, X_b) = q_{ab}^{(i)} \geq \alpha p$ and satisfies $\mathrm{DISC}(q_{ab}^{(i)}, p, \epsilon_0)$. By the counting lemma, provided $c \leq \frac{c_0}{2M}$, we have $G(H) \geq G'_i(H) \geq (1 - \theta)(\alpha p)^{e(H)}(2M)^{-v(H)}$. Therefore, for $c = \min(\frac{c_0}{2M}, c_1, n_0^{-1}, \frac{\eta}{80M^2})$, we see that G contains a copy of H. ■

2.8.3 Graph removal lemma

We prove that the graph removal lemma also holds in sparse jumbled graphs. The proof is much the same as the proof for Turán's theorem, though we include it for completeness.

Theorem 2.1.1 *For every graph H and every $\epsilon > 0$, there exist $\delta > 0$ and $c > 0$ such that if $\beta \leq cp^{d_2(H)+3}n$ then any (p, β)-jumbled graph Γ on n vertices has the following property. Any subgraph of Γ containing at most $\delta p^{e(H)} n^{v(H)}$ copies of H may be made H-free by removing at most ϵpn^2 edges.*

Proof. Suppose that H has vertex set $\{1, 2, \ldots, m\}$, Γ is a (p, β)-jumbled graph on n vertices, where $\beta \leq cp^{d_2(H)+3}n$, and G is a subgraph of Γ containing at most $\delta p^{e(H)} n^{v(H)}$ copies of H.

We will need to apply the one-sided counting lemma, Lemma 2.1.14, with $\alpha = \frac{\epsilon}{16}$

and $\theta = \frac{1}{2}$. We get constants c_0 and $\epsilon_0 > 0$ such that if Γ is $(p, c_0 p^{d_2(H)+3}\sqrt{|X_i|\,|X_j|})$-jumbled and G satisfies $\mathrm{DISC}(q_{ij}, p, \epsilon_0)$, where $\alpha p \le q_{ij} \le p$, between sets X_i and X_j for every $1 \le i < j \le m$ with $ij \in E(H)$, then $G(H) \ge \frac{1}{2} q(H)$.

Apply Lemma 2.8.1 with $\alpha = \epsilon/16$ and ϵ_0. This yields constants c_1 and M such that if Γ is $(p, c_1 pn)$-jumbled then there is a subgraph G' of G with

$$e(G') \ge e(G) - 4\alpha e(\Gamma) \ge e(G) - \frac{\epsilon}{4} e(\Gamma) \ge e(G) - \epsilon p n^2,$$

where we used that $\alpha = \frac{\epsilon}{16}$. Moreover, there is an equitable partition into $k \le M$ pieces $V_1, \dots, V_k$ such that every non-empty subgraph $(V_i, V_j)_{G'}$ has $d(V_i, V_j) = q_{ij} \ge \alpha p$ and satisfies $\mathrm{DISC}(q_{ij}, p, \epsilon_0)$.

Suppose now that there is a copy of H left in G'. If this is the case then we have a collection of vertex subsets $X_1, \dots, X_m$ such that, for every edge $ij \in E(H)$, the induced subgraph $(X_i, X_j)_{G'}$ has $d_{G'}(X_i, X_j) = q_{ij} \ge \alpha p$ and satisfies $\mathrm{DISC}(q_{ij}, p, \epsilon_0)$. By the counting lemma, provided $c \le \frac{c_0}{2M}$, we have $G(H) \ge G'(H) \ge \frac{1}{2}(\alpha p)^{e(H)}(2M)^{-v(H)}$. Therefore, for $c = \min(\frac{c_0}{2M}, c_1)$ and $\delta = \frac{1}{2}\alpha^{e(H)}(2M)^{-v(H)}$, we see that G contains at least $\delta p^{e(H)} n^{v(H)}$ copies of H, contradicting our assumption about G. ∎

2.8.4 Removal lemma for groups

We recall the following removal lemma for groups. Its proof is a straightforward adaption of the proof of the dense version given by Král, Serra and Vena [84].

For the rest of this section, let $k_3 = 3$, $k_4 = 2$, $k_m = 1 + \frac{1}{m-3}$ if $m \ge 5$ is odd, and $k_m = 1 + \frac{1}{m-4}$ if $m \ge 6$ is even.

Theorem 2.1.2 *For each $\epsilon > 0$ and positive integer m, there are $c, \delta > 0$ such that the following holds. Suppose $B_1, \dots, B_m$ are subsets of a group G of order n such that each B_i is (p, β)-jumbled with $\beta \le c p^{k_m} n$. If subsets $A_i \subseteq B_i$ for $i = 1, \dots, m$ are such that there are at most $\delta |B_1| \cdots |B_m|/n$ solutions to the equation $x_1 x_2 \cdots x_m = 1$ with $x_i \in A_i$ for all i, then it is possible to remove at most $\epsilon |B_i|$ elements from each set A_i so as to obtain sets A_i' for which there are no solutions to $x_1 x_2 \cdots x_m = 1$ with*

$x_i \in A'_i$ for all i.

We saw above that the one-sided counting lemma gives the graph removal lemma. For cycles, the removal lemma follows from Proposition 2.7.2. The version we need is stated below.

Proposition 2.8.5. *For every $m \geq 3$ and $\epsilon > 0$, there exist $\delta > 0$ and $c > 0$ so that any graph Γ with vertex subsets $X_1, \dots, X_m$, each of size n, satisfying $(X_i, X_{i+1})_\Gamma$ being (p, β)-jumbled with $\beta \leq cp^{1+k_m}n$ for each $i = 1, \dots, m$ (index taken mod m) has the following property. Any subgraph of Γ containing at most $\delta p^m n^m$ copies of C_m may be made C_m-free by removing at most ϵpn^2 edges, where we only consider embeddings of C_m into Γ where the i-th vertex of C_m embeds into X_i.*

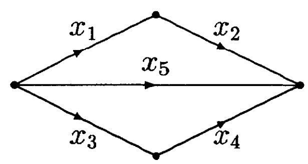
Proof of Theorem 2.1.2. Let Γ denote the graph with vertex set $G \times \{1, \dots, m\}$, the second coordinate taken modulo m, and with vertex (g, i) colored i. Form an edge from (y, i) to $(z, i+1)$ in Γ if and only if $z = yx_i$ for some $x_i \in B_i$, and let G_0 be a subgraph of Γ consisting of those edges with $x_i \in A_i$. Observe that colored m-cycles in the graph G_0 correspond exactly to $(m+1)$-tuples $(y, x_1, x_2, \dots, x_m)$ with $y \in G$ and $x_i \in A_i$ for each i satisfying $x_1 x_2 \dots x_m = 1$. The hypothesis implies that there are at most $\delta |B_1| \cdots |B_m| \leq \delta 2^m p^m n^m$ colored m-cycles in the graph G_0, where we assumed that $c < \frac{1}{2}$ so that $\frac{1}{2}pn \leq |B_i| \leq \frac{3}{2}pn$ by jumbledness. Then by the cycle removal lemma (Proposition 2.8.5) we can choose c and δ so that G_0 can be made C_m-free by removing at most $\frac{\epsilon}{2m}pn^2$ edges.

In A_i, remove the element x_i if at least $\frac{n}{m}$ edges of the form $(y, i)(yx_i, i+1)$ have been removed. Since we removed at most $\frac{\epsilon}{2m}pn^2$ edges, we remove at most $\frac{\epsilon}{2}pn \leq \epsilon |B_i|$ elements from each A_i. Let A'_i denote the remaining elements of A_i. For any solution to $x_1 x_2 \cdots x_m = 1$ for $x_i \in A_i$, consider the n edge-disjoint m-cycles $(g, 1)(gx_1, 2)(gx_1x_2, 3) \cdots (gx_1 \cdots x_m, m)$ in the graph G_0 for $g \in G$. We must have removed at least one edge from each of the n cycles, and so we must have removed at least $\frac{n}{m}$ edges of the form $(y, i)(yx_i, i+1)$ for some i, which implies that $x_i \notin A'_i$. It follows that there is no solution to $x_1 x_2 \cdots x_m = 1$ with $x_i \in A'_i$ for all i. ∎

In [84], the authors also proved removal lemmas for systems of equations which are *graph representable.* For instance, the system

$$x_1 x_2 x_4^{-1} x_3^{-1} = 1$$
$$x_1 x_2 x_5^{-1} = 1$$

can be represented by the graph below, in the sense that solutions to the above system correspond to embeddings of this graph into some larger graph with vertex set $G \times \{1, \ldots, 4\}$, similar to how solutions to $x_1 x_2 \cdots x_n = 1$ correspond to cycles in the proof of Theorem 2.1.2. We refer to the paper [84] for the precise statements. These results can be adapted to the sparse setting in a manner similar to Proposition 2.8.5.



2.9 Concluding remarks

We conclude with discussions on the sharpness of our results, a sparse extension of quasirandom graphs, induced extensions of the various counting and extremal results, other sparse regularity lemmas, algorithmic applications and sparse Ramsey and Turán-type multiplicity results.

2.9.1 Sharpness of results

We have already noted in the introduction that for every H there are (p, β)-jumbled graphs Γ on n vertices, with $\beta = O(p^{d(H)+2)/4} n)$, such that Γ does not contain a copy of H. On the other hand, the results of Section 2.6 tell us that we can always find copies of H in Γ provided that $\beta \leq cp^{d_2(H)+1} n$ and in G provided that $\beta \leq cp^{d_2(H)+3} n$. So, since $d_2(H)$ and $d(H)$ differ by at most a constant factor, our results are sharp up to a multiplicative constant in the exponent for all H. However, we believe that our results are likely to be sharp up to an additive constant on the exponent of p in the jumbledness parameter, with some caveats.

An old conjecture of Erdős [40] asserts that if H is a d-degenerate bipartite graph then there exists $C > 0$ such that every graph G on n vertices with at least $Cn^{2-\frac{1}{d}}$ edges contains a copy of H. This conjecture is known to hold for some bipartite graphs such as $K_{t,t}$ but remains open in general. The best result to date, due to Alon, Krivelevich and Sudakov [5], states that if G has $Cn^{2-\frac{1}{4d}}$ edges then it contains a copy of H.

If Erdős' conjecture is true then this would mean that copies of bipartite H begin to appear already when the density is around $n^{-1/d(H)}$, without any need for a jumbledness condition. If $d_2(H) = d(H) - \frac{1}{2}$ then, even for optimally jumbled graphs, our results only apply down to densities of about $n^{-1/(2d(H)+1)}$.

However, we considered embeddings of H into Γ such that each vertex $\{1, 2, \ldots, m\}$ of H is to be embedded into a separate vertex subset X_i. We believe that in this setting our results are indeed sharp up to an additive constant, even in the case H is bipartite. Without this caveat of embedding each vertex of H into a separate vertex subset in Γ, we still believe that our results should be sharp for many classes of graphs. In particular, we believe the conjecture [47, 86, 119] that there is a $(p, cp^{t-1}n)$-jumbled graph which does not contain a copy of K_t.

One thing which we have left undecided is whether the jumbledness condition for appearance of copies of H in regular subgraphs G of Γ should be the same as that for the appearance of copies of H in Γ alone. For this question, it is natural to consider the case of triangles where we know that there are (p, cp^2n)-jumbled graphs on n vertices which do not contain any triangles. That is, we know the embedding result for Γ is best possible. The result of Sudakov, Szabó and Vu [119] mentioned in the introduction also gives us a sharp result for the (K_3, ϵ)-Turán property. In Section 2.9.6, we will obtain a similar sharp bound for the $(K_3, 2)$-Ramsey property.

While these Turán and Ramsey-type results are suggestive, we believe that the jumbledness condition for counting in G should be stronger than that for counting in Γ. The fact that the results mentioned above are sharp is because there are alternative proofs of Turán's theorem for cliques and Ramsey's theorem for the triangle which only need counting results in Γ rather than within some regular subgraph G. Such a workaround seems unlikely to work for the triangle removal lemma. Kohayakawa, Rödl, Schacht and Skokan [80] conjecture that the jumbledness condition in the sparse triangle removal lemma, Theorem 2.1.2, can be improved from $\beta = o(p^3 n)$ to $\beta = o(p^2 n)$. We conjecture that the contrary holds.

2.9.2 Relative quasirandomness

The study of quasirandom graphs began in the pioneering work of Thomason [127, 128] and Chung, Graham, and Wilson [23]. As briefly discussed in Section 2.1.1, they showed that a large number of interesting graph properties satisfied by random graphs are all equivalent. Perhaps the most surprising aspect of this work is that if the number of cycles of length 4 in a graph is as one would expect in a binomial random graph of the same density, then this is enough to imply that the edges are very well-spread and the number of copies of *any* fixed graph is as one would expect in a binomial random graph of the same density.

While there has been a considerable amount of research aimed at extending quasirandomness to sparse graphs (see [21, 22, 77, 82]), these efforts have been only partially successful. In particular, the key property of counting small subgraphs was absent from previous results in this area. The following theorem bridges this gap and extends these fundamental results to subgraphs of (possibly sparse) pseudorandom graphs. The case where $p = 1$ and Γ is the complete graph corresponds to the original setting. The proof of some of the implications extend easily from the dense case. However, to imply the notable counting properties, we use the counting lemma, Theorem 2.1.12, which acts as a transference principle from the sparse setting to the dense setting.

Such quasirandomness of a structure within a sparse but pseudorandom structure is known as *relative quasirandomness*. This concept has been instrumental in the development of the hypergraph regularity and counting lemma [63, 96, 106, 124]. In the 3-uniform case, for example, one repeatedly has to deal with 3-uniform hypergraphs which are subsets of the triangles of a very pseudorandom graph.

To keep the theorem statement simple, we first describe some notation. The codegree $d_G(v,v')$ of two vertices v, v' in a graph G is the number of vertices which are adjacent to both v and v'. For a graph H, we let $s(H) = \min\left\{\frac{\Delta(L(H))+4}{2}, \frac{d(L(H))+6}{2}\right\}$. For a graph H and another graph G, let $N_H(G)$ denote the number of labeled copies of H (as a subgraph) in G.

Theorem 2.9.1. *Let $k \geq 2$ be a positive integer. For $n \geq 1$, let $\Gamma = \Gamma_n$ be a (p,β)-jumbled graph on n vertices with $p = p(\Gamma)$ and $\beta = \beta(\Gamma) = o(p^k n)$, and $G = G_n$ be a spanning subgraph of Γ_n. The following are equivalent.*

P_1: *For all vertex subsets S and T,*

$$|e_G(S,T) - q|S||T|| = o(pn^2).$$

P_2: *For all vertex subsets S,*

$$\left|e_G(S) - q\frac{|S|^2}{2}\right| = o(pn^2).$$

P_3: *For all vertex subsets S with $|S| = \lfloor \frac{n}{2} \rfloor$,*

$$\left|e_G(S) - q\frac{n^2}{8}\right| = o(pn^2).$$

P_4: *For each graph H with $k \geq s(H)$,*

$$N_H(G) = q^{e(H)}n^{v(H)} + o(p^{e(H)}n^{v(H)}).$$

P_5: $e(G) \geq q\frac{n^2}{2} + o(pn^2)$ *and*

$$N_{C_4}(G) \leq q^4 n^4 + o(p^4 n^4).$$

P_6: $e(G) \geq (1+o(1))q\frac{n^2}{2}$, $\lambda_1 = (1+o(1))qn$, *and* $\lambda_2 = o(pn)$, *where λ_i is the ith largest eigenvalue, in absolute value, of the adjacency matrix of G.*

P_7:

$$\sum_{v,v' \in V(G)} |d_G(v,v') - q^2 n| = o(p^2 n^3).$$

We briefly describe how to prove the equivalences between the various properties in Theorem 2.9.1, with a flow chart shown below.

$$\begin{array}{ccccccccc} & & & & & P_4 & & & \\ & & & & \nearrow & & \searrow & & \\ P_3 & \Leftrightarrow & P_2 & \Leftrightarrow & P_1 & \Longleftarrow & P_5 & \Leftrightarrow & P_7 \\ & & & & \nwarrow & & \swarrow & & \\ & & & & & P_6 & & & \end{array}$$

The equivalence between the discrepancy properties P_1, P_2, P_3 is fairly straightforward and similar to the dense case. Theorem 2.1.12 shows that P_1 implies P_4. As P_5 is a special case of P_4, we have that P_4 implies P_5. Proposition 2.5.4 shows that P_5 implies P_1. The fact P_5 implies P_6 follows easily from the identity that the trace of the fourth power of the adjacency matrix of G is both the number of closed walks in G of length 4, and the sum of the fourth powers of the eigenvalues of the adjacency matrix of G. The fact that P_6 implies P_1 is the standard proof of the expander mixing lemma. The fact P_5 implies P_7 follows easily from the identity

$$N_{C_4}(G) = 4 \sum_{v,v' \in V(G)} \binom{d_G(v,v')}{2}, \tag{2.19}$$

where the sum is over all $\binom{n}{2}$ pairs of distinct vertices, as well as the identity

$$\sum_{v,v'} d_G(v,v') = \sum_{v} \binom{d_G(v)}{2},$$

and two applications of the Cauchy-Schwarz inequality. Finally, we have P_7 implies P_5 for the following reason. From (2.19), we have that P_5 is equivalent to

$$\sum_{v,v'} d_G(v,v')^2 = \frac{1}{2} q^4 n^4 + o(p^4 n^4). \tag{2.20}$$

To verify (2.20), we split up the sum on the left into three sums. The first sum is over pairs v, v' with $|d_G(v,v') - q^2 n| = o(p^2 n)$, the second sum is over pairs v, v' with

$d_G(v,v') > 2p^2n$, and the third sum is over the remaining pairs v, v'. From P_7, almost all pairs v, v' of vertices satisfy $|d_G(v,v') - q^2n| = o(p^2n)$, and so the first sum is $\frac{1}{2}q^4n^4 + o(p^4n^4)$. The second sum satisfies

$$\sum_{v,v':d_G(v,v')>2p^2n} d_G(v,v')^2 \leq \sum_{v,v':d_\Gamma(v,v')>2p^2n} d_\Gamma(v,v')^2 = o(p^4n^4),$$

where the first inequality follows from G is a subgraph of Γ, and the second inequality follows from pseudorandomness in Γ. Finally, as P_7 implies there are $o(n^2)$ pairs v, v' not satisfying $|d_G(v,v') - q^2n| = o(p^2n)$, and the terms in the third sum are at most $2p^2n$, the third sum is $o(p^4n^4)$. This completes the proof sketch of the equivalences between the various properties in Theorem 2.9.1.

2.9.3 Induced extensions of counting lemmas and extremal results

With not much extra effort, we can establish induced versions of the various counting lemmas and extremal results for graphs. We assume that we are in Setup 2.4.1 with the additional condition that, in Setup 2.3.1, the graph Γ satisfies the jumbledness condition for all pairs ab of vertices and not just the sparse edges of H. Define a *strongly induced copy* of H in G to be a copy of H in G such that the nonedges of the copy of H are nonedges of Γ. Since G is a subgraph of Γ, a strongly induced copy of H is an induced copy of H. Define

$$G^*(H) := \int_{x_1\in X_1,\ldots,x_m\in X_m} \prod_{(i,j)\in E(H)} G(x_i,x_j) \prod_{(i,j)\notin E(H)} (1-\Gamma(x_i,x_j))\, dx_1\cdots dx_m$$

and

$$q^*(H) := \prod_{(i,j)\in E(H)} q_{ij} \prod_{(i,j)\notin E(H)} (1-p_{ij}).$$

Note that $G^*(H)$ is the probability that a random compatible map forms a strongly induced copy of H and $q^*(H)$ is the idealized version. Also note that if Γ is (p,β)-jumbled, then its complement $\bar{\Gamma}$ is $(1-p,\beta)$-jumbled. Hence, for p small, we expect that most copies of H guaranteed by Theorem 2.1.14 are strongly induced. This is formalized in the following theorem, which is an induced analogue of the one-sided counting lemma, Theorem 2.1.14.

Theorem 2.9.2. *For every fixed graph H on vertex set $\{1, 2, \ldots, m\}$ and every $\theta > 0$, there exist constants $c > 0$ and $\epsilon > 0$ such that the following holds.*

Let Γ be a graph with vertex sets $X_1, \ldots, X_m$ and suppose that $p \leq \frac{1}{m}$ and the bipartite graph $(X_i, X_j)_\Gamma$ is $(p, cp^{d(H)+\frac{5}{2}}\sqrt{|X_i|\,|X_j|})$-jumbled for every $i < j$. Let G be a subgraph of Γ, with the vertex i of H assigned to the vertex set X_i of G. For each edge ij in H, assume that $(X_i, X_j)_G$ satisfies $\mathrm{DISC}(q_{ij}, p, \epsilon)$. Then

$$G^*(H) \geq (1 - \theta)q^*(H).$$

We next discuss how the proof of Theorem 2.9.2 is a minor modification of the proof of Theorem 2.1.14. As in the proof of Theorem 2.1.14, after $j - 1$ steps in the embedding, we have picked $f(v_1), \ldots, f(v_{j-1})$ and have subsets $T(i, j-1) \subset X_i$ for $j \leq i \leq m$ which consist of the possible vertices for $f(v_i)$ given the choice of the first $j - 1$ embedded vertices. We are left with the task of picking a good set $W(j) \subset T(j, j-1)$ of possible vertices $w = f(v_j)$ to continue the embedding with the desired properties. We will guarantee, in addition to the three properties stated there (which may be maintained since $d_2(H) + 3 \leq d(H) + \frac{5}{2}$), that

4. $|N_{\bar{\Gamma}}(w) \cap T(i, j-1)| \geq (1 - p - \sqrt{c})|T(i, j-1)|$ for each $i > j$ which is not adjacent to j.

As for each such i, if w is chosen for $f(v_j)$, $T(i, j) = N_{\bar{\Gamma}}(w) \cap T(i, j-1)$, this will guarantee that $|T(i, j)| \geq (1 - p - \sqrt{c})|T(i, j-1)|$. As for each such i, the set $T(i, j)$ is only slightly smaller than $T(i, j-1)$, this will affect the discrepancy between each pair of sets by at most a factor $(1 - p - \sqrt{c})^2$. This additional fourth property makes the set $W(j)$ only slightly smaller. Indeed, to guarantee this property, we need that for each of the nonneighbors $i > j$ of j, the vertices w with fewer than $(1 - p - \sqrt{c})|T(i, j-1)|$ nonneighbors in $T(i, j-1)$ in graph $\bar{\Gamma}$ are not in $W(j)$, and there are at most $\frac{\beta^2}{c|T(i,j-1)|}$

such vertices for each i by Lemma 2.3.7. As there are at most m choices of i, and $|T(i,j-1)| \geq \left(1-\frac{\theta_{j-1}}{6}\right) q(i,j-1)|X_i|$, we get that satisfying this additional fourth property requires that the number of additional vertices deleted to obtain $W(j)$ is at most

$$m\frac{\beta^2}{c|T(i,j-1)|} \leq \frac{mcp^{2d(H)+5}}{\left(1-\frac{\theta_{j-1}}{6}\right) q(i,j-1)}|X_j|$$
$$\leq \frac{mcp^{2d(H)+5}}{\left(1-\frac{\theta_{j-1}}{6}\right)^2 q(i,j-1)q(j,j-1)}|T(j,j-1)|,$$

which is neglible since both $q(i,j-1)$ and $q(j,j-1)$ are at most $p^{d(H)}$. We therefore see that, after changing the various parameters in the proof of Theorem 2.1.14 slightly, the simple modification of the proof sketched above completes the proof of Theorem 2.9.2. We remark that the assumption $p \leq \frac{1}{m}$ can be replaced by p is bounded away from 1, which is needed as we must guarantee that the nonedges of the induced copy of H must be nonedges of Γ. We also note that in order to guarantee that nonedges of H map to nonedges of Γ, it is necessary to take the exponent of p in the jumbledness assumption in Theorem 2.9.2 to be $d(H)+\frac{5}{2}$ and not $d_2(H)+3$.

The induced graph removal lemma was proved by Alon, Fischer, Krivelevich, and Szegedy [8]. It states that for each graph H and $\epsilon > 0$ there is $\delta > 0$ such that every graph on n vertices with at most $\delta n^{v(H)}$ induced copies of H can be made induced H-free by adding or deleting at most ϵn^2 edges. This clearly extends the original graph removal lemma. To prove the induced graph removal lemma, they developed the strong regularity lemma, whose proof involves iterating Szemerédi's regularity lemma many times. A new proof of the induced graph removal lemma which gives an improved quantitative estimate was recently obtained in [27].

The first application of Theorem 2.9.2 we discuss is an induced extension of the sparse graph removal, Theorem 2.1.1. It does not imply the induced graph removal lemma above.

Theorem 2.9.3. *For every graph H and every $\epsilon > 0$, there exist $\delta > 0$ and $c > 0$ such* *that if $\beta \le cp^{d(H)+\frac{5}{2}}n$ then any (p,β)-jumbled graph Γ on n vertices with $p \le \frac{1}{v(H)}$ has the following property. Any subgraph of Γ containing at most $\delta p^{e(H)} n^{v(H)}$ (strongly) induced copies of H may be made H-free by removing at most ϵpn^2 edges.*

The proof of Theorem 2.9.3 is the same as the proof of Theorem 2.1.1, except we replace the one-sided counting lemma, Theorem 2.1.14, with its induced variant, Theorem 2.9.2. Note that unlike the standard induced graph removal lemma, here it suffices only to delete edges. Furthermore, all copies of H, not just induced copies, are removed by the deletion of few edges.

The induced Ramsey number $r_{\text{ind}}(H;r)$ is the smallest natural number N for which there is a graph G on N vertices such that in every r-coloring of the edges of G there is an induced monochromatic copy of H. The existence of these numbers was independently proven in the early 1970s by Deuber [36], Erdős, Hajnal and Posa [42], and Rödl [103]. The bounds that these original proofs give on $r_{\text{ind}}(H,r)$ are enormous. However, Trotter conjectured that the induced Ramsey number of bounded degree graphs is at most polynomial in the number of vertices. That is, for each Δ there is $c(\Delta)$ such that $r_{\text{ind}}(H;2) \le n^{c(\Delta)}$. This was proved by Łuczak and Rödl [92], who gave an enormous upper bound on $c(\Delta)$, namely, a tower of twos of height $O(\Delta^2)$. More recently, Fox and Sudakov [48] proved an upper bound on $c(\Delta)$ which is $O(\Delta \log \Delta)$. These proofs giving a polynomial bound on the induced Ramsey number of graphs of bounded degree do not appear to extend to more than two colors.

A graph G is *induced Ramsey (Δ,n,r)-universal* if, for every r-edge-coloring of G, there is a color for which there is a monochromatic induced copy in that color of every graph on n vertices with maximum degree Δ. Clearly, if G is induced Ramsey (Δ,n,r)-universal, then $r_{\text{ind}}(H;r) \le |G|$ for every graph H on n vertices with maximum degree Δ.

Theorem 2.9.4. *For each Δ and r there is $C = C(\Delta,r)$ such that for every n there is an induced Ramsey (Δ,n,r)-universal graph on at most $Cn^{2\Delta+8}$ vertices.*

The exponent of n in the above result is best possible up to a multiplicative factor. This is because even for the much weaker condition that G contains an induced copy of all graphs on n vertices with maximum degree Δ, the number of vertices of G has to be $\Omega(n^{\Delta/2})$ (see, e.g., [18]).

We have the following immediate corollary of Theorem 2.9.4, improving the bound for induced Ramsey numbers of bounded degree graphs. It is also the first polynomial upper bound which works for more than two colors.

Corollary 2.9.5. *For each Δ and r there is $C = C(\Delta, r)$ such that $r_{ind}(H; r) \leq Cn^{2\Delta+8}$ for every n-vertex graph H of maximum degree Δ.*

We next sketch the proof of Theorem 2.9.4. The proof builds on ideas used in the proof of Chvatal, Rödl, Szemerédi, and Trotter [24] that Ramsey numbers of bounded degree graphs grow linearly in the number of vertices. We claim that any graph G on $N = Cn^{2\Delta+8}$ vertices which is (p, β)-jumbled with $p = \frac{1}{n}$ and $\beta = O(\sqrt{pN})$ is the desired induced Ramsey (Δ, n, r)-universal graph. Such a graph exists as almost surely $G(N, p)$ has this jumbledness property. Note that $\beta = cp^{d(H)+\frac{5}{2}}N$ with $c = O(p^2)$. We consider an r-coloring of the edges of G and apply the multicolor sparse regularity lemma so that each color satisfies a discrepancy condition between almost all pairs of parts. Using Turán's theorem and Ramsey's theorem in the reduced graph, we find $\Delta + 1$ parts $X_1, \ldots, X_{\Delta+1}$, each pair of which has density at least $\frac{p}{2r}$ in the same color, say red, and satisfies a discrepancy condition. Let H be a graph on n vertices with maximum degree Δ, so H has chromatic number at most $\Delta + 1$. Assign each vertex a of H to some part so that the vertices assigned to each part form an independent set in H. We then use the induced counting lemma, Theorem 2.9.2, to get an induced monochromatic red copy of H. We make a couple of observations which are vital for this proof to work, and one must look closely into the proof of the induced counting lemma to verify these claims. First, we can choose the constants in the regularity lemma and the counting lemma so that they only depend on the maximum degree Δ and the number of colors r and not on the number n of vertices. Indeed, in addition to the at most 2Δ times that we apply inheritance of regularity, the discrepancy-parameter increases by a factor of at most $(1 - p - \sqrt{c})^{-2n} = (1 - O(p))^{-2n} = (1 - O(\frac{1}{n}))^{-2n} = O(1)$ due to the restrictions imposed by the nonedges of H. So we lose a total of at most a constant factor in the discrepancy, which does not affect the outcome. Second, as we assigned some vertices of H to the same part, they may get embedded to the same vertex. However, one easily checks that almost all the embeddings of H in the proof of the induced counting lemma are one-to-one, and hence there is a monochromatic induced copy of H. Indeed, as there are less than n vertices which are previously embedded at each step of the proof of the induced counting lemma, and $W(j) \gg n$, then there is always a vertex $w \in W(j)$ to pick for $f(v_j)$ to continue the embedding. This completes the proof sketch.

In the proof sketched above, the use of the sparse regularity lemma forces an enormous upper bound on $C(\Delta, r)$, of tower-type. However, all we needed was $\Delta + 1$ parts such that the graph between each pair of parts has density at least $\frac{p}{2r}$ in the same color and satisfies a discrepancy condition. To guarantee this, one does not need the full strength of the regularity lemma, and the sparse version of the Duke-Lefmann-Rödl weak regularity lemma discussed in Subsection 2.9.4 is sufficient. This gives a better bound on $C(\Delta, r)$, which is an exponential tower of constant height.

The last application we mention is an induced extension of the sparse Erdős-Stone-Simonovits theorem, Theorem 2.1.4. We say that a graph Γ is *induced* (H, ϵ)*-Turán* if any subgraph of Γ with at least $(1 - \frac{1}{\chi(H)-1} + \epsilon)e(\Gamma)$ edges contains a strongly induced copy of H.

Theorem 2.9.6. *For every graph H and every $\epsilon > 0$, there exists $c > 0$ such that if $\beta \leq cp^{d(H)+\frac{5}{2}}n$ then any (p, β)-jumbled graph on n vertices with $p \leq \frac{1}{v(H)}$ is induced (H, ϵ)-Turán.*

The proof of Theorem 2.9.6 is the same as the proof of Theorem 2.1.4, except we replace the one-sided counting lemma, Theorem 2.1.14, with its induced variant, Theorem 2.9.2.

2.9.4 Other sparse regularity lemmas

The sparse regularity lemma, in the form due to Scott [112], states that for every $\epsilon > 0$ and positive integer m, there exists a positive integer M such that every graph G has an equitable partition into k pieces $V_1, V_2, \ldots, V_k$ with $m \leq k \leq M$ such that all but ϵk^2 pairs $(V_i, V_j)_G$ satisfy $\mathrm{DISC}(p_{ij}, p_{ij}, \epsilon)$ for some p_{ij}. The additional condition of jumbledness which we imposed in our regularity lemma, Theorem 2.1.11, was there so as to force all of the p_{ij} to be p. If this were not the case, it could easily be that all of the edges of the graph bunch up within a particular bad pair, so the result would tell us nothing.

In our results, we made repeated use of sparse regularity. While convenient, this does have its limitations. In particular, the bounds which the regularity lemma gives on the number of pieces M in the regular partition is (and is necessarily [27, 60]) of tower-type in ϵ. This means that the constants c^{-1} which this method produces for Theorems 2.1.1, 2.1.4, 2.1.5, and 2.1.6 are also of tower-type.

In the dense setting, there are other sparse regularity lemmas which prove sufficient for many of our applications. One such example is the cylinder regularity lemma of Duke, Lefmann and Rödl [38]. This lemma says that for a k-partite graph, between sets $V_1, V_2, \ldots, V_k$, there is an ϵ-regular partition of the cylinder $V_1 \times \cdots \times V_k$ into a relatively small number of cylinders $K = W_1 \times \cdots \times W_k$, with $W_i \subseteq V_i$ for $1 \leq i \leq k$. The definition of an ϵ-regular partition here is that all but an ϵ-fraction of the k-tuples $(v_1, \ldots, v_k) \in V_1 \times \cdots \times V_k$ are in ϵ-regular cylinders, where a cylinder $W_1 \times \cdots \times W_k$ is ϵ-regular if all $\binom{k}{2}$ pairs (W_i, W_j), $1 \leq i < j \leq k$, are ϵ-regular in the usual sense.

For sparse graphs, a similar theorem may be proven by appropriately adapting the proof of Duke, Lefmann and Rödl using the ideas of Scott. Consider a k-partite graph, between sets $V_1, V_2, \ldots, V_k$. We will say that a cylinder $K = W_1 \times \cdots \times W_k$, with $W_i \subseteq V_i$ for $1 \leq i \leq k$, satisfies $\mathrm{DISC}(q_K, p_K, \epsilon)$ with $q_K = (q_{ij})_{1 \leq i < j \leq k}$ and $p_K = (p_{ij})_{1 \leq i < j \leq k}$ if all $\binom{k}{2}$ pairs (W_i, W_j), $1 \leq i < j \leq k$, satisfy $\mathrm{DISC}(q_{ij}, p_{ij}, \epsilon)$. The sparse version of the cylinder regularity lemma is now as follows.

Proposition 2.9.7. *For every $\epsilon > 0$ and positive integer k, there exists $\gamma > 0$ such* *that if $G = (V, E)$ is a k-partite graph with k-partition $V = V_1 \cup \cdots \cup V_k$ then there exists a partition $\mathcal{K}$ of $V_1 \times \cdots \times V_k$ into at most γ^{-1} cylinders such that all but an ϵ-fraction of the k-tuples in $V_1 \times \cdots \times V_k$ are contained in cylinders K satisfying* DISC(q_K, p_K, ϵ) *and, for each $K \in \mathcal{K}$ with $K = W_1 \times \cdots \times W_k$ and $1 \leq i \leq k$, $|W_i| \geq \gamma|V_i|$.*

The constant γ is at most exponential in a power of $k\epsilon^{-1}$. Moreover, this theorem is sufficient for our applications to Turán's theorem and Ramsey's theorem. This results in constants c^{-1} which are at most double exponential in the parameters $|H|$, ϵ and r for Theorems 2.1.4 and 2.1.6.

For the graph removal lemma, we may also make some improvement, but it is of a less dramatic nature. As in the dense case [46], it shows that in Theorem 2.1.1 we may take δ^{-1} and c^{-1} to be a tower of twos of height logarithmic in ϵ^{-1}. The proof essentially transfers to the sparse case using the sparse counting lemma, Theorem 2.1.14.

2.9.5 Algorithmic applications

The algorithmic versions of Szemerédi's regularity lemma and its variants have applications to fundamental algorithmic problems such as max-cut, max-k-sat, and property testing (see [9] and its references). The result of Alon and Naor [6] approximating the cut-norm of a graph via Grothendieck's inequality allows one to obtain algorithmic versions of Szemeredi's regularity lemma [7], the Frieze-Kannan weak regularity lemma [25], and the Duke-Lefmann-Rödl weak regularity lemma. Many of these algorithmic applications can be transferred to the sparse setting using algorithmic versions of the sparse regularity lemmas, allowing one to substantially improve the error approximation in this setting. Our new counting lemmas allows for further sparse extensions. We describe one such extension below.

Suppose we are given a graph H on h vertices, and we want to compute the number of copies of H in a graph G on n vertices. The brute force approach considers all possible h-tuples of vertices and computes the desired number in time $O(n^h)$. The

Duke-Lefmann-Rödl regularity lemma was originally introduced in order to obtain a much faster algorithm, which runs in polynomial time with an absolute constant exponent, at the expense of some error. More precisely, for each $\epsilon > 0$, they found an algorithm which, given a graph on n vertices, runs in polynomial time and approximates the number of copies of H as a subgraph to within ϵn^h. The running time is of the form $C(h, \epsilon)n^c$, where c is an absolute constant and $C(h, \epsilon)$ is exponential in a power of $h\epsilon^{-1}$. We have the following extension of this result to the sparse setting. The proof transfers from the dense setting using the algorithmic version of the sparse Duke-Lefmann-Rödl regularity lemma, Proposition 2.9.7, and the sparse counting lemma, Theorem 2.1.12. For a graph H, we let $s(H) = \min\left\{\frac{\Delta(L(H))+4}{2}, \frac{d(L(H))+6}{2}\right\}$.

Proposition 2.9.8. *Let H be a graph on h vertices with $s(H) \le k$ and $\epsilon > 0$. There is an absolute constant c and another constant $C = C(\epsilon, h)$ depending only exponentially on $h\epsilon^{-1}$ such that the following holds. Given a graph G on n vertices which is known to be a spanning subgraph of a (p, β)-pseudorandom graph with $\beta \le C^{-1}p^k n$, the number of copies of H in G can be computed up to an error $\epsilon p^{e(H)} n^{v(H)}$ in running time Cn^c.*

2.9.6 Multiplicity results

There are many problems and results in graph Ramsey theory and extremal graph theory on the multiplicity of subgraphs. These results can be naturally extended to sparse pseudorandom graphs using the tools developed in this paper. Indeed, by applying the sparse regularity lemma and the new counting lemmas, we get extensions of these results to sparse graphs. In this subsection, we discuss a few of these results.

Recall that Ramsey's theorem states that every 2-edge-coloring of a sufficiently large complete graph K_n contains at least one monochromatic copy of a given graph H. Let $c_{H,n}$ denote the fraction of copies of H in K_n that must be monochromatic in any 2-edge-coloring of G. By an averaging argument, $c_{H,n}$ is a bounded, monotone increasing function in n, and therefore has a positive limit c_H as $n \to \infty$. The constant c_H is known as the *Ramsey multiplicity constant* for the graph H. It is simple to show that $c_H \leq 2^{1-m}$ for a graph H with $m = e(H)$ edges, where this bound comes from considering a random 2-edge-coloring of K_n with each coloring equally likely.

Erdős [39] and, in a more general form, Burr and Rosta [17] suggested that the Ramsey multiplicity constant is achieved by a random coloring. These conjectures are false, as was demonstrated by Thomason [129] even for H being any complete graph K_t with $t \geq 4$. Moreover, as shown in [45], there are graphs H with m edges and $c_H \leq m^{-m/2+o(m)}$, which demonstrates that the random coloring is far from being optimal for some graphs.

For bipartite graphs the situation seems to be very different. The edge density of a graph is the fraction of pairs of vertices that are edges. The conjectures of Erdős-Simonovits [116] and Sidorenko [114] suggest that for any bipartite H the number of copies of H in any graph G on n vertices with edge density p bounded away from 0 is asymptotically at least the same as in the n-vertex random graph with edge density p. This conjecture implies that $c_H = 2^{1-m}$ if H is bipartite with m edges. The most general results on this problem were obtained in [28] and [88], where it is shown that every bipartite graph H which has a vertex in one part complete to the other part satisfies the conjecture.

More generally, let $c_{H,\Gamma}$ denote the fraction of copies of H in Γ that must be monochromatic in any 2-edge-coloring of Γ. For a graph Γ with n vertices, by averaging over all copies of Γ in K_n, we have $c_{H,\Gamma} \leq c_{H,n} \leq c_H$. It is natural to try to find conditions on Γ which imply that $c_{H,\Gamma}$ is close to c_H. The next theorem shows that if Γ is sufficiently jumbled, then $c_{H,\Gamma}$ is close to c_H. The proof follows by noting that the proportion of monochromatic copies of H in the weighted reduced graph R is at least $c_{H,|R|}$. This count then transfers back to Γ using the one-sided counting lemma. We omit the details.

Theorem 2.9.9. *For each $\epsilon > 0$ and graph H, there is $c > 0$ such that if Γ is a (p,β)-jumbled graph on n vertices with $\beta \leq cp^{d_2(H)+3}n$ then every 2-edge-coloring of Γ contains at least $(c_H - \epsilon)p^{e(H)}n^{v(H)}$ labeled monochromatic copies of H.*

Maybe the earliest result on Ramsey multiplicity is Goodman's theorem [59], which determines $c_{K_3,n}$ and, in particular, implies $c_{K_3} = \frac{1}{4}$. The next theorem shows an extension of Goodman's theorem to pseudorandom graphs, giving an optimal jumbledness condition to imply $c_{H,\Gamma} = \frac{1}{4} - o(1)$.

Theorem 2.9.10. *If Γ is a (p,β)-jumbled graph on n vertices with $\beta \leq \frac{1}{10}p^2n$, then every 2-edge-coloring of Γ contains at least $(p^3 - 10p\frac{\beta}{n})\frac{n^3}{24}$ monochromatic triangles.*

The proof of this theorem follows by first noting that $T = A + 2M$, where A denotes the number of triangles in Γ, M the number of monochromatic triangles in Γ, and T the number of ordered triples (a,b,c) of vertices of Γ which form a triangle such that (a,b) and (a,c) are the same color. We then give an upper bound for A and a lower bound for T using the jumbledness conditions and standard inequalities. We omit the precise details.

The previous theorem has the following immediate corollary, giving an optimal jumbledness condition to imply that a graph is $(K_3, 2)$-Ramsey.

Corollary 2.9.11. *If Γ is a (p,β)-jumbled graph on n vertices with $\beta < \frac{p^2}{10}n$, then Γ is $(K_3, 2)$-Ramsey.*

Define the Turán multiplicity $\rho_{H,d,n}$ to be the minimum, over all graphs G on n vertices with edge density at least d, of the fraction of copies of H in K_n which are also in G. Let $\rho_{H,d}$ be the limit of $\rho_{H,d,n}$ as $n \to \infty$. This limit exists by an averaging argument. The conjectures of Erdős-Simonovits [116] and Sidorenko [114] mentioned earlier can be stated as $\rho_{H,d} = d^{e(H)}$ for bipartite H. Recently, Reiher [100], extending work of Razborov [99] and Nikiforov [97] for $t = 3$ and 4, determined $\rho_{K_t,d}$ for all $t \geq 3$.

We can similarly extend these results to the sparse setting. Let $\rho_{H,d,\Gamma}$ be the minimum, over all subgraphs G of Γ with at least $de(\Gamma)$ edges, of the fraction of copies of H in Γ which are also in G. We have the following result, which gives jumbledness conditions on Γ which imply that $\rho_{H,d,\Gamma}$ is close to $\rho_{H,d}$.

Theorem 2.9.12. *For each $\epsilon > 0$ and graph H, there is $c > 0$ such that if Γ is a (p,β)-jumbled graph on n vertices with $\beta \leq cp^{d_2(H)+3}n$ then every subgraph of Γ with at least $de(\Gamma)$ edges contains at least $(\rho_{H,d} - \epsilon)p^{e(H)}n^{v(H)}$ labeled copies of H.*

Chapter 3

A relative Szemerédi theorem

The Green-Tao theorem [69] states that the primes contain arbitrarily long arithmetic progressions. This result, along with their subsequent work [71] on determining the asymptotics for the number of prime k-tuples in arithmetic progression, constitutes one of the great breakthroughs in 21st century mathematics.

The proof of the Green-Tao theorem has two key steps. The first step, which Green and Tao refer to as the "main new ingredient" of their proof, is to establish a relative Szemerédi theorem. Szemerédi's theorem [120] states that any dense subset of the integers contains arbitrarily long arithmetic progressions. More formally, we have the following theorem, which is stated for $\mathbb{Z}_N := \mathbb{Z}/N\mathbb{Z}$ but easily implies an equivalent statement in the set $[N] := \{1, 2, \ldots, N\}$.

Theorem 3.0.1 (Szemerédi's theorem). *For every natural number $k \geq 3$ and every $\delta > 0$, as long as N is sufficiently large, any subset of $\mathbb{Z}_N$ of density at least δ contains an arithmetic progression of length k.*

A relative Szemerédi theorem is a similar statement where the ground set is no longer the set $\mathbb{Z}_N$ but rather a sparse pseudorandom subset of $\mathbb{Z}_N$.

The second step in their proof is to show that the primes are a dense subset of a pseudorandom set of "almost primes", sufficiently pseudorandom that the relative Szemerédi theorem holds. Then, since the primes are a dense subset of this pseudorandom set, an application of the relative Szemerédi theorem implies that the primes contain arbitrarily long arithmetic progressions. This part of the proof uses some ideas from the work of Goldston and Yıldırım [58] (and was subsequently simplified in [122]).

In the work of Green and Tao, the pseudorandomness conditions on the ground set are known as the linear forms condition and the correlation condition. Roughly speaking, both of these conditions say that, in terms of the number of solutions to certain linear systems of equations, the set behaves like a random set of the same density. A natural question is whether these pseudorandomness conditions can be weakened. We address this question by giving a simple proof for a strengthening of the relative Szemerédi theorem, showing that a weak linear forms condition is sufficient for the theorem to hold.

This improvement has two aspects. We remove the correlation condition entirely but we also reduce the set of linear forms for which the correct count is needed. In particular, we remove those corresponding to the dual function condition, a pointwise boundedness condition stated explicitly by Tao [123] in his work on constellations in the Gaussian primes but also used implicitly in [69].

To state the main theorem, we will assume the definition of the k-linear forms condition. The formal definition, which may be found in Section 3.1 below, is stated for measures rather than sets but we will ignore this relatively minor distinction here, reserving a more complete discussion of our terminology for there.

Theorem 3.0.2 (Relative Szemerédi theorem). *For every natural number $k \geq 3$ and every $\delta > 0$, if $S \subset \mathbb{Z}_N$ satisfies the k-linear forms condition and N is sufficiently large, then any subset of S of relative density at least δ contains an arithmetic progression of length k.*

One of the immediate advantages of this theorem is that it simplifies the proof of the Green-Tao theorem. In addition to giving a simple proof of the relative Szemerédi theorem, it removes the need for the number-theoretic estimates involved in establishing the correlation condition for the almost primes. A further advantage is that, by removing the correlation condition, the relative Szemerédi theorem now applies to pseudorandom subsets of $\mathbb{Z}_N$ of density N^{-c_k}. With the correlation condition, one could only hope for such a theorem down to densities of the form $N^{-o(1)}$.

While the relative Szemerédi theorem is the main result of this chapter, the main advance is an approach to regularity in sparse pseudorandom hypergraphs. This allows us to prove analogues of several well-known combinatorial theorems relative to sparse pseudorandom hypergraphs. In particular, we prove a sparse analogue of the hypergraph removal lemma. It is from this that we derive our relative Szemerédi theorem. As always, applying the regularity method has two steps, a regularity lemma and a counting lemma. We provide novel approaches to both.

A counting lemma for subgraphs of sparse pseudorandom graphs was already proved in Chapter 2. In this chapter, we simplify and streamline the approach taken there in order to prove a counting lemma for subgraphs of sparse pseudorandom hypergraphs. This result is the key technical step in our proof and, perhaps, the main contribution of this chapter. Apart from the obvious difficulties in passing from graphs to hypergraphs, the crucial difference between this chapter and Chapter 2 is in the type of pseudorandomness considered. For graphs, we have a long-established notion of pseudorandomness known as jumbledness. The greater part of Chapter 2 is then concerned with optimizing the jumbledness condition which is necessary for counting a particular graph H. For hypergraphs, we use an analogue of the linear forms condition first considered by Tao [123]. It says that our hypergraph is pseudorandom enough for counting H within subgraphs if it contains asymptotically the correct count for the 2-blow-up of H and all its subgraphs.

We also use an alternative approach to regularity in sparse hypergraphs. While it would be natural to use a sparse hypergraph regularity lemma (and, following our approach in Chapter 2, this was how we initially proceeded), it suffices to use a weak sparse hypergraph regularity lemma which is an extension of the weak regularity lemma of Frieze and Kannan [53]. This is also closely related to the transference theorem used by Green and Tao (see, for example, [65] or [102, 131], where it is also referred to as the dense model theorem).

With both a regularity lemma and a counting lemma in place, it is then a straightforward matter to prove a relative extension of the famous hypergraph removal lemma [63, 96, 106, 107, 124]. Such a theorem was first derived by Tao [123] in his work on constellations in the Gaussian primes but, like the Green-Tao relative Szemerédi theorem, needs both a correlation condition and a dual function condition.[1] Our approach removes these conditions. The final step in the proof of the relative Szemerédi theorem is then a standard reduction used to derive Szemerédi's theorem from the hypergraph removal lemma. The details of this reduction already appear in [123] but we include them here for completeness. In fact, the chapter is self-contained apart from assuming the hypergraph removal lemma.

In Section 3.1, we state our results, including the relative Szemerédi theorem and the removal, regularity, and counting lemmas. In Section 3.2, we deduce the relative multidimensional Szemerédi theorem from our relative hypergraph removal lemma. In Section 3.3, we prove the removal lemma from the regularity and counting lemmas. We prove our weak sparse hypergraph regularity lemma in Section 3.4 and the associated counting lemma in Section 3.5. We conclude, in Section 3.6, with some remarks.

3.1 Definitions and results

Notation. *Dependence on N.* We consider functions $\nu = \nu^{(N)}$, where N (usually suppressed) is assumed to be some large integer. We write $o(1)$ for a quantity that tends to zero as $N \to \infty$ along some subset of $\mathbb{Z}$. If the rate at which the quantity tends to zero depends on some other parameters (e.g., k, δ), then we put these parameters in the subscript (e.g., $o_{k,\delta}(1)$).

Expectation. We write $\mathbb{E}[f(x_1, x_2, \dots)|P]$ for the expectation of $f(x_1, x_2, \dots)$ when the variables are chosen uniformly out of all possibilities satisfying P. We write $\mathbb{E}[f(x_1, x_2, \dots)|x_1 \in A_1, x_2 \in A_2, \dots]$ for the expectation of $f(x_1, x_2, \dots)$ when each x_i is chosen uniformly and independently at random from A_i.

[1] The problem of relative hypergraph removal was also recently considered by Towsner [130].

3.1.1 A relative Szemerédi theorem

Here is an equivalent weighted version of Szemerédi's theorem as formulated, for example, in [69, Prop. 2.3].

Theorem 3.1.1 (Szemerédi's theorem, weighted version). *For every $k \geq 3$ and $\delta > 0$, there exists $c > 0$ such that for N sufficiently large and any nonnegative function $f \colon \mathbb{Z}_N \to [0,1]$ satisfying $\mathbb{E}[f] \geq \delta$,*

$$\mathbb{E}[f(x)f(x+d)f(x+2d)\cdots f(x+(k-1)d)|x,d \in \mathbb{Z}_N] \geq c. \tag{3.1}$$

A relative Szemerédi theorem would instead ask for the nonnegative function f to be bounded above by a measure ν instead of the constant function f. For us, a measure will be any nonnegative function on $\mathbb{Z}_N$. We do not explicitly assume the additional condition that

$$\mathbb{E}[\nu(x)|x \in \mathbb{Z}_N] = 1 + o(1),$$

but this property follows from the linear forms condition that we will now assume. Such measures are more general than subsets, as any subset $S \subseteq \mathbb{Z}_N$ (e.g., in Theorem 3.0.2) can be thought of as a measure on $\mathbb{Z}_N$ taking value $N/|S|$ on S and 0 elsewhere. The dense case, as in Theorem 3.1.1, corresponds to taking $\nu = 1$. Our notion of pseudorandomness for measures ν on $\mathbb{Z}_N$ is now as follows.

Definition 3.1.2 (Linear forms condition). A nonnegative function $\nu = \nu^{(N)} : \mathbb{Z}_N \to \mathbb{R}_{\geq 0}$ is said to obey the *k-linear forms condition* if one has

$$\mathbb{E}\Big[\prod_{j=1}^{k} \prod_{\omega \in \{0,1\}^{[k]\setminus\{j\}}} \nu\Big(\sum_{i=1}^{k}(i-j)x_i^{(\omega_i)}\Big)^{n_{j,\omega}} \Big| x_1^{(0)}, x_1^{(1)}, \dots, x_k^{(0)}, x_k^{(1)} \in \mathbb{Z}_N\Big] = 1+o(1) \tag{3.2}$$

for any choices of exponents $n_{j,\omega} \in \{0,1\}$.

Example 3.1.3. For $k = 3$, condition (3.2) says that

$$\mathbb{E}[\nu(y+2z)\nu(y'+2z)\nu(y+2z')\nu(y'+2z')\nu(-x+z)\nu(-x'+z)\nu(-x+z')\nu(-x'+z') \cdot \nu(-2x-y)\nu(-2x'-y)\nu(-2x-y')\nu(-2x'-y')|x,x',y,y',z,z' \in \mathbb{Z}_N] = 1 + o(1)$$

and similar conditions hold if one or more of the twelve ν factors in the expectation are erased.

Our linear forms condition is much weaker than that used in Green and Tao [69]. In particular, Green and Tao need to assume that pointwise estimates such as

$$\mathbb{E}[\nu(a+x)\nu(a+y)\nu(a+x+y)|x,y \in \mathbb{Z}_N] = 1 + o(1)$$

hold uniformly over all $a \in \mathbb{Z}_N$. Such linear forms do not arise in our proof. Moreover, to prove their relative Szemerédi theorem, Green and Tao need to assume a further pseudorandomness condition, which they call the correlation condition. This condition also does not arise in our proofs. Indeed, we prove that a relative Szemerédi theorem holds given only the linear forms condition defined above.

Theorem 3.1.4 (Relative Szemerédi theorem). *For every $k \geq 3$ and $\delta > 0$, there exists $c > 0$ such that if $\nu \colon \mathbb{Z}_N \to \mathbb{R}_{\geq 0}$ satisfies the k-linear forms condition, N is sufficiently large, and $f \colon \mathbb{Z}_N \to \mathbb{R}_{\geq 0}$ satisfies $0 \leq f(x) \leq \nu(x)$ for all $x \in \mathbb{Z}_N$ and $\mathbb{E}[f] \geq \delta$, then*

$$\mathbb{E}[f(x)f(x+d)f(x+2d)\cdots f(x+(k-1)d)|x,d \in \mathbb{Z}_N] \geq c. \tag{3.3}$$

We note that both here and in Theorem 3.0.2, the phrase "N is sufficiently large" indicates not only a dependency on δ and k as in the usual version of Szemerédi's theorem but also a dependency on the $o(1)$ term in the linear forms condition. We will make a similar assumption in many of the theorems stated below.

We prove Theorem 3.1.4 using a new relative hypergraph removal lemma.[2] In the

[2]Green and Tao [69] prove a transference result that allows them to apply the dense version of

next subsection we set up the notation for hypergraphs and state the corresponding pseudorandomness hypothesis.

3.1.2 Hypergraphs

We borrow most of our notation and definitions from Tao [123, 124].

Definition 3.1.5 (Hypergraphs). Let J be a finite set and $r > 0$. Define $\binom{J}{r} = \{e \subseteq J : |e| = r\}$ to be the set of all r-element subsets of J. An *r-uniform hypergraph* on J is defined to be any subset $H \subseteq \binom{J}{r}$.

Definition 3.1.6 (Hypergraph system). A *hypergraph system* is a quadruple $V = (J, (V_j)_{j\in J}, r, H)$, where J is a finite set, $(V_j)_{j\in J}$ is a collection of finite non-empty sets indexed by J, $r \geq 1$ is a positive integer, and $H \subseteq \binom{J}{r}$ is an r-uniform hypergraph. For any $e \subseteq J$, we set $V_e := \prod_{j\in e} V_j$. For any $x = (x_j)_{j\in J} \in V_J$ and any subset $J' \subseteq J$, we write $x_{J'} = (x_j)_{j\in J'} \in V_{J'}$ to mean the natural projection of x onto the coordinates J'. Finally, for any $e \subseteq J$, we write ∂e for the set $\{f \subsetneq e : |f| = |e| - 1\}$, the skeleton of e.

Definition 3.1.7 (Weighted hypergraphs). Let $V = (J, (V_j)_{j\in J}, r, H)$ be a hypergraph system. A *weighted hypergraph* on V is a collection $g = (g_e)_{e\in H}$ of functions $g_e \colon V_e \to \mathbb{R}_{\geq 0}$ indexed by H. We write 0 and 1 to denote the constant-valued weighted hypergraphs of uniform weight 0 and 1, respectively. Given two weighted hypergraphs g and ν on the same hypergraph system, we write $g \leq \nu$ to mean that $g_e \leq \nu_e$ for all e, which in turn means that $g_e(x_e) \leq \nu_e(x_e)$ for all $x_e \in V_e$.

The weighted hypergraph ν plays an analogous role to the ν in Theorem 3.1.4, with $\nu = 1$ again corresponding to the dense case. We have an analogous linear forms condition for ν as a weighted hypergraph. We use the following indexing notation. For a finite set e and $\omega \in \{0,1\}^e$, we write $x_e^{(\omega)}$ to mean the tuple $(x_j^{(\omega_j)})_{j\in e}$. We also write $x_e^{(0)} := (x_j^{(0)})_{j\in e}$ and similarly with $x_e^{(1)}$.

Szemerédi's theorem as a black box. This allows them to show that the optimal c in (3.3) can be taken to be the same as the optimal c in (3.1). The proof in this chapter goes through the hypergraph removal lemma and thus does not obtain the same c. Nevertheless, one can obtain our result with the same c by modifying the argument to an arithmetic setting, as done in Chapter 4.

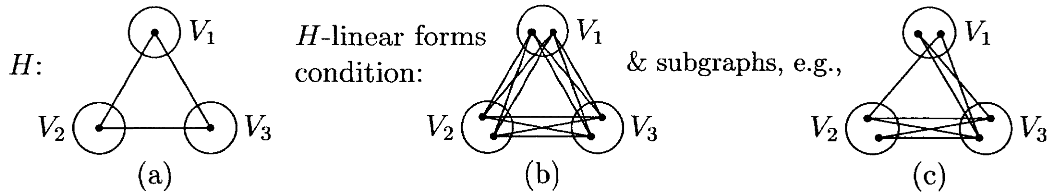


Figure 3-1: Linear forms conditions for $H = K_3$. See Example 3.1.9.

Definition 3.1.8 (Linear forms condition). A weighted hypergraph $\nu = \nu^{(N)}$ on the hypergraph system $V = V^{(N)} = (J, (V_j^{(N)})_{j\in J}, r, H)$ is said to obey the *H-linear forms condition* (or simply the *linear forms condition* if there is no confusion) if one has

$$\mathbb{E}\Big[\prod_{e\in H}\prod_{\omega\in\{0,1\}^e} \nu_e(x_e^{(\omega)})^{n_{e,\omega}} \Big| x_J^{(0)}, x_J^{(1)} \in V_J\Big] = 1 + o(1) \tag{3.4}$$

for any choices of exponents $n_{e,\omega} \in \{0,1\}$.

Example 3.1.9. Let H be the set of all pairs in $J = \{1,2,3\}$. The linear forms condition says that

$$\mathbb{E}\Big[\prod_{ij=12,13,23} \nu_{ij}(x_i,x_j)\nu_{ij}(x_i',x_j)\nu_{ij}(x_i,x_j')\nu_{ij}(x_i',x_j') \\ \Big| x_1,x_1'\in V_1,\ x_2,x_2'\in V_2,\ x_3,x_3'\in V_3\Big] = 1 + o(1)$$

and similarly if one or more of the twelve ν factors are deleted. This expression represents the weighted homomorphism density of $K_{2,2,2}$ in the weighted tripartite graph given by ν, as illustrated in Figure 3-1(b) (the vertices of $K_{2,2,2}$ must map into the corresponding parts). Deleting some ν factors corresponds to considering various subgraphs of $K_{2,2,2}$, e.g., Figure 3-1(c).

In general, the H-linear forms condition says that ν has roughly the expected density for the 2-blow-up[3] of H as well as any subgraph of the 2-blow-up. Our linear

[3] By the 2-blow-up of H we mean the hypergraph consisting of vertices $j^{(0)}, j^{(1)}$ for each $j \in J$, and edges $e^{(\omega)} := \{j^{(\omega_j)} : j \in e\}$ for any $e \in H$ and $\omega \in \{0,1\}^e$. We actually do not need the full strength of this assumption. It suffices to assume that ν has roughly the expected density for any subgraph of a *weak* 2-blow-up of H, where by a weak 2-blow-up we mean the following. Fix some edge $e_1 \in H$ (we will need to assume the condition for all e_1). The weak 2-blow-up of H with respect to e_1 is the subgraph of the usual 2-blow-up consisting of all edges $e^{(\omega)}$ where $\omega_i = \omega_j$ for

forms condition for hypergraphs coincides with the one used by Tao [123, Def. 2.8], although in [123] one assumes additional pseudorandomness hypotheses on ν known as the dual function condition and the correlation condition.

3.1.3 Hypergraph removal lemma

The hypergraph removal lemma was first proved by Gowers [63] and by Nagle, Rödl, Schacht, and Skokan [96, 106, 107]. It states that for every r-uniform hypergraph H on h vertices, every r-uniform hypergraph on n vertices with $o(n^h)$ copies of H can be made H-free by removing $o(n^r)$ edges. As first explicitly stated and proved by Tao [124], the proof of the hypergraph removal lemma further gives that the edges can be removed in a low complexity way (this idea will soon be made formal). We will use a slightly stronger version, where edges are given weights in the interval $[0,1]$. This readily follows from the usual version by a simple rounding argument, as done in [123, Thm. 3.7]. We state this result as Theorem 3.1.11 below.

Definition 3.1.10. For any set e of size r and any $E_e \subseteq V_e = \prod_{j\in e} V_j$, we define the *complexity* of E_e to be the minimum integer T such that there is a partition of E_e into T sets $E_{e,1}, \dots, E_{e,T}$, so that each $E_{e,i}$ is the set of r-cliques of some $(r-1)$-uniform hypergraph, meaning that there exists some $B_{f,i} \subseteq V_f$ for each $f \in \partial e$ so that $1_{E_{e,i}}(x_e) = \prod_{f\in\partial e} 1_{B_{f,i}}(x_f)$ for all $x_e \in V_e$.

Theorem 3.1.11 (Weighted hypergraph removal lemma). *For every $\epsilon > 0$ and finite set J, there exists $\delta > 0$ and $T > 0$ such that the following holds. Let $V = (J, (V_j)_{j\in J}, r, H)$ be a hypergraph system. Let g be a weighted hypergraph on V satisfying $0 \le g \le 1$ and*

$$\mathbb{E}\Big[\prod_{e\in H} g_e(x_e)\Big| x \in V_J\Big] \le \delta.$$

Then for each $e \in H$ there exists a set $E'_e \subseteq V_e$ for which $V_e \setminus E'_e$ has complexity at

any $i, j \in e \setminus e_1$. This weaker version of the H-linear forms condition is all we shall use for the proof, although everything to follow will be stated as in Definition 3.1.8 for clarity.

most T and such that

$$\prod_{e\in H} 1_{E'_e}(x_e) = 0 \text{ for all } x \in V_J$$

and for all $e \in H$ one has

$$\mathbb{E}[g_e(x_e)1_{V_e\setminus E'_e}(x_e)|x_e \in V_e] \le \epsilon.$$

We prove a relativized extension of the hypergraph removal lemma. A relative hypergraph removal lemma was already established by Tao in [123], where he assumed the majorizing measure satisfies three conditions: the linear forms condition, the correlation condition, and the dual function condition. We again show that a linear forms condition is sufficient.

Theorem 3.1.12 (Relative hypergraph removal lemma). *For every $\epsilon > 0$ and finite set J, there exists $\delta > 0$ and $T > 0$ such that the following holds. Let $V = (J, (V_j)_{j\in J}, r, H)$ be a hypergraph system. Let ν and g be weighted hypergraphs on V. Suppose $0 \le g \le \nu$, ν satisfies the H-linear forms condition, and N is sufficiently large. If*

$$\mathbb{E}\Big[\prod_{e\in H} g_e(x_e)\Big|x \in V_J\Big] \le \delta,$$

then for each $e \in H$ there exists a set $E'_e \subseteq V_e$ for which $V_e \setminus E'_e$ has complexity at most T and such that

$$\prod_{e\in H} 1_{E'_e}(x_e) = 0 \text{ for all } x \in V_J$$

and for all $e \in H$ one has

$$\mathbb{E}[g_e(x_e)1_{V_e\setminus E'_e}(x_e)|x_e \in V_e] \le \epsilon.$$

In Section 3.3 we will deduce Theorem 3.1.12 from Theorem 3.1.11 by applying the weak regularity lemma and the counting lemma which are stated in the next two subsections.

3.1.4 Weak hypergraph regularity

The Frieze-Kannan weak regularity lemma [53] allows one to approximate in cut-norm a matrix (or graph) with entries in the interval $[0, 1]$ by another matrix of low complexity. A major advantage over simply applying Szemerédi's regularity lemma is that the complexity has only an exponential dependence on the approximation parameter, as opposed to the tower-type bound that is incurred by Szemerédi's regularity lemma. Unfortunately, these regularity lemmas are not meaningful for sparse graphs as the error term is too large in this setting. Following sparse extensions of Szemerédi's regularity lemma by Kohayakawa [75] and Rödl, a sparse extension of the weak regularity lemma was proved by Bollobás and Riordan [15] and by Coja-Oghlan, Cooper, and Frieze [25]. In [25], they further generalize this to r-dimensional tensors (or r-uniform hypergraphs), but it only gives an approximation which is close in density on all hypergraphs induced by large vertex subsets. In order to prove a relative hypergraph removal lemma, we will need a stronger approximation, which is close in density on all dense r-uniform hypergraphs formed by the clique set of some $(r-1)$-uniform hypergraph. In Section 3.4, we will prove a more general sparse regularity lemma. For now, we state the result in the form that we need.

The weak regularity lemma approximates a weighted hypergraph g on V by another weighted hypergraph $\tilde{g}$ of bounded complexity which satisfies $0 \le \tilde{g} \le 1$. One can think of $\tilde{g}$ as a dense approximation of g. The following definition makes precise in what sense $\tilde{g}$ approximates g.

Definition 3.1.13 (Discrepancy pair). Let e be a finite set and $g_e, \tilde{g}_e \colon \prod_{j \in e} V_j \to \mathbb{R}_{\ge 0}$ be two nonnegative functions. We say that $(g_e, \tilde{g}_e)$ is an *ϵ-discrepancy pair* if for all subsets $B_f \subseteq V_f$, $f \in \partial e$, one has

$$\left| \mathbb{E}\left[(g_e(x_e) - \tilde{g}_e(x_e)) \prod_{f \in \partial e} 1_{B_f}(x_f) \middle| x_e \in V_e \right] \right| \le \epsilon. \tag{3.5}$$

For two weighted hypergraphs g and $\tilde{g}$ on $(J, (V_j)_{j \in J}, r, H)$, we say that $(g, \tilde{g})$ is an *ϵ-discrepancy pair* if $(g_e, \tilde{g}_e)$ is an ϵ-discrepancy pair for all $e \in H$.

One needs an additional hypothesis on g in order to prove a weak regularity lemma. The condition roughly says that g contains "no dense spots."

Definition 3.1.14 (Upper regular). Let e be a finite set, $g_e \colon \prod_{j \in e} V_j \to \mathbb{R}_{\geq 0}$ a nonnegative function, and $\eta > 0$. We say that g_e is *upper η-regular* if for all subsets $B_f \subseteq V_f$, $f \in \partial e$, one has

$$\mathbb{E}\Big[(g_e(x_e) - 1) \prod_{f \in \partial e} 1_{B_f}(x_f) \Big| x_e \in V_e\Big] \leq \eta. \tag{3.6}$$

A hypergraph g on on $(J, (V_j)_{j \in J}, r, H)$ is upper η-regular if g_e is upper η-regular for all $e \in H$.

Note that unlike (3.5), there is no absolute value on the left-hand side of (3.6). The upper regularity hypothesis is needed for establishing the sparse regularity lemma. Fortunately, this mild hypothesis is automatically satisfied in our setting. We will say more about this in Section 3.5.2.

Lemma 3.1.15. *Let $V = (J, (V_j)_{j \in J}, r, H)$ be a hypergraph system. Let ν and g be weighted hypergraphs on V. Suppose $0 \leq g \leq \nu$ and ν satisfies the H-linear forms condition. Then g is upper $o(1)$-regular.*

Define the *complexity* of a function $g \colon V_e \to [0, 1]$ to be the minimum T such that there is a partition of V_e into T subgraphs $S_1, \ldots S_T$, each of which is the set of r-cliques of some $(r-1)$-uniform hypergraph (see Definition 3.1.10), and such that g is constant on each S_i. We state the regularity lemma below with a complexity bound on $\tilde{g}$, although the complexity bound will not actually be needed for our application.

Theorem 3.1.16 (Sparse weak regularity lemma). *For any $\epsilon > 0$ and function $g \colon V_1 \times \cdots \times V_r \to \mathbb{R}_{\geq 0}$ which is upper η-regular with $\eta \leq 2^{-40r/\epsilon^2}$, there exists $\tilde{g} \colon V_1 \times \cdots \times V_r \to [0, 1]$ with complexity at most $2^{20r/\epsilon^2}$ such that $(g, \tilde{g})$ is an ϵ-discrepancy pair.*

The special case $r = 2$ is the sparse extension of the Frieze-Kannan weak regularity lemma.

3.1.5 Counting lemma

Informally, the counting lemma says that if $(g, \tilde{g})$ is an ϵ-discrepancy pair, with the additional assumption that $g \le \nu$ and $\tilde{g} \le 1$, then the density of H in $\tilde{g}$ is close to the density of H in g. This sparse counting lemma is perhaps the most novel ingredient in this thesis.

Theorem 3.1.17 (Counting lemma). *For every $\gamma > 0$ and finite set J, there exists an $\epsilon > 0$ so that the following holds. Let $V = (J, (V_j)_{j\in J}, r, H)$ be a hypergraph system and ν, g, $\tilde{g}$ be weighted hypergraphs on V. Suppose that ν satisfies the H-linear forms condition and N is sufficiently large. Suppose also that $0 \le g \le \nu$, $0 \le \tilde{g} \le 1$, and $(g, \tilde{g})$ is an ϵ-discrepancy pair. Then*

$$\left|\mathbb{E}\Big[\prod_{e\in H} g_e(x_e)\Big| x \in V_J\Big] - \mathbb{E}\Big[\prod_{e\in H} \tilde{g}_e(x_e)\Big| x \in V_J\Big]\right| \le \gamma. \tag{3.7}$$

As a corollary, Theorem 3.1.17 also holds if the hypothesis $0 \le \tilde{g} \le 1$ is replaced by $0 \le \tilde{g} \le \nu$. Indeed, we can use the weak regularity lemma, Theorem 3.1.16, to find a common 1-bounded approximation to g and $\tilde{g}$. The result then follows from Theorem 3.1.17 and the triangle inequality.

To summarize, to get a counting lemma for a fixed hypergraph H in a subgraph of a pseudorandom host hypergraph, it suffices to know that the host hypergraph has approximately the expected count for a somewhat larger family of hypergraphs (namely, subgraphs of the 2-blow-up of H).

3.2 The relative Szemerédi theorem

In this section, we deduce the relative Szemerédi theorem, Theorem 3.1.4, from the relative hypergraph removal lemma, Theorem 3.1.12. We use the relative hypergraph removal lemma to prove a relative arithmetic removal lemma, Theorem 3.2.3. This result then easily implies a relative version of the multidimensional Szemerédi theorem of Furstenberg and Katznelson [54]. This is Theorem 3.2.1 below. The relative Szemerédi theorem, Theorem 3.1.4, follows as a special case of Theorem 3.2.1 by setting $Z = Z' = \mathbb{Z}_N$ and $\phi_j(d) = (j-1)d$. One may easily check that the linear forms condition for the resulting hypergraph is satisfied if $\nu\colon \mathbb{Z}_N \to \mathbb{R}_{\geq 0}$ satisfies the k-linear forms condition.

The statement and proof of Theorem 3.2.1 closely follows the write-up in Tao [123, Thm 2.18], adapted in a straightforward way to our new pseudorandomness conditions as well as to the slightly more general setting of functions instead of subsets. Earlier versions of this type of argument for deducing Szemerédi-type results (in the dense setting) from graph and hypergraph removal lemmas were given by Ruzsa and Szemerédi [108], Frankl and Rödl [51], and Solymosi [118, 117].

Theorem 3.2.1 (Relative multidimensional Szemerédi theorem). *For a finite set J and $\delta > 0$, there exists $c > 0$ so that the following holds. Let Z, Z' be two finite additive groups and let $(\phi_j)_{j\in J}$ be a finite collection of group homomorphisms $\phi_j : Z \to Z'$ from Z to Z'. Assume that the elements $\{\phi_i(d) - \phi_j(d) : i, j \in J, d \in Z\}$ generate Z' as an abelian group. Let $\nu : Z' \to \mathbb{R}_{\geq 0}$ be a nonnegative function with the property that in the hypergraph system $V = (J, (V_j)_{j\in J}, r, H)$, with $V_j := Z$, $r := |J| - 1$, and $H := \binom{J}{r}$, the weighted hypergraph $(\nu_e)_{e\in H}$ defined by*

$$\nu_{J\setminus\{j\}}((x_i)_{i\in J\setminus\{j\}}) := \nu\Big(\sum_{i\in J\setminus\{j\}} (\phi_i(x_i) - \phi_j(x_i))\Big)$$

satisfies the H-linear forms condition. Assume that N is sufficiently large. Then, for any $f\colon Z' \to \mathbb{R}_{\geq 0}$ satisfying $0 \leq f(x) \leq \nu(x)$ for all $x \in Z'$ and $\mathbb{E}[f] \geq \delta$,

$$\mathbb{E}\Big[\prod_{j\in J} f(a + \phi_j(d))\Big| a \in Z', d \in Z\Big] \geq c. \tag{3.8}$$

Example 3.2.2. Let $S \subset \mathbb{Z}_N \times \mathbb{Z}_N$. Suppose the associated measure $\nu = \frac{N}{|S|}1_S$ satisfies

$$\begin{aligned}&\mathbb{E}[\nu(x,y)\nu(x',y)\nu(x,y')\nu(x',y')\nu(x,z-x)\nu(x',z-x')\nu(x,z'-x)\nu(x',z'-x')\\&\quad\cdot\nu(z-y,y)\nu(z-y',y')\nu(z'-y,y)\nu(z'-y',y')|x,x',y,y',z,z'\in\mathbb{Z}_N] = 1 + o(1)\end{aligned}$$

and similar conditions hold if any subset of the twelve ν factors in the expectation are erased. Then any corner-free subset of S has size $o(|S|)$. Here a *corner* in $\mathbb{Z}_N^2$ is a set of the form $\{(x,y),(x+d,y),(x,y+d)\}$ for some $d \neq 0$. This claim follows from Theorem 3.2.1 by setting $Z = \mathbb{Z}_N$, $Z' = \mathbb{Z}_N^2$, $\phi_0(d) = (0,0)$, $\phi_1(d) = (d,0)$, $\phi_2(d) = (0,d)$.

As in [123, Remark 2.19], we note that the hypothesis that $\{\phi_i(d) - \phi_j(d) : i,j \in J, d \in Z\}$ generate Z' can be dropped by foliating Z' into cosets. However, this results in a change to the linear forms hypothesis on ν, namely, that it must be assumed on every coset.

We shall prove Theorem 3.2.1 by proving a somewhat more general removal-type result for arithmetic patterns.

Theorem 3.2.3 (Relative arithmetic removal lemma). *For every finite set J and $\epsilon > 0$, there exists $c > 0$ so that the following holds. Let $Z, Z', (\phi_j)_{j\in J}, \nu$ be the same as in Theorem 3.2.1. For any collection of functions $\{f_j \colon Z' \to \mathbb{R}_{\geq 0}\}_{j\in J}$ satisfying $0 \leq f_j(x) \leq \nu(x)$ for all $x \in Z'$ and $j \in J$, and such that*

$$\mathbb{E}\Big[\prod_{j\in J} f_j(a + \phi_j(d))\Big| a \in Z', d \in Z\Big] \leq c \tag{3.9}$$

one can find $A_j \subseteq Z'$ for each $j \in J$ so that

$$\prod_{j\in J} 1_{A_j}(a + \phi_j(d)) = 0 \quad \text{for all } a \in Z', d \in Z \tag{3.10}$$

and

$$\mathbb{E}[f_j(x) 1_{Z'\setminus A_j}(x) | x \in Z'] \leq \epsilon \quad \text{for all } j \in J. \tag{3.11}$$

Theorem 3.2.1 follows from Theorem 3.2.3 by setting $f_j = f$ for all $j \in J$ and $\epsilon < \delta/(r+1)$. Indeed, if the conclusion (3.8) fails, then Theorem 3.2.3 implies that there exists $A_j \subseteq Z'$ for each $j \in J$ satisfying (3.10) and (3.11). The A_j's cannot have a common intersection, or else (3.10) fails for $d = 0$. It follows that $\{Z' \setminus A_j : j \in J\}$ covers Z', and hence (3.11) implies that $\mathbb{E}[f] \leq \sum_j \mathbb{E}[f_j 1_{Z'\setminus A_j}] \leq (r+1)\epsilon < \delta$, which contradicts the hypothesis $\mathbb{E}[f] \geq \delta$.

Proof of Theorem 3.2.3. Let $V = (J, (V_j), r, H)$ be as in the statement of Theorem 3.2.1. Write $e_j := J \setminus \{j\} \in H$. Define the weighted hypergraph g on V by setting

$$g_{e_j}(x_{e_j}) := f_j(\psi_j(x_{e_j})) \quad \text{for all } j \in J$$

where $\psi_j \colon V_{e_j} \to Z'$ is defined by

$$\psi_j(x_{e_j}) = \sum_{i \in e_j} (\phi_i(x_i) - \phi_j(x_i)) = a + \phi_j(d) \tag{3.12}$$

where

$$a = \sum_{i \in J} \phi_i(x_i) \quad \text{and} \quad d = -\sum_{i \in J} x_i. \tag{3.13}$$

Then, for all $x \in V$ and a, d defined in (3.13), we have

$$\prod_{j \in J} g_{e_j}(x_{e_j}) = \prod_{j \in J} f_j(a + \phi_j(d)). \tag{3.14}$$

The homomorphism $x \mapsto (a, d) \colon V \to Z' \times Z$ given by (3.13) is surjective: the image contains $\{(\phi_i(d) - \phi_j(d), 0) : i, j \in J, d \in Z\}$ and hence all of $Z' \times \{0\}$. Moreover, the image also contains $\{(-\phi_i(d), d) : i \in J, d \in Z\}$. Together, these sets generate all of $Z' \times Z$. It follows that (a, d) varies uniformly over $Z' \times Z$ as x varies uniformly over V_J, and so (3.14) implies that

$$\mathbb{E}\Big[\prod_{j \in J} g_{e_j}(x_{e_j}) \Big| x \in V_J\Big] = \mathbb{E}\Big[\prod_{j \in J} f_j(a + \phi_j(d)) \Big| a \in Z', d \in Z\Big] \leq c.$$

By the relative hypergraph removal lemma, for c small enough (depending on J and ϵ), we can find a subset $E'_j \subset V_{e_j}$ for each $j \in J$ such that

$$\prod_{j \in J} 1_{E'_j}(x_{e_j}) = 0 \quad \text{for all } x \in V_J \tag{3.15}$$

and

$$\mathbb{E}[g_{e_j}(x_{e_j}) 1_{V_{e_j} \setminus E'_j}(x_{e_j}) | x_{e_j} \in V_{e_j}] \leq \epsilon/(r+1) \quad \text{for all } j \in J.$$

For each $j \in J$, define $A_j \subseteq Z'$ by

$$A_j := \{z' \in Z' : |\psi_j^{-1}(z') \cap E'_j| > \tfrac{r}{r+1}|\psi_j^{-1}(z')|\}. \tag{3.16}$$

In other words, A_j contains $z' \in Z'$ if the hypergraph removal lemma removes less than a $1/(r+1)$ fraction of the edges in V_{e_j} representing z' via ψ_j.

For any $z' \in Z' \setminus A_j$, on the fiber $\psi^{-1}(z')$ the function g_{e_j} takes the common value $f_j(z')$. Furthermore, by (3.16), on this fiber, the expectation of $1_{V_{e_j} \setminus E'_j}$ is at least $1/(r+1)$. Hence

$$\mathbb{E}[f_j(x) 1_{Z' \setminus A_j}(x) | x \in Z'] \leq (r+1)\mathbb{E}[g_{e_j}(x_{e_j}) 1_{V_{e_j} \setminus E'_j}(x_{e_j}) | x_{e_j} \in V_{e_j}] \leq \epsilon.$$

This proves (3.11). To prove (3.10), suppose for some $a \in Z', d \in Z$ we have $a + \phi_j(d) \in A_j$ for all $j \in J$. Let $V_J^{a,d} \subset V_J$ consist of all $x \in V_J$ satisfying (3.13). Then $\psi_j(x_{e_j}) = a + \phi_j(d)$ for all $x \in V_J^{a,d}$ by (3.12), and in fact $\psi_j^{-1}(a + \phi_j(d))$ is the projection of $V_J^{a,d}$ onto V_{e_j}. By (3.16), more than an $\frac{r}{r+1}$ fraction of this projection is in E'_j. It follows by the pigeonhole principle (or a union bound on the complement) that there exists some $x \in V_J^{a,d}$ such that $x_{e_j} \in E'_j$ for every $j \in J$. But this contradicts (3.15). Thus (3.10) holds. ■

3.3 The relative hypergraph removal lemma

Proof of Theorem 3.1.12. By Lemma 3.1.15, ν is upper $o(1)$-regular, so we can apply the weak sparse hypergraph regularity lemma (Theorem 3.1.16) to find functions $\tilde{g}_e : V_e \to [0,1]$ for every $e \in H$ so that $(g, \tilde{g})$ is an $o(1)$-discrepancy pair. By the counting lemma (Theorem 3.1.17), we have

$$\mathbb{E}\Big[\prod_{e \in H} \tilde{g}_e(x_e) \Big| x \in V_J\Big] = \mathbb{E}\Big[\prod_{e \in H} g_e(x_e) \Big| x \in V_J\Big] + o(1) \leq \delta + o(1).$$

The dense weighted hypergraph removal lemma (Theorem 3.1.11) tells us that for each $e \in H$ we can choose $E'_e \subset V_e$ for which $V_e \setminus E'_e$ has complexity $O_\delta(1)$ (i.e., at most some constant depending on δ) and such that

$$\prod_{e\in H} 1_{E'_e}(x_e) = 0 \quad \text{for all } x \in V_J$$

and, as long as δ is small enough and N is large enough, we have

$$\mathbb{E}[\tilde{g}_e(x_e)1_{V_e\setminus E'_e}(x_e)|x_e \in V_e] \le \epsilon/2 \quad \text{for all } e \in H. \tag{3.17}$$

As $V_e \setminus E'_e$ has complexity $O_\delta(1)$, there is a partition of $V_e \setminus E'_e$ into $O_\delta(1)$ hypergraphs F_{ei} each of which is the set of r-cliques of some $(r-1)$-uniform hypergraph. We have

$$\begin{aligned}|\mathbb{E}[(\tilde{g}_e - g_e)(x_e)1_{V_e\setminus E'_e}(x_e)|x_e \in V_e]| &\le \sum_i |\mathbb{E}[(\tilde{g}_e - g_e)(x_e)1_{F_{ei}}(x_e)|x_e \in V_e]| \\ &\le \sum_i o(1) = O_\delta(1)o(1) \le \epsilon/2 \quad \text{for all } e \in H.\end{aligned} \tag{3.18}$$

We used that $(g_e, \tilde{g}_e)$ is an $o(1)$-discrepancy pair on each of the terms of the sum, and the final inequality is true as long as N is large enough. Combining (3.17) and (3.18) we obtain

$$\mathbb{E}[g_e(x_e)1_{V_e\setminus E'_e}(x)|x_e \in V_e] \le \epsilon \quad \text{for all } e \in H.$$

This proves the claim. ■

3.4 The weak regularity lemma

Let X be a finite set and $g : X \to \mathbb{R}_{\ge 0}$. Let $\mathcal{F}$ be a family of subsets of X which is closed under intersection, $X \in \mathcal{F}$, all subsets of X of size one are in $\mathcal{F}$, and such that, for every $S \in \mathcal{F}$, there is a partition of X which contains S and consists of members of $\mathcal{F}$. For $t \ge 2$, the family $\mathcal{F}$ is *t-splittable* if for every $S \in \mathcal{F}$ there is a partition P of X into members of $\mathcal{F}$ such that $S \in P$ and $|P| \le t$. The *complexity* $p = p(f)$ of a function $f : X \to \mathbb{R}_{\ge 0}$ is the minimum p for which there is a partition $X = S_1 \cup \cdots \cup S_p$ into p subsets each in $\mathcal{F}$ such that f is constant on each S_i. We call $(g,\tilde{g})$ an *ϵ-discrepancy pair* if for all $A \in \mathcal{F}$,

$$\left|\mathbb{E}[(g-\tilde{g})1_A]\right| \leq \epsilon.$$

All expectations are done with the uniform measure on X. For P a partition of X, let g_P be the function on X given by $g_P(x) = \frac{\mathbb{E}[g1_A]}{\mathbb{E}[1_A]}$ when $x \in A \in P$. That is, $g_P(x)$ is the conditional expectation of $g(x)$ given the partition P and is constant on any part A of the partition.

The function g we call *upper η-regular* if for every $A \in \mathcal{F}$, we have

$$\mathbb{E}[g1_A] \leq \mathbb{E}[1_A] + \eta.$$

If g is upper η-regular, $A, B \in \mathcal{F}$, and $\mathcal{F}$ is t-splittable, then

$$\mathbb{E}[g1_{B\setminus A}] \leq \mathbb{E}[1_{B\setminus A}] + (t-1)\eta. \tag{3.19}$$

Indeed, in this case $B \setminus A$ can be partitioned into $t-1$ sets in $\mathcal{F}$ (we first split with respect to A and then consider the intersections of the parts of the partition with B). Applying the upper η-regularity condition to each of these sets and summing up the inequalities, we arrive at (3.19).

Following Scott [112], let $\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be the convex function given by

$$\phi(u) = \begin{cases} u^2 & \text{if } u \leq 2, \\ 4u-4 & \text{otherwise.} \end{cases}$$

For a partition P of X, let $\phi(P) = \mathbb{E}[\phi(g_P)]$, which is the mean ϕ-density of g with respect to the partition P. As ϕ takes only nonnegative values and $\phi(u) \leq 4u$, we have

$$0 \leq \phi(P) \leq 4\mathbb{E}[g_P] = 4\mathbb{E}[g].$$

Also, by the convexity of ϕ, it follows that if P' is a refinement of P, then $\phi(P') \geq \phi(P)$.

Lemma 3.4.1. *Let X and $\mathcal{F}$ as above be such that $\mathcal{F}$ is t-splittable. Let $0 < \epsilon, \eta < 1$ and $T = t^{20/\epsilon^2}$. For any $g : X \to \mathbb{R}_{\geq 0}$ which is upper η-regular with $\eta \leq \frac{\epsilon}{8tT}$, there is $\tilde{g} : X \to [0,1]$ with complexity at most T such that $(g, \tilde{g})$ is an ϵ-discrepancy pair.*

Proof. Let $\alpha = \frac{\epsilon^2}{4}$. We first find a partition P of X into members of $\mathcal{F}$ with $|P| \leq t^{5/\alpha} = T$ such that for any refinement P' of P into members of $\mathcal{F}$ with $|P'| \leq t|P|$, we have $\phi(P') - \phi(P) < \alpha$. In order to construct P, we first recursively construct a sequence $P_0, P_1, \ldots$ of finer partitions of X into members of $\mathcal{F}$ so that $|P_j| \leq t^j$ and $\phi(P_j) \geq j\alpha$. We begin by considering the trivial partition $P_0 = \{X\}$, which satisfies $\phi(P_0) \geq 0$. At the beginning of step $j+1$, we have a partition P_j of X into members of $\mathcal{F}$ with $|P_j| \leq t^j$ and $\phi(P_j) \geq j\alpha$. If there exists a refinement P_{j+1} of X into members of $\mathcal{F}$ with $|P_{j+1}| \leq t|P_j|$ and $\phi(P_{j+1}) \geq \phi(P_j) + \alpha$, then we continue to step $j+2$. Otherwise, we may pick $P = P_j$ to be the desired partition. Note that this process must stop after at most $5/\alpha$ steps since $5 > 4(1+\eta) \geq 4\mathbb{E}[g] \geq \phi(P_j) \geq j\alpha$, where the second inequality follows from g being upper η-regular. We therefore arrive at the desired partition P.

Let $P : X = S_1 \cup \cdots \cup S_p$. Let $\tilde{g} : X \to [0,1]$, where $\tilde{g} = g_P \wedge 1$ is the minimum of g_P and the constant function 1. We will show that $(g_P, \tilde{g})$ is an $\frac{\epsilon}{4}$-discrepancy pair and (g_P, g) is a $\frac{3\epsilon}{4}$-discrepancy pair, which implies by the triangle inequality that $(g, \tilde{g})$ is an ϵ-discrepancy pair. As $\tilde{g}$ has complexity at most $|P| \leq T$, this will complete the proof.

We first show $(g_P, \tilde{g})$ is an $\frac{\epsilon}{4}$-discrepancy pair. Note that $g_P - \tilde{g}$ is nonnegative and constant on each part of P. If $S_i \in P$ and $g_P - \tilde{g} > 0$ on S_i, then also $g_P > 1$ and $\tilde{g} = 1$ on S_i. As g is upper η-regular, we have $\mathbb{E}[g1_{S_i}] \leq \mathbb{E}[1_{S_i}] + \eta$ and hence $\mathbb{E}[(g - \tilde{g})1_{S_i}] \leq \eta$. Therefore, by summing over all parts in the partition P, we see that if $A \in \mathcal{F}$,

$$0 \leq \mathbb{E}[(g_P - \tilde{g})1_A] \leq \mathbb{E}[(g_P - \tilde{g})] \leq \eta|P| \leq \eta T \leq \frac{\epsilon}{4},$$

and $(g_P, \tilde{g})$ is an $\frac{\epsilon}{4}$-discrepancy pair.

We next show that (g_P, g) is a $\frac{3\epsilon}{4}$-discrepancy pair, which completes the proof.

Suppose for contradiction that there is $A \in \mathcal{F}$ such that

$$|\mathbb{E}[(g_P - g)1_A]| > \frac{3\epsilon}{4}.$$

Let B be the union of all $S_i \cap A$, where $S_i \in P$, for which both $\mathbb{E}[1_{S_i \cap A}] \geq t\eta$ and $\mathbb{E}[1_{S_i \setminus A}] \geq t\eta$.

We claim that for each $S_i \in P$, we have

$$|\mathbb{E}[(g_P - g)(1_{A \cap S_i} - 1_{B \cap S_i})]| \leq 2t\eta. \tag{3.20}$$

Indeed, if $B \cap S_i = A \cap S_i$, then the left hand side of (3.20) is 0. Otherwise, $\mathbb{E}[1_{A \cap S_i}] \leq t\eta$ or $\mathbb{E}[1_{S_i \setminus A}] \leq t\eta$. In the first case, when $\mathbb{E}[1_{A \cap S_i}] \leq t\eta$, we have $1_{B \cap S_i}$ is identically 0, as well as

$$\mathbb{E}[g 1_{A \cap S_i}] \leq \mathbb{E}[1_{A \cap S_i}] + \eta \leq (t+1)\eta$$

and

$$\mathbb{E}[g_P 1_{A \cap S_i}] = \frac{\mathbb{E}[g 1_{S_i}]}{\mathbb{E}[1_{S_i}]}\mathbb{E}[1_{A \cap S_i}] \leq \frac{(\mathbb{E}[1_{S_i}] + \eta)}{\mathbb{E}[1_{S_i}]}\mathbb{E}[1_{A \cap S_i}] \leq \mathbb{E}[1_{A \cap S_i}] + \eta \leq (t+1)\eta,$$

from which (3.20) follows. In the second case, when $\mathbb{E}[1_{S_i \setminus A}] \leq t\eta$, we again have $1_{B \cap S_i}$ is identically 0, so that

$$\begin{aligned} \mathbb{E}[(g - g_P)(1_{A \cap S_i} - 1_{B \cap S_i})] &= \mathbb{E}[(g - g_P)1_{A \cap S_i}] = \mathbb{E}[(g - g_P)(1_{S_i} - 1_{S_i \setminus A})] \\ &= \mathbb{E}[(g - g_P)1_{S_i}] - \mathbb{E}[(g - g_P)1_{S_i \setminus A}] = -\mathbb{E}[(g - g_P)1_{S_i \setminus A}], \end{aligned}$$

and similar to the first case, using (3.19) to estimate $\mathbb{E}[g 1_{S_i \setminus A}]$ and $\mathbb{E}[g_P 1_{S_i \setminus A}]$, we get (3.20).

Notice that

$$|\mathbb{E}[(g_P - g)1_A] - \mathbb{E}[(g_P - g)1_B]| = |\mathbb{E}[(g_P - g)(1_A - 1_B)]| \leq |P| 2t\eta \leq \frac{\epsilon}{4},$$

where the first inequality follows by using (3.20) for each part S_i and the triangle inequality. Hence,

$$|\mathbb{E}[(g_P - g)1_B]| \geq |\mathbb{E}[(g_P - g)1_A]| - |\mathbb{E}[(g_P - g)1_A] - \mathbb{E}[(g_P - g)1_B]| > \frac{3\epsilon}{4} - \frac{\epsilon}{4} = \frac{\epsilon}{2}.$$

Let $\hat{P}$ be the refinement of P where S_i is also in $\hat{P}$ if $B \cap S_i = \emptyset$ and otherwise $S_i \cap B$ and $S_i \setminus B$ are parts of $\hat{P}$, and let P' be a refinement of $\hat{P}$ into at most $t|P|$ members of $\mathcal{F}$. The refinement P' exists as $\mathcal{F}$ is t-splittable and is closed under intersections, P consists of members of $\mathcal{F}$, $A \in \mathcal{F}$, and $S_i \cap B = S_i \cap A \in \mathcal{F}$ if $S_i \cap B \in \hat{P}$. As P' is a refinement of $\hat{P}$ which is a refinement of P, we have $\phi(P') \geq \phi(\hat{P}) \geq \phi(P)$. Let $R \in \{S_i, S_i \cap B, S_i \setminus B\}$, where S_i is a part of P that is refined into two parts in $\hat{P}$, so that $\mathbb{E}[1_R] \geq t\eta$. Letting $u = \frac{\mathbb{E}[g1_R]}{\mathbb{E}[1_R]}$, we see, since g is upper η-regular and using (3.19), that $u \leq 1 + t\eta(t\eta)^{-1} = 2$ and hence $\phi(u) = u^2$. It follows, by considering the functions pointwise, that $\phi(g_{\hat{P}}) - \phi(g_P) = g_{\hat{P}}^2 - g_P^2$. Hence,

$$\begin{aligned}\phi(P') - \phi(P) \geq \phi(\hat{P}) - \phi(P) &= \mathbb{E}[g_{\hat{P}}^2] - \mathbb{E}[g_P^2] = \mathbb{E}[g_{\hat{P}}^2 - g_P^2] = \mathbb{E}[(g_{\hat{P}} - g_P)^2] \\ &\geq \mathbb{E}[(g_{\hat{P}} - g_P)1_B]^2 = \mathbb{E}[(g - g_P)1_B]^2 > \frac{\epsilon^2}{4} = \alpha.\end{aligned}$$

The third equality above is the Pythagorean identity, which uses that $\hat{P}$ is a refinement of P, and the second inequality is an application of the Cauchy-Schwarz inequality. However, since P' is a refinement of P consisting of members of $\mathcal{F}$ with $|P'| \leq t|P|$, this contradicts $\phi(P') - \phi(P) < \alpha$ from the definition of P and completes the proof. ■

To establish the weak hypergraph regularity lemma, Theorem 3.1.16, we use Lemma 3.4.1 with $X = V_1 \times \cdots \times V_r$ and $\mathcal{F}$ being the family of subsets of X which form the r-cliques of some r-partite $(r-1)$-uniform hypergraph with parts $V_1, \ldots, V_r$. Noting that $\mathcal{F}$ is 2^r-splittable in this case, we obtain Theorem 3.1.16.

3.5 The counting lemma

The three main ingredients in our proof of the counting lemma (Theorem 3.1.17) are as follows.

1. A standard telescoping argument [16] in the dense case, i.e., when $\nu = 1$.

2. Repeated applications of the Cauchy-Schwarz inequality. This is a standard technique in this area, e.g., [61, 63, 69, 123].

3. *Densification.* This is the main new ingredient in our proof. At each step, we reduce the problem of counting H in a particular weighted hypergraph to that of counting H in a modified weighted hypergraph. For an edge $e \in H$, we replace the triple $(\nu_e, g_e, \tilde{g}_e)$ by a new triple $(1, g'_e, \tilde{g}'_e)$ with $0 \leq g'_e, \tilde{g}'_e \leq 1$ and such that $(g'_e, \tilde{g}'_e)$ is an ϵ'-discrepancy pair for some $\epsilon' = o_{\epsilon \to 0}(1)$. By repeatedly applying this reduction to all $e \in H$ (we use induction), we reduce the counting lemma to the dense case.

We developed the densification technique in Chapter 2, where we proved a sparse counting lemma in graphs. The proof here significantly simplifies a number of technical steps from Chapter 2 in order to extend the densification technique to hypergraphs.

3.5.1 Telescoping argument

The following argument allows us to prove the counting lemma in the dense case, i.e., when $0 \leq g \leq 1$.

Lemma 3.5.1 (Telescoping argument for dense hypergraphs). *Theorem 3.1.17 holds if we assume that there is some $e_1 \in H$ so that $\nu_e = 1$ for all $e \in H \setminus \{e_1\}$. In fact, in this case,*

$$\left| \mathbb{E}\Big[\prod_{e \in H} g_e(x_e) \Big| x \in V_J\Big] - \mathbb{E}\Big[\prod_{e \in H} \tilde{g}_e(x_e) \Big| x \in V_J\Big] \right| \leq |H|\, \epsilon. \tag{3.21}$$

Lemma 3.5.1 uses only the assumption that $(g_e, \tilde{g}_e)$ is an ϵ-discrepancy pair for every $e \in H$ and nothing about the linear forms condition on ν. Recall that for each fixed $e \in H$, the condition that $(g_e, \tilde{g}_e)$ is an ϵ-discrepancy pair means that for all subsets $B_f \subseteq V_f$, $f \in \partial e$, we have

$$\left|\mathbb{E}\Big[(g_e(x_e) - \tilde{g}_e(x_e)) \prod_{f\in\partial e} 1_{B_f}(x_f)\Big| x_e \in V_e\Big]\right| \leq \epsilon. \tag{3.22}$$

This is equivalent to the condition that for all functions $u_f \colon V_f \to [0,1]$, $f \in \partial e$, we have

$$\left|\mathbb{E}\Big[(g_e(x_e) - \tilde{g}_e(x_e)) \prod_{f\in\partial e} u_f(x_f)\Big| x_e \in V_e\Big]\right| \leq \epsilon. \tag{3.23}$$

Indeed, the expectation is linear in each u_f and hence the extrema occur when the u_f's are $\{0,1\}$-valued, thereby reducing to (3.22).

Proof. Let $h = |H|$ and order the edges of $H \setminus \{e_1\}$ arbitrarily as $e_2, \ldots, e_h$. We can write the left-hand side of (3.21), without the absolute values, as a telescoping sum

$$\sum_{t=1}^{h} \mathbb{E}\Big[\Big(\prod_{s=1}^{t-1} \tilde{g}_{e_s}(x_{e_s})\Big)(g_{e_t}(x_{e_t}) - \tilde{g}_{e_t}(x_{e_t}))\Big(\prod_{s=t+1}^{h} g_{e_s}(x_{e_s})\Big)\Big| x \in V_J\Big]. \tag{3.24}$$

For the t-th term in the sum, when we fix the value of $x_{J\setminus e_t} \in V_{J\setminus e_t}$, the expectation has the form

$$\mathbb{E}\Big[(g_{e_t}(x_{e_t}) - \tilde{g}_{e_t}(x_{e_t})) \prod_{f\in\partial e_t} u_f(x_f)\Big| x_{e_t} \in V_{e_t}\Big] \tag{3.25}$$

for some functions $u_f \colon V_f \to [0,1]$ (here we used the key fact that $g_{e_s} \leq 1$ for all $s > 1$ and $\tilde{g}_{e_s} \leq 1$ for all s). Since $(g_{e_t}, \tilde{g}_{e_t})$ is an ϵ-discrepancy pair, (3.23) implies that (3.25) is bounded in absolute value by ϵ. The same bound holds after we vary $x_{J\setminus e_t} \in V_{J\setminus e_t}$. So every term in (3.24) is bounded by ϵ in absolute value, and hence (3.24) is at most $h\epsilon$ in absolute value. ■

3.5.2 Strong linear forms

The main result of this subsection tells us that ν can be replaced by the constant function 1 in counting expressions. Though somewhat technical in detail, the main idea of the proof is quite simple and may be summarized as follows: we use the

Cauchy-Schwarz inequality to double each vertex j of a certain edge in turn, at each step majorizing those edges which do not contain j. This method is quite standard in the field. In the work of Green and Tao, it is used to prove generalized von Neumann theorems [69, Prop. 5.3], [123, Thm. 3.8], although the statement of our lemma is perhaps more similar to the uniform distribution property [69, Prop. 6.2], [123, Prop. 5.1].

We begin by using a similar method to prove a somewhat easier result. It shows that if ν satisfies the H-linear forms condition then $(\nu, 1)$ is an $o(1)$-discrepancy pair, which implies Lemma 3.1.15.

Lemma 3.5.2. *Let e be a finite set, V_j a finite set for each $j \in e$, and $V_e = \prod_{j\in e} V_j$. Then, for any function $\nu\colon V_e \to \mathbb{R}$ and any collection of $B_f \subseteq V_f$ for $f \in \partial e$,*

$$\left|\mathbb{E}\Big[(\nu_e(x_e) - 1) \prod_{f\in\partial e} 1_{B_f}(x_f) \Big| x_e \in V_e\Big]\right| \le \mathbb{E}\Big[\prod_{\omega\in\{0,1\}^e} (\nu_e(x_e^{(\omega)}) - 1) \Big| x_e^{(0)}, x_e^{(1)} \in V_e\Big]^{1/2^{|e|}}. \tag{3.26}$$

Lemma 3.5.2 follows from a direct application of the Gowers-Cauchy-Schwarz [61] inequality for hypergraphs (see [29]). We include the proof here for completeness.

Proof. For $\emptyset \subseteq d \subseteq e$, let

$$X_d := \prod_{\omega\in\{0,1\}^d} (v_e(x_{e\setminus d}, x_d^{(\omega)}) - 1), \qquad Y_d := \prod_{\substack{f\in\partial e\\ f\supseteq d}} \prod_{\omega\in\{0,1\}^d} 1_{B_f}(x_{f\setminus d}, x_d^{(\omega)}),$$

and

$$Q_d := \mathbb{E}[X_d Y_d | x_{e\setminus d} \in V_{e\setminus d},\ x_d^{(0)}, x_d^{(1)} \in V_d].$$

Then (3.26) can be written as $|Q_\emptyset| \le Q_e^{1/2^{|e|}}$. By induction, it suffices to show that $Q_d^2 \le Q_{d\cup\{j\}}$ whenever $j \in e \setminus d$. Let $Y_d = Y_d^{\ni j} Y_d^{\not\ni j}$ where $Y_d^{\ni j}$ consists of all the factors in Y_d that contain x_j in the argument, and $Y_d^{\not\ni j}$ consists of all other factors. By the Cauchy-Schwarz inequality, we have

$$Q_d^2 = \mathbb{E}[\mathbb{E}[X_d Y_d^{\ni j} | x_j \in V_j] Y_d^{\not\ni j}]^2 \le \mathbb{E}[\mathbb{E}[X_d Y_d^{\ni j} | x_j \in V_j]^2] \mathbb{E}[(Y_d^{\not\ni j})^2] \le Q_{d\cup\{j\}},$$

since $Q_{d\cup\{j\}} = \mathbb{E}[\mathbb{E}[X_d Y_d^{\ni j} | x_j \in V_j]^2]$ and $0 \le Y_d^{\not\ni j} \le 1$, where the outer expectations are taken over all free variables. This shows that $Q_d^2 \le Q_{d\cup\{j\}}$. Hence, $|Q_\emptyset| \le Q_e^{1/2^{|e|}}$, as desired. ■

The next lemma is very similar, except that now we need to invoke the linear forms condition.

Lemma 3.5.3 (Strong linear forms). *Let $V = (J, (V_j)_{j\in J}, r, H)$ be a hypergraph system and let ν be a weighted hypergraph on V satisfying the linear forms condition. Let $e_1 \in H$. For each $\iota \in \{0,1\}$ and $e \in H \setminus \{e_1\}$, let $g_e^{(\iota)} \colon V_e \to \mathbb{R}_{\ge 0}$ be a function so that either $g_e^{(\iota)} \le 1$ or $g_e^{(\iota)} \le \nu_e$ holds. Then*

$$\mathbb{E}\Big[(\nu_{e_1}(x_{e_1}) - 1) \prod_{\iota\in\{0,1\}} \Big(\prod_{e\in H\setminus\{e_1\}} g_e^{(\iota)}(x_e^{(\iota)})\Big) \Big| x_J^{(0)}, x_J^{(1)} \in V_J; x_{e_1}^{(0)} = x_{e_1}^{(1)} = x_{e_1}\Big] = o(1). \tag{3.27}$$

In (3.27) the notation $x_{e_1}^{(0)} = x_{e_1}^{(1)} = x_{e_1}$ means that $x_j^{(0)}, x_j^{(1)}, x_j$ are taken to be the same for all $j \in e_1$. Recall that we write $o(1)$ for a quantity that tends to zero as $N \to \infty$.

Proof. For each $\iota \in \{0,1\}$ and $e \in H \setminus \{e_1\}$, let $\bar{g}_e^{(\iota)}$ be either 1 or ν_e so that $g_e^{(\iota)} \le \bar{g}_e^{(\iota)}$ holds. For $\emptyset \subseteq d \subseteq e_1$, define

$$X_d := \prod_{\omega\in\{0,1\}^d} (\nu_{e_1}(x_{e_1\setminus d}, x_d^{(\omega)}) - 1),$$

$$Y_d := \prod_{\iota\in\{0,1\}} \prod_{e\in H\setminus\{e_1\}} \prod_{\omega\in\{0,1\}^{e\cap d}} \begin{Bmatrix} g_e^{(\iota)}(x_{e\setminus e_1}^{(\iota)}, x_d^{(\omega)}, x_{e\cap e_1\setminus d}) & \text{if } e \supseteq d \\ \bar{g}_e^{(\iota)}(x_{e\setminus e_1}^{(\iota)}, x_{e\cap d}^{(\omega)}, x_{e\cap e_1\setminus d}) & \text{if } e \not\supseteq d \end{Bmatrix},$$

and

$$Q_d := \mathbb{E}\big[X_d Y_d \big| x_{(J\setminus e_1)\cup d}^{(0)}, x_{(J\setminus e_1)\cup d}^{(1)} \in V_{(J\setminus e_1)\cup d},\ x_{e_1\setminus d} \in V_{e_1\setminus d}\big].$$

We observe that $Q_\emptyset$ is equal to the left-hand side of (3.27) and

$$Q_{e_1} = \mathbb{E}\Big[\prod_{\omega \in \{0,1\}^{e_1}} (\nu_{e_1}(x_{e_1}^{(\omega)}) - 1) \prod_{\iota \in \{0,1\}} \prod_{e \in H \setminus \{e_1\}} \prod_{\omega \in \{0,1\}^{e \cap e_1}} \bar{g}_e^{(\iota)}(x_{e \setminus e_1}^{(\iota)}, x_{e \cap e_1}^{(\omega)}) \Big| x_J^{(0)}, x_J^{(1)} \in V_J \Big] = o(1)$$

by the linear forms condition (3.4).[4] Indeed, after we expand $\prod_{\omega \in \{0,1\}^{e_1}} (\nu_{e_1}(x_{e_1}^{(\omega)}) - 1)$, every term in Q_{e_1} has the form of (3.4) (since $\bar{g}_e^{(\iota)}$ is 1 or ν_e). Thus Q_{e_1} is the sum of $2^{|e_1|}$ terms, each of which is $\pm(1 + o(1))$ by the linear forms condition, and they cancel accordingly to $o(1)$.

We claim that if $j \in e_1 \setminus d$ then

$$|Q_d| \leq (1 + o(1)) Q_{d \cup \{j\}}^{1/2}, \tag{3.28}$$

from which it would follow by induction that

$$|\text{LHS of (3.27)}| = |Q_\emptyset| \leq (1 + o(1)) Q_{e_1}^{1/2^r} = o(1).$$

Now we prove (3.28). Let $Y_d = Y_d^{\ni j} Y_d^{\not\ni j}$ where $Y_d^{\ni j}$ consists of all the factors in Y_d that contain x_j in the argument, and $Y_d^{\not\ni j}$ consists of all other factors. Using the Cauchy-Schwarz inequality and $Y_d^{\not\ni j} \leq \overline{Y}_d^{\not\ni j}$ one has

$$\begin{aligned} Q_d^2 &= \mathbb{E}[\mathbb{E}[X_d Y_d^{\ni j} | x_j \in V_j] Y_d^{\not\ni j}]^2 \leq \mathbb{E}[\mathbb{E}[X_d Y_d^{\ni j} | x_j \in V_j]^2 Y_d^{\not\ni j}] \ \mathbb{E}[Y_d^{\not\ni j}] \\ &\leq \mathbb{E}[\mathbb{E}[X_d Y_d^{\ni j} | x_j \in V_j]^2 \overline{Y}_d^{\not\ni j}] \ \mathbb{E}[\overline{Y}_d^{\not\ni j}] = Q_{d \cup \{j\}} \ \mathbb{E}[\overline{Y}_d^{\not\ni j}] \end{aligned} \tag{3.29}$$

where the outer expectations are taken over all free variables. The second factor in (3.29) is $1 + o(1)$ by the linear forms condition (3.4) as $\overline{Y}_d^{\not\ni j}$ consists only of ν factors. This proves (3.28). ■

[4]This is where the weak 2-blow-up of H arises, since the estimate $Q_{e_1} = o(1)$ only relies upon knowing that ν has roughly the expected density for certain subgraphs of the weak 2-blow-up.

3.5.3 Counting lemma proof

As already mentioned, the main idea of the following proof is a process called densification, where we reduce the problem of counting H in a sparse hypergraph to that of counting H in a dense hypergraph by replacing sparse edges with dense edges one at a time. Several steps are needed to densify a given edge e_1. The first step is to double all vertices outside of e_1 and to majorize g_{e_1} by ν_{e_1}. We then use the strong linear forms condition to remove the edge corresponding to e_1 entirely. This leaves us with the seemingly harder problem of counting the graph H' consisting of two copies of $H\backslash\{e_1\}$ joined along the vertices of e_1. However, an inductive hypothesis tells us that we can count copies of $H\backslash\{e_1\}$. The core of the proof is in showing that this allows us to replace one of the copies of $H\backslash\{e_1\}$ in H' by a dense edge, thus reducing our problem to that of counting H with one edge replaced by a dense edge.

Proof of Theorem 3.1.17. We use induction on $|\{e \in H : \nu_e \neq 1\}|$. For the base case, when $|\{e \in H : \nu_e \neq 1\}| = 0$ or 1, the result follows from Lemma 3.5.1. Now take $e_1 \in H$ so that $\nu_{e_1} \neq 1$.

We assume that $|J|$ is a fixed constant. We write $o(1)$ for a quantity that tends to zero as $N \to \infty$ and $o_{\epsilon\to 0}(1)$ for a quantity that tends to zero as $N \to \infty$ and $\epsilon \to 0$. We need to show that the following quantity is $o_{\epsilon\to 0}(1)$:

$$\begin{aligned}
\mathbb{E}\Big[\prod_{e\in H} g_e(x_e)\Big| x \in V_J\Big] &- \mathbb{E}\Big[\prod_{e\in H} \tilde{g}_e(x_e)\Big| x \in V_J\Big] \\
&= \mathbb{E}\Big[g_{e_1}(x_{e_1})\Big(\prod_{e\in H\backslash\{e_1\}} g_e(x_e) - \prod_{e\in H\backslash\{e_1\}} \tilde{g}_e(x_e)\Big)\Big| x \in V_J\Big] \\
&\quad + \mathbb{E}\Big[(g_{e_1}(x_{e_1}) - \tilde{g}_{e_1}(x_{e_1}))\Big(\prod_{e\in H\backslash\{e_1\}} \tilde{g}_e(x_e)\Big)\Big| x \in V_J\Big]. \qquad (3.30)
\end{aligned}$$

The second term in (3.30) is at most ϵ in absolute value since $(g_{e_1}, \tilde{g}_{e_1})$ is an ϵ-discrepancy pair and $\tilde{g} \leq 1$ (e.g., see proof of Lemma 3.5.1). It remains to show that the first term in (3.30) is $o_{\epsilon\to 0}(1)$.

Define functions $\nu'_{e_1}, g'_{e_1}, \tilde{g}'_{e_1} : V_{e_1} \to \mathbb{R}_{\geq 0}$ by

$$\nu'_{e_1}(x_{e_1}) := \mathbb{E}\Big[\prod_{e \in H \setminus \{e_1\}} \nu_e(x_e) \Big| x_{J \setminus e_1} \in V_{J \setminus e_1} \Big], \tag{3.31}$$

$$g'_{e_1}(x_{e_1}) := \mathbb{E}\Big[\prod_{e \in H \setminus \{e_1\}} g_e(x_e) \Big| x_{J \setminus e_1} \in V_{J \setminus e_1} \Big], \tag{3.32}$$

$$\tilde{g}'_{e_1}(x_{e_1}) := \mathbb{E}\Big[\prod_{e \in H \setminus \{e_1\}} \tilde{g}_e(x_e) \Big| x_{J \setminus e_1} \in V_{J \setminus e_1} \Big]. \tag{3.33}$$

We have $g'_{e_1} \leq \nu'_{e_1}$ and $\tilde{g}_{e_1} \leq 1$ (pointwise). In the rest of this proof, unless otherwise specified, expectations are for functions on V_{e_1} with arguments varying uniformly over V_{e_1}. The linear forms condition (3.4) implies that $\mathbb{E}[\nu'_{e_1}] = 1 + o(1)$ and $\mathbb{E}[(\nu'_{e_1})^2] = 1 + o(1)$, so that[5]

$$\mathbb{E}[(\nu'_{e_1} - 1)^2] = o(1). \tag{3.34}$$

The square of the first term in (3.30) equals

$$\begin{aligned} \mathbb{E}[g_{e_1}(g'_{e_1} - \tilde{g}'_{e_1})]^2 \leq \mathbb{E}[g_{e_1}(g'_{e_1} - \tilde{g}'_{e_1})^2]\ \mathbb{E}[g_{e_1}] &\leq \mathbb{E}[\nu_{e_1}(g'_{e_1} - \tilde{g}'_{e_1})^2]\ \mathbb{E}[\nu_{e_1}] \\ &= (\mathbb{E}[(g'_{e_1} - \tilde{g}'_{e_1})^2] + o(1))(1 + o(1)). \end{aligned} \tag{3.35}$$

The first inequality above is due to the Cauchy-Schwarz inequality. In the final step, both factors are estimated using Lemma 3.5.3 (for the first factor, expand the square $(g'_{e_1} - \tilde{g}'_{e_1})^2$ and apply Lemma 3.5.3 term by term). Continuing (3.35) it suffices to show that the following quantity is $o_{\epsilon \to 0}(1)$:

$$\mathbb{E}[(g'_{e_1} - \tilde{g}'_{e_1})^2] = \mathbb{E}[(g'_{e_1} - \tilde{g}'_{e_1})(g'_{e_1} - g'_{e_1} \wedge 1)] + \mathbb{E}[(g'_{e_1} - \tilde{g}'_{e_1})(g'_{e_1} \wedge 1 - \tilde{g}'_{e_1})] \tag{3.36}$$

(here $a \wedge b := \min\{a, b\}$). That is, we are capping the weighted hypergraph g'_{e_1} by 1. Since ν'_{e_1} is very close to 1 by (3.34), this should not result in a large loss. Indeed,

[5] In fact, the only assumptions on ν needed for the proof of Theorem 3.1.17 are (3.34) and the strong linear forms condition, Lemma 3.5.3, as well as analogous conditions for other choices of $e_1 \in H$ and allowing some subset of the functions ν_e to be replaced by 1.

since $0 \le g'_{e_1} \le \nu'_{e_1}$, we have

$$0 \le g'_{e_1} - g'_{e_1} \wedge 1 = \max\{g'_{e_1} - 1, 0\} \le \max\{\nu'_{e_1} - 1, 0\} \le |\nu'_{e_1} - 1|. \tag{3.37}$$

Using (3.37), $g'_{e_1} \le \nu'_{e_1}$, and $\tilde{g}'_{e_1} \le 1$, we bound the magnitude of the first term on the right-hand side of (3.36) by

$$\begin{aligned}\mathbb{E}[(\nu'_{e_1} + 1)\,|\nu'_{e_1} - 1|] &= \mathbb{E}[(\nu'_{e_1} - 1)\,|\nu'_{e_1} - 1|] + 2\mathbb{E}[|\nu'_{e_1} - 1|] \\ &\le \mathbb{E}[(\nu'_{e_1} - 1)^2] + 2\mathbb{E}[(\nu'_{e_1} - 1)^2]^{1/2} = o(1)\end{aligned}$$

by the triangle inequality, the Cauchy-Schwarz inequality, and (3.34). To estimate the second term on the right-hand side of (3.36), we need the following claim.

Claim. *$(g'_{e_1} \wedge 1, \tilde{g}'_{e_1})$ is an ϵ'-discrepancy pair with $\epsilon' = o_{\epsilon\to 0}(1)$.*

Proof of Claim. We need to show that, whenever $B_f \subseteq V_f$ for all $f \in \partial e_1$, we have

$$\mathbb{E}\Big[(g'_{e_1}(x_{e_1}) \wedge 1 - \tilde{g}'_{e_1}(x_{e_1})) \prod_{f\in\partial e_1} 1_{B_f}(x_f)\Big| x_{e_1} \in V_{e_1}\Big] = o_{\epsilon\to 0}(1). \tag{3.38}$$

Define $g''_{e_1} \colon V_{e_1} \to \mathbb{R}_{\ge 0}$ by $g''_{e_1}(x_{e_1}) = \prod_{f\in\partial e_1} 1_{B_f}(x_f)$. So the left-hand side of (3.38) is equal to

$$\mathbb{E}[(g'_{e_1} \wedge 1 - g'_{e_1})g''_{e_1}] + \mathbb{E}[(g'_{e_1} - \tilde{g}'_{e_1})g''_{e_1}]. \tag{3.39}$$

Using $0 \le g''_{e_1} \le 1$, (3.37), the Cauchy-Schwarz inequality, and (3.34), we can bound the magnitude of the first term in (3.39) by

$$\mathbb{E}[|\nu'_{e_1} - 1|] \le \mathbb{E}[(\nu'_{e_1} - 1)^2]^{1/2} = o(1).$$

The second term on the right-hand side of (3.39) is equal to

$$\mathbb{E}\Big[\Big(\prod_{e\in H\setminus\{e_1\}} g_e(x_e) - \prod_{e\in H\setminus\{e_1\}} \tilde{g}_e(x_e)\Big) g''_{e_1}(x_{e_1}) \Big| x \in V_J\Big].$$

This is $o_{\epsilon\to 0}(1)$ by the induction hypothesis applied to new weighted hypergraphs where the old $(\nu_{e_1}, g_{e_1}, \tilde{g}_{e_1})$ gets replaced by $(1, g''_{e_1}, g''_{e_1})$, thereby decreasing $|\{e \in H : \nu_e \neq 1\}|$. Note that the linear forms condition continues to hold. Thus (3.38) holds, so $(g'_{e_1} \wedge 1, \tilde{g}'_{e_1})$ is an ϵ'-discrepancy pair with $\epsilon' = o_{\epsilon\to 0}(1)$. □

We expand the second term of (3.36) as

$$\mathbb{E}[(g'_{e_1}-\tilde{g}'_{e_1})(g'_{e_1}\wedge 1-\tilde{g}'_{e_1})] = \mathbb{E}[g'_{e_1}(g'_{e_1}\wedge 1)]-\mathbb{E}[g'_{e_1}\tilde{g}'_{e_1}]-\mathbb{E}[\tilde{g}'_{e_1}(g'_{e_1}\wedge 1)]+\mathbb{E}[(\tilde{g}'_{e_1})^2]. \quad (3.40)$$

We claim that each expectation on the right-hand side of (3.40) is $\mathbb{E}[(\tilde{g}'_{e_1})^2]+o_{\epsilon\to 0}(1)$. Indeed, by (3.32) and (3.33) we have

$$\mathbb{E}[g'_{e_1}(g'_{e_1}\wedge 1)]-\mathbb{E}[(\tilde{g}'_{e_1})^2] = \mathbb{E}\Big[\Big((g'_{e_1}(x_{e_1})\wedge 1)\prod_{e\in H\setminus\{e_1\}} g_e(x_e) - \tilde{g}'_{e_1}(x_{e_1})\prod_{e\in H\setminus\{e_1\}}\tilde{g}_e(x_e)\Big)\Big|x\in V_J\Big],$$

which is $o_{\epsilon\to 0}(1)$ by the induction hypothesis applied to new weighted hypergraphs where the old $(\nu_{e_1}, g_{e_1}, \tilde{g}_{e_1})$ is replaced by $(1, g'_{e_1}\wedge 1, \tilde{g}'_{e_1})$. This is allowed as $(g'_{e_1}\wedge 1, \tilde{g}'_{e_1})$ is an ϵ'-discrepancy pair with $\epsilon' = o_{\epsilon\to 0}(1)$, the new ν still satisfies the linear forms condition, and $|\{e \in H : \nu_e \neq 1\}|$ has decreased. The claims that the other terms on the right-hand side of (3.40) are each $\mathbb{E}[(\tilde{g}'_{e_1})^2] + o_{\epsilon\to 0}(1)$ are similar (in fact, easier). It follows that (3.40) is $o_{\epsilon\to 0}(1)$, so (3.36) is $o_{\epsilon\to 0}(1)$ and we are done. ■

3.6 Concluding remarks

Conditions for counting lemmas. In this chapter, we determined sufficient conditions for establishing a relative Szemerédi theorem and, more generally, a counting lemma for sparse hypergraphs. We have assumed that the hypergraph we want to count within is a subgraph of a pseudorandom hypergraph. The main question then is to determine a good notion of pseudorandomness which is suffficient to establish a counting lemma.

There is a marked difference between this chapter and the Chapter 2 on graphs in terms of the type of pseudorandom condition assumed for the majorizing hypergraph. In this chapter, we prove a counting lemma for a given hypergraph H by assuming that the underlying pseudorandom hypergraph contains approximately the correct count for each hypergraph in a certain collection of hypergraphs $\mathcal{H}$ derived from H. That is, for each $H' \in \mathcal{H}$, we assume that our pseudorandom hypergraph contains $(1+o(1))p^{e(H')}n^{v(H')}$ labeled copies of H', where p is the edge density of the pseudorandom hypergraph.

The approach used in Chapter 2 is equivalent, up to some polynomial loss in ϵ, to assuming that the number of labeled cycles of length 4 in our pseudorandom graph is $(1+\epsilon)p^4n^4$, where ϵ is now a carefully controlled term and the question of whether H can be embedded in our pseudorandom graph depends on whether ϵ is sufficiently small with respect to H and p. It is possible to adapt the methods of this chapter so that the notion of pseudorandomness used for hypergraphs is more closely related to this latter notion. However, for the purposes of applying the results to a relative Szemerédi theorem, the current formulation seemed more appropriate.

Gowers uniformity norms. For a function $f : \mathbb{Z}_N \to \mathbb{R}$, the Gowers U^r-norm of f is defined to be

$$\|f\|_{U^r} = \mathbb{E}\Big[\prod_{\omega\in\{0,1\}^r} f(x_0+\omega\cdot\mathbf{x})\Big|x_0,x_1,\dots,x_r\in\mathbb{Z}_N\Big]^{1/2^r},$$

where $\mathbf{x} = (x_1, \dots, x_r)$. The following inequality, referred to as a generalized von Neumann theorem, bounds the weighted count of $(r+1)$-term arithmetic progressions from functions $f_0, \dots, f_r$ in terms of the Gowers uniformity norm:

$$\Big|\mathbb{E}\Big[f_0(x)f_1(x+d)f_2(x+2d)\cdots f_r(x+rd)\Big|x,d\in\mathbb{Z}_N\Big]\Big| \le \|f_j\|_{U^r}\prod_{i\neq j}\|f_i\|_\infty.$$

This fundamental fact is an important starting point for Gowers' celebrated proof [61] of Szemerédi's theorem as well as many later developments in additive combinatorics. For a sparse set $S \subseteq \mathbb{Z}_N$ of density p, this inequality implies the correct count of $(r+1)$-term arithmetic progressions in S as long as $\|\nu - 1\|_{U^r} = o(p^r)$, where $\nu = p^{-1}1_S$ (a more careful analysis shows that it suffices to assume $\|\nu - 1\|_{U^r} = o(p^{r/2})$).

Gowers [65][6] and Green [67] asked if $\|\nu - 1\|_{U^s} = o(1)$ for some large $s = s(r)$ is sufficient for ν to satisfy a relative Szemerédi theorem for $(r+1)$-term arithmetic progressions. Note that this is precisely a linear forms condition and we proved in this chapter that a different linear forms condition is sufficient. However, we do not even know if such a condition implies the existence of $(r+1)$-term arithmetic progressions in ν. Clearly $s(r)$ cannot be too small and indeed we know from the recent work of Bennett and Bohman [12] on the random AP-free process that one can find a 3-AP-free $S \subset \mathbb{Z}_N$ such that $\nu = (N/|S|)1_S$ satisfies $\|\nu - 1\|_{U^2} = o(1)$. Therefore, if $s(2)$ exists, it must be greater than 2. More generally, they show that $s(r) > 1 + \log_2 r$. We can show (details can be found in [29]) that if a measure ν satisfies the stronger condition $\|\nu - 1\|_{U^r} = o(p^r)$, where $p = \|\nu\|_\infty^{-1}$, then the relative Szemerédi theorem holds with respect to ν for $(r+1)$-term arithmetic progressions. This strengthens the consequence of the generalized von Neumann theorem discussed above.

Corners in products of pseudorandom sets. Example 3.2.2 illustrates the relative multidimensional Szemerédi theorem applied to a pseudorandom set $S \subset \mathbb{Z}_N^2$. However, the situation is quite different for $S \times S \subset \mathbb{Z}_N^2$ with some pseudorandom set $S \subset \mathbb{Z}_N$. Indeed, $S \times S \subset \mathbb{Z}_N^2$ does not satisfy the linear forms condition in Example 3.2.2. Intuitively, this is because the events $(x, y) \in S \times S$ and $(x, y') \in S \times S$ are correlated as both involve $x \in S$.

However, we may still deduce the following result using our relative triangle removal lemma. Recall that a corner in $\mathbb{Z}_N^2$ is a set of the form $\{(x, y), (x+d, y), (x, y+d)\}$, where $d \neq 0$.

Proposition 3.6.1. *If $S \subset \mathbb{Z}_N$ is such that $\nu = \frac{N}{|S|} 1_S$ satisfies*

$$\begin{aligned}\mathbb{E}[&\nu(x)\nu(x')\nu(z-x)\nu(z-x')\nu(z'-x)\nu(z'-x') \\ &\cdot \nu(y)\nu(y')\nu(z-y)\nu(z-y')\nu(z'-y)\nu(z'-y') | x, x', y, y', z, z' \in \mathbb{Z}_N] = 1 + o(1)\end{aligned} \tag{3.41}$$

and similar conditions hold if any subset of the ν factors are erased, then any corner-

[6]This question can be found in the penultimate paragraph in §4 of the arXiv version of [65].

free subset of $S \times S$ *has size* $o(|S|^2)$.

Proof (sketch). Let A be a corner-free subset of $S \times S$. We build two tripartite graph Γ and G on the same vertex set $X \cup Y \cup Z$ with $X = Y = S$ and $Z = \mathbb{Z}_N$ (note that unlike the proof of Theorem 3.2.1 we do not take X and Y to be the whole of $\mathbb{Z}_N$ here). In Γ, we place a complete bipartite graph between X and Y; between Y and Z the edge $(y, z) \in Y \times Z$ is present if and only if $z - y \in S$; and between X and Z the edge $(x, z) \in X \times Z$ is present if and only if $z - x \in S$. In G, between X and Y the edge $(x, y) \in (X, Y)$ is present if and only if $(x, y) \in A$; between Y and Z the edge $(y, z) \in Y \times Z$ is present if and only if $(z - y, y) \in A$; and between X and Z the edge $(x, z) \in X \times Z$ is present if and only if $(x, z - x) \in A$.

The vertices $(x, y, z) \in X \times Y \times Z$ form a triangle if and only if $(x, y), (z - y, y), (x, z - x) \in A$. These three points form a corner, which is degenerate only when $x + y = z$. Since A is corner-free, every edge of G is contained in exactly one triangle (namely the one that completes the equation $x + y = z$). In particular, G contains exactly $|A|$ triangles. After checking some hypotheses, we can apply our relative triangle removal lemma (as a special case of Theorem 3.1.12) to conclude that it is possible to remove all triangles from G by deleting $o(|S|^2)$ edges. Since every edge of G is contained in exactly one triangle, and $|G|$ has $3|A|$ edges, we have $|A| = o(|S|^2)$, as desired. ∎

One can easily generalize the above Proposition to $S^m \subset \mathbb{Z}_N^m$ (as before, $S \subset \mathbb{Z}_N$). Here a corner is a set of the form $\{\mathbf{x}, \mathbf{x}+d\mathbf{e}_1, \ldots, \mathbf{x}+d\mathbf{e}_m\}$, where $\mathbf{x} \in \mathbb{Z}_N$, $0 \neq d \in \mathbb{Z}_N$, and $\mathbf{e}_i$ is the i-th coordinate vector. Then, for any fixed m, any corner-free subset of S^m must have size $o(|S|^m)$, provided that $\nu = \frac{N}{|S|}1_S$ satisfies the linear forms condition

$$\mathbb{E}\left[\prod_{i=1}^{m}\Big(\nu(x_i^{(0)})^{n_{i,0}}\nu(x_i^{(1)})^{n_{i,1}} \prod_{\omega \in \{0,1\}^{\{0\}\cup[m]\setminus\{i\}}} (x_0^{(\omega_0)} - \sum_{j \in [m]\setminus\{i\}} x_j^{(\omega_j)})^{n_{i,\omega}}\Big) \,\middle|\, x_0^{(0)}, x_0^{(1)}, \ldots, x_m^{(0)}, x_m^{(1)} \in \mathbb{Z}_N\right] = 1 + o(1)$$

for any choices of exponents $n_{i,0}, n_{i,1}, n_{i,\omega} \in \{0, 1\}$.

A more general result concerning the existence of arbitrarily shaped constellations in S^m is known, provided that S satisfies certain stronger linear forms hypotheses. We refer the readers to [35, 50, 125] for further details. In particular, the multidimensional relative Szemerédi theorem holds in P^m, where P is the primes.

Sparse graph limits. The regularity method played a fundamental role in the development of the theory of dense graph limits [16, 89]. However, no satisfactory theory of graph limits is known for graphs with edge density $o(1)$. Bollobás and Riordan [15] asked a number of questions and made explicit conjectures on suitable conditions for sparse graph limits and counting lemmas. Our work gives some natural sufficient conditions for obtaining a counting lemma in a sequence of sparse graphs G_N. The new counting lemma allows us to transfer the results of Lovász and Szegedy [89, 90] on the existence of the limit graphon, as well as the results of Borgs, Chayes, Lovász, Sós, and Vesztergombi [16] on the equivalence of left-convergence (i.e., convergence in homomorphism densities) and convergence in cut distance. The famous quasirandomness results of Chung, Graham, and Wilson [23] also transfer, namely, that an appropriate relationship between edge density and C_4-density (of homomorphisms) determines the asymptotic F-density for every graph F.

Existing applications of the Green-Tao method. Though our discussion has focused on the relative Szemerédi theorem, we have proved a relative version of the stronger multidimensional Szemerédi theorem. Following Tao [123], this may be used to prove that the Gaussian primes contain arbitrarily shaped constellations, though without the need to verify either the correlation condition or the dual function condition. It seems likely that our method could also be useful for simplifying several other papers where the machinery of Green and Tao is used [34, 71, 87, 94, 95, 126]. In some cases it should be possible to use our results verbatim but in others, such as the paper of Tao and Ziegler [126] proving that there are arbitrarily long polynomial progressions in the primes, it will probably require substantial additional work.

Sparse hypergraph regularity. In proving a hypergraph removal lemma for subgraphs of pseudorandom hypergraphs, we have developed a general approach to regularity and counting in sparse pseudorandom hypergraphs which has the potential for much broader application. It is, for example, quite easy to use our results to prove analogues of well-known combinatorial theorems such as Ramsey's theorem and Turán's theorem relative to sparse pseudorandom hypergraphs of density N^{-c_H}. We omit the details. In the graph case, a number of further applications were discussed in Chapter 2. We expect that hypergraph versions of many of these applications should be an easy corollary of our results.

Counting in random hypergraphs. There has been much recent work on counting lemmas and relative versions of combinatorial theorems within random graphs and hypergraphs [10, 32, 33, 110, 111]. Surprisingly, there are a number of disparate approaches to these problems, each having its own strengths and weaknesses. We believe that our results can be used to give an alternative framework for one of these approaches, due to Conlon and Gowers [32].[7] Their proof relies heavily upon an application of the Green-Tao transference theorem, which we believe can be replaced with an application of the sparse Frieze-Kannan regularity lemma and our densification technique. However, the key technical step in [32], which in our language is to verify that the strong linear forms condition, Lemma 3.5.3, holds when ν is a random measure, would remain unchanged.

Sparse arithmetic removal. In Theorem 3.2.3, we proved an arithmetic removal lemma for linear patterns such as arithmetic progressions. More generally, an arithmetic removal lemma claims that if a system of linear equations $Ma = b$ over the integers has a small number of solutions $a = (a_1, a_2, \ldots, a_n)$ with $a_i \in A_i$ for all $i = 1, 2, \ldots, n$ then one may remove a small number of elements from each A_i to find subsets A_i' such that there are no solutions $a' = (a_1', a_2', \ldots, a_n')$ to $Ma' = b$ with $a_i' \in A_i'$ for all $i = 1, 2, \ldots, n$. Such a result was conjectured by Green [68] and proved by Král', Serra, and Vena [85] and, independently, Shapira [113]. Both of these proofs are based upon representing a system of linear equations by a hypergraph and deducing the arithmetic removal lemma from a hypergraph removal lemma. Such an idea

[7]This should at least be true for theorems regarding graphs and hypergraphs, though we feel that a similar approach should also be possible for subsets of the integers.

was first used by Král', Serra, and Vena [84] with graphs (instead of hypergraphs). In Chapter 2, we adapted the arguments of [84] to sparse pseudorandom subsets of the integers using the removal lemma in sparse pseudorandom graphs. Likewise, our results on hypergraph removal in this chapter may be used to prove a sparse pseudorandom generalization of the arithmetic removal lemma [85, 113] for all systems of linear equations.

Chapter 4

Arithmetic transference

In Chapter 3, we obtained a relative extension of the hypergraph removal lemma from which we deduced our relative Szemerédi theorem via standard arguments. In this chapter, we give an alternative approach to proving the relative Szemerédi theorem. Instead of going through the hypergraph removal lemma, we use Szemerédi's theorem directly as a black box. To transfer Szemerédi's theorem to the sparse setting, we apply the dense model theorem of Green-Tao [69] and Tao-Zieger [126], which was subsequently simplified by Gowers [65], and independently Reingold, Trevisan, Tulsiani, and Vadhan [102]. This tool lets us model a subset of a sparse pseudorandom set of integers by a dense subset. The dense model is a good approximation of the original set with respect to a discrepancy-type norm (similar to the cut metric for graphs). This contrasts previous proofs the Green-Tao theorem [69, 65, 102] where the dense model theorem is applied with respect to the Gowers uniformity norm, which gives a stronger notion of approximation.

We make key use of the relative counting lemma, Theorem 3.1.17, which implies that the dense model behaves similarly to the original set in the number of arithmetic progressions.

The arithmetic transference approach presented here establishes the new relative Szemerédi theorem in a more direct fashion compared to Chapter 3, and it also gives better quantitative bounds. Indeed, instead of going through the hypergraph removal lemma, which currently has an Ackermann-type dependence on the bounds (due to the application of the hypergraph regularity lemma), we can now use Szemerédi's theorem as a black box and automatically transfer the best quantitative bounds available (currently the state-of-art is [109] for 3-term APs, [70] for 4-term APs, and [61] for longer APs). The approach presented here, however, is less general compared to Chapter 3, since it does not provide a relative hypergraph removal lemma, nor does it give a more general sparse regularity approach to hypergraphs.

We shall use, as a black box, the following weighted version of Szemerédi's theorem as formulated, for example, in [69, Prop. 2.3]. Compared to Theorem 3.1.1, we emphasize the constant $c(k, \delta)$. It may be helpful to think of f as the indicator function 1_A of some set $A \subseteq \mathbb{Z}_N$. It will be easier for us to work in $\mathbb{Z}_N := \mathbb{Z}/N\mathbb{Z}$ as opposed to $[N] := \{1, \ldots, N\}$, although these two settings are easily seen to be equivalent.

Theorem 4.0.1 (Szemerédi's theorem, weighted version). *Let $k \geq 3$ and $0 < \delta \leq 1$ be fixed. Let $f \colon \mathbb{Z}_N \to [0, 1]$ be a function satisfying $\mathbb{E}[f] \geq \delta$. Then*

$$\mathbb{E}[f(x)f(x+d)f(x+2d)\cdots f(x+(k-1)d)|x, d \in \mathbb{Z}_N] \geq c(k, \delta) - o_{k,\delta}(1)$$

for some constant $c(k, \delta) > 0$ which does not depend on f or N.

Gowers' results [61] (along with a Varnavides-type [133] averaging argument) imply that Theorem 4.0.1 holds with $c(k, \delta) = \exp(-\exp(\delta^{-c_k}))$ with $c_k = 2^{2^{k+9}}$ (see [109] and [70] for the current best bounds for $k = 3$ and 4 respectively).

The main result of this chapter is the following theorem. Recall the *k-linear forms condition* from Definition 3.1.2.

Theorem 4.0.2 (Relative Szemerédi theorem). *Let $k \geq 3$ and $0 < \delta \leq 1$ be fixed. Let $\nu \colon \mathbb{Z}_N \to \mathbb{R}_{\geq 0}$ satisfy the k-linear forms condition. Assume that N is sufficiently large and relatively prime to $(k-1)!$. Let $f \colon \mathbb{Z}_N \to \mathbb{R}_{\geq 0}$ satisfy $0 \leq f(x) \leq \nu(x)$ for all $x \in \mathbb{Z}_N$ and $\mathbb{E}[f] \geq \delta$. Then*

$$\mathbb{E}[f(x)f(x+d)f(x+2d)\cdots f(x+(k-1)d)|x, d \in \mathbb{Z}_N] \geq c(k, \delta) - o_{k,\delta}(1),$$

where $c(k,\delta)$ is the same constant which appears in Theorem 4.0.1. The rate at which the $o_{k,\delta}(1)$ term goes to zero depends not only on k and δ but also the rate of convergence in the k-linear forms condition for ν.

This theorem was proved in Chapter 3 without the additional conclusion that $c(k,\delta)$ can be taken to be the same as in Theorem 4.0.1. Indeed, the proof in Chapter 3 uses the hypergraph removal lemma as a black box, so that the constants $c(k,\delta)$ there are much worse, with an Ackermann-type dependence due to the use of hypergraph regularity. In [69], Green and Tao also transfered Szemerédi's theorem directly to obtain the same constants $c(k,\delta)$ as in Theorem 4.0.1, but under stronger pseudorandomness hypotheses for ν. So Theorem 4.0.2 combines the conclusions of the two relative Szemerédi theorems in Chapter 3 and [69].

In Section 4.1, we apply a dense model theorem to find a dense approximation of the original set. In Section 4.2, we apply a counting lemma to show that the dense model has approximately the same number of k-term APs as the original set. Finally in Section 4.3, we put everything together and apply Szemerédi's theorem as a black box to conclude the proof.

4.1 Dense model theorem

In this section, we show that the f in Theorem 4.0.2 can be modeled by a function $\tilde{f}\colon \mathbb{Z}_N \to [0,1]$. We state our results in terms of a finite abelian group G (written additively), but there is no loss in thinking $G = \mathbb{Z}_N$. For $x = (x_1, \dots, x_r) \in G^r$, and $I \subseteq [r]$, we write $x_I = (x_i)_{i \in I}$.

Definition 4.1.1. Let G be a finite abelian group, r be a positive integer, $\psi\colon G^r \to G$ be a surjective homomorphism, and $f, \tilde{f}\colon G \to \mathbb{R}_{\geq 0}$ be two functions. We say that $(f, \tilde{f})$ is an (r,ϵ)*-discrepancy pair with respect to* ψ if

$$\left|\mathbb{E}\left[(f(\psi(x)) - \tilde{f}(\psi(x))) \prod_{i=1}^{r} u_i(x_{[r]\setminus\{i\}}) \middle| x \in G^r\right]\right| \leq \epsilon \tag{4.1}$$

for all collections of functions $u_1, \dots, u_r\colon G^{r-1} \to [0,1]$.

Example 4.1.2. When $r = 2$ and $\psi(x, y) = x + y$, (4.1) says

$$|\mathbb{E}[(f(x + y) - \tilde{f}(x + y))u_1(y)u_2(x)|x, y \in G]| \leq \epsilon.$$

In other words, this says that the two weighted graphs $g, \tilde{g}\colon G \times G \to \mathbb{R}_{\geq 0}$ given by $g(x, y) = f(x + y)$ and $\tilde{g}(x, y) = \tilde{f}(x + y)$ satisfy $\|g - \tilde{g}\|_\square \leq \epsilon$, where $\|\cdot\|_\square$ is the cut norm for bipartite graphs.

When $r = 3$ and $\psi(x, y, z) = x + y + z$, (4.1) says

$$|\mathbb{E}[(f(x + y + z) - \tilde{f}(x + y + z))u_1(y, z)u_2(x, z)u_3(x, y)|x, y, z \in G]| \leq \epsilon.$$

The following key lemma says that any $0 \leq f \leq \nu$ can be approximated by a $0 \leq \tilde{f} \leq 1$ in the above sense.

Lemma 4.1.3. *For every $\epsilon > 0$ there is an $\epsilon' = \exp(-\epsilon^{-O(1)})$ such that the following holds:*

Let G be a finite abelian group, r be a positive integer, and $\psi\colon G^r \to G$ be a surjective homomorphism. Let $f, \nu\colon G \to \mathbb{R}_{\geq 0}$ be such that $0 \leq f \leq \nu$, $\mathbb{E}[f] \leq 1$, and $(\nu, 1)$ is an (r, ϵ')-discrepancy pair with respect to ψ. Then there exists a function $\tilde{f}\colon G \to [0, 1]$ so that $\mathbb{E}[\tilde{f}] = \mathbb{E}[f]$ and $(f, \tilde{f})$ is an (r, ϵ)-discrepancy pair with respect to ψ.

The proof of Lemma 4.1.3 uses the dense model theorem of Green-Tao [69] and Tao-Ziegler [126], which was later simplified in [65] and [102]. The expository note [101] has a nice and short write-up of the proof of the dense model theorem, and we quote the statement from there.

Let X be a finite set. For any two functions $f, g\colon X \to \mathbb{R}$, we write $\langle f, g\rangle = \mathbb{E}[f(x)g(x)|x \in X]$. For $\mathcal{F}$ a collection of functions $\varphi\colon X \to [-1, 1]$, we write $\mathcal{F}^k$ to mean the collections of all functions of the form $\prod_{i=1}^{k'} \varphi_i$, where $\varphi_i \in \mathcal{F}$ and $k' \leq k$. In particular, if $\mathcal{F}$ is closed under multiplication, then $\mathcal{F}^k = \mathcal{F}$.

Lemma 4.1.4 (Green-Tao-Ziegler dense model theorem). *For every $\epsilon > 0$, there is a $k = (1/\epsilon)^{O(1)}$ and an $\epsilon' = \exp(-(1/\epsilon)^{O(1)})$ such that the following holds:*

Suppose that $\mathcal{F}$ is a collection of functions $\varphi\colon X \to [-1,1]$ on a finite set X, $\nu\colon X \to \mathbb{R}_{\geq 0}$ satisfies

$$|\langle \nu - 1, \varphi\rangle| \leq \epsilon' \text{ for all } \varphi \in \mathcal{F}^k,$$

and $f\colon X \to \mathbb{R}_{\geq 0}$ satisfies $f \leq \nu$ and $\mathbb{E}[f] \leq 1$. Then there is a function $\tilde{f}\colon X \to [0,1]$ such that $\mathbb{E}[\tilde{f}] = \mathbb{E}[f]$, and

$$|\langle f - \tilde{f}, \varphi\rangle| \leq \epsilon \text{ for all } \varphi \in \mathcal{F}.$$

We shall use Lemma 4.1.4 with $\mathcal{F}$ closed under multiplication, so that k plays no role. This is an important point in our simplification over previous approaches using the dense model theorem.

Proof of Lemma 4.1.3. For any collection of functions $u_1, \ldots, u_r\colon G^{r-1} \to \mathbb{R}$, define a generalized convolution $(u_1, \ldots, u_r)^*_\psi\colon G \to \mathbb{R}$ by

$$(u_1, \ldots, u_r)^*_\psi(x) = \mathbb{E}\Big[\prod_{i=1}^{r} u_i(y_{[r]\setminus\{i\}})\Big| y \in G^r,\ \psi(y) = x\Big].$$

Then the left-hand side of (4.1) can be written as $|\langle f - \tilde{f}, (u_1, \ldots, u_r)^*_\psi\rangle|$. Let $\mathcal{F}$ be the set of functions which can be obtained by convex combinations of functions of the form $(u_1, \ldots, u_r)^*_\psi$, varying over all combinations of functions $u_1, \ldots, u_r\colon G^{r-1} \to [0,1]$ (but ψ is fixed). Then $(f, \tilde{f})$ being an (r, ϵ')-discrepancy pair with respect to ψ is equivalent to $|\langle f - \tilde{f}, \varphi\rangle| \leq \epsilon$ for all $\varphi \in \mathcal{F}$. The desired claim would then follow from Lemma 4.1.4 and the triangle inequality provided we can show that $\mathcal{F}$ is closed under multiplication. It suffices to show that for $u_1, \ldots, u_r, u_1', \ldots, u_r'\colon G^{r-1} \to [0,1]$, the product of $(u_1, \dots, u_r)^*_\psi$ and $(u'_1, \dots, u'_r)^*_\psi$ still lies in $\mathcal{F}$. Indeed, we have

$$\begin{aligned}
&(u_1, \dots, u_r)^*_\psi(x)(u'_1, \dots, u'_r)^*_\psi(x) \\
&= \mathbb{E}\Big[\prod_{i=1}^r u_i(y_{[r]\setminus\{i\}})u'_i(y'_{[r]\setminus\{i\}})\Big| y, y' \in G^r,\ \psi(y) = \psi(y') = x\Big] \\
&= \mathbb{E}\Big[\prod_{i=1}^r u_i(y_{[r]\setminus\{i\}})u'_i(y_{[r]\setminus\{i\}} + z_{[r]\setminus\{i\}})\Big| y, z \in G^r,\ \psi(y) = x, \psi(z) = 0\Big] \\
&= \mathbb{E}[(v_{1,z_{[r]\setminus\{1\}}}, v_{2,z_{[r]\setminus\{2\}}}, \dots, v_{r,z_{[r]\setminus\{r\}}})^*_\psi(x) | z \in G^r,\ \psi(z) = 0]
\end{aligned}$$

where $v_{i,z_{[r]\setminus\{i\}}} \colon G^{r-1} \to [0,1]$ is defined by $v_{i,z_{[r]\setminus\{i\}}}(y_{[r]\setminus\{i\}}) = u_i(y_{[r]\setminus\{i\}})u'_i(y_{[r]\setminus\{i\}} + z_{[r]\setminus\{i\}})$. This shows that the product of two such generalized convolutions is a convex combination of generalized convolutions, so that $\mathcal{F}$ is closed under multiplication. ■

4.2 Counting lemma

Next we show that if $(f, \tilde{f})$ is a $(k-1, \epsilon)$-discrepancy pair, with $f \le \nu$ and $\tilde{f} \le 1$, then f and $\tilde{f}$ have similar number of (weighted) k-term APs. This is a special case of the counting lemma for sparse hypergraphs, Theorem 3.1.17.

Lemma 4.2.1 (k-AP counting lemma). *For every $k \ge 3$ and $\gamma > 0$, there exists an $\epsilon > 0$ so that the following holds.*

Let $\nu, f, \tilde{f} \colon \mathbb{Z}_N \to \mathbb{R}_{\ge 0}$ be functions. Suppose that ν satisfies the k-linear forms condition and N is sufficiently large. Suppose also that $0 \le f \le \nu$, $0 \le \tilde{f} \le 1$, and $(f, \tilde{f})$ is a $(k-1, \epsilon)$-discrepancy pair with respect to each of $\psi_1, \dots, \psi_k$, where $\psi_j \colon \mathbb{Z}_N^{k-1} \to \mathbb{Z}_N$ is defined by

$$\psi_j(x_1, \dots, x_{j-1}, x_{j+1}, \cdots, x_k) := \sum_{i \in [k]\setminus\{j\}} (i-j)x_i.$$

Then

$$\left|\mathbb{E}\Big[\prod_{i=0}^{k-1} f(a+id)\Big| a, d \in \mathbb{Z}_N\Big] - \mathbb{E}\Big[\prod_{i=0}^{k-1} \tilde{f}(a+id)\Big| a, d \in \mathbb{Z}_N\Big]\right| \le \gamma. \tag{4.2}$$

Let us explain why Lemma 4.2.1 is a special case of Theorem 3.1.17. We use the hypergraph notation from Chapter 3. Let $V = (J, (V_j)_{j\in J}, k-1, H)$ be a hypergraph system, where $J = [k]$, $V_j = \mathbb{Z}_N$ for every $j \in J$, and $H = \binom{J}{k-1}$ (corresponding to a simplex). Let $(\nu_e)_{e\in H}$, $(g_e)_{e\in H}$, and $(\tilde{g}_e)_{e\in H}$ be weighted hypergraphs on V defined by

$$\begin{aligned}
\nu_{[k]\setminus\{j\}}(x_{[k]\setminus\{j\}}) &= \nu(\psi_j(x_{[k]\setminus\{j\}})) \\
g_{[k]\setminus\{j\}}(x_{[k]\setminus\{j\}}) &= f(\psi_j(x_{[k]\setminus\{j\}})) \\
\tilde{g}_{[k]\setminus\{j\}}(x_{[k]\setminus\{j\}}) &= \tilde{f}(\psi_j(x_{[k]\setminus\{j\}}))
\end{aligned}$$

for $j \in [k]$ and $x_{[k]\setminus\{j\}} \in V_{[k]\setminus\{j\}} = \mathbb{Z}_N^{k-1}$. Then the weighted hypergraph $(\nu_e)_{e\in H}$ satisfies the H-linear forms condition, Definition 3.1.8 (which is equivalent to $\nu\colon \mathbb{Z}_N \to \mathbb{R}_{\geq 0}$ satisfying the k-linear forms condition). That $(f, \tilde{f})$ is a $(k-1, \epsilon)$-discrepancy pair with respect to ψ_j is equivalent to $(g_{[k]\setminus\{j\}}, \tilde{g}_{[k]\setminus\{j\}})$ being an ϵ-discrepancy pair as weighted hypergraphs (Definition 3.1.13). Note that

$$\mathbb{E}\Big[\prod_{i=0}^{k-1} f(a+id)\Big| x, d \in \mathbb{Z}_N\Big] = \mathbb{E}\Big[\prod_{e\in H} g_e(x_e)\Big| x \in V_J\Big]$$

(to see this, let $a = \psi_1(x_2, \ldots, x_k)$ and $d = -(x_1 + \cdots + x_k)$) and similarly with $\tilde{f}$ and $\tilde{g}_e$. Then the relative hypergraph counting lemma, Theorem 3.1.17, reduces to Lemma 4.2.1.

4.3 Proof of the relative Szemerédi theorem

Proof of Theorem 4.0.2. We begin with the following simple observation, that for any $g, g'\colon \mathbb{Z}_N \to \mathbb{R}_{\geq 0}$, if (g, g') is a $(k-1, \epsilon)$-discrepancy pair with respect to one ψ_j from Lemma 4.2.1, then it is so with respect to all ψ_j. This is simply because $1, 2, \ldots, k-1$ all have multiplicative inverses in $\mathbb{Z}_N$, as N is coprime to $(k-1)!$, and a scaling of variables in (4.1) allows one to convert one linear form ψ_j to another $\psi_{j'}$.

The linear forms condition on ν implies that $(\nu, 1)$ is a $(k-1, o(1))$-discrepancy pair with respect to ψ_1 from Lemma 4.2.1. Indeed, we have the following inequality

$$\begin{aligned} &\left|\mathbb{E}\left[(\nu(\psi_1(x)) - 1)\prod_{i=1}^{r} u_i(x_{[r]\setminus\{i\}})\middle| x \in G^r\right]\right| \\ &\qquad \le \mathbb{E}\left[\prod_{\omega\in\{0,1\}^r} (\nu(\psi_1(x^{(\omega)})) - 1)\middle| x^{(0)}, x^{(1)} \in G^r\right]^{1/2^r} \end{aligned} \tag{4.3}$$

which is proved by a sequence of Cauchy-Schwarz inequalities, similar to Lemma 3.5.2. The right-hand side of (4.3) is $o(1)$ by the linear forms condition (expand the product so that each term is $\pm 1 + o(1)$ by (3.2), and everything cancels accordingly).

Since $(\nu, 1)$ is a $(k-1, o(1))$-discrepancy pair with respect to ψ_1, Lemma 4.1.3 implies that there exists $\tilde{f}\colon G \to [0,1]$ so that $\mathbb{E}[\tilde{f}] = \mathbb{E}[f] \ge \delta$ (if $\mathbb{E}[f] > 1$, then replace f by $\delta f/\mathbb{E}[f]$) and $(f, \tilde{f})$ is a $(k-1, o(1))$-discrepancy pair with respect to ψ_1, and hence with respect to all ψ_j, $1 \le j \le k$. So

$$\begin{aligned} \mathbb{E}\left[\prod_{i=0}^{k-1} f(x+id)\middle| x, d \in \mathbb{Z}_N\right] &\ge \mathbb{E}\left[\prod_{i=0}^{k-1} \tilde{f}(x+id)\middle| x, d \in \mathbb{Z}_N\right] - o(1) \\ &\ge c(k,\delta) - o_{k,\delta}(1), \end{aligned}$$

where the first inequality is by Lemma 4.2.1 and the second inequality is by Theorem 4.0.1. ■

Chapter 5

Multidimensional Szemerédi theorem in the primes

Let $\mathcal{P}_N$ denote the set of primes at most N, and let $[N] := \{1, 2, \ldots, N\}$. Tao [123] conjectured the following result as a natural extension of the Green-Tao theorem [69] on arithmetic progressions in the primes and the Furstenberg-Katznelson [54] multidimensional generalization of Szemerédi's theorem. Special cases of this conjecture were proven in [34] and [93]. The conjecture was very recently resolved by Cook, Magyar, and Titichetrakun [35] and independently by Tao and Ziegler [125].

Theorem 5.0.1. *Let d be a positive integer, $v_1, \ldots, v_k \in \mathbb{Z}^d$, and $\delta > 0$. Then, if N is sufficiently large, every subset A of $\mathcal{P}_N^d$ of cardinality $|A| \geq \delta |\mathcal{P}_N|^d$ contains a set of the form $a + tv_1, \ldots, a + tv_k$, where $a \in \mathbb{Z}^d$ and t is a positive integer.*

In this chapter we give a short alternative proof of the theorem, using the landmark result of Green and Tao [71] (which is conditional on results later proved in [72] and with Ziegler in [73]) on the asymptotics for the number of primes satisfying certain systems of linear equations, as well as the following multidimensional generalization of Szemerédi's theorem established by Furstenberg and Katznelson [54].

Theorem 5.0.2 (Multidimensional Szemerédi theorem [54]). *Let d be a positive integer, $v_1, \ldots, v_k \in \mathbb{Z}^d$, and $\delta > 0$. If N is sufficiently large, then every subset A of*

$[N]^d$ *of cardinality* $|A| \geq \delta N^d$ *contains a set of the form* $a + tv_1, \ldots, a + tv_k$*, where* $a \in \mathbb{Z}^d$ *and* t *is a positive integer.*

To prove Theorem 5.0.1, we begin by fixing $d, v_1, \ldots, v_k, \delta$. Using Theorem 5.0.2, we can fix a large integer $m > 2d/\delta$ so that any subset of $[m]^d$ with at least $\delta m^d/2$ elements contains a set of the form $a+tv_1, \ldots, a+tv_k$, where $a \in \mathbb{Z}^d$ and t is a positive integer.

We next discuss a sketch of the proof idea. The Green-Tao theorem [69] tells us that there are arbitrarily long arithmetic progressions in the primes. It follows that for N large, $\mathcal{P}_N^d$ contains homothetic copies of $[m]^d$. We use a Varnavides-type argument [133] and consider a random homothetic copy of the grid $[m]^d$ inside $\mathcal{P}_N^d$. In expectation, the set A should occupy at least a $\delta/2$ fraction of the random homothetic copy of $[m]^d$. This follows from a linearity of expectation argument. Indeed, the Green-Tao-Ziegler result [71, 72, 73] and a second moment argument imply that most points of $\mathcal{P}_N^d$ appear in about the expected number of such copies of the grid $[m]^d$. Once we find a homothetic copy of $[m]^d$ containing at least $\delta m^d/2$ elements of A, we obtain by Theorem 5.0.2 a subset of A of the form $a + tv_1, \ldots, a + tv_k$, as desired.

To make the above idea actually work, we first apply the W-trick as described below. This is done to avoid certain biases in the primes. We also only consider homothetic copies of $[m]^d$ with common difference $r \leq N/m^2$ in order to guarantee that almost all elements of $\mathcal{P}_N^d$ are in about the same number of such homothetic copies of $[m]^d$.

Remarks. This argument also produces a relative multidimensional Szemerédi theorem, where the complexity of the linear forms condition on the majorizing measure depends on $d, v_1, \ldots, v_k$ and δ. It seems plausible that the dependence on δ is unnecessary; this was shown for the one-dimensional case in Chapters 3 and 4. Our arguments share some features with those of Tao and Ziegler [125], who also use the results in [71, 72, 73]. However, the proof in [125] first establishes a relativized version of the Furstenberg correspondence principle and then proceeds in the ergodic theoretic setting, whereas we go directly to the multidimensional Szemerédi theorem. Cook, Magyar, and Titichetrakun [35] take a different approach and develop a relative hypergraph removal lemma from scratch, and they also require a linear forms condition whose complexity depend on δ.

Conditional on a certain polynomial extension of the Green-Tao-Ziegler result (c.f. the Bateman-Horn conjecture [11]), one can also combine this sampling argument with the polynomial extension of Szemerédi's theorem by Bergelson and Leibman [13] to obtain a polynomial extension of Theorem 5.0.1.

The hypothesis that $|A| \geq \delta |\mathcal{P}_N|^d$ implies that

$$\sum_{n_1,\dots,n_d \in [N]} 1_A(n_1,\dots,n_d)\Lambda'(n_1)\cdots\Lambda'(n_d) \geq (\delta - o(1))N^d, \tag{5.1}$$

where 1_A is the indicator function of A, and $o(1)$ denotes some quantity that goes to zero as $N \to \infty$, and $\Lambda'(p) = \log p$ for prime p and $\Lambda'(n) = 0$ for nonprime n.

Next we apply the W-trick [71, Sec. 5]. Fix some slowly growing function $w = w(N)$; the choice $w := \log\log\log N$ will do. Define $W := \prod_{p \leq w} p$ to be the product of all primes at most w. For each $b \in [W]$ with $\gcd(b, W) = 1$, define

$$\Lambda'_{b,W}(n) := \frac{\phi(W)}{W}\Lambda'(Wn + b)$$

where $\phi(W) = \#\{b \in [W] : \gcd(b, W) = 1\}$ is the Euler totient function. Also define

$$1_{A_{b_1,\dots,b_d,W}}(n_1,\dots,n_d) := 1_A(Wn_1 + b_1,\dots,Wn_d + b_d).$$

By (5.1) and the pigeonhole principle, we can choose $b_1,\dots,b_d \in [W]$ all coprime to W so that

$$\sum_{1 \leq n_1,\dots,n_d \leq N/W} 1_{A_{b_1,\dots,b_d,W}}(n_1,\dots,n_d)\Lambda'_{b_1,W}(n)\Lambda'_{b_2,W}(n)\cdots\Lambda'_{b_d,W}(n) \geq (\delta - o(1))\left(\frac{N}{W}\right)^d, \tag{5.2}$$

We shall write

$$\widetilde{N} := \lfloor N/W \rfloor, \qquad R := \lfloor \widetilde{N}/m^2 \rfloor, \qquad \widetilde{A} := 1_{A_{b_1,\ldots,b_d,W}} \qquad \text{and} \qquad \widetilde{\Lambda}_j := \Lambda'_{b_j,W}$$

(all depending on N). So (5.2) reads

$$\sum_{n_1,\ldots,n_d \in [\widetilde{N}]} \widetilde{A}(n_1,\ldots,n_d)\widetilde{\Lambda}_1(n_1)\widetilde{\Lambda}_2(n_2)\cdots\widetilde{\Lambda}_d(n_d) \geq (\delta - o(1))\widetilde{N}^d \tag{5.3}$$

The Green-Tao result [71] (along with [72, 73]) says that $\Lambda'_{b_j,W}$ acts pseudorandomly with average value about 1 in terms of counts of linear forms. The statement below is an easy corollary of [71, Thm. 5.1].

Theorem 5.0.3 (Pseudorandomness of the W-tricked primes). *Fix a linear map $\Psi = (\psi_1,\ldots,\psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ (in particular $\Psi(0) = 0$) where no two ψ_i, ψ_j are linearly dependent. Let $K \subseteq [-\widetilde{N},\widetilde{N}]^d$ be any convex body. Then, for any $b_1,\ldots,b_t \in [W]$ all coprime to W, we have*

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{j \in [t]} \Lambda'_{b_j,W}(\psi_j(n)) = \#\{n \in K \cap \mathbb{Z}^d : \psi_j(n) > 0 \ \forall j\} + o(\widetilde{N}^d).$$

where $o(\widetilde{N}^d) := o(1)\widetilde{N}^d$. Note that the error term does not depend on $b_1,\ldots,b_t$ (although it does depend on Ψ).

The next lemma shows that A in expectation contains a considerable fraction of a random homothetic copy of $[m]^d$ with common difference at most $R = \lfloor N/m^2 \rfloor$ in the W-tricked subgrid of $\mathcal{P}_N^d$.

Lemma 5.0.4. *If $\widetilde{A}$ satisfies* (5.3)*, then*

$$\sum_{\substack{n_1,\ldots,n_d \in [\widetilde{N}] \\ r \in [R]}} \left(\sum_{i_1,\ldots,i_d \in [m]} \widetilde{A}(n_1 + i_1 r, \ldots, n_d + i_d r) \right) \prod_{j \in [d]} \prod_{i \in [m]} \widetilde{\Lambda}_j(n_j + ir) \geq (\delta m^d - dm^{d-1} - o(1))R\widetilde{N}^d. \tag{5.4}$$

Proof of Theorem 5.0.1 (assuming Lemma 5.0.4) . By Theorem 5.0.3 we have

$$\sum_{\substack{n_1,\ldots,n_d\in[\widetilde{N}]\\ r\in[R]}} \prod_{j\in[d]} \prod_{i\in[m]} \widetilde{\Lambda}_j(n_j+ir) = (1+o(1))R\widetilde{N}^d,$$

So by (5.4), for sufficiently large N, there exists some choice of $n_1,\ldots,n_d \in [\widetilde{N}]$ and $r \in [R]$ so that

$$\sum_{i_1,\ldots,i_d\in[m]} \widetilde{A}(n_1+i_1r,\ldots,n_d+i_dr) \geq \frac{1}{2}\delta m^d.$$

This means that a certain dilation of the grid $[m]^d$ contains at least $\delta m^d/2$ elements of A, from which it follows by the choice of m that it must contain a set of the form $a+tv_1,\ldots,a+tv_k$. ■

Lemma 5.0.4 follows from the next lemma by summing over all choices of $i_1,\ldots,i_d \in [m]$.

Lemma 5.0.5. *Suppose $\widetilde{A}$ satisfies* (5.3). *Fix $i_1,\ldots,i_d \in [m]$. Then we have*

$$\sum_{\substack{n_1,\ldots,n_d\in[\widetilde{N}]\\ r\in[R]}} \widetilde{A}(n_1+i_1r,\ldots,n_d+i_dr) \prod_{j\in[d]} \prod_{i\in[m]} \widetilde{\Lambda}_j(n_j+ir) \geq \left(\delta - \frac{d}{m} - o(1)\right) R\widetilde{N}^d. \quad (5.5)$$

Proof. By a change of variables $n'_j = n_j + i_jr$ for each j, we write the LHS of (5.5) as

$$\sum_{r\in[R]} \sum_{\substack{n'_1,\ldots,n'_d\in\mathbb{Z}\\ n'_j-i_jr\in[\widetilde{N}]\ \forall j}} \widetilde{A}(n'_1,\ldots,n'_d) \prod_{j\in[d]} \prod_{i\in[m]} \widetilde{\Lambda}_j(n'_j+(i-i_j)r). \quad (5.6)$$

Recall that $R = \lfloor \widetilde{N}/m^2 \rfloor$. Note that (5.6) is at least

$$\sum_{r\in[R]} \sum_{\widetilde{N}/m<n'_1,\ldots,n'_d\leq\widetilde{N}} \widetilde{A}(n'_1,\ldots,n'_d) \prod_{j\in[d]} \prod_{i\in[m]} \widetilde{\Lambda}_j(n'_j+(i-i_j)r). \quad (5.7)$$

By (5.3) and Theorem 5.0.3 we have

$$\sum_{\widetilde{N}/m<n_1,\ldots,n_d\leq\widetilde{N}} \widetilde{A}(n_1,\ldots,n_d)\widetilde{\Lambda}_1(n_1)\widetilde{\Lambda}_2(n_2)\cdots\widetilde{\Lambda}_d(n_d) \geq \left(\delta-\frac{d}{m}-o(1)\right)\widetilde{N}^d \quad (5.8)$$

(the difference between the left-hand side sums of (5.3) and (5.8) consists of terms with $(n_1,\ldots,n_d)$ in some box of the form $[\widetilde{N}]^{j-1}\times[\widetilde{N}/m]\times[\widetilde{N}]^{d-j}$, which can be upper bounded by using $\widetilde{A}\leq 1$, applying Theorem 5.0.3, and then taking the union bound over $j\in[d]$). It remains to show that

$$(5.7) - R\cdot(\text{LHS of }(5.8)) = o(\widetilde{N}^{d+1}).$$

We have

$$\begin{aligned}
&(5.7) - R\cdot(\text{LHS of }(5.8))\\
&= \sum_{\substack{\widetilde{N}/m<n'_1,\ldots,n'_d\leq\widetilde{N}\\ r\in[R]}} \widetilde{A}(n'_1,\ldots,n'_d)\left(\prod_{j\in[d]}\prod_{i\in[m]}\widetilde{\Lambda}_j(n'_j+(i-i_j)r) - \prod_{j\in[d]}\widetilde{\Lambda}_j(n'_j)\right)\\
&= \sum_{\widetilde{N}/m<n'_1,\ldots,n'_d\leq\widetilde{N}} \widetilde{A}(n'_1,\ldots,n'_d)\left(\prod_{j\in[d]}\widetilde{\Lambda}_j(n'_j)\right)\cdot\\
&\qquad\cdot\left(\sum_{r\in[R]}\left(\prod_{j\in[d]}\prod_{i\in[m]\setminus\{i_j\}}\widetilde{\Lambda}_j(n'_j+(i-i_j)r)\ -1\right)\right).
\end{aligned}$$

By the Cauchy-Schwarz inequality and $0\leq\widetilde{A}\leq 1$, the above expression can be bounded in absolute value by $\sqrt{ST}$, where

$$\begin{aligned}
S &= \sum_{\widetilde{N}/m<n'_1,\ldots,n'_d\leq\widetilde{N}}\prod_{j\in[d]}\widetilde{\Lambda}_j(n'_j),\\
T &= \sum_{\widetilde{N}/m<n'_1,\ldots,n'_d\leq\widetilde{N}}\left(\prod_{j\in[d]}\widetilde{\Lambda}_j(n'_j)\right)\left(\sum_{r\in[R]}\left(\prod_{j\in[d]}\prod_{i\in[m]\setminus\{i_j\}}\widetilde{\Lambda}_j(n'_j+(i-i_j)r)\ -1\right)\right)^2\\
&= T_1 - 2T_2 + T_3,
\end{aligned}$$

and

$$T_1 = \sum_{\substack{\widetilde{N}/m < n'_1, \ldots, n'_d \le \widetilde{N} \\ r, r' \in [R]}} \prod_{j \in [d]} \widetilde{\Lambda}_j(n'_j) \prod_{i \in [m] \setminus \{i_j\}} \widetilde{\Lambda}_j(n'_j + (i - i_j)r) \widetilde{\Lambda}_j(n'_j + (i - i_j)r'),$$

$$T_2 = \sum_{\substack{\widetilde{N}/m < n'_1, \ldots, n'_d \le \widetilde{N} \\ r, r' \in [R]}} \prod_{j \in [d]} \widetilde{\Lambda}_j(n'_j) \prod_{i \in [m] \setminus \{i_j\}} \widetilde{\Lambda}_j(n'_j + (i - i_j)r),$$

$$T_3 = \sum_{\substack{\widetilde{N}/m < n'_1, \ldots, n'_d \le \widetilde{N} \\ r, r' \in [R]}} \prod_{j \in [d]} \widetilde{\Lambda}_j(n'_j).$$

By Theorem 5.0.3 we have $S = O(\widetilde{N}^d)$, and T_1, T_2, T_3 pairwise differ by $o(\widetilde{N}^{d+2})$, so that $T = o(\widetilde{N}^{d+2})$. Thus $\sqrt{ST} = o(\widetilde{N}^{d+1})$, as desired. ∎

Bibliography

[1] E. Aigner-Horev, H. Hàn, and M. Schacht, *Extremal results for odd cycles in sparse pseudorandom graphs*, Combinatorica **34** (2014), 379–406.

[2] N. Alon, *Explicit Ramsey graphs and orthonormal labelings*, Electron. J. Combin. **1** (1994), Research Paper 12, approx. 8 pp. (electronic).

[3] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster, *The algorithmic aspects of the regularity lemma*, J. Algorithms **16** (1994), 80–109.

[4] N. Alon and N. Kahale, *Approximating the independence number via the θ-function*, Math. Program. **80** (1998), 253–264.

[5] N. Alon, M. Krivelevich, and B. Sudakov, *Turán numbers of bipartite graphs and related Ramsey-type questions*, Combin. Probab. Comput. **12** (2003), 477–494, Special issue on Ramsey theory.

[6] N. Alon and A. Naor, *Approximating the cut-norm via Grothendieck's inequality*, SIAM J. Comput. **35** (2006), 787–803 (electronic).

[7] N. Alon, A. Coja-Oghlan, H. Hàn, M. Kang, V. Rödl, and M. Schacht, *Quasi-randomness and algorithmic regularity for graphs with general degree distributions*, SIAM J. Comput. **39** (2010), 2336–2362.

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

[9] N. Alon, A. Shapira, and B. Sudakov, *Additive approximation for edge-deletion problems*, Ann. of Math. (2) **170** (2009), 371–411.

[10] J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, J. Amer. Math. Soc., to appear.

[11] P. T. Bateman and R. A. Horn, *A heuristic asymptotic formula concerning the distribution of prime numbers*, Math. Comp. **16** (1962), 363–367.

[12] P. Bennett and T. Bohman, *A note on the random greedy independent set algorithm*, arXiv:1308.3732.

[13] V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc. **9** (1996), 725–753.

[14] Y. Bilu and N. Linial, *Lifts, discrepancy and nearly optimal spectral gap*, Combinatorica **26** (2006), 495–519.

[15] B. Bollobás and O. Riordan, *Metrics for sparse graphs*, Surveys in combinatorics 2009, London Math. Soc. Lecture Note Ser., vol. 365, Cambridge Univ. Press, Cambridge, 2009, pp. 211–287.

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

[17] S. A. Burr and V. Rosta, *On the Ramsey multiplicities of graphs—problems and recent results*, J. Graph Theory **4** (1980), 347–361.

[18] S. Butler, *Induced-universal graphs for graphs with bounded maximum degree*, Graphs Combin. **25** (2009), 461–468.

[19] F. R. K. Chung, *A spectral Turán theorem*, Combin. Probab. Comput. **14** (2005), 755–767.

[20] F. R. K. Chung and R. L. Graham, *Quasi-random set systems*, J. Amer. Math. Soc. **4** (1991), 151–196.

[21] F. R. K. Chung and R. L. Graham, *Sparse quasi-random graphs*, Combinatorica **22** (2002), 217–244, Special issue: Paul Erdős and his mathematics.

[22] F. R. K. Chung and R. L. Graham, *Quasi-random graphs with given degree sequences*, Random Structures Algorithms **32** (2008), 1–19.

[23] F. R. K. Chung, R. L. Graham, and R. M. Wilson, *Quasi-random graphs*, Combinatorica **9** (1989), 345–362.

[24] V. Chvátal, V. Rödl, E. Szemerédi, and W. T. Trotter, Jr., *The Ramsey number of a graph with bounded maximum degree*, J. Combin. Theory Ser. B **34** (1983), 239–243.

[25] A. Coja-Oghlan, C. Cooper, and A. Frieze, *An efficient sparse regularity concept*, SIAM J. Discrete Math. **23** (2009/10), 2000–2034.

[26] D. Conlon, *A new upper bound for diagonal Ramsey numbers*, Ann. of Math. (2) **170** (2009), 941–960.

[27] D. Conlon and J. Fox, *Bounds for graph regularity and removal lemmas*, Geom. Funct. Anal. **22** (2012), 1191–1256.

[28] D. Conlon, J. Fox, and B. Sudakov, *An approximate version of Sidorenko's conjecture*, Geom. Funct. Anal. **20** (2010), 1354–1366.

[29] D. Conlon, J. Fox, and Y. Zhao, *Linear forms from the Gowers uniformity norm*, unpublished companion note, arXiv:1305.5565.

[30] D. Conlon, J. Fox, and Y. Zhao, *A relative Szemerédi theorem.*, Geom. Funct. Anal., to appear.

[31] D. Conlon, J. Fox, and Y. Zhao, *Extremal results in sparse pseudorandom graphs*, Adv. Math. **256** (2014), 206–290.

[32] D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, Submitted.

[33] D. Conlon, W. T. Gowers, W. Samotij, and M. Schacht, *On the KŁR conjecture in random graphs*, Israel J. Math. **203** (2014), 535–580.

[34] B. Cook and A. Magyar, *Constellations in* $\mathbb{P}^d$, Int. Math. Res. Not. IMRN **2012** (2012), 2794–2816.

[35] B. Cook, A. Magyar, and T. Titichetrakun, *A multidimensional Szemerédi theorem in the primes*, arXiv:1306.3025.

[36] W. Deuber, *Generalizations of Ramsey's theorem*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. I, North-Holland, Amsterdam, 1975, pp. 323–332. Colloq. Math. Soc. János Bolyai, Vol. 10.

[37] A. Dudek and V. Rödl, *On the Folkman number* $f(2,3,4)$, Experiment. Math. **17** (2008), 63–67.

[38] R. A. Duke, H. Lefmann, and V. Rödl, *A fast approximation algorithm for computing the frequencies of subgraphs in a given graph*, SIAM J. Comput. **24** (1995), 598–620.

[39] P. Erdős, *On the number of complete subgraphs contained in certain graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **7** (1962), 459–464.

[40] P. Erdős, *Some recent results on extremal problems in graph theory. Results*, Theory of Graphs (Internat. Sympos., Rome, 1966), Gordon and Breach, New York, 1967, pp. 117–123 (English); pp. 124–130 (French).

[41] P. Erdős, M. Goldberg, J. Pach, and J. Spencer, *Cutting a graph into two dissimilar halves*, J. Graph Theory **12** (1988), 121–131.

[42] P. Erdős, A. Hajnal, and L. Pósa, *Strong embeddings of graphs into colored graphs*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. I, North-Holland, Amsterdam, 1975, pp. 585–595. Colloq. Math. Soc. János Bolyai, Vol. 10.

[43] P. Erdős and J. Spencer, *Imbalances in k-colorations*, Networks **1** (1971/72), 379–385.

[44] P. Erdős and A. H. Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc. **52** (1946), 1087–1091.

[45] J. Fox, *There exist graphs with super-exponential Ramsey multiplicity constant*, J. Graph Theory **57** (2008), 89–98.

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

[47] J. Fox, C. Lee, and B. Sudakov, *Chromatic number, clique subdivisions, and the conjectures of Hajós and Erdős-Fajtlowicz*, Combinatorica **33** (2013), 181–197.

[48] J. Fox and B. Sudakov, *Induced Ramsey-type theorems*, Adv. Math. **219** (2008), 1771–1800.

[49] J. Fox and B. Sudakov, *Two remarks on the Burr-Erdős conjecture*, European J. Combin. **30** (2009), 1630–1645.

[50] J. Fox and Y. Zhao, *A short proof of the multidimensional Szemerédi theorem in the primes*, Amer. J. Math., to appear.

[51] P. Frankl and V. Rödl, *Extremal problems on set systems*, Random Structures Algorithms **20** (2002), 131–164.

[52] E. Friedgut, V. Rödl, and M. Schacht, *Ramsey properties of random discrete structures*, Random Structures Algorithms **37** (2010), 407–436.

[53] A. Frieze and R. Kannan, *Quick approximation to matrices and applications*, Combinatorica **19** (1999), 175–220.

[54] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291.

[55] S. Gerke, Y. Kohayakawa, V. Rödl, and A. Steger, *Small subsets inherit sparse ϵ-regularity*, J. Combin. Theory Ser. B **97** (2007), 34–56.

[56] S. Gerke, M. Marciniszyn, and A. Steger, *A probabilistic counting lemma for complete graphs*, Random Structures Algorithms **31** (2007), 517–534.

[57] S. Gerke and A. Steger, *The sparse regularity lemma and its applications*, Surveys in Combinatorics 2005, London Math. Soc. Lecture Note Ser., vol. 327, Cambridge Univ. Press, Cambridge, 2005, pp. 227–258.

[58] D. A. Goldston and C. Y. Yıldırım, *Higher correlations of divisor sums related to primes. I. Triple correlations*, Integers **3** (2003), A5, 66.

[59] A. W. Goodman, *On sets of acquaintances and strangers at any party*, Amer. Math. Monthly **66** (1959), 778–783.

[60] W. T. Gowers, *Lower bounds of tower type for Szemerédi's uniformity lemma*, Geom. Funct. Anal. **7** (1997), 322–337.

[61] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465–588.

[62] W. T. Gowers, *Quasirandomness, counting and regularity for 3-uniform hypergraphs*, Combin. Probab. Comput. **15** (2006), 143–184.

[63] W. T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. (2) **166** (2007), 897–946.

[64] W. T. Gowers, *Quasirandom groups*, Combin. Probab. Comput. **17** (2008), 363–387.

[65] W. T. Gowers, *Decompositions, approximate structure, transference, and the Hahn-Banach theorem*, Bull. Lond. Math. Soc. **42** (2010), 573–606.

[66] R. L. Graham, V. Rödl, and A. Ruciński, *On graphs with linear Ramsey numbers*, J. Graph Theory **35** (2000), 176–192.

[67] B. Green, Personal communication.

[68] B. Green, *A Szemerédi-type regularity lemma in abelian groups, with applications*, Geom. Funct. Anal. **15** (2005), 340–376.

[69] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), 481–547.

[70] B. Green and T. Tao, *New bounds for Szemerédi's theorem. II. A new bound for $r_4(N)$*, Analytic number theory, Cambridge Univ. Press, Cambridge, 2009, pp. 180–204.

[71] B. Green and T. Tao, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), 1753–1850.

[72] B. Green and T. Tao, *The Möbius function is strongly orthogonal to nilsequences*, Ann. of Math. (2) **175** (2012), 541–566.

[73] B. Green, T. Tao, and T. Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Ann. of Math. (2) **176** (2012), 1231–1372.

[74] P. E. Haxell, Y. Kohayakawa, and T. Łuczak, *Turán's extremal problem in random graphs: forbidding even cycles*, J. Combin. Theory Ser. B **64** (1995), 273–287.

[75] Y. Kohayakawa, *Szemerédi's regularity lemma for sparse graphs*, Foundations of computational mathematics (Rio de Janeiro, 1997), Springer, Berlin, 1997, pp. 216–230.

[76] Y. Kohayakawa, T. Łuczak, and V. Rödl, *On K^4-free subgraphs of random graphs*, Combinatorica **17** (1997), 173–213.

[77] Y. Kohayakawa and V. Rödl, *Regular pairs in sparse random graphs. I*, Random Structures Algorithms **22** (2003), 359–434.

[78] Y. Kohayakawa and V. Rödl, *Szemerédi's regularity lemma and quasi-randomness*, Recent advances in algorithms and combinatorics, CMS Books Math./Ouvrages Math. SMC, vol. 11, Springer, New York, 2003, pp. 289–351.

[79] Y. Kohayakawa, V. Rödl, M. Schacht, P. Sissokho, and J. Skokan, *Turán's theorem for pseudo-random graphs*, J. Combin. Theory Ser. A **114** (2007), 631–657.

[80] Y. Kohayakawa, V. Rödl, M. Schacht, and J. Skokan, *On the triangle removal lemma for subgraphs of sparse pseudorandom graphs*, An Irregular Mind (Szemerédi is 70), Bolyai Soc. Math. Stud., vol. 21, Springer Berlin, 2010, pp. 359–404.

[81] Y. Kohayakawa, V. Rödl, M. Schacht, and E. Szemerédi, *Sparse partition universal graphs for graphs of bounded degree*, Adv. Math. **226** (2011), 5041–5065.

[82] Y. Kohayakawa, V. Rödl, and P. Sissokho, *Embedding graphs with bounded degree in sparse pseudorandom graphs*, Israel J. Math. **139** (2004), 93–137.

[83] J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc. Math. Stud., vol. 2, János Bolyai Math. Soc., Budapest, 1996, pp. 295–352.

[84] D. Král', O. Serra, and L. Vena, *A combinatorial proof of the removal lemma for groups*, J. Combin. Theory Ser. A **116** (2009), 971–978.

[85] D. Král', O. Serra, and L. Vena, *A removal lemma for systems of linear equations over finite fields*, Israel J. Math. **187** (2012), 193–207.

[86] M. Krivelevich and B. Sudakov, *Pseudo-random graphs*, More sets, graphs and numbers, Bolyai Soc. Math. Stud., vol. 15, Springer, Berlin, 2006, pp. 199–262.

[87] T. H. Lê, *Green-Tao theorem in function fields*, Acta Arith. **147** (2011), 129–152.

[88] J. L. X. Li and B. Szegedy, *On the logarithmic calculus and Sidorenko's conjecture*, Combinatorica, to appear.

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

[90] L. Lovász and B. Szegedy, *Szemerédi's lemma for the analyst*, Geom. Funct. Anal. **17** (2007), 252–270.

[91] L. Lu, *Explicit construction of small Folkman graphs*, SIAM J. Discrete Math. **21** (2007), 1053–1060.

[92] T. Łuczak and V. Rödl, *On induced Ramsey numbers for graphs with bounded maximum degree*, J. Combin. Theory Ser. B **66** (1996), 324–333.

[93] A. Magyar and T. Titichetrakun, *Corners in dense subsets of* $\mathbf{P}^d$, arXiv:1306.3026.

[94] L. Matthiesen, *Correlations of the divisor function*, Proc. Lond. Math. Soc. **104** (2012), 827–858.

[95] L. Matthiesen, *Linear correlations amongst numbers represented by positive definite binary quadratic forms*, Acta Arith. **154** (2012), 235–306.

[96] B. Nagle, V. Rödl, and M. Schacht, *The counting lemma for regular k-uniform hypergraphs*, Random Structures Algorithms **28** (2006), 113–179.

[97] V. Nikiforov, *The number of cliques in graphs of given order and size*, Trans. Amer. Math. Soc. **363** (2011), 1599–1618.

[98] F. P. Ramsey, *On a problem of formal logic*, Proc. London Math. Soc. Ser. 2 **30** (1930), 264–286.

[99] A. A. Razborov, *On the minimal density of triangles in graphs*, Combin. Probab. Comput. **17** (2008), 603–618.

[100] C. Reiher, *The clique density theorem*, Submitted.

[101] O. Reingold, L. Trevisan, M. Tulsiani, and S. Vadhan, *New proofs of the Green-Tao-Ziegler dense model theorem: An exposition*, arXiv:0806.0381.

[102] O. Reingold, L. Trevisan, M. Tulsiani, and S. Vadhan, *Dense subsets of pseudorandom sets*, 49th Annual IEEE Symposium on Foundations of Computer Science, IEEE Computer Society, 2008, pp. 76–85.

[103] V. Rödl, *The dimension of a graph and generalized Ramsey theorems*, Master's thesis, Charles University, 1973.

[104] V. Rödl and A. Ruciński, *Threshold functions for Ramsey properties*, J. Amer. Math. Soc. **8** (1995), 917–942.

[105] V. Rödl and M. Schacht, *Regularity lemmas for graphs*, Fete of Combinatorics and Computer Science, Bolyai Soc. Math. Stud., vol. 20, János Bolyai Math. Soc., Budapest, 2010, pp. 287–325.

[106] V. Rödl and J. Skokan, *Regularity lemma for k-uniform hypergraphs*, Random Structures Algorithms **25** (2004), 1–42.

[107] V. Rödl and J. Skokan, *Applications of the regularity lemma for uniform hypergraphs*, Random Structures Algorithms **28** (2006), 180–194.

[108] I. Z. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, Colloq. Math. Soc. János Bolyai, vol. 18, North-Holland, Amsterdam, 1978, pp. 939–945.

[109] T. Sanders, *On Roth's theorem on progressions*, Ann. of Math. (2) **174** (2011), 619–636.

[110] D. Saxton and A. Thomason, *Hypergraph containers*, Invent. Math., to appear.

[111] M. Schacht, *Extremal results for random discrete structures*, Submitted.

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

[113] A. Shapira, *A proof of Green's conjecture regarding the removal properties of sets of linear equations*, J. Lond. Math. Soc. **81** (2010), 355–373.

[114] A. Sidorenko, *A correlation inequality for bipartite graphs*, Graphs Combin. **9** (1993), 201–204.

[115] M. Simonovits, *A method for solving extremal problems in graph theory, stability problems*, Theory of Graphs (Proc. Colloq., Tihany, 1966), Academic Press, New York, 1968, pp. 279–319.

[116] M. Simonovits, *Extremal graph problems, degenerate extremal problems, and supersaturated graphs*, Progress in graph theory (Waterloo, Ont., 1982), Academic Press, Toronto, ON, 1984, pp. 419–437.

[117] J. Solymosi, *Note on a generalization of Roth's theorem*, Discrete and computational geometry, Algorithms Combin., vol. 25, Springer, Berlin, 2003, pp. 825–827.

[118] J. Solymosi, *A note on a question of Erdős and Graham*, Combin. Probab. Comput. **13** (2004), 263–267.

[119] B. Sudakov, T. Szabó, and V. H. Vu, *A generalization of Turán's theorem*, J. Graph Theory **49** (2005), 187–195.

[120] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

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

[122] T. Tao, *A remark on Goldston-Yıldırım correlation estimates*, unpublished.

[123] T. Tao, *The Gaussian primes contain arbitrarily shaped constellations*, J. Anal. Math. **99** (2006), 109–176.

[124] T. Tao, *A variant of the hypergraph removal lemma*, J. Combin. Theory Ser. A **113** (2006), 1257–1280.

[125] T. Tao and T. Ziegler, *A multi-dimensional Szemerédi theorem for the primes via a correspondence principle*, Israel J. Math., to appear.

[126] T. Tao and T. Ziegler, *The primes contain arbitrarily long polynomial progressions*, Acta Math. **201** (2008), 213–305.

[127] A. Thomason, *Pseudorandom graphs*, Random graphs '85 (Poznań, 1985), North-Holland Math. Stud., vol. 144, North-Holland, Amsterdam, 1987, pp. 307–331.

[128] A. Thomason, *Random graphs, strongly regular graphs and pseudorandom graphs*, Surveys in Combinatorics 1987 (New Cross, 1987), London Math. Soc. Lecture Note Ser., vol. 123, Cambridge Univ. Press, Cambridge, 1987, pp. 173–195.

[129] A. Thomason, *A disproof of a conjecture of Erdős in Ramsey theory*, J. London Math. Soc. (2) **39** (1989), 246–255.

[130] H. Towsner, *An analytic approach to sparse hypergraphs: hypergraph removal*, arXiv:1204.1884.

[131] L. Trevisan, M. Tulsiani, and S. Vadhan, *Regularity, boosting, and efficiently simulating every high-entropy distribution*, 24th Annual IEEE Conference on Computational Complexity, IEEE Computer Society, 2009, pp. 126–136.

[132] P. Turán, *Eine Extremalaufgabe aus der Graphentheorie*, Mat. Fiz. Lapok **48** (1941), 436–452 (Hungarian, with German summary).

[133] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. **34** (1959), 358–360.

[134] Y. Zhao, *An arithmetic transference proof of a relative Szemerédi theorem*, Math. Proc. Cambridge Philos. Soc. **156** (2014), 255–261.