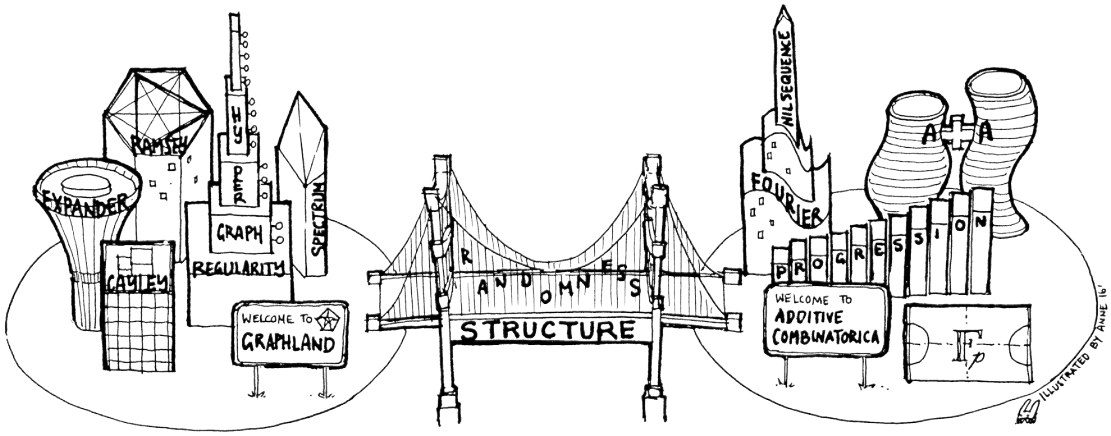
GRAPH THEORY AND ADDITIVE COMBINATORICS



NOTES FOR MIT 18.217 (FALL 2019)

LECTURER: YUFEI ZHAO

http://yufeizhao.com/gtac/

About this document

This document contains the course notes for Graph Theory and Additive Combinatorics, a graduate-level course taught by Prof. Yufei Zhao at MIT in Fall 2019.

The notes were written by the students of the class based on the lectures, and edited with the help of the professor.

The notes have not been thoroughly checked for accuracy, especially attributions of results. They are intended to serve as study resources and not as a substitute for professionally prepared publications. We apologize for any inadvertent inaccuracies or misrepresentations.

More information about the course, including problem sets and lecture videos (to appear), can be found on the course website:

http://yufeizhao.com/gtac/

Contents

1
Introduction

1.1 Schur's theorem

In the 1910's, Schur attempted to prove Fermat's Last Theorem by reducing the equation $X^n + Y^n = Z^n$ modulo a prime p. However, he was unsuccessful. It turns out that, for every positive integer n, the equation has nontrivial solutions mod p for all sufficiently large primes p, which Schur established by proving the following classic result.

Schur (1916)

Theorem 1.1 (Schur's theorem)**.** *If the positive integers are colored with finitely many colors, then there is always a monochromatic solution to $x + y = z$ (i.e., x, y, z all have the same color).*

We will prove Schur's theorem shortly. But first, let us show how to deduce the existence of solutions to $X^n + Y^n \equiv Z^n \pmod{p}$ using Schur's theorem.

Schur's theorem is stated above in its "infinitary" (or qualitative) form. It is equivalent to a "finitary" (or quantitative) formulation below.

We write $[N] := \{1, 2, \dots, N\}$.

Theorem 1.2 (Schur's theorem, finitary version)**.** *For every positive integer r, there exists a positive integer $N = N(r)$ such that if the elements of $[N]$ are colored with r colors, then there is a monochromatic solution to $x + y = z$ with $x, y, z \in [N]$.*

With the finitary version, we can also ask quantitative questions such as how big does $N(r)$ have to be as a function of r. For most questions of this type, we do not know the answer, even approximately.

Let us show that the two formulations, Theorem 1.1 and Theorem 1.2, are equivalent. It is clear that the finitary version of Schur's theorem implies the infinitary version. To see that the infinitary version implies the finitary version, fix r, and suppose that for every

N there is some coloring $\phi_N\colon [N] \to [r]$ that avoids monochromatic solutions to $x + y = z$. We can take an infinite subsequence of (ϕ_N) such that, for every $k \in \mathbb{N}$, the value of $\phi_N(k)$ stabilizes as N increases along this subsequence. Then the ϕ_N's, along this subsequence, converges pointwise to some coloring $\phi\colon \mathbb{N} \to [r]$ avoiding monochromatic solutions to $x + y = z$, but this contradicts the infinitary statement.

Let us now deduce Schur's claim about $X^n + Y^n \equiv Z^n \pmod p$.

Theorem 1.3. *Let n be a positive integer. For all sufficiently large primes p, there are $X, Y, Z \in \{1, \dots, p-1\}$ such that $X^n + Y^n \equiv Z^n \pmod p$.*

Schur (1916)

Proof of Theorem 1.3 assuming Schur's theorem (Theorem 1.2). We write $(\mathbb{Z}/p\mathbb{Z})^\times$ for the group of nonzero residues mod p under multiplication. Let H be the subgroup of n-th powers in $(\mathbb{Z}/p\mathbb{Z})^\times$. The index of H in $(\mathbb{Z}/p\mathbb{Z})^\times$ is at most n. So the cosets of H partition $\{1, 2, \dots, p-1\}$ into at most n sets. By the finitary statement of Schur's theorem (Theorem 1.2), for p large enough, there is a solution to

$$x + y = z \quad \text{in } \mathbb{Z}$$

in one of the cosets of H, say aH for some $a \in (\mathbb{Z}/p\mathbb{Z})^\times$. Since H consists of n-th powers, we have $x = aX^n$, $y = aY^n$, and $z = aZ^n$ for some $X, Y, Z \in (\mathbb{Z}/p\mathbb{Z})^\times$. Thus

$$aX^n + aY^n \equiv aZ^n \pmod p.$$

Hence

$$X^n + Y^n \equiv Z^n \pmod p$$

as desired. □

Now let us prove Theorem 1.2 by deducing it from a similar sounding result about coloring the edges of a complete graph. The next result is a special case of Ramsey's theorem.

Theorem 1.4. *For every positive integer r, there is some integer $N = N(r)$ such that if the edges of K_N, the complete graph on N vertices, are colored with r colors, then there is always a monochromatic triangle.*

Ramsey (1929)

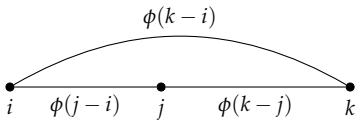
FRANK RAMSEY (1903–1930) had made major contributions to mathematical logic, philosophy, and economics, before his untimely death at age 26 after suffering from chronic liver problems.

Proof. We use induction on r. Clearly $N(1) = 3$ works for $r = 1$. Let $r \geq 2$ and suppose that the claim holds for $r - 1$ colors with $N = N'$. We will prove that taking $N = r(N' - 1) + 2$ works for r colors..

Suppose we color the edges of a complete graph on $r(N' - 1) + 2$ vertices using r colors. Pick an arbitrary vertex v. Of the $r(N' - 1) + 1$ edges incident to v, by the pigeonhole principle, at least N' edges incident to v have the same color, say red. Let V_0 be the vertices joined to v by a red edge. If there is a red edge inside V_0, we obtain a red

triangle. Otherwise, there are at most $r-1$ colors appearing among $|V_0| \geq N'$ vertices, and we have a monochromatic triangle by induction. □

We are now ready to prove Schur's theorem by setting up a graph whose triangles correspond to solutions to $x+y=z$, thereby allowing us to "transfer" the above result to the integers.



Proof of Schur's theorem (Theorem 1.2). Let $\phi\colon [N] \to [r]$ be a coloring. Color the edges of a complete graph with vertices $\{1, \dots, N+1\}$ by giving the edge $\{i,j\}$ with $i<j$ the color $\phi(j-i)$. By Theorem 1.4, if N is large enough, then there is a monochromatic triangle, say on vertices $i<j<k$. So $\phi(j-i) = \phi(k-j) = \phi(k-i)$. Take $x=j-i$, $y=k-j$, and $z=k-i$. Then $\phi(x)=\phi(y)=\phi(z)$ and $x+y=z$, as desired. □

Notice how we solved a number theory problem by moving over to a graph theoretic setup. The Ramsey theorem argument in Theorem 1.4 is difficult to do directly inside the integers. Thus we gained greater flexibility by considering graphs. Later on we will see other more sophisticated examples of this idea, where taking a number theoretic problem to the land of graph theory gives us a new perspective.

1.2 *Highlights from additive combinatorics*

Schur's theorem above is one of the earliest examples of an area now known as ***additive combinatorics***, which is a term coined by Terry Tao in the early 2000's to describe a rapidly growing body of mathematics motivated by simple-to-state questions about addition and multiplication of integers. The problems and methods in additive combinatorics are deep and far-reaching, connecting many different areas of mathematics such as graph theory, harmonic analysis, ergodic theory, discrete geometry, and model theory. The rest of this section highlights some important developments in additive combinatorics in the past century.

Green (2009)

In the 1920's, van der Waerden proved the following result about monochromatic arithmetic progressions in the integers.

Theorem 1.5 (van der Waerden's theorem)**.** *If the integers are colored with finitely many colors, then one of the color classes must contain arbitrarily long arithmetic progressions.*

B. L. van der Waerden, Beweis einer Baudetschen Vermutung. *Nieuw Arch. Wisk.* **15**, 212–216, 1927.

Remark 1.6. Having arbitrarily long arithmetic progressions is very different from having infinitely long arithmetic progressions. As an exercise, show that one can color the integers using just two colors so

that there are no infinitely long monochromatic arithmetic progressions.

In the 1930's, Erdős and Turán conjectured a stronger statement, that any subset of the integers with positive density contains arbitrarily long arithmetic progressions. To be precise, we say that $A \subseteq \mathbb{Z}$ has ***positive upper density*** if

$$\limsup_{N\to\infty} \frac{|A \cap \{-N, \dots, N\}|}{2N+1} > 0.$$

(There are several variations of this definition—the exact formulation is not crucial.)

Erdős and Turán (1936)

In the 1950's, Roth proved the conjecture for 3-term arithmetic progression using Fourier analytic methods. In the 1970's, Szemerédi fully settled the conjecture using combinatorial techniques. These are landmark theorems in the field. Much of what we will discuss are motivated by these results and the developments around them.

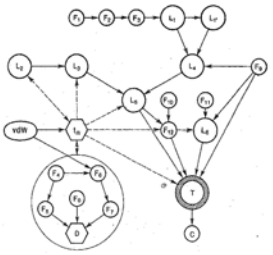
ENDRE SZEMERÉDI (1940–) received the prestigious Abel Prize in 2012 "for his fundamental contributions to discrete mathematics and theoretical computer science, and in recognition of the profound and lasting impact of these contributions on additive number theory and ergodic theory."

Theorem 1.7 (Roth's theorem)**.** *Every subset of the integers with positive upper density contains a 3-term arithmetic progression.*

Roth (1953)

Theorem 1.8 (Szemerédi's theorem)**.** *Every subset of the integers with positive upper density contains arbitrarily long arithmetic progressions.*

Szemerédi (1975)



Szemerédi's proof was a combinatorial tour de force. This figures is taken from the introduction of his paper showing the logical dependencies of his argument.

Szemerédi's theorem is deep and intricate. This important work led to many subsequent developments in additive combinatorics. Several different proofs of Szemerédi's theorem have since been discovered, and some of them have blossomed into rich areas of mathematical research. Here are some the most influential modern proofs of Szemerédi's theorem (in historical order):

- The ergodic theoretic approach (Furstenberg)
- Higher-order Fourier analysis (Gowers)
- Hypergraph regularity lemma (Rödl et al./Gowers)

Furstenberg (1977)

Gowers (2001)

Rödl et al. (2005)

Gowers (2007)

Another modern proof of Szemerédi's theorem results from the *density Hales–Jewett theorem*, which was originally proved by Furstenberg and Katznelson using ergodic theory, and subsequently a new combinatorial proof was found in the first successful Polymath Project, an online collaborative project initiated by Gowers.

Furstenberg and Katznelson (1991)

Polymath (2012)

All subsequent Polymath Project papers are written under the pseudonym D. H. J. Polymath, whose initials stand for "density Hales–Jewett."

The relationships between these disparate approaches are not yet completely understood, and there are many open problems, especially regarding quantitative bounds. A unifying theme underlying all known approaches to Szemerédi's theorem is the ***dichotomy between structure and pseudorandomness***. We will later see different

Tao (2007)

facets of this dichotomy both in the context of graph theory as well as in number theory.

Here are a few other important subsequent developments to Szemerédi's theorem.

Instead of working over subsets of integers, let us consider subsets of a higher dimensional lattice $\mathbb{Z}^d$. We say that $A \subset \mathbb{Z}^d$ has positive upper density if

$$\limsup_{N\to\infty} \frac{|A \cap [-N, N]^d|}{(2N+1)^d} > 0$$

(as before, other similar definitions are possible). We say that A ***contains arbitrary constellations*** if for every finite set $F \subset \mathbb{Z}^d$, there is some $a \in \mathbb{Z}^d$ and $t \in \mathbb{Z}_{>0}$ such that $a + t \cdot F = \{a + tx : x \in F\}$ is contained in A. In other words, A contains every finite pattern, each consisting of some finite subset of the integer grid allowing dilation and translation. The following multidimensional generalization of Szemerédi's theorem was proved by Furstenberg and Katznelson initially using ergodic theory, though a combinatorial proof was later discovered as a consequence of the hypergraph regularity method mentioned earlier.

Theorem 1.9 (Multidimensional Szemerédi theorem)**.** *Every subset of $\mathbb{Z}^d$ of positive upper density contains arbitrary constellations.*

Furstenberg and Katznelson (1978)

For example, the theorem implies that every subset of $\mathbb{Z}^d$ of positive upper density contains a 10×10 set of points that form an axis-aligned square grid.

There is also a polynomial extension of Szemerédi's theorem. Let us first state a special case, originally conjectured by Lovász and proved independently by Furstenberg and Sárkőzy.

Theorem 1.10. *Any subset of the integers with positive upper density contains two numbers differing by a square.*

Furstenberg (1977)
Sárközy (1978)

In other words, the set always contains $\{x, x + y^2\}$ for some $x \in \mathbb{Z}$ and $y \in \mathbb{Z}_{>0}$. What about other polynomial patterns? The following polynomial generalization was proved by Bergelson and Leibman.

Theorem 1.11 (Polynomial Szemerédi theorem)**.** *Suppose $A \subset \mathbb{Z}$ has positive upper density. If $P_1, \dots, P_k \in \mathbb{Z}[X]$ are polynomials with $P_1(0) = \cdots = P_k(0) = 0$, then there exist $x \in \mathbb{Z}$ and $y \in \mathbb{Z}_{>0}$ such that $x + P_1(y), \dots, x + P_k(y) \in A$.*

Bergelson and Leibman (1996)

We leave it as an exercise to formulate a common extension of the above two theorems (i.e., a multidimensional polynomial Szemerédi theorem). Such a theorem was also proved by Bergelson and Leibman.

We will not cover the proof of Theorems 1.9 and 1.11. In fact, currently the only known general proof of the polynomial Szemerédi theorem uses ergodic theory, though for special cases there are some recent exciting developments. Peluse (2019+)

Building on Szemerédi's theorem as well as other important developments in number theory, Green and Tao proved their famous theorem that settled an old folklore conjecture about prime numbers. Their theorem is considered one of the most celebrated mathematical results this century.

Theorem 1.12 (Green–Tao theorem)**.** *The primes contain arbitrarily long arithmetic progressions.* Green and Tao (2008)

We will discuss many central ideas behind the proof of the Green–Tao theorem. See the reference on the right for a modern exposition of the Green–Tao theorem emphasizing the graph theoretic perspective, and incorporating some simplifications of the proof that have been found since the original work. Conlon, Fox, and Zhao (2014)

1.3 What's next?

One of our goals is to understand two different proofs of Roth's theorem, which can be rephrased as:

Theorem 1.13 (Roth's theorem)**.** *Every subset of $[N]$ that does not contain 3-term arithmetic progressions has size $o(N)$.*

Roth originally proved his result using Fourier analytic techniques, which we will see in the second half of this book. In the 1970's, leading up to Szemerédi's proof of his landmark result, Szemerédi developed an important tool known as the ***graph regularity lemma***. Szemerédi (1978)

Ruzsa and Szemerédi used the graph regularity lemma to give a new graph theoretic proof of Roth's theorem. One of our first goals is to understand this graph theoretic proof. Ruzsa and Szemerédi (1978)

As in the proof of Schur's theorem, we will formulate a graph theoretic problem whose solution implies Roth's theorem. This topic fits nicely in an area of combinatorics called ***extremal graph theory***. A starting point (historically and also pedagogically) in extremal graph theory is the following question:

Question 1.14. What is the maximum number of edges in a triangle-free graph on n vertices?

This question is relatively easy, and it was answered by Mantel in the early 1900's (and subsequently rediscovered and generalized by Turán). It will be the first result that we shall prove next. However, even though it sounds similar to Roth's theorem, it cannot be used to

deduce Roth's theorem. Later on, we will construct a graph that corresponds to Roth's theorem, and it turns out that the right question to ask is:

Question 1.15. What is the maximum number of edges in an n-vertex graph where every edge is contained in a unique triangle?

This innocent looking question turns out to be incredible mysterious. We are still far from knowing the truth. We will later prove, using Szemerédi's regularity lemma, that any such graph must have $o(n^2)$ edges, and we will then deduce Roth's theorem from this graph theoretic claim.

Part I

Graph theory

2
Forbidding subgraphs

2.1 *Mantel's theorem: forbidding a triangle*

We begin our discussion of extremal graph theory with the following basic question.

Question 2.1. What is the maximum number of edges in an n-vertex graph that does not contain a triangle?

Bipartite graphs are always triangle-free. A complete bipartite graph, where the vertex set is split equally into two parts (or differing by one vertex, in case n is odd), has $\lfloor n^2/4 \rfloor$ edges. Mantel's theorem states that we cannot obtain a better bound:

Theorem 2.2 (Mantel)**.** *Every triangle-free graph on n vertices has at most $\lfloor n^2/4 \rfloor$ edges.*

W. Mantel, "Problem 28 (Solution by H. Gouwentak, W. Mantel, J. Teixeira de Mattes, F. Schuh and W. A. Wythoff). *Wiskundige Opgaven* **10**, 60 —61, 1907.

We will give two proofs of Theorem 2.2.

Proof 1. Let $G = (V, E)$ a triangle-free graph with n vertices and m edges. Observe that for distinct $x, y \in V$ such that $xy \in E$, x and y must not share neighbors by triangle-freeness.

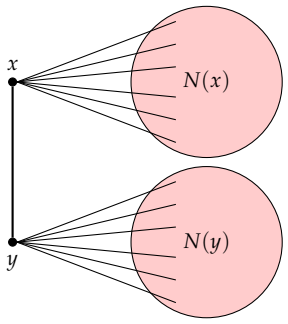
Therefore, $d(x) + d(y) \leq n$, which implies that

$$\sum_{x \in V} d(x)^2 = \sum_{xy \in E} (d(x) + d(y)) \leq mn.$$

On the other hand, by the handshake lemma, $\sum_{x \in V} d(x) = 2m$. Now by the Cauchy–Schwarz inequality and the equation above,

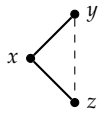
$$4m^2 = \left(\sum_{x \in V} d(x) \right)^2 \leq n \left(\sum_{x \in V} d(x)^2 \right) \leq mn^2;$$

hence $m \leq n^2/4$. Since m is an integer, this gives $m \leq \lfloor n^2/4 \rfloor$. □



Adjacent vertices have disjoint neighborhoods in a triangle-free graph.

Proof 2. Let $G = (V, E)$ be as before. Since G is triangle-free, the neighborhood $N(x)$ of every vertex $x \in V$ is an independent set.



An edge within $N(x)$ creates a triangle

Let $A \subseteq V$ be a maximum independent set. Then $d(x) \leq |A|$ for all $x \in V$. Let $B = V \setminus A$. Since A contains no edges, every edge of G intersects B. Therefore,

$$e(G) \leq \sum_{x \in B} d(x) \leq |A||B| \underset{\text{AM-GM}}{\leq} \left\lfloor \left(\frac{|A| + |B|}{2} \right)^2 \right\rfloor = \left\lfloor \frac{n^2}{4} \right\rfloor.$$

□

Remark 2.3. For equality to occur in Mantel's theorem, in the above proof, we must have

- $e(G) = \sum_{x \in B} d(x)$, which implies that no edges are strictly in B.
- $\sum_{x \in B} d(x) = |A||B|$, which implies that every vertex in B is complete in A.
- The equality case in AM-GM must hold (or almost hold, when n is odd), hence $||A| - |B|| \leq 1$.

Thus a triangle-free graph on n vertices has exactly $\lfloor n^2/4 \rfloor$ edges if and only if it is the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

2.2 *Turán's theorem: forbidding a clique*

Motivated by Theorem 2.2, we turn to the following more general question.

Question 2.4. What is the maximum number of edges in a K_{r+1}-free graph on n vertices?

Extending the bipartite construction earlier, we see that an r-partite graph does not contain any copy of K_{r+1}.

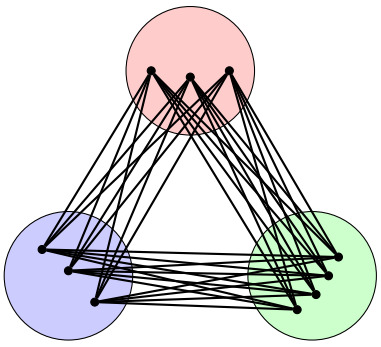
Definition 2.5. The ***Turán graph*** $T_{n,r}$ is defined to be the complete, n-vertex, r-partite graph, with part sizes either $\lfloor \frac{n}{r} \rfloor$ or $\lceil \frac{n}{r} \rceil$.

In this section, we prove that $T_{n,r}$ does, in fact, maximize the number of edges in a K_r-free graph:

Theorem 2.6 (Turán). *If G is an n-vertex K_{r+1}-free graph, then $e(G) \leq e(T_{n,r})$.*

When $r = 2$, this is simply Theorem 2.2.

We now give three proofs of Theorem 2.6. The first two are in the same spirit as the proofs of Theorem 2.2.



The Turán graph $T_{10,3}$

P. Turán, On an extremal problem in graph theory. *Math. Fiz. Lapok* **48**, 436—452, 1941.

Proof 1. Fix r. We proceed by induction on n. Observe that the statement is trivial if $n \leq r$, as K_n is K_{r+1}-free. Now, assume that $n > r$ and that Turán's theorem holds for all graphs on fewer than n vertices. Let G be an n-vertex, K_{r+1}-free graph with the maximum possible number of edges. Note that G must contain K_r as a subgraph, or else we could add an edge in G and still be K_{r+1}-free. Let A be the vertex set of an r-clique in G, and let $B := V \backslash A$. Since G is K_{r+1}-free, every $v \in B$ has at most $r-1$ neighbors in A. Therefore

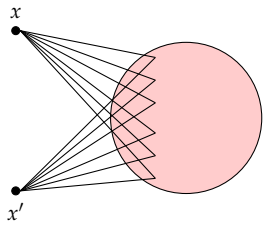
$$\begin{aligned} e(G) &\leq \binom{r}{2} + (r-1)|B| + e(B) \\ &\leq \binom{r}{2} + (r-1)(n-r) + e(T_{n-r,r}) \\ &= e(T_{n,r}). \end{aligned}$$

The first inequality follows from counting the edges in A, B, and everything in between. The second inequality follows from the inductive hypothesis. The last equality follows by noting removing one vertex from each of the r parts in $T_{n,r}$ would remove a total of $\binom{r}{2} + (r-1)(n-r)$ edges. □

Proof 2 (Zykov symmetrization). As before, let G be an n-vertex, K_{r+1}-free graph with the maximum possible number of edges.

We claim that the non-edges of G form an equivalence relation; that is, if $xy, yz \notin E$, then $xz \notin E$. Symmetry and reflexivity are easy to check. To check transitivity, Assume for purpose of contradiction that there exists $x, y, z \in V$ for which $xy, yz \notin E$ but $xz \in E$.

If $d(y) < d(x)$, we may replace y with a "clone" of x. That is, we delete y and add a new vertex x' whose neighbors are precisely the as the neighbors of x (and no edge between x and x'). (See figure on the right.)



x and its clone x'

Then, the resulting graph G' is also K_{r+1}-free since x was not in any K_{r+1}. On the other hand, G' has more edges than G, contradicting maximality.

Therefore we have that $d(y) \geq d(x)$ for all $xy \notin E$. Similarly, $d(y) \geq d(z)$. Now, replace both x and z by "clones" of y. The new graph G' is K_{r+1}-free since y was not in any K_{r+1}, and

$$e(G') = e(G) - (d(x) + d(z) - 1) + 2d(y) > e(G),$$

contradicting maximality of $e(G)$. Therefore such a triple (x, y, z) cannot exist in G, and transitivity holds.

The equivalence relation shows that the complement of G is a union of cliques. Therefore G is a complete multipartite graph with at most r parts. One checks that increasing the number of parts increases the number of edges in G. Similarly, one checks that if the

number of vertices in two parts differ by more than 1, moving one vertex from the larger part to the smaller part increases the number of edges in G. It follows that the graph that achieves the maximum number of edges is $T_{n,r}$. □

Our third and final proof uses a technique called the ***probabilistic method***. In this method, one introduces randomness to a deterministic problem in a clever way to obtain deterministic results.

Proof 3. Let $G = (V, E)$ be an n-vertex, K_{r+1}-free graph. Consider a uniform random ordering σ of the vertices. Let

$$X = \{v \in V : v \text{ is adjacent to all earlier vertices in } \sigma\}.$$

Observe that the set of vertices in X form a clique. Since the permutation was chosen uniformly at random, we have

$$\mathbb{P}(v \in X) = \mathbb{P}(v \text{ appears before all non-neighbors}) = \frac{1}{n - d(v)}.$$

Therefore,

$$r \geq \mathbb{E}|X| = \sum_{v \in V} \mathbb{P}(v \in X) = \sum_{v \in V} \frac{1}{n - d(v)} \overset{\text{convexity}}{\geq} \frac{n}{n - 2m/n}.$$

Rearranging gives $m \leq \left(1 - \frac{1}{r}\right) \frac{n^2}{2}$ (a bound that is already good for most purposes). Note that if n is divisible by r, then the bound immediately gives a proof of Turán's theorem. When n is not divisible by r, one needs to a bit more work and use convexity to argue that the $d(v)$ should be as close as possible. We omit the details. □

2.3 *Hypergraph Turán problem*

The short proofs given in the previous sections make problems in extremal graph theory seem deceptively simple. In reality, many generalizations of what we just discussed remain wide open.

Here we discuss one notorous open problem that is a hypergraph generalization of Mantel/Turán.

An r-uniform hypergraph consists of a vertex set V and an edge set, where every edge is now an r-element subset of V. Graphs correspond to $r = 2$.

Question 2.7. What is the maximum number of triples in an n vertex 3-uniform hypergraph without a tetrahedron?

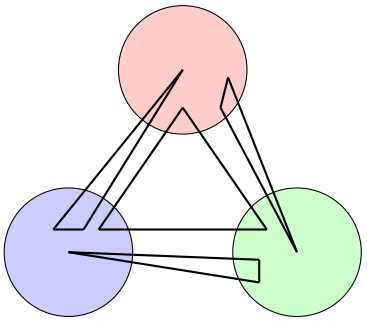
Turán proposed the following construction, which is conjectured to be optimal.

Example 2.8 (Turán)**.** Let V be a set of n vertices. Partition V into 3 (roughly) equal sets V_1, V_2, V_3. Add a triple $\{x, y, z\}$ to $e(G)$ if it satisfies one of the four following conditions:

- x, y, z are in different partitions
- $x, y \in V_1$ and $z \in V_2$
- $x, y \in V_2$ and $z \in V_3$
- $x, y \in V_3$ and $z \in V_1$

where we consider x, y, z up to permutation (See Example 2.8). One checks that the 3-uniform hypergraph constructed is tetrahedron-free, and that it has edge density $5/9$.

On the other hand, the best known upper bound is approximately 0.562 , obtained recently using the technique of flag algebras.



Turán's construction of a tetrahedron-free 3-uniform hypergraph

Keevash (2011)

Baber and Talbot (2011)

Razborov (2010)

2.4 *Erdős–Stone–Simonovits theorem (statement): forbidding a general subgraph*

One might also wonder what happens if K_{r+1} in Theorem 2.6 were replaced with an arbitrary graph H:

Question 2.9. Fix some graph H. If G is an n vertex graph in which H does not appear as a subgraph, what is the maximum possible number of edges in G?

Notice that we only require H to be a ***subgraph***, not necessarily an ***induced subgraph***. An induced subgraph H' of G must contain all edges present between the vertices of H', while there is no such restriction for arbitrary subgraphs.

Definition 2.10. For a graph H and $n \in \mathbb{N}$, define $\mathrm{ex}(n, H)$ to be the maximum number of edges in an n-vertex H-free graph.

For example, Theorem 2.6 tells us that for any given r,

$$\mathrm{ex}(n, K_{r+1}) = e(T_{n,r}) = \left(1 - \frac{1}{r} + o(1)\right)\binom{n}{2}$$

where $o(1)$ represents some quantity that goes to zero as $n \to \infty$.

At a first glance, one might not expect a clean answer to Question 2.9. Indeed, the solution would seem to depend on various characteristics of H (for example, its diameter or maximum degree). Surprisingly, it turns out that a single parameter, the chromatic number of H, governs the growth of $\mathrm{ex}(n, H)$.

Definition 2.11. The ***chromatic number*** of a graph G, denoted $\chi(G)$, is the minimal number of colors needed to color the vertices of G such that no two adjacent vertices have the same color.

Example 2.12. $\chi(K_{r+1}) = r + 1$ and $\chi(T_{n,r}) = r$.

Observe that if $H \subseteq G$, then $\chi(H) \leq \chi(G)$. Indeed, any proper coloring of G restricts to a proper coloring of H. From this, we gather that if $\chi(H) = r+1$, then $T_{n,r}$ is H-free. Therefore,

$$\mathrm{ex}(n,H) \geq e(T_{n,r}) = \left(1 - \frac{1}{r} + o(1)\right)\binom{n}{2}.$$

Is this the best we can do? The answer turns out to be affirmative.

Theorem 2.13 (Erdős–Stone–Simonovits)**.** *For all graphs H, we have*

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n,H)}{\binom{n}{2}} = 1 - \frac{1}{\chi(H)-1}.$$

Erdős and Stone (1946)
Erdős and Simonovits (1966)

We'll skip the proof for now.

Remark 2.14. Later in the book we will show how to deduce Theorem 2.13 from Theorem 2.6 using the ***Szemerédi regularity lemma***.

Example 2.15. When $H = K_3$, Theorem 2.13 tells us that

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n,H)}{\binom{n}{2}} = \frac{1}{2},$$

in agreement with Theorem 2.6.

When $H = K_4$, we get

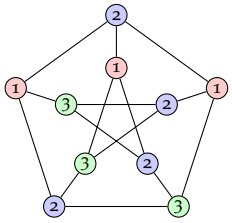
$$\lim_{n\to\infty} \frac{\mathrm{ex}(n,H)}{\binom{n}{2}} = \frac{2}{3},$$

also in agreement with Theorem 2.6.

When H is the Peterson graph, Theorem 2.13 tells us that

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n,H)}{\binom{n}{2}} = \frac{1}{2},$$

which is the same answer as for $H = K_3$! This is surprising since the Peterson graph seems much more complicated than the triangle.

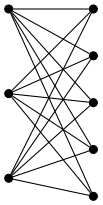


The Peterson graph with a proper 3-coloring.

2.5 *Kővári–Sós–Turán theorem: forbidding a complete bipartite graph*

The Erdős–Stone–Simonovits Theorem (Theorem 2.13) gives a first-order approximation of $\mathrm{ex}(n,H)$ when $\chi(H) > 2$. Unfortunately, Theorem 2.13 does not tell us the whole story. When $\chi(H) = 2$, i.e. H is bipartite, the theorem implies that $\mathrm{ex}(n,H) = o(n^2)$, which compels us to ask if we may obtain more precise bounds. For example, if we write $\mathrm{ex}(n,H)$ as a function of n, what its growth with respect to n? This is an open problem for most bipartite graphs (for example, $K_{4,4}$) and the focus of the remainder of the chapter.

Let $K_{s,t}$ be the complete bipartite graph where the two parts of the bipartite graph have s and t vertices respectively. In this section, we consider $\mathrm{ex}(n,K_{s,t})$, and seek to answer the following main question:


An example of a complete bipartite graph $K_{3,5}$.

Question 2.16 (Zarankiewicz problem)**.** For some $r, s \geq 1$, what is the maximum number of edges in an n-vertex graph which does not contain $K_{s,t}$ as a subgraph.

Every bipartite graph H is a subgraph of some complete bipartite graph $K_{s,t}$. If $H \subseteq K_{s,t}$, then $\mathrm{ex}(n, H) \leq \mathrm{ex}(n, K_{s,t})$. Therefore, by understanding the upper bound on the extremal number of complete bipartite graphs, we obtain an upper bound on the extremal number of general bipartite graphs as well. Later, we will give improved bounds for several specific biparite graphs.

Kővári, Sós and Turán gave an upper bound on $K_{s,t}$:

Theorem 2.17 (Kővári–Sós–Turán)**.** *For every integers $1 \leq s \leq t$, there exists some constant C, such that*

Kővári, Sós, and Turán (1954)

$$\mathrm{ex}(n, K_{s,t}) \leq Cn^{2-\frac{1}{s}}.$$

There is an easy way to remember the name of this theorem: "KST", the initials of the authors, is also the letters for the complete bipartite graph $K_{s,t}$.

Proof. Let G be a $K_{s,t}$-free n-vertex graph with m edges.

First, we repeatedly remove all vertices $v \in V(G)$ where $d(v) < s-1$. Since we only remove at most $(s-2)n$ edges this way, it suffices to prove the theorem assuming that all vertices have degree at least $s-1$.

We denote the number of copies of $K_{s,1}$ in G as $\#K_{s,1}$. The proof establishes an upper bound and a lower bound on $\#K_{s,1}$, and then gets a bound on m by combining the upper bound and the lower bound.

Since $K_{s,1}$ is a complete bipartite graph, we can call the side with s vertices the 'left side', and the side with 1 vertices the 'right side'.

On the one hand, we can count $\#K_{s,1}$ by enumerating the 'left side'. For any subset of s vertices, the number of $K_{s,1}$ where these s vertices form the 'left side' is exactly the number of common neighbors of these s vertices. Since G is $K_{s,t}$-free, the number of common neighbors of any subset of s vertices is at most $t-1$. Thus, we establish that $\#K_{s,1} \leq \binom{n}{s}(t-1)$.

On the other hand, for each vertex $v \in V(G)$, the number of copies of $K_{s,1}$ where v is the 'right side' is exactly $\binom{d(v)}{s}$. Therefore,

Here we regard $\binom{x}{s}$ as a degree s polynomial in x, so it makes sense for x to be non-integers. The function $\binom{x}{s}$ is convex when $x \geq s-1$.

$$\#K_{s,1} = \sum_{v \in V(G)} \binom{d(v)}{s} \geq n \binom{\frac{1}{n}\sum_{v \in V(G)} d(v)}{s} = n\binom{2m/n}{s},$$

where the inequality step uses the convextiy of $x \mapsto \binom{x}{s}$.

Combining the upper bound and lower bound of $\#K_{s,1}$, we obtain that $n\binom{2m/n}{s} \leq \binom{n}{s}(t-1)$. For constant s, we can use $\binom{x}{s} = (1+o(1))\frac{n^s}{s!}$ to get $n\left(\frac{2m}{n}\right)^s \leq (1+o(1))n^s(t-1)$. The above inequality simplifies to

$$m \leq \left(\frac{1}{2} + o(1)\right)(t-1)^{1/s} n^{2-\frac{1}{s}}.$$

□

Let us discuss a geometric application of Theorem 2.17.

Question 2.18 (Unit distance problem)**.** What is the maximum number of unit distances formed by n points in $\mathbb{R}^2$?

Erdős (1946)

For small values of n, we precisely know the answer to the unit distance problem. The best configurations are listed in Figure 2.1.

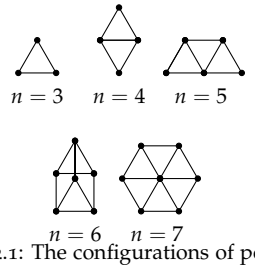
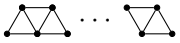


Figure 2.1: The configurations of points for small values of n with maximum number of unit distances. The edges between vertices mean that the distance is 1. These constructions are unique up to isomorphism except when $n = 6$.

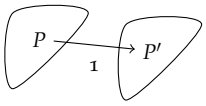
It is possible to generalize some of these constructions to arbitrary n.

- A line graph has $(n-1)$ unit distances.

- A chain of triangles has $(2n-3)$ unit distances for $n \geq 3$.



- There is also a recursive construction. Given a configuration P with $n/2$ points that have $f(n/2)$ unit distances, we can copy P and translate it by an arbitrary unit vector to get P'. The configuration $P \cup P'$ have at least $2f(n/2) + n/2$ unit distances. We can solve the recursion to get $f(n) = \Omega(n \log n)$.



The current best lower bound on the maximum number of unit distances is given by Erdős.

Erdős (1946)

Proposition 2.19. *There exists a set of n points in $\mathbb{R}^2$ that have at least $n^{1+c/\log\log n}$ unit distances for some constant c.*

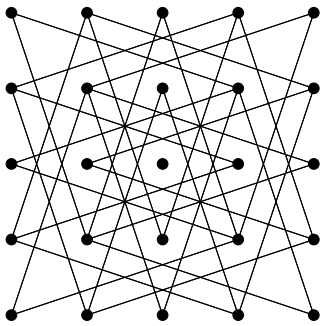


Figure 2.2: An example grid graph where $n = 25$ and $r = 10$.

Proof sketch. Consider a square grid with $\lfloor\sqrt{n}\rfloor \times \lfloor\sqrt{n}\rfloor$ vertices. We can scale the graph arbitrarily so that $\sqrt{r}$ becomes the unit distance for some integer r. We can pick r so that r can be represented as a sum of two squares in many different ways. One candidate of such r is a product of many primes that are congruent to 1 module 4. We can use some number-theoretical theorems to analyze the best r, and get the $n^{1+c/\log\log n}$ bound. □

Theorem 2.17 can be used to prove an upper bound on the number of unit distances.

Theorem 2.20. *Every set of n points in $\mathbb{R}^2$ has at most $O(n^{3/2})$ unit distances.*

Proof. Given any set of points $S \subset \mathbb{R}^2$, we can create the *unit distance graph* G as follows:

- The vertex set of G is S,
- For any point p, q where $d(p, q) = 1$, we add an edge between p and q.

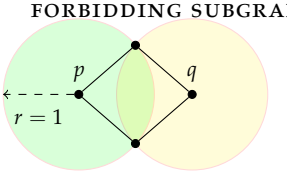


Figure 2.3: Two vertices p, q can have at most two common neighbors in the unit distance graph.

The graph G is $K_{2,3}$-free since for every pair of points p, q, there are at most 2 points that have unit distances to both of them. By applying Theorem 2.17, we obtain that $e(G) = O(n^{3/2})$. □

Remark 2.21. The best known upper bound on the number of unit distances is $O(n^{4/3})$. The proof is a nice application of the crossing number inequality which will be introduced later in this book.

Spencer, Szemerédi and Trotter (1984)

Here is another problem that is strongly related to the unit distance problem:

Question 2.22 (Distinct distance problem)**.** What is the minimum number of distinct distances formed by n points in $\mathbb{R}^2$?

Example 2.23. Consider n points on the x-axis where the i-th point has coordinate $(i, 0)$. The number of distinct distances for these points is $n - 1$.

The current best construction for minimum number of distinct distances is also the grid graph. Consider a square grid with $\lfloor\sqrt{n}\rfloor \times \lfloor\sqrt{n}\rfloor$ vertices. Possible distances between two vertices are numbers that can be expressed as a sum of the squares of two numbers that are at most $\lfloor\sqrt{n}\rfloor$. Using number-theoretical methods, we can obtain that the number of such distances: $\Theta(n/\sqrt{\log n})$.

The maximum number of unit distances is also the maximum number that each distance can occur. Therefore, we have the following relationship between distinct distances and unit distances:

$$\#\text{distinct distances} \geq \frac{\binom{n}{2}}{\max \#\text{unit distances}}.$$

If we apply Theorem 2.20 to the above inequality, we immediately get an $\Omega(n^{0.5})$ lower bound for the number of distinct distances. Many mathematicians successively improved the exponent in this lower bound over the span of seven decades. Recently, Guth and Katz gave the following celebrated theorem, which almost matches the upper bound (only off by an $O(\sqrt{\log n})$) factor).

Theorem 2.24 (Guth–Katz)**.** *Every set of n points in $\mathbb{R}^2$ has at least $cn/\log n$ distinct distances for some constant c.*

Guth and Katz (2015)

The proof of Theorem 2.24 is quite sophisticated: it uses tools ranging from polynomial method to algebraic geometry. We won't cover it in this book.

2.6 *Lower bounds: randomized constructions*

It is conjectured that the bound proven in Theorem 2.17 is tight. In other words, $\text{ex}(n, K_{s,t}) = \Theta(n^{2-1/s})$. Although this still remains

open for arbitrary $K_{s,t}$, it is already proven for a few small cases, and in cases where t is way larger than s. In this and the next two sections, we will show techniques for constructing H-free graphs. Here are the three main types of constructions that we will cover:

- **Randomized construction.** This method is powerful and general, but introducing randomness means that the constructions are usually *not tight*.
- **Algebraic construction.** This method uses tools in number theory or algebra to assist construction. It gives tighter results, but they are usually 'magical', and only works in a small set of cases.
- **Randomized algebraic construction.** This method is the hybrid of the two methods above and combines the advantages of both.

This section will focus on randomized constructions. We start with a general lower bound for extremal numbers.

Theorem 2.25. *For any graph H with at least 2 edges, there exists a constant $c > 0$, such that for any $n \in \mathbb{N}$, there exists an H-free graph on n vertices with at least $cn^{2-\frac{v(H)-2}{e(H)-1}}$ edges. In other words,*

$$\mathrm{ex}(n, H) \geq cn^{2-\frac{v(H)-2}{e(H)-1}}.$$

Proof. The idea is to use the *alteration method*: we can construct a graph that has few copies of H in it, and delete one edge from each copy to eliminate the occurrences of H.

Consider $G = G(n, p)$ as a random graph with n vertices where each edge appears with probability p (p to be determined). Let $\#H$ be the number of copies of H in G. Then,

$$\mathbb{E}[\#H] = \frac{n(n-1)\cdots(n-v(H)+1)}{|\operatorname{Aut}(H)|} p^{e(H)} \leq p^{e(H)} n^{v(H)},$$

where $\operatorname{Aut}(H)$ is the automorphism group of graph H, and

$$\mathbb{E}[e(G)] = p\binom{n}{2}.$$

Let $p = \frac{1}{2} n^{-\frac{v(H)-2}{e(H)-1}}$, chosen so that

$$\mathbb{E}[\#H] \leq \frac{1}{2}\mathbb{E}[e(G)],$$

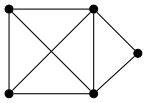
which further implies

$$\mathbb{E}[e(G) - \#H] \geq \frac{1}{2} p\binom{n}{2} \geq \frac{1}{16} n^{2-\frac{v(H)-2}{e(H)-1}}.$$

Thus, there exists a graph G, such that the value of $(e(G) - \#H)$ is at least the expectation. Remove one edge from each copy of H in G, and we get an H-free graph with enough edges. □

The random graph $G(n, p)$ is called the ***Erdős–Rényi random graph***, which appears in many randomized constructions.

Remark 2.26. For example, if H is the following graph



then applying Theorem 2.25 directly gives

$$\mathrm{ex}(n, H) \gtrsim n^{11/7}.$$

However, if we forbid H's subgraph K_4 instead (forbidding a subgraph will automatically forbid the original graph), Theorem 2.25 actually gives us a better bound:

$$\mathrm{ex}(n, H) \geq \mathrm{ex}(n, K_4) \gtrsim n^{8/5}.$$

For a general H, we apply Theorem 2.25 to the subgraph of H with the maximum $(e-1)/(v-2)$ value. For this purpose, define the ***2-density of*** H as

$$m_2(H) := \max_{\substack{H' \subseteq H \\ v(H') \geq 3}} \frac{e(H') - 1}{v(H') - 2}.$$

We have the following corollary.

Corollary 2.27. *For any graph H with at least two edges, there exists constant $c = c_H > 0$ such that*

$$\mathrm{ex}(n, H) \geq cn^{2-1/m_2(H)}.$$

Example 2.28. We present some specific examples of Theorem 2.25. This lower bound, combined with the upper bound from the Kővári–Sós–Turán theorem (Theorem 2.17), gives that for every $2 \leq s \leq t$,

$$n^{2-\frac{s+t-2}{st-1}} \lesssim \mathrm{ex}(n, K_{s,t}) \lesssim n^{2-1/s}.$$

When t is large compared to s, the exponents in the two bounds above are close to each other (but never equal).

When $t = s$, the above bounds specialize to

$$n^{2-\frac{2}{s+1}} \lesssim n^{2-\frac{s+t-2}{st-1}} \lesssim\lesssim n^{2-1/s}.$$

In particular, for $s = 2$, we obtain

$$n^{4/3} \lesssim \mathrm{ex}(n, K_{2,2}) \lesssim n^{3/2}.$$

It turns out what the upper bound is close to tight, as we show next a different, algebraic, construction of a $K_{2,2}$-free graph.

2.7 *Lower bounds: algebraic constructions*

In this section, we use algebraic constructions to find $K_{s,t}$-free graphs, for various values of (s,t), that match the upper bound in the Kővári–Sós–Turán theorem (Theorem 2.17) up to a constant factor.

The simplest example of such an algebraic construction is the following construction of $K_{2,2}$-free graphs with many edges.

Theorem 2.29 (Erdős–Rényi–Sós)**.**

Erdős, Rényi and Sós (1966)

$$\mathrm{ex}(n, K_{2,2}) \geq \left(\frac{1}{2} - o(1)\right) n^{3/2}.$$

Proof. Suppose $n = p^2 - 1$ where p is a prime. Consider the following graph G (called *polarity graph*):

- $V(G) = \mathbb{F}_p^2 \setminus \{(0,0)\}$,
- $E(G) = \{(x,y) \sim (a,b) | ax + by = 1 \text{ in } \mathbb{F}_p\}$.

For any two distinct vertices $(a,b) \neq (a',b') \in V(G)$, there is at most one solution (common neighbour) $(x,y) \in V(G)$ satisfying both $ax + by = 1$ and $a'x + b'y = 1$. Therefore, G is $K_{2,2}$-free.

Moreover, every vertex has degree p or $p-1$, so the total number of edges

$$e(G) = \left(\frac{1}{2} - o(1)\right) p^3 = \left(\frac{1}{2} - o(1)\right) n^{3/2},$$

which concludes our proof.

If n does not have the form $p^2 - 1$ for some prime, then we let p be the largest prime such that $p^2 - 1 \leq n$. Then $p = (1 - o(1))n$ and constructing the same graph G_{p^2-1} with $n - p^2 + 1$ isolated vertices. □

Why is it called a polarity graph? It may be helpful to first think about the partite version of the construction, where one vertex set is the set of points of of a (projective) plane over $\mathbb{F}_p$, and the other vertex set is the set of lines in the same plane, and one has an edge between point p and line ℓ if $p \in \ell$. This graph is C_4-free since no two lines intersect in two distinct points.

The construction in the proof of Theorem 2.29 has one vertex set that identifies points with lines. This duality pairing between points and lines is known in projective geometry a polarity.

Most vertices have degree p because the equation $ax + by = 1$ has exactly p solutions (x,y). Sometimes we have to subtract 1 because one of the solutions might be (a,b) itself, which forms a self-loop.

Here we use that the smallest prime greater than n has size $n + o(n)$. The best result of this form says that there exists a prime in the interval $[n - n^{0.525}, n]$ for every sufficiently large n.

Baker, Harman and Pintz (2001)

A natural question to ask here is whether the construction above can be generalized. The next construction gives us a construction for $K_{3,3}$-free graphs.

Theorem 2.30 (Brown)**.**

Brown (1966)

$$\mathrm{ex}(n, K_{3,3}) \geq \left(\frac{1}{2} - o(1)\right) n^{5/3}$$

It is known that the constant $1/2$ in Theorem 2.30 is the best constant possible.

Proof sketch. Let $n = p^3$ where p is a prime. Consider the following graph G:

- $V(G) = \mathbb{F}_p^3$
- $E(G) = \{(x,y,z) \sim (a,b,c) | (a-x)^2 + (b-y)^2 + (c-z)^2 = u \text{ in } \mathbb{F}_p\}$, where u is some carefully-chosen fixed nonzero element in $\mathbb{F}_p$

One needs to check that it is possible to choose u so that the above graph is $K_{3,3}$. We omit the proof but give some intuition. Had we used points in $\mathbb{R}^3$ instead of $\mathbb{F}_p^3$, the $K_{3,3}$-freeness is equivalent to the statement that three unit spheres have at most two common points. This statement about unit spheres in $\mathbb{R}^3$, and it can be proved rigorously by some algebraic manipulation. One would carry out a similar algebraic manipulation over $\mathbb{F}_p$ to verify that the graph above is $K_{3,3}$-free.

Moreover, each vertex has degree around p^2 since the distribution of $(a-x)^2+(b-y)^2+(c-z)^2$ is almost uniform across $\mathbb{F}_p$ as (x,y,z) varies randomly over $\mathbb{F}_p^3$, and so we expect roughly a $1/p$ fraction of (x,y,z) to have $(a-x)^2+(b-y)^2+(c-z)^2=u$. Again we omit the details. □

Although the case of $K_{2,2}$ and $K_{3,3}$ are fully solved, the corresponding problem for $K_{4,4}$ is a central open problem in extremal graph theory.

Open problem 2.31. What is the order of growth of $\mathrm{ex}(n, K_{4,4})$? Is it $\Theta(n^{7/4})$, matching the upper bound in Theorem 2.17?

We have obtained the Kővári–Sós–Turán bound up to a constant factor for $K_{2,2}$ and $K_{3,3}$. Now we present a construction that matches the Kővári–Sós–Turán bound for $K_{s,t}$ whenever t is sufficiently large compared to s.

Theorem 2.32 (Alon, Kollár, Rónyai, Szabó). *If $t \geq (s-1)!+1$ then*

$$ex(n, K_{s,t}) = \Theta(n^{2-\frac{1}{s}}).$$

Kollár, Rónyai, and Szabó (1996)
Alon, Rónyai, and Szabó (1999)

We begin by proving a weaker version for $t \geq s!+1$. This will be similar in spirit and later we will make an adjustment to achieve the desired bound. Take a prime p and $n = p^s$ with $s \geq 2$. Consider the norm map $N\colon \mathbb{F}_{p^s} \to \mathbb{F}_p$ defined by

$$N(x) = x \cdot x^p \cdot x^{p^2} \cdots x^{p^{s-1}} = x^{\frac{p^s-1}{p-1}}.$$

Notice that we said the image of N lies in $\mathbb{F}_p$ rather than $\mathbb{F}_{p^s}$. We can easily check this is indeed the case as $N(x)^p = N(x)$.

Define the graph $\mathsf{NormGraph}_{p,s} = (V, E)$ with

$$V = \mathbb{F}_{p^s} \text{ and } E = \{\{a,b\} | a \neq b, N(a+b) = 1\}.$$

Proposition 2.33. *In $\mathsf{NormGraph}_{p,s}$ defined as above, letting $n = p^s$ be the number of vertices,*

$$|E| \geq \frac{1}{2} n^{2-\frac{1}{s}}.$$

Proof. Since $\mathbb{F}_{p^s}^\times$ is a cyclic group of order p^s-1 we know that

$$|\{x \in \mathbb{F}_{p^s} | N(x) = 1\}| = \frac{p^s-1}{p-1}.$$

Thus for every vertex x (the minus one accounts for vertices with $N(x+x)=1$)

$$\deg(x) \geq \frac{p^s-1}{p-1} - 1 \geq p^{s-1} = n^{1-\frac{1}{s}}.$$

This gives us the desired lower bound on the number of edges. □

Proposition 2.34. $\mathsf{NormGraph}_{p,s}$ *is* $K_{s,s!+1}$*-free.*

We wish to upper bound the number of common neighbors to a set of s vertices. We quote without proof the following result, which can be proved using algebraic geometry.

Theorem 2.35. *Let* $\mathbb{F}$ *be any field and* $a_{ij}, b_i \in \mathbb{F}$ *such that* $a_{ij} \neq a_{i'j}$ *for all* $i \neq i'$*. Then the system of equations* Kollár, Rónyai, and Szabó (1996)

$$\begin{aligned}(x_1 - a_{11})(x_2 - a_{12})\cdots(x_s - a_{1s}) &= b_1\\ (x_1 - a_{21})(x_2 - a_{22})\cdots(x_s - a_{2s}) &= b_2\\ &\vdots\\ (x_1 - a_{s1})(x_2 - a_{s2})\cdots(x_s - a_{ss}) &= b_s\end{aligned}$$

has at most $s!$ *solutions in* $\mathbb{F}^s$*.*

Remark 2.36. Consider the special case when all the b_i are 0. In this case, since the a_{ij} are distinct for a fixed j, we are picking an i_j for which $x_j = a_{i_j j}$. Since all the i_j are distinct, this is equivalent to picking a permutation on $[s]$. Therefore there are exactly $s!$ solutions.

We can now prove Proposition 2.34.

Proof of Proposition 2.34. Consider distinct $y_1, y_2, \ldots, y_s \in \mathbb{F}_{p^s}$. We wish to bound the number of common neighbors x. We can use the fact that in a field with characteristic p we have $(x+y)^p = x^p + y^p$ to obtain

$$\begin{aligned}1 = N(x+y_i) &= (x+y_i)(x+y_i)^p \ldots (x+y_i)^{p^{s-1}}\\ &= (x+y_i)(x^p + y_i^p)\ldots(x^{p^{s-1}} + y_i^{p^{s-1}})\end{aligned}$$

for all $1 \leq i \leq s$. By Theorem 2.35 these s equations have at most $s!$ solutions in x. Notice we do in fact satisfy the hypothesis since $y_i^p = y_j^p$ if and only if $y_i = y_j$ in our field. □

Now we introduce the adjustment to achieve the bound $t \geq (s-1)! + 1$ in Theorem 2.32. We define the graph $\mathsf{ProjNormGraph}_{p,s} = (V, E)$ with $V = \mathbb{F}_{p^{s-1}} \times \mathbb{F}_p^\times$ for $s \geq 3$. Here $n = (p-1)p^{s-1}$. Define the edge relation as $(X, x) \sim (Y, y)$ if and only if

$$N(X+Y) = xy.$$

Proposition 2.37. *In* $\mathsf{ProjNormGraph}_{p,s}$ *defined as above, letting* $n = (p-1)p^{s-1}$ *denote the number of vertices,*

$$|E| = \left(\frac{1}{2} - o(1)\right) n^{2-\frac{1}{s}}.$$

Proof. It follows from that every vertex (X, x) has degree $p^{s-1} - 1 = (1 - o(1))n^{1-1/s}$ since its neighbors are $(Y, N(X+Y)/x)$ as Y ranges over elements of $\mathbb{F}_{p^{s-1}}$ over than $-X$. □

Now that we know we have a sufficient amount of edges we just need our graph to be $K_{s,(s-1)!+1}$-free.

Proposition 2.38. $\mathsf{ProjNormGraph}_{p,s}$ *is* $K_{s,(s-1)!+1}$*-free.*

Proof. Once again we fix distinct $(Y_i, y_i) \in V$ for $1 \le i \le s$ and we wish to find all common neighbors (X, x). Then

$$N(X + Y_i) = xy_i.$$

Assume this system has at least one solution. Then if $Y_i = Y_j$ with $i \neq j$ we must have that $y_i = y_j$. Therefore all the Y_i are distinct. For each $i < s$ we can take $N(X + Y_i) = xy_i$ and divide by $N(X + Y_s) = xy_s$ to obtain

$$N\left(\frac{X + Y_i}{X + Y_s}\right) = \frac{y_i}{y_s}.$$

Dividing both sides by $N(Y_i - Y_s)$ we obtain

$$N\left(\frac{1}{X + Y_s} + \frac{1}{Y_i - Y_s}\right) = \frac{y_i}{N(Y_i - Y_s)y_s}$$

for all $1 \le i \le s-1$. Now applying Theorem 2.35 there are at most $(s-1)!$ choices for X, which also determines $x = N(X + Y_1)/y_1$. Thus there are at most $(s-1)!$ common neighbors. □

Now we are ready to prove Theorem 2.32.

Proof of Theorem 2.32. By Proposition 2.37 and Proposition 2.38 we know that $\mathsf{ProjNormGraph}_{p,s}$ is $K_{s,(s-1)!+1}$-free and therefore $K_{s,t}$-free and has $\left(\frac{1}{2} - o(1)\right) n^{2-\frac{1}{s}}$ edges as desired. □

2.8 *Lower bounds: randomized algebraic constructions*

So far we have seen both constructions using random graphs and algebraic constructions. In this section we present an alternative construction of $K_{s,t}$-free graphs due to Bukh with $\Theta(n^{2-\frac{1}{s}})$ edges provided $t > t_0(s)$ for some function t_0. This is an algebraic construction with some randomness added to it.

Bukh (2015)

First fix $s \geq 4$ and take a prime power q. Let $d = s^2 - s + 2$ and $f \in \mathbb{F}_q[x_1, x_2, \ldots, x_s, y_1, y_2, \ldots, y_s]$ be a polynomial chosen uniformly at random among all polynomials with degree at most d in each of $X = (x_1, x_2, \ldots, x_s)$ and $Y = (y_1, y_2, \ldots, y_s)$. Take G bipartite with vertex parts $n = L = R = \mathbb{F}_q^s$ and define the edge relation as $(X, Y) \in L \times R$ when $f(X, Y) = 0$.

Lemma 2.39. *For all $u, v \in \mathbb{F}_q^s$ and f chosen randomly as above*

$$\mathbb{P}[f(u, v) = 0] = \frac{1}{q}.$$

Proof. Notice that if g is a uniformly random constant in $\mathbb{F}_q$, then $f(u, v)$ and $f(u, v) + g$ are identically distributed. Hence each of the q possibilities are equally likely to the probability is $1/q$. □

Now the expected number of edges is the order we want as $\mathbb{E}[e(G)] = \frac{n^2}{q}$. All that we need is for the number of copies of $K_{s,t}$ to be relatively low. In order to do so, we must answer the following question. For a set of vertices in L of size s, how many common neighbors can it have?

Lemma 2.40. *Suppose $r, s \leq min(\sqrt{q}, d)$ and $U, V \subset \mathbb{F}_q^s$ with $|U| = s$ and $|V| = r$. Furthermore let $f \in \mathbb{F}_q[x_1, x_2, \ldots, x_s, y_1, y_2, \ldots, y_s]$ be a polynomial chosen uniformly at random among all polynomials with degree at most d in each of $X = (x_1, x_2, \ldots, x_s)$ and $Y = (y_1, y_2, \ldots, y_s)$. Then*

$$\mathbb{P}[f(u, v) = 0 \text{ for all } u \in U, v \in V] = q^{-sr}.$$

Proof. First let us consider the special case where the first coordinates of points in U and V are all distinct. Define

$$g(X_1, Y_1) = \sum_{\substack{0 \leq i \leq s-1 \\ 0 \leq j \leq r-1}} a_{ij} X_1^i Y_1^j$$

with a_{ij} each uniform iid random variables over $\mathbb{F}_q$. We know that f and $f + g$ have the same distribution, so it suffices to show for all $b_{uv} \in \mathbb{F}_q$ where $u \in U$ and $v \in V$ there exists a_{ij} for which $g(u, v) = b_{uv}$ for all $u \in U, v \in V$. The idea is to apply Lagrange Interpolation twice. First for all $u \in U$ we can find a single variable polynomial $g_u(Y_1)$ with degree at most $r - 1$ such that $g_u(v) = b_{uv}$ for all $v \in V$. Then we can view $g(X_1, Y_1)$ as a polynomial in Y_1 with coefficients being polynomials in X_1, i.e.,

$$g(X_1, Y_1) = \sum_{0 \leq j \leq r-1} a_j(X_1) Y_1^j.$$

Applying the Lagrange interpolation theorem for a second time we can find polynomials $a_0, a_1, \ldots, a_{r-1}$ such that for all $u \in U$, $g(u, Y_1) = g_u(Y_1)$ as polynomials in Y_1.

Now suppose the first coordinates are not necessarily distinct. It suffices to find linear maps $T, S\colon \mathbb{F}_q^s \to \mathbb{F}_q^s$ such that TU and SV have all their first coordinates different. Let us prove that such a map T exists. If we find a linear map $T_1\colon \mathbb{F}_q^s \to \mathbb{F}_q$ that sends the elements of U to distinct elements, then we can extend T_1 to T by using T_1 for the first coordinate. To find T_1 pick T_1 uniformly among all linear maps. Then for every pair in U the probability of collision is $\frac{1}{q}$. So by union bounding we have the probability of success is at least $1 - \binom{|U|}{2}\frac{1}{q} > 0$, so such a map T exists. Similarly S exists. □

Fix $U \subset \mathbb{F}_q^s$ with $|U| = s$. We wish to upper bound the number of instances of U having many common neighbors. In order to do this, we will use the method of moments. Let $I(v)$ represent the indicator variable which is 1 exactly when v is a common neighbor of U and set X to be the number of common neighbors of U. Then using Lemma 2.40,

$$\begin{aligned}\mathbb{E}[X^d] = \mathbb{E}[(\sum_{v\in\mathbb{F}_q^s} I(v))^d] &= \sum_{v_1,\ldots,v_d\in\mathbb{F}_q^s} \mathbb{E}[I(v_1)\cdots I(v_d)] \\ &= \sum_{r\le d}\binom{q^s}{r} q^{-rs} M_r \le \sum_{r\le d} M_r = M,\end{aligned}$$

where M_r is defined as the number of surjections from $[d]$ to $[r]$ and $M = \sum_{r\le d} M_r$. Using Markov's inequality we get

$$\mathbb{P}(X \ge \lambda) \le \frac{\mathbb{E}[X^d]}{\lambda^d} \le \frac{M}{\lambda^d}.$$

Now even if the expectation of X is low, we cannot be certain that the probability of X being large is low. For example if we took the random graph with $p = n^{-\frac{1}{s}}$ then X will have low expectation but a long, smooth-decaying tail and therefore it is likely that X will be large for some U.

It turns out what algebraic geometry prevents the number of common neighbors X from taking arbitrary values. The common neighbors are determined by the zeros of a set of polynomial equations, and hence form an algebraic variety. The intuition is that either we are in a "zero-dimensional" case where X is very small or a "positive dimensional" case where X is at least on the order of q.

Lemma 2.41. *For all s, d there exists a constant C such that if $f_1(Y), \ldots, f_s(Y)$ are polynomials on $\mathbb{F}_q^s$ of degree at most d then* Bukh (2015)

$$\{y \in \mathbb{F}_q^s | f_1(y) = \ldots f_s(y) = 0\}$$

has size either at most C at least $q - C\sqrt{q}$.

The lemma can be deduced from the following important result from algebraic geometry known as the Lang–Weil bound, which says that the number of points of an r-dimensional algebraic variety in $\mathbb{F}_q^s$ is roughly q^r, as long as certain irreducibility hypotheses are satisfied.

Theorem 2.42 (Lang–Weil bound). *If $V = \{y \in \overline{\mathbb{F}}_q^s | g_1(y) = g_2(y) = \ldots = g_m(y)\}$ is irreducible and g_i has degree at most d, then* Lang and Weil (1954)

$$|V \cap \mathbb{F}_q^s| = q^{\dim V}(1 + O_{s,m,d}(q^{-\frac{1}{2}})).$$

Now we can use our bound from Markov's Inequality along with Lemma 2.41. Let the s polynomials $f_1(Y), \ldots, f_s(Y)$ in Lemma 2.41 be the s polynomials $f(u, Y)$ as u ranges over the s elements of U. Then for large enough q there exists a constant C from Lemma 2.41 such that having $X > C$ would imply $X \geq q - C\sqrt{q} > q/2$, so that

$$\mathbb{P}(X > C) = \mathbb{P}\left(X > \frac{q}{2}\right) \leq \frac{M}{(q/2)^d}.$$

Thus the number of subsets of L or R with size s and more than C common neighbors is at most

$$2\binom{n}{s}\frac{M}{(q/2)^d} = O(q^{s-2})$$

in expectation. Take G and remove a vertex from every such subset to create G'. First we have that G' is $K_{s,C+1}$-free. Then

$$\mathbb{E}[e(G')] \geq \frac{n^2}{q} - O(nq^{s-2}) = (1 - o(1))\frac{n^2}{q} = (1 - o(1))n^{2-\frac{1}{s}}$$

and $v(G') \leq 2n$. So there exists an instance of G' that obtains the desired bound.

2.9 *Forbidding a sparse bipartite graph*

For any bipartite graph H, it is always contained in $K_{s,t}$ for some s, t. Therefore by Theorem 2.17,

$$\mathrm{ex}(n, H) \leq \mathrm{ex}(n, K_{s,t}) \lesssim n^{2-\frac{1}{s}}.$$

The first inequality is not tight in general when H is some sparse bipartite graph. In this section, we will see some techniques that give a better upper bound on $\mathrm{ex}(n, H)$ for sparse bipartite graphs H.

The first result we are going to see is an upper bound on $\mathrm{ex}(n, H)$ when H is bipartite and the degrees of vertices in one part are bounded above.

Theorem 2.43. *Let H be a bipartite graph whose vertex set is $A \cup B$ such that every vertex in A has degree at most r. Then there exists a constant $C = C_H$ such that*

$$\mathrm{ex}(n, H) \leq Cn^{2-\frac{1}{r}}$$

Füredi (1991)
Alon, Krivelevich and Sudakov (2003)

Remark 2.44. Theorem 2.32 shows that the exponent $2 - \frac{1}{r}$ is the best possible as function of r since we can take $H = K_{r,t}$ for some $t \leq (r-1)! + 1$.

To show this result, we introduce the following powerful probabilistic technique called dependent random choice. The main idea of this lemma is the following: if G has many edges, then there exists a large subset U of $V(G)$ such that all small subsets of vertices in U have many common neighbors.

Lemma 2.45 (Dependent random choice)**.** *Let $u, n, r, m, t \in \mathbb{N}, \alpha > 0$ be numbers that satisfy the inequality*

$$n\alpha^t - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq u.$$

Alon, Krivelevich and Sudakov (2003)

Then every graph G with n vertices and at least $\alpha n^2/2$ edges contains a subset U of vertices with size at least u such that every r-element subset S of U has at least m common neighbors.

Proof. Let T be a list of t vertices chosen uniformly at random from $V(G)$ with replacement (allowing repetition). Let A be the common neighborhood of T. The expected value of $|A|$ is

$$\begin{aligned}
\mathbb{E}|A| &= \sum_{v \in V} \mathbb{P}(v \in A) \\
&= \sum_{v \in V} \mathbb{P}(T \subseteq N(v)) \\
&= \sum_{v \in V} \left(\frac{d(v)}{n}\right)^t \\
&\geq n\left(\frac{1}{n}\sum_{v \in V}\frac{d(v)}{n}\right)^t && \text{(convexity)} \\
&\geq n\alpha^t.
\end{aligned}$$

For every r-element subset S of V, the event of A containing S occurs if and only if T is contained in the common neighborhood of S, which occurs with probability

$$\left(\frac{\#\text{common neighbors of } S}{n}\right)^t.$$

Call a set S *bad* if it has less than m common neighbors. Then each bad r-element subset $S \subset V$ is contained in A with probability less

than $(m/n)^t$. Therefore by linearity of expectation,

$$\mathbb{E}[\text{the number bad } r\text{-element subset of } A] < \binom{n}{r}\left(\frac{m}{n}\right)^t.$$

To make sure that there are no bad subsets, we can get rid of one element in each bad subset. The number of remaining elements is at least $|A| - (\#\text{bad } r\text{-element subset of } A)$, whose expected value is at least

$$n\alpha^t - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq u.$$

Consequently, there exists a T such that there are at least u elements in A remaining after getting rid of all bad r-element subsets. The set U of the remaining u elements satisfies the desired properties. □

Setting the parameters of Lemma 2.45 to what we need for proving Theorem 2.43, we get the following corollary.

Corollary 2.46. *For any bipartite graph H with vertex set $A \cup B$ where each vertex in A has degree at most r, there exists C such that the following holds: Every graph with at least $Cn^{2-\frac{1}{r}}$ edges contains a vertex subset U with $|U| = |B|$ such that every r-element subset in U has at least $|A| + |B|$ common neighbors.*

Proof. By Lemma 2.45 with $u = |B|$, $m = |A| + |B|$, and $t = r$, it suffices to check that there exists C so that

$$n\left(2Cn^{-\frac{1}{r}}\right)^r - \binom{n}{r}\left(\frac{|A|+|B|}{n}\right)^r \geq |B|.$$

The first term evaluates to $(2C)^r$, and the second term is $O_H(1)$. Therefore we can choose C large enough to make this inequality hold. □

Now we are ready to show Theorem 2.43.

Proof of Theorem 2.43. Let G be a graph with n vertices and at least $Cn^{2-\frac{1}{r}}$ edges, where C is chosen as in Corollary 2.46. First embed B into $V(G)$ using U from Corollary 2.46. The plan is to extend this embedding furthermore to $A \cup B \hookrightarrow V(G)$. To do this, assume that we have an embedding $\phi : A' \cup B \hookrightarrow V(G)$ already where $A' \subseteq A$, and we want to extend ϕ to an arbitrary $v \in A \backslash A'$. We have to make sure that $\phi(v)$ is a common neighbor of $\phi(N(v))$ in G. Note that by assumption, $|\phi(N(v))| = |N(v)| \leq r$, and so by the choice of B, the set $\phi(N(v))$ has at least $|A| + |B|$ common neighbors. $\phi(v)$ can then be any of those common neighbors, with an exception that $\phi(v)$ cannot be the same as $\phi(u)$ for any other $u \in A' \cup B$. This eliminates $|A'| + |B| \leq |A| + |B| - 1$ possibilities for $\phi(v)$. Since there are at least $|A| + |B|$ vertices to choose from, we can just extend ϕ by setting $\phi(v)$

to be one of the remaining choices. With this process, we can extend the embedding to $A \cup B \hookrightarrow V(G)$, which shows that there is a copy of H in G. □

This is a general result that can be applied to all bipartite graphs. However, for some specific bipartite graph H, there could be room for improvement. For example, from this technique, the bound we get for C_6 is the same as C_4, which is $O(n^{3/2})$. This is nonetheless not tight.

Theorem 2.47 (Even cycles)**.** *For all integer $k \geq 2$, there exists a constant C so that*

$$\mathrm{ex}(n, C_{2k}) \leq Cn^{1+\frac{1}{k}}.$$

Bondy and Simonovits (1974)

Remark 2.48. It is known that $\mathrm{ex}(n, C_{2k}) = \Theta\left(n^{1+1/k}\right)$ for $k = 2, 3, 5$. However, it is open whether the same holds for other values of k.

Benson (1966)

Instead of this theorem, we will show a weaker result:

Theorem 2.49. *For any integer $k \geq 2$, there exists a constant C so that every graph G with n vertices and at least $Cn^{1+1/k}$ edges contains an even cycle of length at most $2k$.*

To show this theorem, we will first "clean up" the graph so that the minimum degree of the graph is large enough, and also the graph is bipartite. The following two lemmas will allow us to focus on a subgraph of G that satisfies those nice properties.

Lemma 2.50. *Let $t \in \mathbb{R}$ and G a graph with average degree $2t$. Then G contains a subgraph with minimum degree greater than r.*

Proof. We have $e(G) = v(G)t$. Removing a vertex of degree at most t cannot decrease the average degree. We can keep removing vertices of degree at most t until every vertex has degree more than t. This algorithm must terminate before reaching the empty subgraph since every graph with at most $2t$ vertices has average degree less than $2t$. The remaining subgraph when the algorithm terminates is then a subgraph whose minimum degree is more than t. □

Lemma 2.51. *Every G has a bipartite subgraph with at least $e(G)/2$ edges.*

Proof. Color every vertex with one of two colors uniformly at random. Then the expected value of non-monochromatic edges is $e(G)/2$. Hence there exists a coloring that has at least $e(G)/2$ non-monochromatic edges. □

Proof of Theorem 2.49. Suppose that G contains no even cycles of length at most $2k$. By Lemma 2.50 and Lemma 2.51 there exists a bipartite subgraph G' with minimum degree at least $\delta := Cn^{1/k}/2$.

Let $A_0 = \{u\}$ where u is an arbitrary vertex in $V(G')$. Let $A_{i+1} = N_{G'}(A_i) \backslash A_{i-1}$. Then A_i is the set of vertices that are distance exactly i away from the starting vertex u since G' is bipartite.

Now for every two different vertices v, v' in A_{i-1} for some $1 \leq i \leq k$, if they have a common neighbor w in A_i, then there are two different shortest paths from u to w. The union two distinct paths (even if they overlap) contains an even-cycle of length at most $2i \leq 2k$, which is a contradiction. Therefore the common neighbors of any two vertices in A_{i-1} can only lie in A_{i-2}, which implies that $|A_i| \geq (\delta - 1)|A_{i-1}|$. Hence $|A_k| \geq (\delta - 1)^k \geq (Cn^{1/k} - 1)^k$. If C is chosen large enough then we get $|A_k| > n$, which is a contradiction. □

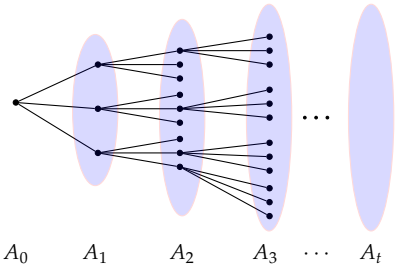


Figure 2.4: Diagram for Proof of Theorem 2.49

If H is a bipartite graph with vertex set $A \cup B$ and each vertex in A has degree at most 2, then $\mathrm{ex}(n, H) = O(n^{3/2})$. The exponent $3/2$ is optimal since $\mathrm{ex}(n, K_{2,2}) = \Theta(n^{3/2})$ and hence the same holds whenever H contains $K_{2,2}$. It turns out that this exponent can be improved whenever H does not contain any copy of $K_{2,2}$.

Colon and Lee (2019+)

Theorem 2.52. *Let H be a bipartite graph with vertex bipartition $A \cup B$ such that each vertex in A has degree at most 2, and H does not contain $K_{2,2}$. Then there exist $c, C > 0$ dependent on H such that*

$$\mathrm{ex}(n, H) \leq Cn^{\frac{3}{2}-c}.$$

To prove this theorem, we show an equivalent statement formulated using the notion of subdivisons. For a graph H, the 1-subdivision $H^{1\text{-sub}}$ of H is obtained by adding an extra vertex in the middle of every edge in H. Notice that every H in the setting of Theorem 2.52 is a subgraph of some $K_t^{1\text{-sub}}$. Therefore we can consider the following alternative formulation of Theorem 2.52.

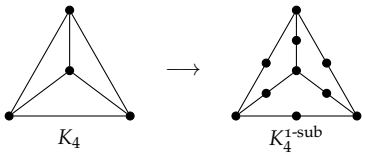


Figure 2.5: 1-subdivision of K_4

Theorem 2.53. *For all $t \geq 3$, there exists $c_t > 0$ such that*

$$\mathrm{ex}(n, K_t^{1\text{-sub}}) = O(n^{\frac{3}{2}-c_t}).$$

Janzer (2018)

Now we present a proof of Theorem 2.53 by Janzer. As in Theorem 2.49, it is helpful to pass the entire argument to a subgraph where we have a better control of the degrees of the vertices. To do so, we are going to use the following lemma (proof omitted) to find an *almost regular* subgraph.

Colon and Lee (2019+)

Lemma 2.54. *For all $0 < \alpha < 1$, there exist constants $\beta, k > 0$ such that for all $C > 0$, n sufficiently large, every n-vertex graph G with $\geq Cn^{1+\alpha}$ edges has a subgraph G' such that*

(a) $v(G') \geq n^{\beta}$,

(b) $e(G') \geq \frac{1}{10} C v(G')^{1+\alpha}$,

(c) $\max \deg(G') \leq K \min \deg(G')$,

(d) *G' is bipartite with two parts of sizes differing by factor* ≤ 2.

From now on, we treat t as a constant. For any two vertices $u, v \in A$, we say that the pair uv is *light* if the number of common neighbors of u and v is at least 1 and less than $\binom{t}{2}$; moreover, we say that the pair uv is *heavy* if the number of common neighbors of u and v is at least $\binom{t}{2}$. Note that pairs $u, v \in A$ without any common neighbors are neither light nor heavy. The following lemma gives a lower bound on the number of light pairs.

Lemma 2.55. *Let G be a $K_t^{1\text{-sub}}$-free bipartite graph with bipartition $U \cup B$, $d(x) \geq \delta$ for all $x \in U$, and $|U| \geq 4|B|t/\delta$. Then there exists $u \in U$ in $\Omega(\delta^2|U|/|B|)$ light pairs in U.*

Proof. Let S be the set of $\{(\{u, v\}, x) | u, v \in U, x \in B\}$ where $\{u, v\}$ is an unordered pair of vertices in U and x is a common neighbor of $\{u, v\}$. We can count this by choosing $x \in B$ first:

$$|S| = \sum_{x \in B} \binom{d(x)}{2} \geq |B| \binom{e(G)/|B|}{2} \geq \frac{|B|}{4} \left(\frac{\delta|U|}{|B|}\right)^2 = \frac{\delta^2|U|^2}{4|B|}.$$

Notice that the low-degree vertices in B contributes very little since

$$\sum_{\substack{x \in B \\ d(x) < 2t}} \binom{d(x)}{2} \leq 2t^2|B| \leq \frac{\delta^2|U|^2}{8|B|}.$$

Therefore

$$\sum_{\substack{x \in B \\ d(x) \geq 2t}} \binom{d(x)}{2} \geq \frac{\delta^2|U|^2}{8|B|}.$$

Note that if there are t mutually heavy vertices in U, then we can choose a common neighbor u_{ij} for every pair $\{v_i, v_j\}$ with $i < j$. Since there are at least $\binom{t}{2}$ such neighbors for each pair $\{v_i, v_j\}$, one can make choices so that all u_{ij} are distinct. This then produces a $K_t^{1\text{-sub}}$ subgraph, which is a contradiction. Therefore there do not exist t mutually heavy vertices in U, and by Turán's Theorem, the number of heavy pairs in $N(x)$ for $x \in B$ is at most $e(T_{d(x),t-1})$. Since any two vertices in $N(x)$ have at least one common neighbor x, they either form a light pair or a heavy pair. This shows that there are at least

$\binom{d(x)}{2} - e(T_{d(x),t-1})$ light pairs among $N(x)$. If $d(x) \geq 2t$, then

$$\begin{aligned}
&\binom{d(x)}{2} - e(T_{d(x),t-1}) \\
&\geq \binom{d(x)}{2} - \binom{t-1}{2}\left(\frac{d(x)}{t-1}\right)^2 \\
&= \frac{1}{2(t-1)}d(x)^2 - \frac{1}{2}d(x) \\
&\gtrsim d(x)^2.
\end{aligned}$$

If we sum over $x \in B$, then each light pair is only going to be counted for at most $\binom{t}{2}$ times according to the definition. This is constant since we view t as a constant. Therefore

$$\#\text{light pairs in } U \gtrsim \sum_{x\in B} d(x)^2 \gtrsim |S| \gtrsim \frac{\delta^2|U|^2}{|B|},$$

and by pigeon hole principle there exists a vertex $u \in U$ that is in $\Omega(\delta^2|U|/|B|)$ light pairs. □

With these lemmas, we are ready to prove Theorem 2.53.

Proof of Theorem 2.53. Let G be any $K_t^{1\text{-sub}}$-free graph. First pick G' by Lemma 2.54 with $\alpha = (t-2)/(2t-3)$, and say that the two parts are A and B. Set δ to be the minimum degree of G'. We will prove that $\delta \leq Cv(G')^{(t-2)/(2t-3)}$ for some sufficiently large constant C by contradiction. Suppose that $\delta > Cv(G')^{(t-2)/(2t-3)}$. Our plan is to pick $v_1, v_2, \ldots, v_t$ such that v_iv_j are light for all $i < j$, and no three of $v_1, \ldots, v_t$ have common neighbors. This will give us a $K_t^{1\text{-sub}}$ and hence a contradiction.

We will do so by repeatedly using Lemma 2.55 and induction on a stronger hypothesis: For each $1 \leq i \leq t$, there exists $A = U_1 \supseteq U_2 \supseteq \cdots \supseteq U_i$ and $v_j \in U_j$ such that

(a) v_j is in at least $\Theta(\delta^2|U_j|/v(G'))$ light pairs in U_j for all $1 \leq j \leq i-1$,

(b) v_j is light to all vertices in U_{j+1} for all $1 \leq j \leq i-1$.

(c) no three of $v_1, \ldots, v_i$ have common neighbors,

(d) $|U_{j+1}| \gtrsim \delta^2|U_j|/v(G')$ for all $1 \leq j \leq i-1$,

This statement clearly holds when $i = 1$ by choosing v_1 to be the vertex found by Lemma 2.55. Now suppose that we have constructed $A = U_1 \supseteq \cdots \supseteq U_{i-1}$ with $v_j \in U_j$ for all $j = 1, \ldots, i-1$. To construct U_i, let U_i' be the set of vertices that form light pairs with v_{i-1}. Then $|U_i'| \gtrsim \delta^2|U_{i-1}|/v(G')$ by the inductive hypothesis (a). Now we get rid of all the vertices in U_i' that violate (c) to get U_i. It suffices to look

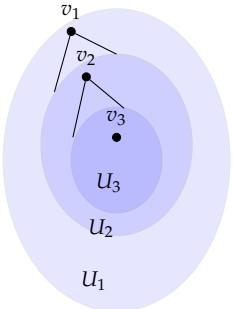


Figure 2.6: Repeatedly applying Lemma 2.55 to obtain v_i's and U_i's

at each pair $v_j v_k$, look at their common neighbors u and delete all the neighbors of u from U_i'. There are $\binom{i-1}{2}$ choices $v_j v_k$, and they have at most $\binom{t}{2}$ common neighbors since they form a light pair, and each such neighbors has degree at most $K\delta$. Therefore the number of vertices removed is at most

$$\binom{i-1}{2}\binom{t}{2}K\delta = O(\delta)$$

since t and K are constants. Therefore after this alteration, (d) will still hold as long as $|U_i'| = \Omega(\delta)$ and C is chosen sufficiently large. This is true since

$$|U_i'| \gtrsim \left(\frac{\delta^2}{V(G')}\right)^{i-1} |A| \gtrsim \delta^{2t-2} V(G')^{t-2} = \Theta(\delta)$$

given that $i \leq t$. Therefore (d) holds for i, and we just need to choose a vertex v_i from Lemma 2.55 in U_i and (a), (b), (c) follow directly. Therefore by induction, this also holds for $i = t$. Now by (b) and (c), there exists a copy of $K_t^{1\text{-sub}}$ in G', which is a contradiction.

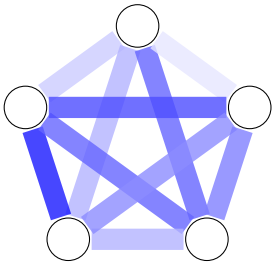
The above argument shows that $\delta \leq Cv(G')^{(t-2)/(2t-3)}$, and so the maximum degree is at most $KCv(G')^{(t-2)/(2t-3)}$. Hence $e(G') \leq KCv(G')^{1+\alpha}$, and by the choice of G', we know that $e(G) \leq 10KCn^{1+\alpha}$, as desired. □

3
Szemerédi's regularity lemma

3.1 Statement and proof

Szemerédi's regularity lemma is one of the most important results in graph theory, particularly the study of large graphs. Informally, the lemma states that for all large dense graphs G, we can partition the vertices of G into a bounded number of parts so that edges between most different parts behave "random-like."

To give a notion of "random-like," we first state some definitions.



The edges between parts behave in a "random-like" fashion.

Definition 3.1. Let X and Y be sets of vertices in a graph G. Let $e_G(X,Y)$ be the number of edges between X and Y; that is,

$$e_G(X,Y) = |\{(x,y) \in X \times Y \mid xy \in E(G)\}|.$$

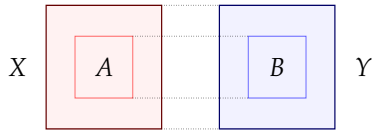
From this, we can define the ***edge density*** between X and Y to be

$$d_G(X,Y) = \frac{e_G(X,Y)}{|X||Y|}.$$

We will drop the subscript G if context is clear.

Definition 3.2 (ϵ-regular pair)**.** Let G be a graph and $X, Y \subseteq V(G)$. We call (X,Y) an ***ϵ-regular pair*** (in G) if for all $A \subset X, B \subset Y$ with $|A| \geq \epsilon|X|, |B| \geq \epsilon|Y|$, one has

$$|d(A,B) - d(X,Y)| \leq \epsilon.$$



The subset pairs of an ϵ-regular pair are similar in edge density to the main pair.

Remark 3.3. The different ϵ in Definition 3.2 play different roles, but it is not important to distinguish them. We use only one ϵ for convenience of notation.

Suppose (X,Y) is not ϵ-regular. Then their irregularity is "witnessed" by some $A \subset X, B \subset Y$ with $A \geq \epsilon|X|, |B| \geq \epsilon|Y|$, and $|d(A,B) - d(X,Y)| > \epsilon$.

Definition 3.4 (ϵ-regular partition)**.** A partition $\mathcal{P} = \{V_1, \ldots, V_k\}$ of $V(G)$ is an ***ϵ-regular partition*** if

$$\sum_{\substack{(i,j)\in[k]^2 \\ (V_i,V_j) \text{ not } \epsilon\text{-regular}}} |V_i||V_j| \leq \epsilon |V(G)|^2.$$

Note that this definition allows a few irregular pairs as long as their total size is not too big.

We can now state the regularity lemma.

Theorem 3.5 (Szemerédi's regularity lemma)**.** *For every $\epsilon > 0$, there exists a constant M such that every graph has an ϵ-regular partition into at most M parts.*

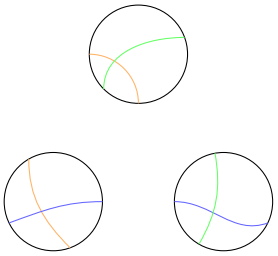
Szemerédi (1978)

A stronger version of the lemma allows us to find an equitable partition — that is, every part of the partition has size either $\lfloor \frac{n}{k} \rfloor$ or $\lceil \frac{n}{k} \rceil$ where the graph has n vertices and the partition has k parts.

Theorem 3.6 (Equitable Szemerédi's regularity lemma)**.** *For all $\epsilon > 0$ and m_0, there exists a constant M such that every graph has an ϵ-regular equitable partition of its vertex set into k parts with $m_0 \leq k \leq M$.*

We start with a sketch of the proof. We will generate the partition according to the following algorithm:

- Start with the trivial partition (1 part).
- While the partition is not ϵ-regular:
  - For each (V_i, V_j) that is not ϵ-regular, find $A^{i,j} \subset V_i$ and $A^{j,i} \subset V_j$ witnessing the irregularity of (V_i, V_j).
  - Simultaneously refine the partition using all $A^{i,j}$.



The boundaries of irregular witnesses refine each part of the partition.

If this process stops after a bounded number of steps, the regularity lemma would be successfully proven. To show that we will stop in a bounded amount of time, we will apply a technique called the ***energy increment argument***.

Definition 3.7 (Energy)**.** Let $U, W \subseteq V(G)$ and $n = |V(G)|$. Define

$$q(U, W) = \frac{|U||W|}{n^2} d(U, W)^2.$$

For partitions $\mathcal{P}_U = \{U_1, \ldots, U_k\}$ of U and $\mathcal{P}_W = \{W_1, \ldots, W_l\}$ of W, define

$$q(\mathcal{P}_U, \mathcal{P}_W) = \sum_{i=1}^{k} \sum_{j=1}^{l} q(U_i, W_j).$$

Finally, for a partition $\mathcal{P} = \{V_1, \ldots, V_k\}$ of $V(G)$, define the ***energy*** of $\mathcal{P}$ to be $q(\mathcal{P}, \mathcal{P})$. Specifically,

$$q(\mathcal{P}) = \sum_{i=1}^{k} \sum_{j=1}^{k} q(V_i, V_j) = \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2.$$

This is a mean-square quantity, so it is an L^2 quantity. Borrowing from physics, this motivates the name "energy".

Observe that energy is between 0 and 1 because edge density is bounded above by 1:

$$q(\mathcal{P}) = \sum_{i=1}^{k}\sum_{j=1}^{k}\frac{|V_i||V_j|}{n^2}d(V_i,V_j)^2 \leq \sum_{i=1}^{k}\sum_{j=1}^{k}\frac{|V_i||V_j|}{n^2} = 1.$$

We proceed with a sequence of lemmas that culminate in the main proof. These lemmas will show that energy cannot decrease upon refinement, but can increase substantially if the partition we refine is irregular.

Lemma 3.8. *For any partitions $\mathcal{P}_U$ and $\mathcal{P}_W$ of vertex sets U and W, $q(\mathcal{P}_U,\mathcal{P}_W) \geq q(U,W)$.*

Proof. Let $\mathcal{P}_U = \{U_1,\ldots,U_k\}$ and $\mathcal{P}_W = \{W_1,\ldots,W_l\}$. Choose vertices x uniformly from U and y uniformly from W. Let U_i be the part of $\mathcal{P}_U$ that contains x and W_j be the part of $\mathcal{P}_W$ that contains y. Then define the random variable $Z = d(U_i,W_j)$. Let us look at properties of Z. The expectation is

$$\mathbb{E}[Z] = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{|U_i|}{|U|}\frac{|W_j|}{|W|}d(U_i,W_j) = \frac{e(U,W)}{|U||W|} = d(U,W).$$

The second moment is

$$\mathbb{E}[Z^2] = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{|U_i|}{|U|}\frac{|W_j|}{|W|}d(U_i,W_j)^2 = \frac{n^2}{|U||W|}q(\mathcal{P}_U,\mathcal{P}_W).$$

By convexity, $\mathbb{E}[Z^2] \geq \mathbb{E}[Z]^2$, which implies the lemma. □

Lemma 3.9. *If $\mathcal{P}'$ refines $\mathcal{P}$, then $q(\mathcal{P}') \geq q(\mathcal{P})$.*

Proof. Let $\mathcal{P} = \{V_1,\ldots,V_m\}$ and apply Lemma 3.8 to every (V_i,V_j). □

Lemma 3.10 (Energy boost lemma)**.** *If (U,W) is not ϵ-regular as witnessed by $U_1 \subset U$ and $W_1 \subset W$, then*

$$q\left(\{U_1,U\backslash U_1\},\{W_1,W\backslash W_1\}\right) > q(U,W) + \epsilon^4\frac{|U||W|}{n^2}.$$

This is the Red Bull Lemma, giving an energy boost if you are feeling irregular.

Proof. Define Z as in the proof of Lemma 3.8. Then

$$\begin{aligned}\mathrm{Var}(Z) &= \mathbb{E}[Z^2] - \mathbb{E}[Z]^2\\ &= \frac{n^2}{|U||W|}\left(q\left(\{U_1,U\backslash U_1\},\{W_1,W\backslash W_1\}\right) - q(U,W)\right).\end{aligned}$$

But observe that $|Z - \mathbb{E}[Z]| = |d(U_1,W_1) - d(U,W)|$ with probability $\frac{|U_1|}{|U|}\frac{|W_1|}{|W|}$ (corresponding to $x \in U_1$ and $y \in W_1$), so

$$\begin{aligned}\mathrm{Var}(Z) &= \mathbb{E}[(Z-\mathbb{E}[Z])^2]\\ &\geq \frac{|U_1|}{|U|}\frac{|W_1|}{|W|}(d(U_1,W_1) - d(U,W))^2\\ &> \epsilon\cdot\epsilon\cdot\epsilon^2\end{aligned}$$

as desired. □

Lemma 3.11. *If a partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of $V(G)$ is not ϵ-regular, then there exists a refinement $\mathcal{Q}$ of $\mathcal{P}$ where every V_i is partitioned into at most 2^k parts such that*

$$q(\mathcal{Q}) \geq q(\mathcal{P}) + \epsilon^5.$$

Proof. For all (i, j) such that (V_i, V_j) is not ϵ-regular, find $A^{i,j} \subset V_i$ and $A^{j,i} \subset V_j$ that witness irregularity (do this simultaneously for all irregular pairs). Let $\mathcal{Q}$ be a common refinement of $\mathcal{P}$ by $A^{i,j}$'s. Each V_i is partitioned into at most 2^k parts as desired.

Then

$$\begin{aligned} q(\mathcal{Q}) &= \sum_{(i,j)\in[k]^2} q(\mathcal{Q}_{V_i}, \mathcal{Q}_{V_j}) \\ &= \sum_{\substack{(i,j)\in[k]^2 \\ (V_i,V_j)\ \epsilon\text{-regular}}} q(\mathcal{Q}_{V_i}, \mathcal{Q}_{V_j}) + \sum_{\substack{(i,j)\in[k]^2 \\ (V_i,V_j)\text{ not }\epsilon\text{-regular}}} q(\mathcal{Q}_{V_i}, \mathcal{Q}_{V_j}) \end{aligned}$$

where $\mathcal{Q}_{V_i}$ is the partition of V_i given by $\mathcal{Q}$. By Lemma 3.8, the above quantity is at least

$$\sum_{\substack{(i,j)\in[k]^2 \\ (V_i,V_j)\ \epsilon\text{-regular}}} q(V_i, V_j) + \sum_{\substack{(i,j)\in[k]^2 \\ (V_i,V_j)\text{ not }\epsilon\text{-regular}}} q(\{A^{i,j}, V_i \backslash A^{i,j}\}, \{A^{j,i}, V_j \backslash A^{j,i}\})$$

since V_i is cut by $A^{i,j}$ when creating $\mathcal{Q}$, so $\mathcal{Q}_{V_i}$ is a refinement of $\{A^{i,j}, V_i \backslash A^{i,j}\}$. By Lemma 3.10, the above sum is at least

$$\sum_{(i,j)\in[k]^2} q(V_i, V_j) + \sum_{\substack{(i,j)\in[k]^2 \\ (V_i,V_j)\text{ not }\epsilon\text{-regular}}} \epsilon^4 \frac{|V_i||V_j|}{n^2}.$$

But the second sum is at least ϵ^5 since $\mathcal{P}$ is not ϵ-regular, so we deduce the desired inequality. □

Now we can prove Szemerédi's regularity lemma.

Proof of Theorem 3.5. Start with a trivial partition. Repeatedly apply Lemma 3.11 whenever the current partition is not ϵ-regular. By the definition of energy, $0 \leq q(\mathcal{P}) \leq 1$. However, by Lemma 3.11, $q(\mathcal{P})$ increases by at least ϵ^5 at each iteration. So we will stop after at most ϵ^{-5} steps, resulting in an ϵ-regular partition. □

An interesting question is that of how many parts this algorithm provides. If $\mathcal{P}$ has k parts, Lemma 3.11 refines $\mathcal{P}$ into at most $k2^k \leq 2^{2^k}$ parts. Iterating this ϵ^{-5} times produces an upper bound of $\underbrace{2^{2^{\cdot^{\cdot^{\cdot^2}}}}}_{2\epsilon^{-5}\ 2\text{'s}}$. One might think that a better proof could produce a better bound, as we take no care in minimizing the number of parts we refine to. Surprisingly, this is essentially the best bound.

Theorem 3.12 (Gowers). *There exists a constant $c > 0$ such that for all $\epsilon > 0$ small enough, there exists a graph all of whose ϵ-regular partitions require at least $\underbrace{2^{2^{\cdot^{\cdot^{2}}}}}_{\geq \epsilon^{-c}\ 2\text{'s}}$ parts.*

Gowers (1997)

Another question which stems from this proof is how we can make the partition equitable. Here is a modification to the algorithm above which proves Theorem 3.6:

- Start with an arbitrary equitable partition of the graph into m_0 parts.
- While the partition is not ϵ-regular:
  - Refine the partition using pairs that witness irregularity.
  - Refine further and rebalance to make the partition equitable. To do this, move and merge sets with small numbers of vertices.

There is a wrong way to make the partition equitable. Suppose you apply the regularity lemma *and then* try to refine further and rebalance. You may lose ϵ-regularity in the process. One must directly modify the algorithm in the proof of Szemerédi's regularity lemma to get an equitable partition.

The refinement steps increase energy by at least ϵ^5 as before. The energy might go down in the rebalancing step, but it turns out that the decrease does not affect the end result. In the end, the increase is still $\Omega(\epsilon^5)$, which allows the process to terminate after $O(\epsilon^{-5})$ steps.

3.2 *Triangle counting and removal lemmas*

Szemerédi's regularity lemma is a powerful tool for tackling problems in extremal graph theory and additive combinatorics. In this section, we apply the regularity lemma to prove Theorem 1.7, Roth's theorem on 3-term arithmetic progressions. We first establish the triangle counting lemma, which provides one way of extracting information from regular partitions, and then use this result to prove the triangle removal lemma, from which Roth's theorem follows.

As we noted in the previous section, if two subsets of the vertices of a graph G are ϵ-regular, then intuitively the bipartite graph between those subsets behaves random-like with error ϵ. One interpretation of random-like behavior is that the number of instances of "small patterns" should be roughly equal to the count we would see in a random graph with the same edge density. Often, these patterns correspond to fixed subgraphs, such as triangles.

If a graph G with subsets of vertices X, Y, Z is random-like, we would expect that the number of triples $(x, y, z) \in X \times Y \times Z$ such that x, y, z form a triangle in G is roughly

Note that the sets X, Y, Z are not necessarily disjoint.

$$d(X,Y)d(X,Z)d(Y,Z) \cdot |X||Y||Z|. \tag{3.1}$$

The triangle counting lemma makes this intuition precise.

Theorem 3.13 (Triangle counting lemma)**.** *Let G be a graph and X, Y, Z be subsets of the vertices of G such that (X, Y), (Y, Z), (Z, X) are all ϵ-regular pairs some $\epsilon > 0$. Let d_{XY}, d_{XZ}, d_{YZ} denote the edge densities $d(X, Y), d(X, Z), d(Y, Z)$ respectively. If $d_{XY}, d_{XZ}, d_{YZ} \geq 2\epsilon$, then the number of triples $(x, y, z) \in X \times Y \times Z$ such that x, y, z form a triangle in G is at least*

$$(1 - 2\epsilon)(d_{XY} - \epsilon)(d_{XZ} - \epsilon)(d_{YZ} - \epsilon) \cdot |X||Y||Z|.$$

Remark 3.14. The lower bound given in the theorem for the number of triples in $X \times Y \times Z$ that are triangles is similar to the expression in (3.1), except that we have introduced additional error terms that depend on ϵ, since the graph is not perfectly random.

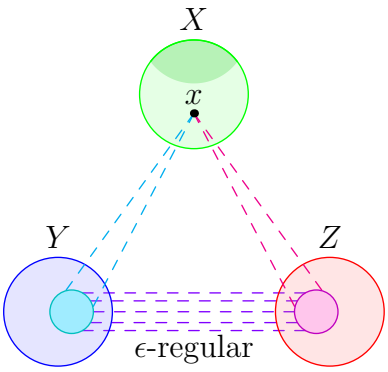
Proof. By assumption, (X, Y) is an ϵ-regular pair. This implies that fewer than $\epsilon|X|$ of the vertices in X have fewer than $(d_{XY} - \epsilon)|Y|$ neighbors in Y. If this were not the case, then we could take Y together with the subset consisting of all vertices in X that have fewer than $(d_{XY} - \epsilon)|Y|$ neighbors in Y and obtain a pair of subsets witnessing the irregularity of (X, Y), which would contradict our assumption. Intuitively these bounds make sense, since if the edges between X and Y were random-like we would expect most vertices in X to have about $d_{XY}|Y|$ neighbors in Y, meaning that not too many vertices in X can have very small degree in Y.

Applying the same argument to the ϵ-regular pair (X, Z) proves the analogous result that fewer than $\epsilon|X|$ of the vertices in X have fewer than $(d_{XZ} - \epsilon)|Z|$ neighbors in Z. Combining these two results, we see that we can find a subset X' of X of size at least $(1 - 2\epsilon)|X|$ such that every vertex $x \in X'$ is adjacent to at least $(d_{XY} - \epsilon)|Y|$ of the elements in Y and $(d_{XZ} - \epsilon)|Z|$ of the elements in Z. Using the hypothesis that $d_{XY}, d_{XZ} \geq 2\epsilon$ and the fact that (Y, Z) is ϵ-regular, we see that for any $x \in X'$, the edge density between the neighborhoods of x in Y and Z is at least $(d_{YZ} - \epsilon)$.

Now, for each vertex $x \in X'$, of which there are at least $(1 - 2\epsilon)|X|$, and choice of edge between the neighborhoods of x in Y and x in Z, of which there are at least $(d_{XY} - \epsilon)(d_{XZ} - \epsilon)(d_{YZ} - \epsilon)|Y||Z|$, we get a unique (X, Y, Z)-triangle in G. It follows that the number of such triangles is at least

$$(1 - 2\epsilon)(d_{XY} - \epsilon)(d_{XZ} - \epsilon)(d_{YZ} - \epsilon) \cdot |X||Y||Z|$$

as claimed. □



For all but a 2ϵ fraction of the $x \in X$, we can get large neighborhoods that yield many (X, Y, Z)-triangles.

Our next step is to use Theorem 3.13 to prove the triangle removal lemma, which states that a graph with few triangles can be made triangle-free by removing a small number of edges. Here, "few" and

"small" refer to a subcubic number of triangles and a subquadratic number of edges respectively.

Theorem 3.15 (Triangle removal lemma). *For all $\epsilon > 0$, there exists $\delta > 0$ such that any graph on n vertices with less than or equal to δn^3 triangles can be made triangle-free by removing at most ϵn^2 edges.*

Ruzsa and Szemerédi (1976)

Remark 3.16. An equivalent, but lazier, way to state the triangle removal lemma would be to say that

> Any graph on n vertices with $o(n^3)$ triangles can be made triangle-free by removing $o(n^2)$ edges.

This statement is a useful way to think about Theorem 3.15, but is a bit opaque due to the use of asymptotic notation. One way to interpret the statement that it asserts

> For any function $f(n) = o(n^3)$, there exists a function $g(n) = o(n^2)$ such that whenever a graph on n vertices has less than or equal to $f(n)$ triangles, we can remove at most $g(n)$ edges to make the graph triangle-free.

Another way to formalize the initial statement is to view it as a result about sequences of graphs, which claims

> Given a sequence of graphs $\{G_n\}$ with the property that for every natural n the graph G_n has n vertices and $o(n^3)$ triangles, we can make all of the graphs in the sequence triangle-free by removing $o(n^2)$ edges from each graph G_n.

It is a worthwhile exercise to verify that all of these versions of the triangle removal lemma are really the same.

The proof of Theorem 3.15 invokes the Szemerédi regularity lemma, and works as a nice demonstration of how to apply the regularity lemma in general. Our recipe for employing the regularity lemma proceeds in three steps.

1. **Partition** the vertices of a graph by applying Theorem 3.5 to obtain an ϵ-regular partition for some $\epsilon > 0$.

2. **Clean** the graph by removing edges that behave poorly with the structure imposed by the regularity lemma. Specifically, remove edges between irregular pairs, pairs with low edge density, and pairs where one of the parts is small. By design, the total number of edges removed in this step is small.

3. **Count** the number of instances of a specific pattern in the cleaned graph, and apply a counting lemma (e.g. Theorem 3.13 when the pattern is triangles) to find many patterns.

We prove the triangle removal lemma using this procedure. We first **partition** the vertices into a regular partition and then **clean** up the partition by following the recipe and removing various edges. We then show that this edge removal process eliminates all the triangles in the graph, which establishes the desired result. This last step is a proof by contradiction that uses the triangle counting lemma to show that if the graph still has triangles after the cleanup stage, the total **count** of triangles must have been large to begin with.

Proof of Theorem 3.15. Suppose we are given a graph on n vertices with fewer than δn^3 triangles, for some parameter δ we will choose later. Begin by taking an $\epsilon/4$-regular partition of the graph with parts $V_1, V_2, \cdots, V_M$. Next, for each ordered pair of parts (V_i, V_j), remove all edges between V_i and V_j if

(a) (V_i, V_j) is an irregular pair,

(b) the density $d(V_i, V_j)$ is less than $\epsilon/2$, or

(c) either V_i or V_j has at most $(\epsilon/4M)n$ vertices (is "small").

How many edges are removed in this process? Well, since we took an $\epsilon/4$-regular partition, by definition

$$\sum_{\substack{i,j \\ (V_i,V_j) \text{ not } (\epsilon/4)\text{-regular}}} |V_i||V_j| \leq \frac{\epsilon}{4} n^2.$$

so at most $(\epsilon/4)n^2$ edges are removed between irregular pairs in **(a)**. The number of edges removed from low-density pairs in **(b)** is

$$\sum_{\substack{i,j \\ d(V_i,V_j)<\epsilon/2}} d(V_i, V_j)|V_i||V_j| \leq \frac{\epsilon}{2} \sum_{i,j} |V_i||V_j| = \frac{\epsilon}{2} n^2$$

where the intermediate sum is taken over all ordered pairs of parts. The number of edges removed between small parts in **(c)** is at most

$$n \cdot \frac{\epsilon}{4M} n \cdot M = \frac{\epsilon}{4} n^2$$

since each of the n vertices is adjacent to at most $(\epsilon/4M)n$ vertices in each small part, and there are at most M small parts.

As expected, cleaning up the graph by removing edges between badly behaving parts does not remove too many edges. We claim that after this process, for some choice of δ, the graph is triangle-free. The removal lemma follows from this claim, since the previous step removed less than ϵn^2 edges from the graph.

Indeed, suppose that after following the above procedure and (possibly) removing some edges the resulting graph still has some triangle. Then we can find parts V_i, V_j, V_k (not necessarily distinct) containing each of the vertices of this triangle. Because edges between

the pairs described in **(a)** and **(b)** were removed, V_i, V_j, V_k satisfy the hypotheses of the triangle counting lemma. Applying Theorem 3.13 to this triple of subsets implies that the graph still has at least

$$\left(1-\frac{\epsilon}{2}\right)\left(\frac{\epsilon}{4}\right)^3 \cdot |V_i||V_j||V_k|$$

such triangles. By **(c)** each of these parts has size at least $(\epsilon/4M)n$, so in fact the number of (V_i, V_j, V_k)-triangles after removal is at least

$$\left(1-\frac{\epsilon}{2}\right)\left(\frac{\epsilon}{4}\right)^3\left(\frac{\epsilon}{4M}\right)^3 \cdot n^3.$$

Then by choosing positive

$$\delta < \frac{1}{6}\left(1-\frac{\epsilon}{2}\right)\left(\frac{\epsilon}{4}\right)^3\left(\frac{\epsilon}{4M}\right)^3$$

we obtain a contradiction, since the original graph has less than δn^3 triangles by assumption, but the triangle counting lemma shows that we have strictly more than this many triangles after removing some edges in the graph. The factor of $1/6$ is included here to deal with overcounting that may occur (e.g. when $V_i = V_j = V_k$). Since δ only depends on ϵ and the constant M from Theorem 3.5, this completes our proof. □

Remark 3.17. In the proof presented above, δ depends on M, the constant from Theorem 3.5. As noted in Theorem 3.12, the constant M can grow quite quickly. In particular, our proof only shows that we can pick δ so that $1/\delta$ is bounded below by a tower of twos of height $\epsilon^{-O(1)}$. It turns out that as long as we pick δ such that $1/\delta$ is bounded below by a tower of twos with height $O(\log(1/\epsilon))$, the statement of the triangle removal lemma holds. In contrast, the best known "lower bound" result in this context is that if δ satisfies the conditions of Theorem 3.15, then $1/\delta$ is bounded above by $\epsilon^{-O(\log(1/\epsilon))}$ (this bound will follow from the construction of 3-AP-free sets that we will discuss soon). The separation between these upper and lower bounds is large, and closing this gap is a major open problem in graph theory.

Fox (2012)

Historically, a major motivation for proving Theorem 3.15 was the lemma's connection with Roth's theorem. This connection comes from looking at a special type of graph, mentioned previously in Question 1.15. The following corollary of the triangle removal lemma is helpful in investigating such graphs.

Corollary 3.18. *Suppose G is a graph on n vertices such that every edge of G lies in a unique triangle. Then G has $o(n^2)$ edges.*

Proof. Let G have m edges. Because each edge lies in one triangle, the number of triangles in G is $m/3$. Since $m < n^2$, this means that G has $o(n^3)$ triangles. By Remark 3.16, we can remove $o(n^2)$ edges to make G triangle-free. However, deleting an edge removes at most one triangle from the graph by assumption, so the number of edges removed in this process is at least $m/3$. It follows that m is $o(n^2)$ as claimed. □

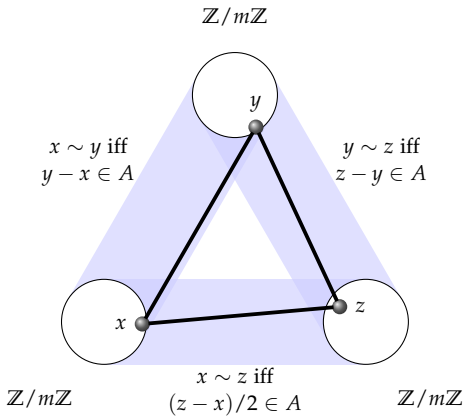
3.3 *Roth's theorem*

Theorem 3.19 (Roth's theorem)**.** *Every subset of the integers with positive upper density contains a 3-term arithmetic progression.*

Roth (1953)

Proof. Take a subset A of $[N]$ that has no 3-term arithmetic progressions. We will show that A has $o(N)$ elements, which will prove the theorem. To make our lives easier and avoid dealing with edge cases involving large elements in A, we will embed A into a cyclic group. Take $M = 2N + 1$ and view $A \subseteq \mathbb{Z}/M\mathbb{Z}$. Since we picked M large enough so that the sum of any two elements in A is less than M, no wraparound occurs and A has no 3-term arithmetic progressions (with respect to addition modulo M) in $\mathbb{Z}/M\mathbb{Z}$.

Now, we construct a tripartite graph G whose parts X, Y, Z are all copies of $\mathbb{Z}/M\mathbb{Z}$. Connect a vertex $x \in X$ to a vertex $y \in Y$ if $y - x \in A$. Similarly, connect $z \in Z$ with $y \in Y$ if $z - y \in A$. Finally, connect $x \in X$ with $z \in Z$ if $(z - x)/2 \in A$. Because we picked M to be odd, 2 is invertible modulo M and this last step makes sense.



This construction is set up so that if x, y, z form a triangle, then we get elements

$$y - x, \frac{z - x}{2}, z - y$$

that all belong to A. These numbers form an arithmetic progression in the listed order. The assumption on A then tells us this progression must be trivial: the elements listed above are all equal. But this condition is equivalent to the assertion that x, y, z is an arithmetic progression in $\mathbb{Z}/M\mathbb{Z}$.

Consequently, every edge of G lies in exactly one triangle. This is because given an edge (i.e. two elements of $\mathbb{Z}/M\mathbb{Z}$), there is a a unique way to extend that edge to a triangle (add another element of the group to form an arithmetic progression in the correct order).

Then Corollary 3.18 implies that G has $o(M^2)$ edges. But by construction G has precisely $3M|A|$ edges. Since $M = 2N + 1$, it follows that $|A|$ is $o(N)$ as claimed. □

Later in the book we discuss a Fourier-analytic proof of Roth's theorem which, although it uses different methods, has similar themes

to the above proof.

If we pay attention to the bounds implied by the triangle removal lemma, our proof here yields an upper bound of $N/(\log^* N)^c$ for $|A|$, where $\log^* N$ denotes the number of times the logarithm must be applied to N to make it less than 1 and c is some constant. This is the inverse of the tower of twos function we have previously seen. The current best upper bound on A asserts that if A has no 3-term arithmetic progressions, then

The $\log^*$ function grows incredibly slowly. It is sometimes said that although $\log^* n$ tends to infinity, it has "never been observed to do so."

Sanders (2011)
Bloom (2016)

$$|A| \leq \frac{N}{(\log N)^{1-o(1)}}.$$

In the next section, we will prove a lower bound on the size of the large subset of $[N]$ without any 3-term arithmetic progressions. It turns out that there exist $A \subseteq [N]$ with size $N^{1-o(1)}$ that contains no 3-term arithmetic progression. Actually, we will provide an example where $|A| \geq Ne^{-C\sqrt{\log N}}$ for some constant C.

Remark 3.20. Beyond the result presented in Corollary 3.18, not much is known about the answer to Question 1.15. In the proof of Roth's theorem we showed that, given any subset A of $[N]$ with no 3-term arithmetic progressions, we can construct a graph on $O(N)$ vertices that has on the order of $N|A|$ edges such that each of its edges is contained in a unique triangle. This is more or less the only known way to construct relatively dense graphs with the property that each edge is contained in a unique triangle.

3.4 *Constructing sets without 3-term arithmetic progressions*

One way to construct a subset $A \subseteq [N]$ free of 3-term arithmetic progressions is to greedily construct a sub-sequence of the natural numbers with such property. This would produce the following sequence, which is known as a Stanley sequence:

$$0 \quad 1 \quad 3 \quad 4 \quad 9 \quad 10 \quad 12 \quad 13 \quad 27 \quad 28 \quad 30 \quad 31 \quad \cdots$$

Observe that this sequence consist of all natural numbers whose ternary representations have only the digits 0 and 1. Up to $N = 3^k$, the subset $A \subseteq [N]$ so constructed has size $|A| = 2^k = N^{\log_3 2}$. For quite some time, people thought this example was close to the optimal. But in the 1940s, Salem and Spencer found a much better construction. Their proof was later simplified and improved by Behrend, whose version we present below. Surprisingly, this lower bound has hardly been improved since the 40s.

Indeed, given any three distinct numbers a, b, c whose ternary representations do not contain the digit 2, we can add up the ternary representations of any two numbers digit by digit without having any "carryover". Then, each digit in the ternary representation of $2b = b + b$ is either 0 or 2, whilst the ternary representation of $a + c$ would have the digit 1 appearing in those positions at which a and c differ. Hence, $a + c \neq 2b$, or in other words, $b - a \neq c - b$.

Salem and Spencer (1942)
Behrend (1946)

Theorem 3.21. *There exists a constant $C > 0$ such that for every positive integer N, there exists a subset $A \subseteq [N]$ with size $|A| \geqslant Ne^{-C\sqrt{\log N}}$ that contains no 3-term arithmetic progression.*

Proof. Let m and d be two positive integers depending on N to be specified later. Consider the box of lattice points in d dimensions $X := [m]^d$, and its intersections with spheres of radius $\sqrt{L}$ $(L \in \mathbb{N})$

$$X_L := \left\{(x_1, \ldots, x_d) \in X : x_1^2 + \cdots + x_d^2 = L\right\}.$$

Set $M := dm^2$. Then, $X = X_1 \sqcup \cdots \sqcup X_M$, and by the pigeonhole principle, there exists an $L_0 \in [M]$ such that $|X_{L_0}| \geqslant m^d/M$. Consider the base $2m$ expansion $\varphi : X \to \mathbb{N}$ defined by

$$\varphi(x_1, \ldots, x_d) := \sum_{i=1}^{d} x_i (2m)^{i-1}.$$

Clearly, φ is injective. Moreover, since each entry of $(x_1, \ldots, x_d)$ is in $[m]$, any three distinct $\vec{x}, \vec{y}, \vec{z} \in X$ are mapped to a three-term arithmetic progression in $\mathbb{N}$ if and only if $\vec{x}, \vec{y}, \vec{z}$ form a three-term arithmetic progression in X. Being a subset of a sphere, the set X_{L_0} is free of three-term arithmetic progressions. Then, the image $\varphi\left(X_{L_0}\right)$ is also free of three-term arithmetic progressions. Therefore, taking $m = \frac{1}{2}\left\lfloor e^{\sqrt{\log N}} \right\rfloor$ and $d = \left\lfloor \sqrt{\log N} \right\rfloor$ we find a subset of $[N]$, namely $A = \varphi\left(X_{L_0}\right)$, which contains no three-term arithmetic progression and has size

$$|A| = \left|X_{L_0}\right| \geqslant \frac{m^d}{dm^2} \geqslant N e^{-C\sqrt{\log N}},$$

where C is some absolute constant. □

Next, let's study some variations of Roth's theorem. We will start with a higher dimensional version of Roth's theorem, which is a special case of the multidimensional Szemerédi theorem mentioned back in Chapter 1.

Definition 3.22. A ***corner*** in $\mathbb{Z}^2$ is a three-element set of the form $\{(x, y), (x + d, y), (x, y + d)\}$ with $d > 0$.

Theorem 3.23. *If a subset $A \subseteq [N]^2$ is free of corners, then $|A| = o\left(N^2\right)$.* Ajtai and Szemerédi (1975)

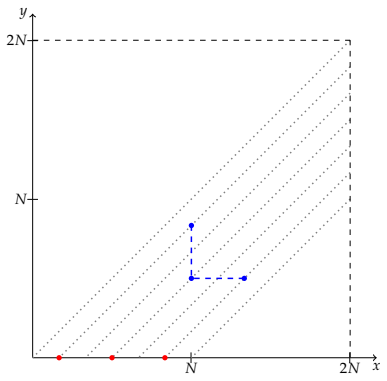
Proof. Consider the sum set $A + A \subseteq [2N]^2$. By the pigeonhole principle, there exists a point $z \in [2N]^2$ such that there are at least $\frac{|A|^2}{(2N)^2}$ pairs of $(a, b) \in A \times A$ satisfying $a + b = z$. Put $A' = A \cap (z - A)$. Then, the size of A' is exactly the number of ways to write z as a sum of two elements of A. So, $|A'| \geqslant \frac{|A|^2}{(2N)^2}$, and it suffices to show that $|A'| = o\left(N^2\right)$. The set A' is free of corners because A is. Moreover, since $A' = z - A'$, no 3-subset of A' is of the form $\{(x, y), (x + d, y), (x, y + d)\}$ with $d \neq 0$. Solymosi (2003)

Now, build a tripartite graph G with parts $X = \{x_1, \ldots, x_N\}$, $Y = \{y_1, \ldots, y_N\}$ and $Z = \{z_1, \ldots, z_{2N}\}$, where each vertex x_i corresponds to a vertical line $\{x = i\} \subseteq \mathbb{Z}^2$, each vertex y_j corresponds to a

horizontal line $\{y = j\}$, and each vertex z_k corresponds to a slanted line $\{y = -x + k\}$ with slope -1. Join two distinct vertices of G with an edge if and only if the corresponding lines intersect at a point belonging to A'. Then, each triangle in the graph G corresponds to a set of three lines such that each pair of lines meet at a point of A'. Since A' has no corners with $d \neq 0$, three vertices x_i, y_j, z_k induces a triangle in G if and only if the three corresponding lines pass through the same point of A' and form a trivial corner with $d = 0$. Since there are exactly one vertical line, one horizontal line and one line with slope -1 passing through each point of A', it follows that each edge of G belongs to exactly one triangle. Thus, by Corollary 3.18,

$$3|A'| = e(G) = o\left(N^2\right). \qquad \square$$

Note that we can deduce Roth's theorem from the corners theorem in the following way.



Corollary 3.24. *Let $r_3(N)$ be the size of the largest subset of $[N]$ which contains no 3-term arithmetic progression, and $r_\llcorner(N)$ be the size of the largest subset of $[N]^2$ which contains no corner. Then, $r_3(N)N \leqslant r_\llcorner(2N)$.*

Proof. Given any set $A \subseteq [N]$, define a set

$$B := \left\{(x, y) \in [2N]^2 : x - y \in A\right\}.$$

Because for each $a \in [N]$ there are at least n pairs of $(x, y) \in [2N]^2$ such that $x - y = a$, we have that $|B| \geqslant N|A|$. In addition, since each corner $\{(x, y), (x + d, y), (x, y + d)\}$ in B would be projected onto a 3-term arithmetic progression $\{x - y - d, x - y, x - y + d\}$ in A via $(x, y) \overset{\pi}{\mapsto} x - y$, if A is free of 3-term arithmetic progressions, then B is free of corners. Thus, $r_3(N)N \leqslant r_\llcorner(2N)$. $\square$

So, any upper bound on corner-free sets will induce an upper bound on 3-AP-free sets, and any lower bound on 3-AP-free sets will induce a lower bound on corner-free sets. In particular, Behrend's construction of 3-AP-free sets easily extends to the construction of large corner-free sets. The best upper bound on the size of corner-free subsets of $[N]^2$ that we currently have is $N^2(\log\log N)^{-C}$, with $C > 0$ an absolute constant, which was proven by Shkredov using Fourier analytic methods.

Shkredov (2006)

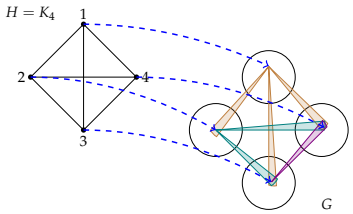
3.5 *Graph embedding, counting and removal lemmas*

As seen in the proof of the triangle removal lemma Theorem 3.15, one key stepping stone to removal lemmas are counting lemmas. Thus, we would like to generalize the triangle counting lemma to

general graphs. To reach our goal, we have two strategies: one is to embed the vertices of a fixed graph one by one in a way that the yet-to-be embedded vertices have lots of choices left, and the other is to analytically remove one edge at a time.

Theorem 3.25 (Graph embedding lemma)**.** *Let H be an r-partite graph with vertices of degree no more than Δ. Let G be a graph, and $V_1, \dots, V_r \subseteq V(G)$ be vertex sets of size at least $\frac{1}{\epsilon}v(H)$. If every pair (V_i, V_j) is ϵ-regular and has density $d(V_i, V_j) \geqslant 2\epsilon^{1/\Delta}$. Then, G contains a copy of H.*

Remark 3.26. The vertex sets $V_1, \dots, V_r$ in the theorem need not be disjoint or even distinct.



Let us illustrate some ideas of the proof and omit the details. The proof of Theorem 3.25 is an extension of the proof the proof of Theorem 3.13 for counting triangles.

Suppose that we trying to embed $H = K_4$, where each vertex of the K_4 goes into its own part, where the four parts are pairwise ϵ-regular with edge density not too small. Let us embed the vertices sequentially. The choice of the first vertex limits the choices for the sequences vertices. Most choices of the first vertex will not reduce the possibilities for the remaining vertices by a factor much more than what one should expect based on the edge densities. One the first vertex has been embedded, we move on the second vertex, and again, choose an embedding so that lots of choices remain for the third and fourth vertices, and so on.

Next, let's use our second strategy to prove a counting lemma.

Theorem 3.27 (Graph counting lemma)**.** *Let H be a graph with $V(H) = [k]$, and let $\epsilon > 0$. Let G be an n-vertex graph with vertex subsets $V_1, \dots, V_k \subseteq V(G)$ such that (V_i, V_j) is ϵ-regular whenever $\{i, j\} \in E(H)$. Then, the number of tuples $(v_1, \dots, v_k) \in V_1 \times \cdots \times, V_k$ such that $\{v_i, v_j\} \in E(G)$ whenever $\{i, j\} \in E(H)$ is within $e(H)\epsilon|V_1| \cdots |V_k|$ of*

$$\left(\prod_{\{i,j\} \in E(H)} d(V_i, V_j) \right) \left(\prod_{i=1}^{k} |V_i| \right).$$

Remark 3.28. The theorem can be rephrased into the following probabilistic form: Choose $v_1 \in V_1, \dots, v_k \in V_k$ uniformly and independently at random. Then,

$$\left| \mathbb{P}\left(\{v_i, v_j\} \in E(G) \text{ for all } \{i, j\} \in E(H) \right) - \prod_{\{i,j\} \in E(H)} d(V_i, V_j) \right| \leqslant e(H)\,\epsilon. \tag{3.2}$$

Proof. After relabelling if necessary, we may assume that $\{1, 2\}$ is an edge of H. To simplify notation, set

$$P = \mathbb{P}\left(\{v_i, v_j\} \in E(G) \text{ for all } \{i, j\} \in E(H) \right).$$

We will show that

$$\left|P - d(V_1, V_2)\mathbb{P}\left(\{v_i, v_j\} \in E(G) \text{ for all } \{i,j\} \in E(H) \setminus \{\{1,2\}\}\right)\right| \leqslant \epsilon \tag{3.3}$$

Couple the two random processes of choosing v_i's. It suffices to show that (3.3) holds when $v_3, \ldots, v_k$ are fixed arbitrarily and only v_1 and v_2 are random. Define

$$A_1 := \left\{v_1 \in V_1 : \{v_1, v_i\} \in E(G) \text{ whenever } i \in N_H(1) \setminus \{2\}\right\},$$
$$A_2 := \left\{v_2 \in V_2 : \{v_2, v_i\} \in E(G) \text{ whenever } i \in N_H(2) \setminus \{1\}\right\}.$$

If $|A_1| \leqslant \epsilon|V_1|$ or $|A_2| \leqslant \epsilon|V_2|$, then

$$\frac{e(A_1, A_2)}{|V_1||V_2|} \leqslant \frac{|A_1||A_2|}{|V_1||V_2|} \leqslant \epsilon$$

and

$$d(V_1, V_2)\frac{|A_1||A_2|}{|V_1||V_2|} \leqslant d(V_1, V_2)\frac{|A_1||A_2|}{|V_1||V_2|} \leqslant \epsilon,$$

so we have

$$\left|\frac{e(A_1, A_2)}{|V_1||V_2|} - d(V_1, V_2)\frac{|A_1||A_2|}{|V_1||V_2|}\right| \leqslant \epsilon.$$

Else if $|A_1| > \epsilon|V_1|$ and $|A_2| > \epsilon|V_2|$, then by the ϵ-regularity of (V_1, V_2), we also have

$$\begin{aligned}&\left|\frac{e(A_1, A_2)}{|V_1||V_2|} - d(V_1, V_2)\frac{|A_1||A_2|}{|V_1||V_2|}\right| \\ &\quad = \left|\frac{e(A_1, A_2)}{|A_1||A_2|} - d(V_1, V_2)\right| \cdot \frac{|A_1||A_2|}{|V_1||V_2|} < \epsilon.\end{aligned}$$

So, in either case, (3.3) holds when $v_3, \ldots, v_k$ are viewed as fixed vertices in $V_3, \ldots, V_k$, respectively.

To complete the proof of the counting lemma, do induction on $e(H)$. Let H' denote the graph obtained by removing the edge $\{1,2\}$ from H, and assume that (3.2) holds when H is replaced by H' throughout. Then,

$$\begin{aligned}&\left|P - \prod_{\{i,j\} \in E(H)} d(V_i, V_j)\right| \\ &\leqslant d(V_1, V_2)\left|\mathbb{P}\left(\{v_i, v_j\} \in E(G) \text{ for all } \{i,j\} \in E(H')\right) - \prod_{\{i,j\} \in E(H')} d(V_i, V_j)\right| \\ &\quad + \left|P - d(V_1, V_2)\mathbb{P}\left(\{v_i, v_j\} \in E(G) \text{ for all } \{i,j\} \in E(H')\right)\right| \\ &\leqslant d(V_1, V_2)e(H')\,\epsilon + \epsilon \\ &\leqslant \left(e(H') + 1\right)\epsilon = e(H)\,\epsilon.\end{aligned}$$

□

Theorem 3.29 (Graph removal lemma)**.** *For each graph H and each constant $\epsilon > 0$, there exists a constant $\delta > 0$ such that every n-vertex graph G with fewer than $\delta n^{v(H)}$ copies of H can be made H-free by removing no more than ϵn^2 edges.*

To prove the graph removal lemma, we adopt the proof of Theorem 3.15 as follows:

Partition the vertex set using the graph regularity lemma.

Remove all edges that belong to low-density or irregular pairs or are adjacent to small vertex sets.

Count the number of remaining edges, and show that if the resulting graph still contains any copy of H, then it would contains lots of copies of H, which would be a contradiction.

We are now ready to prove Theorem 2.13 which we recall below.

Theorem 3.30 (Erdős–Stone–Simonovits)**.** *For every fixed graph H, we have*

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

Proof. Fix a constant $\epsilon > 0$. Let $r + 1$ denote the chromatic number of H, and G be any n-vertex graph with at least $\left(1 - \frac{1}{r} + \epsilon\right) \frac{n^2}{2}$ edges. We claim that if $n = n(\epsilon, H)$ is sufficiently large, then G contains a copy of H.

Let $V(G) = V_1 \sqcup \cdots \sqcup V_m$ be an η-regular partition of the vertex set of G, where $\eta := \frac{1}{2e(H)} \left(\frac{\epsilon}{8}\right)^{e(H)}$. Remove an edge $(x, y) \in V_i \times V_j$ if

(a) (V_i, V_j) is not η-regular, or

(b) $d(V_i, V_j) < \frac{\epsilon}{8}$, or

(c) $|V_i|$ or $|V_j|$ is less than $\frac{\epsilon}{8m} n$.

Then, the number of edges that fall into case **(a)** is no more than ηn^2, the number of edges that fall into case **(b)** is no more than $\frac{\epsilon}{8} n^2$, and the number of edges that fall into case **(c)** is no more than $mn \frac{\epsilon}{8m} n = \frac{\epsilon}{8} n^2$. Thus, the total number of edges removed is no more than $\eta n^2 + \frac{\epsilon}{8} n^2 + \frac{\epsilon}{8} n^2 \leqslant \frac{3\epsilon}{8} n^2$. Therefore, the resulting graph G' has at least $\left(1 - \frac{1}{r} + \frac{\epsilon}{4}\right) \frac{n^2}{2}$ edges. So, by Turán's theorem, we know that G' contains a copy of K_{r+1}. Let's label the vertices of this copy of K_{r+1} with the numbers $1, 2, \ldots, r + 1$. Suppose the vertices of K_{r+1} lie in $V_{i_1}, \cdots, V_{i_{r+1}}$, respectively, with the indices $i_1, \ldots, i_{r+1}$ possibly repeated. Then, every pair (V_{i_r}, V_{i_s}) is η-regular. Since $\chi(H) = r + 1$, there exists a proper coloring $c : V(H) = [k] \to [r + 1]$. Set $\tilde{V}_j := V_{c(j)}$ for each $j \in [k]$. Then, we can apply the graph counting lemma

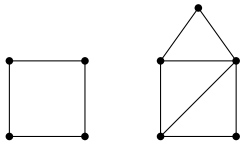
Theorem 3.27 to $\{\tilde{V}_j : j \in [k]\}$, and find that the number of graph homomorphisms from H to G' is at least

$$\left(\prod_{\{i,j\}\in E(H)} d(\tilde{V}_i, \tilde{V}_j)\right)\left(\prod_{i=1}^{k}|\tilde{V}_i|\right) - e(H)\eta\left(\prod_{i=1}^{k}|\tilde{V}_i|\right)$$
$$\geqslant \left(\left(\frac{\epsilon}{8}\right)^{e(H)} - e(H)\eta\right)\left(\frac{\epsilon n}{8m}\right)^{v(H)}.$$

Given that the there are only $O_H(n^{v(H)-1})$ non-injective maps $V(H) \to V(G)$, for n sufficiently large, G contains a copy of H. □

3.6 *Induced graph removal lemma*

We will now consider a different version of the graph removal lemma. Instead of copies of H, we will now consider induced copies of H. As a reminder, we say H is an ***induced subgraph*** of G if one can obtain H from G by deleting vertices of G. Accordingly, G is ***induced-H-free*** if G contains no induced subgraph isomorphic to H.



H is a subgraph but not an induced subgraph of G.

Theorem 3.31 (Induced graph removal lemma)**.** *For any graph H and constant $\epsilon > 0$, there exists a constant $\delta > 0$ such that if an n-vertex graph has fewer than $\delta n^{v(H)}$ copies of H, then it can be made induced H-free by adding and/or deleting fewer than ϵn^2 edges.*

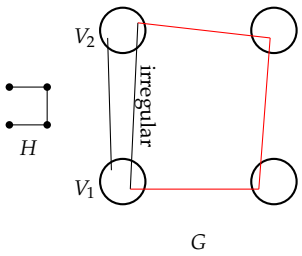
Alon, Fischer, Krivelevich, and Szegedy (2000)

The number of edges added and/or deleted is also known as the ***edit distance***. The analogous statement where we are only allowed to delete edges would be false. For a sequence of graphs giving a counterexample, let H be the 3-vertex graph with no edges and G_n be the complete graph on n vertices with a triangle missing.

Let us first attempt to apply the proof strategy from the proof of the graph removal lemma (Theorem 3.29).

Partition. Pick a regular partition of the vertex set using Szemerédi's regularity lemma.

Clean. Remove all edges between low density pairs (density less than ϵ), and add all edges between high density pairs (density more than $1-\epsilon$). However, it is not clear what to do with irregular pairs. Earlier, we just removed all edges between irregular pairs. The problem is that this may create many induced copies of H that were not present previously (note that this is not true for usual subgraphs), and in this case we would have no hope of showing that there are no (or only a few) copies of H left in the **counting** step. The same is true if we were to add all edges between irregular pairs.



Removing all edges between the irregular pair (V_1, V_2) would create induced copies of H.

This prompts the question whether there is a way to partition which guarantees that there are no irregular pairs. The answer is no, as can be seen in the case of the half-graph H_n, which is the bipartite graph on vertices $\{a_1, \ldots, a_n, b_1, \ldots, b_n\}$ with edges $\{a_i b_j : i \leq j\}$. Our strategy will be to instead prove that there is another good way of partitioning, i.e., another regularity lemma. Let us first note that the induced graph removal lemma is a special case of the following theorem.

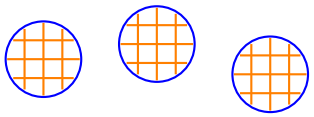
Theorem 3.32 (Colorful graph removal lemma)**.** *For all positive integers k, r, and constant $\epsilon > 0$, there exists a constant $\delta > 0$ so that if $\mathcal{H}$ is a set of r-edge-colorings of K_k, then every r-edge coloring of K_n with less than a δ fraction of its k-vertex subgraphs belonging to $\mathcal{H}$ can be made $\mathcal{H}$-free by recoloring (using the same r colors) a smaller than ϵ fraction of the edges.*

Note that the induced graph removal lemma is the special case with $r = 2$ and the blue-red colorings of K_k being those in which the graph formed by the blue edges is isomorphic to H (and the graph formed by the red edges is its complement). We will not prove the colorful graph removal lemma. However, we will prove the induced graph removal lemma, and there is an analogous proof of the colorful graph removal lemma.

To prove the induced graph removal lemma, we will rely on a new regularity lemma. Recall that for a partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of $V(G)$ with $n = |V(G)|$, we defined the energy

$$q(\mathcal{P}) = \sum_{i,j=1}^{n} \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2.$$

In the proof of Szemerédi's regularity lemma (Theorem 3.5), we used an energy increment argument, namely that if $\mathcal{P}$ is not ϵ-regular, then there exists a refinement $\mathcal{Q}$ of $\mathcal{P}$ so that $|\mathcal{Q}| \leq |\mathcal{P}|2^{|\mathcal{P}|}$ and $q(\mathcal{Q}) \geq q(\mathcal{P}) + \epsilon^5$. The new regularity lemma is the following.



The partition $\mathcal{Q}$ in orange refines the partition $\mathcal{P}$ in blue.
Alon, Fischer, Krivelevich, and Szegedy (2000)

Theorem 3.33 (Strong regularity lemma)**.** *For all sequences of constants $\epsilon_0 \geq \epsilon_1 \geq \epsilon_2 \dots > 0$, there exists an integer M so that every graph has two vertex partitions $\mathcal{P}, \mathcal{Q}$ so that $\mathcal{Q}$ refines $\mathcal{P}$, $|\mathcal{Q}| \leq M$, $\mathcal{P}$ is ϵ_0-regular, $\mathcal{Q}$ is $\epsilon_{|\mathcal{P}|}$-regular, and $q(\mathcal{Q}) \leq q(\mathcal{P}) + \epsilon_0$.*

For a refinement $\mathcal{Q}$ of a partition $\mathcal{P}$, we say $\mathcal{Q}$ is ***extremely regular*** if it is $\epsilon_{|\mathcal{P}|}$-regular. Theorem 3.33 says that there exists a partition with an extremely regular refinement.

Proof. We repeatedly apply the following version of Szemerédi's regularity lemma (Theorem 3.5):

> For all $\epsilon > 0$, there exists an integer $M_0 = M_0(\epsilon)$ so that for all partitions $\mathcal{P}$ of $V(G)$, there exists a refinement $\mathcal{P}'$ of $\mathcal{P}$ with each part in $\mathcal{P}$ refined into $\leq M_0$ parts so that $\mathcal{P}'$ is ϵ-regular.

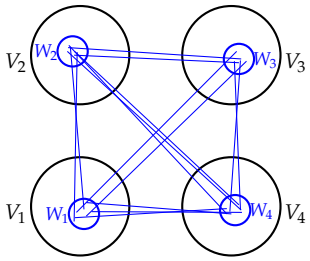
The above version has the same proof as the proof we gave for Theorem 3.5, except instead of starting from the trivial partition, we start from the partition $\mathcal{P}$.

By iteratively applying the above lemma, we obtain a sequence of partitions $\mathcal{P}_0, \mathcal{P}_1, \dots$ of $V(G)$ starting with $\mathcal{P}_0$ being a trivial partition so that each $\mathcal{P}_{i+1}$ refines $\mathcal{P}_i$, $\mathcal{P}_{i+1}$ is $\epsilon_{|\mathcal{P}_i|}$-regular, and $|\mathcal{P}_{i+1}| \leq |\mathcal{P}_i| M_0(\epsilon_{|\mathcal{P}_i|})$.

Since $0 \leq q(i) \leq 1$, there exists $i \leq \epsilon_0^{-1}$ so that $q(\mathcal{P}_{i+1}) \leq q(\mathcal{P}_i) + \epsilon_0$. Set $\mathcal{P} = \mathcal{P}_i$, $Q = \mathcal{P}_{i+1}$. Since we are iterating at most ϵ_0^{-1} times and each refinement is into a bounded number of parts (depending only on the corresponding $\epsilon_{\mathcal{P}_i}$), we have $|\mathcal{Q}| = O_{\vec{\epsilon}}(1)$. □

What bounds does this proof give on the constant M? This depends on the sequence ϵ_i. For instance, if $\epsilon_i = \frac{\epsilon}{i+1}$, then M is essentially M_0 applied in succession $\frac{1}{\epsilon}$ times. Note that M_0 is a tower function, and this makes M a tower function iterated i times. In other words, we are going one step up in the Ackermann hierarchy. This iterated tower function is called the wowzer function.

In fact, the same result can also be proved with the extra assumption that $\mathcal{P}$ and $\mathcal{Q}$ are equitable partitions, and this is the result we will assume.



A partition with regular subsets.

Corollary 3.34. *For all sequences of constants $\epsilon_0 \geq \epsilon_1 \geq \epsilon_2 \ldots > 0$, there exists a constant $\delta > 0$ so that every n-vertex graph has an equitable vertex partition $V_1, \ldots, V_k$ and $W_i \subset V_i$ so that*

(a) $|W_i| \geq \delta n$

(b) (W_i, W_j) is ϵ_k-regular for all $1 \leq i \leq j \leq k$

(c) $|d(V_i, V_j) - d(W_i, W_j)| \leq \epsilon_0$ for all but fewer than $\epsilon_0 k^2$ pairs $(i, j) \in [k]^2$.

Proof sketch. Let us first explain how to obtain a partition that almost satisfies (b). Note that without requiring (W_i, W_i) to be regular, one can obtain $W_i \subseteq V_i$ by picking a uniformly random part of $\mathcal{Q}$ inside each part of $\mathcal{P}$ in the strong regularity lemma. This follows from $\mathcal{Q}$ being extremely regular. So all (W_i, W_j) for $i \neq j$ are regular with high probability. It is possible to also make each (W_i, W_i) be regular, and this is left as an exercise to the reader.

With this construction, part (c) is a consequence of $q(\mathcal{Q}) \leq q(\mathcal{P}) + \epsilon_0$. Recall from the proof of Lemma 3.8 that the energy q is the expectation of the square of a random variable Z, namely $Z_{\mathcal{P}} = d(V_i, V_j)$ for random i, j. So $q(\mathcal{Q}) - q(\mathcal{P}) = \mathbb{E}[Z_{\mathcal{Q}}^2] - \mathbb{E}[Z_{\mathcal{P}}^2] = \mathbb{E}[(Z_{\mathcal{Q}} - Z_{\mathcal{P}})^2]$, where the last equality can be thought of as a Pythagorean identity. To prove the last equality, expand the expectation as a sum over all pairs of parts of $\mathcal{P}$. On each pair, $Z_{\mathcal{P}}$ is constant and $Z_{\mathcal{Q}}$ averages to it, so the equality follows for the pair, and also for the sum. Then, (c) follows by reinterpreting the random variables as densities.

Finally, part (a) follows from a bound on $|\mathcal{Q}|$. □

We will now prove the induced graph removal lemma using Corollary 3.34.

Proof of the induced graph removal lemma. We have the usual 3 steps.

Partition. We apply the corollary to get a partition $V_1 \cup \ldots \cup V_k$ with $W_1 \subset V_1, \ldots, W_k \subset V_k$, so that the following hold.

- (W_i, W_j) is $\frac{1}{\binom{v(H)}{2}} \left(\frac{\epsilon}{4}\right)^{\binom{v(H)}{2}}$-regular for all $i \leq j$.
- $|d(V_i, V_j) - d(W_i, W_j)| \leq \frac{\epsilon}{2}$ for all but fewer than $\frac{\epsilon k^2}{2}$ pairs $(i, j) \in [k]^2$

- $|W_i| \geq \delta_0 n$, with $\delta_0 = \delta_0(\epsilon, H) > 0$.

Clean. For all $i \leq j$ (including $i = j$):

- If $d(W_i, W_j) \leq \frac{\epsilon}{2}$, we remove all edges between (V_i, V_j).
- If $d(W_i, W_j) \geq 1 - \frac{\epsilon}{2}$, then we add all edges between (V_i, V_j).

By construction, the total number of edges added/removed from G is less than $2\epsilon n^2$.

Count. Now we are done if we show that there are no induced copies of H left. Well, suppose there is some induced H left. Let $\phi\colon V(H) \to [k]$ be the function that indexes which part V_i each vertex of this copy of H is in. In other words, the function ϕ is such that for our copy of H, the vertex $v \in V(H)$ is in the part $V_{\phi(v)}$. The goal now is to apply the counting lemma to show that there are actually many such copies of H in G where $v \in V(H)$ is mapped to a vertex in $W_{\phi(v)}$. We will make use of the following trick: instead of considering copies of H in our graph G, we modify G to get a graph G' for which a complete graph on $v(H)$ vertices with the vertices coming from the parts given by ϕ is present if and only if restricting to the same vertices in G gives rise to an induced copy of H. We construct G' in the following way. For each vertex v in our copy of H in G, we take a different copy of $V_{\phi(v)}$. Edges between two copies of the same vertex will never be present in G'. For all other pairs of vertices in G', whether there is an edge between them is determined in the following way: if uv is an edge, then the edges between $V_{\phi(v)}$ and $V_{\phi(u)}$ in G' are taken to be the same as in G. If uv is not an edge, then the edges $V_{\phi(v)}$ and $V_{\phi(u)}$ in G' are taken to be those in the complement of G.

Note that this G' indeed satisfies the desired property – if there is a complete subgraph in G' on vertices from these parts $V_{\phi(v)}$, then G has an induced copy of H at the same vertices. Now by the graph counting lemma (Theorem 3.27), the number of $K_{v(H)}$ with each vertex $u \in V(H)$ coming from $W_{\phi(u)}$ is within

$$\left(\frac{\epsilon}{4}\right)^{\binom{v(H)}{2}} \prod_{u \in V(H)} |W_{\phi(u)}|$$

of

$$\prod_{uv \in E(H)} d\left(W_{\phi(u)}, W_{\phi(v)}\right) \prod_{uv \in E\left(\overline{H}\right)} \left(1 - d\left(W_{\phi(u)}, W_{\phi(v)}\right)\right) \prod_{u \in V(H)} \left|W_{\phi(u)}\right|.$$

Hence, the number of induced H in G is also at least

$$\left(\left(\frac{\epsilon}{2}\right)^{\binom{v(H)}{2}} - \left(\frac{\epsilon}{4}\right)^{\binom{v(H)}{2}}\right) \delta_0^{v(H)} n^{v(H)}. \qquad \square$$

Note that the strong regularity lemma was useful in that it allowed us to get rid of irregular parts in a restricted sense without actually having to get rid of irregular pairs.

Theorem 3.35 (Infinite removal lemma)**.** *For each (possibly infinite) set of graphs $\mathcal{H}$ and $\epsilon > 0$, there exists h_0 and $\delta > 0$ so that every n-vertex graph with fewer than $\delta n^{v(H)}$ induced copies of H for all $H \in \mathcal{H}$ with $v(H) \leq h_0$ can be made induced-$\mathcal{H}$-free by adding or removing fewer than ϵn^2 edges.* Alon and Shapira (2008)

This theorem has a similar proof as the induced graph removal lemma, where ϵ_k from the corollary depends on k and $\mathcal{H}$.

3.7 *Property testing*

We are looking for an efficient randomized algorithm to distinguish large graphs that are triangle-free from graphs that are ϵ-far from triangle-free. We say a graph is ***ϵ-far*** from a property $\mathcal{P}$ if the minimal number of edges one needs to change (add or remove) to get to a graph that has the property $\mathcal{P}$ is greater than ϵn^2. We propose the following.

Algorithm 3.36. Sample a random triple of vertices, and check if these form a triangle. Repeat $C(\epsilon)$ times, and if no triangle is found, return that the graph is triangle-free. Else, return that the graph is ϵ-far from triangle-free.

Theorem 3.37. *For all constants $\epsilon > 0$, there exists a constant $C(\epsilon)$ so that Algorithm 3.36 outputs the correct answer with probability greater than $\frac{2}{3}$.* Alon and Shapira (2008)

Proof. If the graph G is triangle-free, the algorithm is always successful, since no sampled triple ever gives a triangle. If G is ϵ-far from triangle-free, then by the triangle removal lemma, G has at least δn^3 triangles, where $\delta = \delta(\epsilon)$ comes from the triangle removal lemma (Theorem 3.15). We set the constant number of samples to be $C(\epsilon) = \frac{1}{\delta}$. The probability that the algorithm fails is equal to the probability that we nevertheless sample no triangles, and since each sample is picked independently, this probability is $\left(1 - \frac{\delta n^3}{\binom{n}{3}}\right)^{1/\delta} \leq (1 - 6\delta)^{1/\delta} \leq e^{-6}$. □

So far, we have seen that there is a sampling algorithm that tests whether a graph is triangle-free or ϵ-far from triangle-free. Can we find any other properties that are testable? More formally, for which properties $\mathcal{P}$ is there an algorithm such that if we input a graph G that either has property $\mathcal{P}$ or is ϵ-far from having property $\mathcal{P}$, the

algorithm determines which of the two cases the graph is in? In particular, for which graphs can this can be done using only an oblivious tester, or in other words by only sampling $k = O(1)$ vertices?

A property is ***hereditary*** if it is closed under vertex-deletion. Some examples of hereditary properties are H-freeness, planarity, induced-H-freeness, 3-colorability, and being a perfect graph. The infinite removal lemma (Theorem 3.35) implies that every hereditary property is testable with one sided-error by an oblivious tester. Namely, we pick $\mathcal{H}$ to be the family of all graphs that do not have the property $\mathcal{P}$, and note that for a hereditary property $\mathcal{P}$, not having $\mathcal{P}$ is equivalent to not containing any graph that has property $\mathcal{P}$. This also explains why this approach would not work for properties that are not hereditary. In fact, properties that are not (almost) hereditary cannot be tested by an oblivious tester.

For example, if a graph is planar, then so is any induced subgraph. Hence, planarity is a hereditary property.

Alon and Shapira (2008)

3.8 Hypergraph removal lemma

For every interesting fact about graphs, the question of how that fact can be generalized to hypergraphs, if at all, naturally arises. We now state that generalization for Theorem 3.29, the graph removal lemma. Recall that an ***r-uniform hypergraph***, called an ***r-graph*** for short, is a pair (V, E), where $E \subset \binom{V}{r}$, i.e. the edges are r-element subsets of V.

Theorem 3.38 (Hypergraph removal lemma)**.** *For all r-graphs H and all $\epsilon > 0$, there exists $\delta > 0$ such that, if G is an n-vertex graph with fewer than $\delta n^{v(H)}$ copies of H, then G can be made H-free by removing fewer than ϵn^r edges from G.*

Rödl et al. (2005)
Gowers (2007)

Why do we care about this lemma? Recall that we deduced Roth's Theorem (Theorem 3.19) from a corollary of the triangle removal lemma, namely that every graph in which ever edge lies in exactly one triangle has $o(n^2)$ edges. We can do the same here, using Theorem 3.38, to prove the natural generalization of Roth's Theorem, namely Szemerédi's Theorem (Theorem 1.8), which states that, for fixed k, if $A \subset [N]$ is k-AP-free, then $|A| = o(N)$.

You may ask: couldn't we do the same thing with ordinary graphs? In fact, no! The reason is deeply seated in an idea called complexity of a linear pattern, which we will not elaborate on here. It turns out that a 4-AP has complexity 2, whereas a 3-AP has complexity 1. The techniques that we have developed so far work well for complexity 1 patterns, but higher complexity patterns are much more difficult to handle.

Green and Tao (2010)

We now state a corollary of Theorem 3.38 that is highly reminiscent of Corollary 3.18:

Corollary 3.39. *If G is a 3-graph such that every edge is contained in a unique tetrahedron, then G has $o(n^3)$ edges.*

Recall that a tetrahedron is $K_4^{(3)}$, i.e. a complete 3-graph on 4 vertices.

This corollary follows immediately from the hypergraph removal lemma. We now use this corollary to prove Szemerédi's Theorem:

Proof of Theorem 1.8. We will illustrate the proof for $k = 4$. Larger values of k are analogous. Let $M = 6N + 1$ (what is important here is that $M > 3N$ and that M is coprime to 6). Build a 4-partite 3-graph G with parts X, Y, Z, W, all of which are M-element sets with vertices indexed by the elements of $\mathbb{Z}/M\mathbb{Z}$. We will define edges as follows (assume that x, y, z, w represent elements of X, Y, Z, W, respectively):

$$\begin{aligned} xyz \in E(G) &\text{ if and only if } 3x + 2y + z \in A, \\ xyw \in E(G) &\text{ if and only if } 2x + y - w \in A, \\ xzw \in E(G) &\text{ if and only if } x - z - 2w \in A, \\ yzw \in E(G) &\text{ if and only if } -y - 2z - 3w \in A. \end{aligned}$$

Observe that the i^{th} linear form does not include the i^{th} variable.

For the sake of clarity, M needs to be coprime to 6 because we want to always have exactly one solution for the fourth variable given the other three and given a value for any of the above linear forms.

Notice that $xyzw$ is a tetrahedron if and only if $3x + 2y + z, 2x + y - w, x - z - 2w, -y - 2z - 3w \in A$. However, these values form a 4-AP with common difference $-x - y - z - w$. Since A is 4-AP-free, the only tetrahedra in A are trivial 4-APs. Thus every edge lies in exactly one tetrahedron. By the Corollary above, the number of edges is $o(M^3)$. But the number of edges is $4M^2|A|$, so we can deduce that $|A| = o(M) = o(N)$. □

A similar argument to the one above can be used to show Theorem 1.9, which guarantees that every subset of $\mathbb{Z}^d$ of positive density contains arbitrary constellations. An example of this is the square in $\mathbb{Z}^2$, composed of points $(x, y), (x + d, y), (x, y + d), (x + d, y + d)$ for some $x, y \in \mathbb{Z}$ and positive integer d.

3.9 *Hypergraph regularity*

Hypergraph regularity is a more difficult concept than ordinary graph regularity. We will not go into details but simple discuss some core ideas. See Gowers for an excellent exposition of one of the approaches.

Gowers 2006

A naïve attempt at defining hypergraph regularity would be to define it analogously to ordinary graph regularity, something like this:

Definition 3.40 (Naïve definition of 3-graph regularity)**.** Given a 3-graph $G^{(3)}$ and three subsets $V_1, V_2, V_3 \subset V(G^{(3)})$, we say that (V_1, V_2, V_3) is ϵ-regular if, for all $A_i \subset V_i$ such that $|A_i| \geq \epsilon |V_i|$, we

have $|d(V_1, V_2, V_3) - d(A_1, A_2, A_3)| \leq \epsilon$. Here, $d(X, Y, Z)$ denotes the fraction of elements of $X \times Y \times Z$ that are in $E(G^{(3)})$.

If you run through the proof of the Szemerédi Regularity Lemma with this notion, you can construct a very similar proof for hypergraphs that shows that, for all $\epsilon > 0$, there exists $M = M(\epsilon)$ such that every graph has a partition into at most M parts so that the fraction of triples of parts that are not ϵ-regular is less than ϵ. In fact, one can even make the partition equitable if one wishes.

So what's wrong with what we have? Recall that our proofs involving the Szemerédi Regularity Lemma typically have three steps: Partition, Clean, and Count. It turns out that the Count step is what will give us trouble.

Recall that regularity is supposed to represent pseudorandomness. Because of this, why don't we try truly random hypergraphs and see what happens? Let us consider two different random 3-graph constructions:

1. First pick constants $p, q \in [0, 1]$. Build a random graph $G^{(2)} = G(n, p)$, an ordinary Erdős-Renyi graph. Then make $G^{(3)}$ by including each triangle of $G^{(2)}$ as an edge of $G^{(3)}$ with probability q. Call this 3-graph A.

2. For each possible edge (i.e. triple of vertices), include the edge with probability p^3q, independent of all other edges. Call this 3-graph B.

Both A and B have each triple appear independently with probability p^3q, and both graphs satisfy our above notion of ϵ-regularity with high probability. However, we can compute the densities of $K_4^{(3)}$ (tetrahedra) in both of these graphs and see that they do not match. In graph B, each edge occurs with probability p^3q, and the edges appear independently, so the probability of an tetrahedron appearing is $(p^3q)^4$. However, in graph A, a tetrahedron requires the existence of K_4 in $G^{(2)}$. Since K_4 has 6 edges, it appears in $G^{(2)}$ with probability p^6, and then each triangle that makes up the tetrahedron occurs independently with probability q. Thus, the probability of any given tetrahedron appearing in A is p^6q^4, which is clearly not the same as $(p^3q)^4$. It follows that the above notion of hypergraph regularity does not appropriately constrain the frequency of subgraphs.

This notion of hypergraph regularity is still far from useless, however. It turns out that there is a counting lemma for hypergraphs H if H is ***linear***, meaning that every pair of edges intersects in at most 1 vertex. The proof is similar to that of Theorem 3.27, the graph counting lemma. But for now, let us move on to the better notion of hypergraph regularity, which will give us what we want.

Definition 3.41 (Triple density on top of 2-graphs)**.** Given $A, B, C \subset E(K_n)$ (think of A, B, C as subgraphs) and a 3-graph G, $d_G(A, B, C)$ is defined to be the fraction of triples $\{xyz \mid yz \in A, xz \in B, xy \in C\}$ that are triples of G.

Using the above definition, we can then define a regular triple of edge subsets and a regular partition, both of which we describe here informally. Consider a partition $E(K_n) = G_1^{(2)} \cup \cdots \cup G_l^{(2)}$ such that for most triples (i, j, k), there are a lot of triangles on top of $\left(G_i^{(2)}, G_j^{(2)}, G_k^{(2)}\right)$. We say that $\left(G_i^{(2)}, G_j^{(2)}, G_k^{(2)}\right)$ is regular in the sense that for all subgraphs $A_i^{(2)} \subset G_i^{(2)}$ with not too few triangles on top of $\left(A_i^{(2)}, A_j^{(2)}, A_k^{(2)}\right)$, we have

$$\left| d\left(G_i^{(2)}, G_j^{(2)}, G_k^{(2)}\right) - d\left(A_i^{(2)}, A_j^{(2)}, A_k^{(2)}\right) \right| \leq \epsilon.$$

We then subsequently define a regular partition as a partition in which the triples of parts that are not regular constitute at most an ϵ fraction of all triples of parts in the partition.

In addition to this, we need to further regularize $G_1^{(2)}, \ldots, G_l^{(2)}$ via a partition of the vertex set. As a result, we have the total data of hypergraph regularity as follows:

1. a partition of $E(K_n)$ into graphs such that $G^{(3)}$ sits pseudorandomly on top;
2. a partition of $V(G)$ such that the graphs in the above step are extremely pseudorandom (in a fashion resembling Theorem 3.33).

Note that many versions of hypergraph regularity exist in the literature, and not all of them are obviously equivalent. In fact, in some cases, it takes a lot of work to show that they are equivalent. We still are not quite sure which notion of hypergraph regularity, if any, is the most "natural."

In a similar vein to ordinary graph regularity, we can ask what bounds we get for hypergraph regularity, and the answers are equally horrifying. For a 2-uniform hypergraph, i.e. a normal graph, the bounds required a TOWER function (repeated exponentiation), also known as tetration. For a 3-uniform hypergraph, the bounds require us to go one step up the Ackermann hierarchy, to the WOWZER function (repeated applications of TOWER), also known as pentation. For 4-uniform hypergraphs, we must move one more step up the Ackermann hierarchy, and so on. As a result, applications of hypergraph regularity tend to give us very poor quantitative bounds involving the inverse Ackermann function. In fact, the best known bounds for k-APs are as follows:

Theorem 3.42 (Gowers). *For every $k \geq 3$ there is some $c_k > 0$ such that every k-AP-free subset of $[N]$ has at most $N(\log\log N)^{-c_k}$ elements.*

Gowers (2001)

This is the best known bound for $k \geq 5$, although for $k = 3, 4$ there are better known bounds.

For the multidimensional Szemerédi theorem (Theorem 1.9), the best known bounds generally come from the hypergraph regularity lemma. The first known proof came from ergodic theory, which actually gives no quantitative bounds due to its reliance on compactness arguments. A major motivation for working with hypergraph regularity was getting quantitative bounds for Theorem 1.9.

3.10 *Spectral proof of Szemerédi regularity lemma*

We previously proved the Szemerédi regularity lemma using the energy increment argument. We now explain another method of proof using the spectrum of a graph. Like the above discussion on hypergraph regularity, this discussion will skim over a number of details.

Tao (2012)

Given an n-vertex graph G, the adjacency matrix, denoted A_G, is an $n \times n$ matrix that has a 1 as the ij-entry (which we will denote $A_G(i,j)$) if vertices i and j are attached by an edge and 0 otherwise.

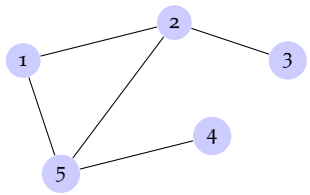
The adjacency matrix is always a real symmetric matrix. As a result, it always has real eigenvalues, and one can find an orthonormal basis of eigenvectors. Suppose that A_G has eigenvalues λ_i for $1 \leq i \leq n$, where the ordering is based on decreasing magnitude: $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. This gives us a spectral decomposition

$$A_G = \sum_{i=1}^{n} \lambda_i u_i u_i^T,$$

where u_i is a unit eigenvector with $A_G u_i = \lambda_i u_i$. One can additionally observe that

$$\begin{aligned} \sum_{i=1}^{n} \lambda_i^2 &= \operatorname{tr}(A^2) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} A_G(i,j)^2 \\ &= 2e(G) \\ &\leq n^2, \end{aligned}$$

where the second equality follows from the fact that A is symmetric.



For example, the graph G above has the following adjacency matrix:

$$A_G = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Lemma 3.43. $|\lambda_i| \leq \frac{n}{\sqrt{i}}$

Proof. If $|\lambda_k| > \frac{n}{\sqrt{k}}$ for some k, then $\sum_{i=1}^{k} \lambda_i^2 > n^2$, a contradiction. □

Lemma 3.44. *Let $\epsilon > 0$ and $F : \mathbb{N} \to \mathbb{N}$ be an arbitrary "growth function" such that $f(j) \geq j$ for all j. Then there exists $C = C(\epsilon, F)$ such*

that for all G, A_G as above, there exists $J < C$ such that

$$\sum_{J \leq i < F(J)} \lambda_i^2 \leq \epsilon n^2.$$

Proof. Let $J_1 = 1$ and $J_{i+1} = F(J_i)$ for all $i \geq 1$. One cannot have $\sum_{J_k \leq i < J_{k+1}} \lambda_i^2 > \epsilon n^2$ for all $k \leq \frac{1}{\epsilon}$, or else the total sum is greater than n^2. Therefore, the desired inequality above holds for some $J = J(k)$, where $k \leq \frac{1}{\epsilon}$. Therefore, J is bounded; in particular, $J < F(F(\dots F(1)\dots))$, where F is applied $\frac{1}{\epsilon}$ times. □

Notice the analogy of the above fact with the energy increment step of our original proof of the Szemerédi Regularity Lemma.

We now introduce the idea of regularity decompositions, which were popularized by Tao. Pick J as in the Lemma above. We can decompose A_G as

$$A_G = A_{\text{str}} + A_{\text{sml}} + A_{\text{psr}},$$

where "str" stands for "structured," "sml" stands for "small," and "psr" stands for "pseudorandom." We define these terms as follows:

$$A_{\text{str}} = \sum_{i<J} \lambda_i u_i u_i^T$$
$$A_{\text{sml}} = \sum_{J \leq i < F(J)} \lambda_i u_i u_i^T$$
$$A_{\text{psr}} = \sum_{i \geq F(J)} \lambda_i u_i u_i^T$$

Here, A_{str} corresponds roughly to the bounded partition, A_{sml} corresponds roughly to the irregular pairs, and A_{psr} corresponds roughly to the pseudorandomness between pairs.

Here we define two notions of the norm of a matrix. The spectral radius (or spectral norm) of a matrix A is defined as $\max |\lambda_i(A)|$ over all possible eigenvalues λ_i. Alternatvely, the operator norm is defined by

$$\|A\| = \max_{v \neq 0} \frac{|Av|}{|v|} = \max_{u,v \neq 0} \frac{|u^T Av|}{|u|\,|v|}.$$

It is important to note that, for real symmetric matrices, the spectral norm and operator norm are equal.

Notice that A_{str} has eigenvectors $u_1, \dots, u_{J-1}$. These are the eigenvectors with the largest eigenvalues of A_G. Let us pretend that $u_i \in \{-1, 1\}^n$ for all $i = 1, \dots, J-1$. This is most definitely false, but let us pretend that this is the case for the sake of illustration. By taking these coordinate values, we see that the level sets of $u_1, \dots, u_{J-1}$ partition $V(G)$ into $P = O_{\epsilon,J}(1)$ parts $V_1, \dots, V_P$ such that A_{str} is roughly constant on each cell of the matrix defined by this partition. (The dependence on ϵ comes from the rounding of the coordinate

values; in reality, we let the eigenvectors vary by a small amount.) However, for two vertex subsets $U \subset V_k$ and $W \subset V_l$, we have:

$$\begin{aligned}\left|1_U^T A_{\text{psr}} 1_W\right| | &\leq |1_U|\,|1_W|\,\|A_{\text{psr}}\| \\ &\leq \sqrt{n} \cdot \sqrt{n} \cdot \frac{n}{\sqrt{F(J)}}.\end{aligned}$$

By choosing $F(J)$ large compared to P, we can guarantee that the above quantity is small. In particular, we can show that it is much less than $\epsilon \left(\frac{n}{P}\right)^2$. The significance of the quantity $1_U^T A_{\text{psr}} 1_W$ is that it equals $e(U,W) - d_{kl}|U||W|$, where d_{kl} is the average of the entries in the $V_k \times V_l$ block of A_{str}. Therefore, the fact that this quantity is small implies regularity.

We can also obtain a bound on the sum of the squares of the entries (known as the Frobenius norm) of A_{sml}. For real symmetric matrices, this equals the Hilbert–Schmidt norm, which equals the sum of the squares of the eigenvalues:

$$\begin{aligned}\|A_{\text{sml}}\|_{\text{F}} &= \|A_{\text{sml}}\|_{\text{HS}} \\ &= \sum_{J \leq i \leq F(J)} \lambda_i^2 \\ &\leq \epsilon n^2.\end{aligned}$$

Therefore, A_{sml} might destroy ϵ-regularity for roughly an ϵ fraction of pairs of parts, but the partition will still be regular.

It is worth mentioning that there are ways to massage this method to get our various desired modifications of the Szemerédi Regularity Lemma, such as the desire for an equitable partition. We will not attempt to discuss those here.

4
Pseudorandom graphs

The term "pseudorandom" refers to a wide range of ideas and phenomenon where non-random objects behave in certain ways like genuinely random objects. For example, while the prime numbers are not random, their distribution among the integers have many properties that resemble random sets. The famous Riemann Hypothesis is a notable conjecture about the pseudorandomness of the primes in a certain sense.

When used more precisely, we can ask whether some given objects behaves in some *specific* way similar to a typical random object? In this chapter, we examine such questions for graphs, and study ways that a non-random graph can have properties that resemble a typical random graph.

4.1 Quasirandom graphs

The next theorem is a foundational result in the subject. It lists several seemingly different pseudorandomness properties that a graph can have (with some seemingly easier to verify than others), and asserts, somewhat surprisingly, that these properties turn out to be all equivalent to each other.

Theorem 4.1. *Let $\{G_n\}$ be a sequence of graphs with G_n having n vertices and $(p+o(1))\binom{n}{2}$ edges, for fixed $0 < p < 1$. Denote G_n by G. The following properties are equivalent:*

1. ***DISC** ("discrepancy"): $|e(X,Y) - p|X||Y|| = o(n^2)$ for all $X, Y \subset V(G)$.*
2. ***DISC'**: $|e(X) - p\binom{|X|}{2}| = o(n^2)$ for all $X \subset V(G)$.*
3. ***COUNT**: For all graphs H, the number of **labeled copies** of H in G (i.e. vertices in H are distinguished) is $(p^{e(H)} + o(1))n^{v(H)}$. The $o(1)$ term may depend on H.*

Chung, Graham, and Wilson (1989)

Theorem 4.1 should be understood as a theorem about *dense graphs*, i.e., graphs with constant order edge density. Sparser graphs can have very different behavior and will be discussed in later sections.

4. ***C4****: The number of labeled copies of C_4 is at most $(p^4 + o(1))n^4$.*

5. ***CODEG*** *(codegree): If $\mathrm{codeg}(u, v)$ is the number of common neighbors of u and v, then $\sum_{u,v \in V(G)} |\mathrm{codeg}(u, v) - p^2 n| = o(n^3)$.*

6. ***EIG*** *(eigenvalue): If $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{v(G)}$ are the eigenvalues of the adjacency matrix of G, then $\lambda_1 = pn + o(n)$ and $\max_{i \neq 1} |\lambda_i| = o(n)$.*

Remark 4.2. In particular, for a d-regular graph, the largest eigenvalue is d, with corresponding eigenvector the all-1 vector, and **EIG** states that $\lambda_2, \lambda_{v(G)} = o(n)$.

We can equivalently state the conditions in the theorem in terms of some ϵ: for instance, **DISC** can be reformulated as

DISC(ϵ): For all $X, Y \subset V(G)$, $|e(X, Y) - p|X||Y|| < \epsilon n^2$.

Then we will see from the proof of Theorem 4.1 that the conditions in the theorem are equivalent up to at most polynomial change in ϵ, i.e. **Prop1**$(\epsilon) \implies$ **Prop2**(ϵ^c) for some c.

Since we will use the Cauchy–Schwarz inequality many times in this proof, let's begin with an exercise.

Lemma 4.3. *If G is a graph with n vertices, $e(G) \geq pn^2/2$, then the number of labeled copies of C_4 is $\geq (p^4 - o(1))n^4$.*

Proof. We want to count the size of $S = \mathrm{Hom}(C_4, G)$, the set of graph homomorphisms from C_4 to G. We also include in S some non-injective maps, i.e. where points in C_4 may map to the same point in G, since there are only $O(n^3)$ of them anyway. It is equal to $\sum_{u,v \in V(G)} \mathrm{codeg}(u, v)^2$, by considering reflections across a diagonal of C_4. Using Cauchy–Schwarz twice, we have

$$\begin{aligned}
|\mathrm{Hom}(C_4, G)| &= \sum_{u,v \in V(G)} \mathrm{codeg}(u, v)^2 \\
&\geq \frac{1}{n^2} \left(\sum_{u,v \in V(G)} \mathrm{codeg}(u, v) \right)^2 \\
&= \frac{1}{n^2} \left(\sum_{x \in G} \deg(x)^2 \right)^2 \\
&\geq \frac{1}{n^2} \left(\frac{1}{n} \left(\sum_{x \in G} \deg(x) \right)^2 \right)^2 \\
&= \frac{1}{n^2} \left(\frac{1}{n} (pn^2)^2 \right)^2 \\
&= p^4 n^4
\end{aligned}$$

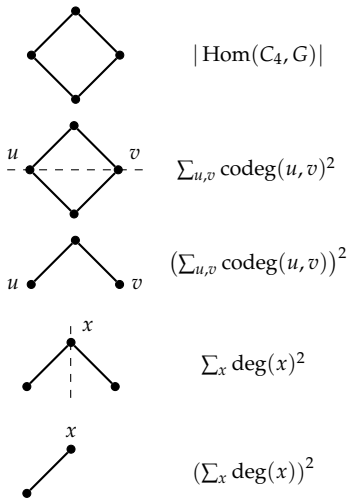


Figure 4.1: Visualization of Cauchy–Schwarz

where in the second line we have $\sum_{u,v\in V(G)} \operatorname{codeg}(u,v) = \sum_{x\in G} \deg(x)^2$ by counting the number of paths of length 2 in two ways. □

Remark 4.4. We can keep track of our Cauchy–Schwarz manipulations with a "visual anchor": see Figure 4.1. We see that Cauchy–Schwarz bounds exploit symmetries in the graph.

Now we prove the theorem.

Proof. **DISC** $\implies$ **DISC′**: Take $Y = X$ in **DISC**.

DISC′ $\implies$ **DISC**: By categorizing the types of edges counted in $e(X,Y)$ (see Figure 4.2), we can write $e(X,Y)$ in terms of the edge counts of individual vertex sets:

$$e(X,Y) = e(X\cup Y) + e(X\cap Y) - e(X\setminus Y) - e(Y\setminus X).$$

Then we can use **DISC′** to get that this is

$$\begin{aligned} &p\left(\binom{|X\cup Y|}{2} + \binom{|X\cap Y|}{2} + \binom{|X\setminus Y|}{2} + \binom{|Y\setminus X|}{2} + o(n^2)\right)\\ =&p|X||Y| + o(n^2). \end{aligned}$$

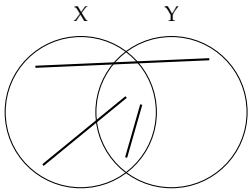


Figure 4.2: Visualization of the expression for $e(X,Y)$

DISC $\implies$ **COUNT**: This follows from the graph counting lemma (Theorem 3.27), taking $V_i = G$ for $i = 1,\dots,v(H)$.

COUNT $\implies$ **C4**: **C4** is just a special case of **COUNT**.

C4 $\implies$ **CODEG**: Given **C4**, we have

$$\sum_{u,v\in G} \operatorname{codeg}(u,v) = \sum_{x\in G} \deg(x)^2 \ge n\left(\frac{2e(G)}{n}\right)^2 = \left(p^2+o(1)\right)n^3.$$

We also have

$$\begin{aligned} \sum_{u,v} \operatorname{codeg}(u,v)^2 &= \text{Number of labeled copies of } C_4 + o(n^4)\\ &\le \left(p^4+o(1)\right)n^4. \end{aligned}$$

Therefore, we can use Cauchy–Schwarz to find

$$\begin{aligned} \sum_{u,v\in G} |\operatorname{codeg}(u,v) - p^2n| &\le n\left(\sum_{u,v\in G}\left(\operatorname{codeg}(u,v)-p^2n\right)^2\right)^{1/2}\\ &= n\left(\sum_{u,v\in G}\operatorname{codeg}(u,v)^2 - 2p^2n\sum_{u,v\in G}\operatorname{codeg}(u,v) + p^4n^4\right)^{1/2}\\ &\le n\left(p^4n^4 - 2p^2n\cdot p^2n^3 + p^4n^4 + o(n^4)\right)^{1/2}\\ &= o(n^3), \end{aligned}$$

as desired.

Remark 4.5. This technique is similar to the ***second moment method*** in probabilistic combinatorics: we want to show that the variance of $\operatorname{codeg}(u, v)$ is not too large.

CODEG $\implies$ **DISC**: First, note that we have

$$\begin{aligned}
\sum_{u \in G} |\deg u - pn| &\le n^{1/2} \left(\sum_{u \in G} (\deg u - pn)^2 \right)^{1/2} \\
&= n^{1/2} \left(\sum_{u \in G} (\deg u)^2 - 2pn \sum_{u \in G} \deg u + p^2 n^3 \right)^{1/2} \\
&= n^{1/2} \left(\sum_{u,v \in G} \operatorname{codeg}(u, v) - 4pn \cdot e(G) + p^2 n^3 \right)^{1/2} \\
&= n^{1/2} \left(p^2 n^3 - 2p^2 n^3 + p^2 n^3 + o(n^3) \right)^{1/2} \\
&= o(n^2).
\end{aligned}$$

Then we can write

$$\begin{aligned}
|e(X, Y) - p|X||Y|| &= \left| \sum_{x \in X} (\deg(x, Y) - p|Y|) \right| \\
&\le n^{1/2} \left(\sum_{x \in X} (\deg(x, Y) - p|Y|)^2 \right)^{1/2}.
\end{aligned}$$

Since the summand is nonnegative, we can even enlarge the domain of summation from X to $V(G)$. So we have

$$\begin{aligned}
|e(X, Y) - p|X||Y|| &\le n^{1/2} \left(\sum_{x \in V} \deg(x, Y)^2 - 2p|Y| \sum_{x \in V} \deg(x, Y) + p^2 n |Y|^2 \right)^{1/2} \\
&= n^{1/2} \left(\sum_{y, y' \in Y} \operatorname{codeg}(y, y') - 2p|Y| \sum_{y \in Y} \deg y + p^2 n |Y|^2 \right)^{1/2} \\
&= n^{1/2} \left(|Y|^2 p^2 n - 2p|Y| \cdot |Y| pn + p^2 n |Y|^2 + o(n^3) \right)^{1/2} \\
&= o(n^2).
\end{aligned}$$

Now that we have proven the "C_4" between the statements **DISC** $\implies$ **COUNT** $\implies$ **C4** $\implies$ **CODEG** $\implies$ **DISC**, we relate the final condition, **EIG**, to the **C4** condition.

EIG $\implies$ **C4**: The number of labeled C_4s is within $O(n^3)$ of the number of closed walks of length 4, which is $\operatorname{tr}(A_G^4)$, where A_G is the adjacency matrix of G. From linear algebra, $\operatorname{tr}(A_G^4) = \sum_{i=1}^n \lambda_i^4$. The main term is λ_1: by assumption, $\lambda_1^4 = p^4 n^4 + o(n^4)$. Then we want to make sure that the sum of the other λ_i^4s is not too big. If you bound them individually, you just get $o(n^5)$, which is not enough. Instead,

we can write

$$\sum_{i\geq 2} \lambda_i^4 \leq \max_{i\neq 2} |\lambda_i|^2 \sum_{i\geq 1} \lambda_i^2$$

and note that $\sum_{i\geq 1} \lambda_i^2 = \operatorname{tr}(A_G^2) = 2e(G)$, so

$$\sum_{i=1}^{n} \lambda_i^4 = p^4 n^4 + o(n^4) + o(n^2)n^2 = p^4 n^4 + o(n^4).$$

C4 $\implies$ **EIG**: We use the *Courant–Fischer theorem* (also called the *min-max theorem*): for a real symmetric matrix A, the largest eigenvalue is

$$\lambda_1 = \sup_{x\neq 0} \frac{x^T A x}{x^T x}.$$

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of A_G, and let $\mathbb{1}$ be the all-1 vector in $\mathbb{R}^{V(G)}$. Then we have

$$\lambda_1 \geq \frac{\mathbb{1}^T A_G \mathbb{1}}{\mathbb{1}^T \mathbb{1}} = \frac{2e(G)}{n} = (p + o(1))\, n.$$

But from **C4**, we have

$$\lambda_1^4 \leq \sum_{i=1}^{n} \lambda_i^4 = \operatorname{tr} A_G^4 \leq p^4 n^4 + o(n^4),$$

which implies $\lambda_1 \leq pn + o(n)$. Hence, $\lambda_1 = pn + o(n)$.

We also have

$$\max_{i\neq 1} |\lambda_i|^4 \leq \operatorname{tr}(A_G^4) - \lambda_1^4 \leq p^4 n^4 - p^4 n^4 + o(n^4) = o(n^4),$$

as desired. □

What is most remarkable about Theorem 4.1 that the **C4** condition, seemingly the weakest of all the conditions, actually implies all the other conditions.

Remember that this theorem is about dense graphs (i.e. p is constant). We can write some analogs of the conditions for sparse graphs, where $p = p_n \to 0$ as $n \to \infty$. For example, in **DISC**, we need to change the $o(n^2)$ to $o(pn^2)$ to capture the idea that the number of edges of the quasirandom graph should be close to the expected number of edges of a truly random graph. Analogously, in **COUNT**, the number of labeled copies of H is $(1 + o(1))p^{e(H)}n^{v(H)}$. However, these conditions are *not* equivalent for sparse graphs. In particular, the counting lemma fails. For instance, here is a graph that satisfies the sparse analog of **DISC**, but does not even have any C_3.

Example 4.6. Take $p = o(n^{-1/2})$. The number of C_3s should be around $\binom{n}{3}p^3$, and the number of edges is $\binom{n}{2}p$. But by choice of p, the number of C_3s is now asymptotically smaller than the number

of edges, so we can just remove an edge from each triangle in this $G(n,p)$. We will those have removed $o(n^2p)$ edges, so the sparse analog of **DISC** still holds, but now the graph is triangle-free. This graph is pseudorandom in one sense, in that it still satisfies the discrepancy condition, but not in another sense, in that it has zero triangles.

4.2 *Expander mixing lemma*

Now we talk about a certain class of graphs, *expander graphs*, with a particularly strong discrepancy property.

Theorem 4.7 (Expander mixing lemma)**.** *Let G be an n-vertex, d-regular graph, with adjacency matrix having eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Let $\lambda = \max\{|\lambda_2|, |\lambda_n|\}$. Then for all $X, Y \subset V(G)$,*

$$\left| e(X,Y) - \frac{d}{n}|X||Y| \right| \leq \lambda \sqrt{|X||Y|}.$$

Proof. Let J be the all-1 matrix. We have

$$\begin{aligned} \left| e(X,Y) - \frac{d}{n}|X||Y| \right| &= \left| 1_X^T \left(A_G - \frac{d}{n} J \right) 1_Y \right| \\ &\leq \left\| A_G - \frac{d}{n} J \right\| |1_X|\,|1_Y| \\ &= \left\| A_G - \frac{d}{n} J \right\| \sqrt{|X||Y|}. \end{aligned}$$

It suffices to prove that the largest eigenvalue of $A_G - \frac{d}{n}J$ is at most λ.

Let v be an eigenvector of A_G. Since G is d-regular, one possibility for $v = (v_1, \ldots, v_n)$ is $\mathbb{1}$, which has corresponding eigenvalue d in A_G. Then $\mathbb{1}$ is also an eigenvector of $A_G - \frac{d}{n}J$, with corresponding eigenvalue 0. If $v \neq \mathbb{1}$, then it is orthogonal to $\mathbb{1}$, i.e. $v \cdot \mathbb{1} = \sum_{i=1}^n v_i = 0$. Therefore, $Jv = 0$, so v is also an eigenvector of $A_G - \frac{d}{n}J$ with same eigenvalue as in A_G. Thus, $A_G - \frac{d}{n}J$ has eigenvalues $0, \lambda_2, \lambda_3, \ldots, \lambda_n$, so its largest eigenvalue is λ, as desired. □

Expanders are related to pseudorandom graphs: when you have some small subset of vertices, you can expect them to have many neighbors. These kinds of graphs are called expanders because many vertices of the graph can be quickly reached via neighbors.

We now restrict our attention to a special class of graphs.

Definition 4.8. An ***(n,d,λ)-graph*** is an n-vertex, d-regular graph whose adjacency matrix has eigenvalues $d = \lambda_1 \geq \cdots \geq \lambda_n$ satisfying $\max\{|\lambda_2|, |\lambda_n|\} \leq \lambda$.

The expander mixing lemma (Theorem 4.7) can be rephrased as saying that if G is an (n, d, λ)-graph, then

$$\left| e(X, Y) - \frac{d}{n}|X||Y| \right| \leq \lambda \sqrt{|X||Y|}$$

for all $X, Y \subseteq V(G)$.

A random graph is pseudorandom with high probability. However, we would like to give deterministic constructions that have pseudorandom properties. The following is an example of such a construction.

Definition 4.9. Let Γ be a finite group, and let $S \subseteq \Gamma$ be a subset with $S = S^{-1}$. The ***Cayley graph*** $\mathrm{Cay}(\Gamma, S) = (V, E)$ is defined by $V = \Gamma$ and

$$E = \{(g, gs) \colon g \in \Gamma, s \in S\}.$$

Example 4.10. The ***Paley graph*** is a graph $\mathrm{Cay}(\mathbb{Z}/p\mathbb{Z}, S)$ for $p \equiv 1 \pmod 4$ a prime, and S the set of nonzero quadratic residues in $\mathbb{Z}/p\mathbb{Z}$.

Unfortunately, Raymond Paley was killed by an avalanche at the age of 26. His contributions include Paley graphs, the Paley–Wiener theorem, and Littlewood–Paley theory.

Proposition 4.11. *The Paley graph $G = \mathrm{Cay}(\mathbb{Z}/p\mathbb{Z}, S)$ satisfies $|\lambda_2|, |\lambda_p| \leq \frac{\sqrt{p}+1}{2}$, where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of its adjacency matrix.*

Proof. We simply write down a list of eigenvectors. Let the vertex 0 correspond to the first coordinate, the vertex 1 correspond to the second coordinate, etc. Let

$$\begin{aligned} v_1 &= (1, \ldots, 1) \\ v_2 &= (1, \omega, \omega^2, \ldots, \omega^{p-1}) \\ v_3 &= (1, \omega^2, \omega^4, \ldots, \omega^{2(p-1)}) \\ &\vdots \\ v_p &= (1, \omega^{p-1}, \ldots, \omega^{(p-1)(p-1)}), \end{aligned}$$

where ω is a primitive p-th root of unity.

We first check that these are eigenvectors. The all 1's vector v_1 has eigenvalue $d = \lambda_1$. We compute that the j-th coordinate of $A_G v_2$ is

$$\sum_{s \in S} \omega^{j+s} = \omega^j \sum_{s \in S} \omega^s.$$

Since ω_j is the j-th coordinate of v_2, and this holds for all j, the sum is the eigenvalue. In general, for $0 \leq k \leq p-1$,

$$\lambda_{k+1} = \sum_{s \in S} \omega^{ks}.$$

Note that this is a generic fact about Cayley graphs on $\mathbb{Z}/p\mathbb{Z}$, and the eigenvectors do not depend on S. Now we compute the sizes of the λ_i. For $k > 0$, we have

$$2\lambda_{k+1} + 1 = \sum_{a \in \mathbb{Z}/p\mathbb{Z}} \omega^{ka^2}.$$

Here, we used that S is the set of nonzero quadratic residues. The sum on the right is known as a ***Gauss sum***. It is evaluated as follows. We square the sum to get

$$\left| \sum_{a \in \mathbb{Z}/p\mathbb{Z}} \omega^{ka^2} \right|^2 = \sum_{a,b \in \mathbb{Z}/p\mathbb{Z}} \omega^{k((a+b)^2 - a^2)} = \sum_{a,b \in \mathbb{Z}/p\mathbb{Z}} \omega^{k(2ab+b^2)}.$$

For $b \neq 0$, the sum

$$\sum_{a \in \mathbb{Z}/p\mathbb{Z}} \omega^{k(2ab+b^2)} = 0,$$

since $k(2ab + b^2)$ for $a \in \mathbb{Z}/p\mathbb{Z}$ is a permutation of $\mathbb{Z}/p\mathbb{Z}$. For $b = 0$,

$$\sum_{a} \omega^{k(2ab+b^2)} = p.$$

Thus, the square of the Gauss sum is equal to p, so $\lambda_{k+1} = \frac{\pm\sqrt{p}-1}{2}$ for all $k > 0$. □

You might recognize $\sum_{s \in S} \omega^{ks}$ as a Fourier coefficient of the indicator function of S, viewed as a function on $\mathbb{Z}/p\mathbb{Z}$. Indeed, there is an intimate connection between the eigenvalues of a Cayley graph of an abelian group and the Fourier transform of a function on the group. In fact, the two spectra are identical up to scaling (partly the reason why we use the name "spectrum" for both eigenvalues and Fourier). There is a similar story for non-abelian groups, though Fourier analysis on non-abelian groups involves representation theory.

4.3 *Quasirandom Cayley graphs*

We saw that the Chung–Graham–Wilson theorem fails to hold for sparse analogs of the pseudorandomness conditions. However, it turns out, somewhat surprisingly, that if we restrict to Cayley graphs of groups (including non-abelian), no matter at what edge-density, the sparse analogs of DISC and EIG are equivalent.

For sparse graphs in general, the sparse analog of DISC does not imply the sparse analog of EIG. Consider the disjoint union of a large random d-regular graph and a K_{d+1}. This graph satisfies the sparse analog of DISC because the large random d-regular graph does. However, the top two eigenvalues are both $\lambda_1 = \lambda_2 = d$, because the all 1's vectors on each of the components is an eigenvector

with eigenvalue d, where as the sparse analog of EIG would give $\lambda_2 = o(d)$.

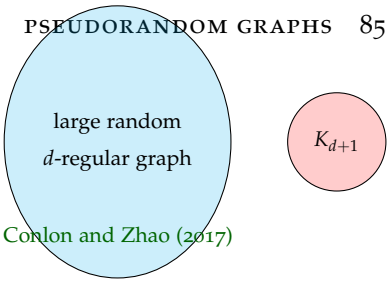


Figure 4.3: DISC does not imply EIG for a general graph.

Theorem 4.12 (Conlon–Zhao). *Let Γ be a finite group and $S \subseteq \Gamma$ a subset with $S = S^{-1}$. Let $G = \mathrm{Cay}(\Gamma, S)$. Let $n = |\Gamma|$ and $d = |S|$. For $\epsilon > 0$, we say that G has the property*

- *$DISC(\epsilon)$ if for all $X, Y \subseteq G$, we have $|e(X,Y) - \frac{d}{n}|X||Y|| \leq \epsilon dn$, and*
- *$EIG(\epsilon)$ if G is an (n,d,λ)-graph with $\lambda \leq \epsilon d$.*

Then if G satisfies $EIG(\epsilon)$, it also satisfies $DISC(\epsilon)$, and if it satisfies $DISC(\epsilon)$, then it also satisfies $EIG(8\epsilon)$.

Conlon and Zhao (2017)

The proof of Theorem 4.12 uses ***Grothendieck's inequality***.

Theorem 4.13 (Grothendieck's inequality). *There exists an absolute constant $K > 0$ such that for all matrices $A = (a_{i,j}) \in \mathbb{R}^{n \times n}$,*

$$\sup_{\substack{x_i \in B \\ y_i \in B}} \sum_{i,j} a_{i,j} \langle x_i, y_i \rangle \leq K \sup_{\substack{x_i \in \{\pm 1\} \\ y_i \in \{\pm 1\}}} \sum_{i,j} a_{i,j} x_i y_j.$$

In the left hand side, the supremum is taken over all unit balls B in some $\mathbb{R}^m$.

Grothendieck (1953)

The right hand side of Grothendieck's inequality is the supremum of the bilinear form $\langle x, Ay \rangle$ over a discrete set. It is important combinatorially, but hard to evaluate. The left hand side is a "semidefinite relaxation" of the right hand side. There exist efficient methods to evaluate it, it is always at least the right hand side, and Grothiendieck's inequality tells us that we don't lose more than a constant factor when using it as an approximation for the right hand side.

Remark 4.14. It is known that $K = 1.78$ works. The optimal value, known as the "real Grothendieck constant," is unknown.

Krivine (1979)

Proof of Theorem 4.12. The fact that $EIG(\epsilon)$ implies $DISC(\epsilon)$ follows from the expander mixing lemma. Specifically, it tells us that

$$\left| e(X,Y) - \frac{d}{n}|X||Y| \right| \leq \lambda \sqrt{|X||Y|} \leq \epsilon dn$$

for any $X, Y \subseteq G$, which is what we want.

To prove the other implication, suppose $DISC(\epsilon)$ holds. For all $x, y \in \{\pm 1\}^\Gamma$, let $x^+, x^-, y^+, y^- \in \{0,1\}^\Gamma$ be such that

$$x_g^+ = \begin{cases} 1 & \text{if } x_g = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad x_g^- = \begin{cases} 1 & \text{if } x_g = -1 \\ 0 & \text{otherwise.} \end{cases}$$

Then $x = x^+ - x^-$. Similarly define y^+ and y^-.

Consider the matrix $A \in \mathbb{R}^{\Gamma\times\Gamma}$ with $A_{g,h} = 1_S(g^{-1}h) - \frac{d}{n}$ (here 1_S is the indicator function of S). Then

$$\langle x, Ay\rangle = \langle x^+, Ay^+\rangle - \langle x^-, Ay^+\rangle - \langle x^+, Ay^-\rangle + \langle x^-, Ay^-\rangle.$$

Each term in this sum is controlled by DISC. For example,

$$\langle x^+, Ay^+\rangle = e(X^+, Y^+) - \frac{d}{n}|X^+||Y^+|,$$

where $X^+ = \{g \in \Gamma : x_g = 1\}$, and $Y^+ = \{g \in \Gamma : y_g = 1\}$. Thus, $|\langle x^+, Ay^+\rangle| \le \epsilon dn$. This holds for the other terms as well, so

$$|\langle x, Ay\rangle| \le 4\epsilon dn \quad \text{for all} \quad x, y, \in \{\pm 1\}^\Gamma. \tag{4.1}$$

By the min-max characterization of the eigenvalue,

$$\max\{|\lambda_2|, |\lambda_n|\} = \sup_{\substack{|x|,|y|=1\\ x,y\in\mathbb{R}^\Gamma}} \langle x, Ay\rangle.$$

For all $x \in \mathbb{R}^\Gamma$, define $x^g \in \mathbb{R}^\Gamma$ by setting the coordinate $x^g_s = x_{sg}$ for all $s \in \Gamma$. Then $|x| = |x^g|$ since x^g simply permutes the coordinates of x. Then for all $x, y \in \mathbb{R}^\Gamma$ with $|x|, |y| = 1$,

$$\begin{aligned}
\langle x, Ay\rangle &= \sum_{g,h} A_{g,h} x_g y_h \\
&= \frac{1}{n} \sum_{g,h,s} A_{sg,sh} x_{sg} y_{sh} \\
&= \frac{1}{n} \sum_{g,h,s} A_{g,h} x_{sg} y_{sh} \\
&= \frac{1}{n} \sum_{g,h} A_{g,h} \langle x^g, y^h\rangle \le 8\epsilon d.
\end{aligned}$$

The inequality comes from Grothendieck's inequality with $K < 2$ combined with (4.1). Thus, $EIG(8\epsilon)$ is true. □

4.4 *Alon–Boppana bound*

In an (n, d, λ) graph, the smaller λ is, the more pseudorandom the graph is. A natural question to ask is, for fixed d, how small can λ be? We have the ***Alon–Boppana bound***.

Theorem 4.15 (Alon–Boppana bound). *Fix d. If G is an n-vertex graph whose adjacency matrix A_G has eigenvalues $\lambda_1 \ge \cdots \ge \lambda_n$, then* Alon (1986)

$$\lambda_2 \ge 2\sqrt{d-1} - o(1),$$

where $o(1) \to 0$ as $n \to \infty$.

Proof. Let $V = V(G)$. By Courant–Fischer, it suffices to exhibit a vector $z \in \mathbb{R}^V - \{0\}$ such that $\langle z, 1\rangle = 0$ and

Nilli (1991)

$$\frac{z^T A z}{z^T z} \geq 2\sqrt{d-1}.$$

Let $r \in \mathbb{N}$. Pick $v \in V$, and let V_i be the set of vertices at distance i from v. For example, $V_0 = \{v\}$ and $V_1 = N(v)$. Let $x \in \mathbb{R}^V$ be the vector with

$$x_u = w_i := (d-1)^{-i/2} \quad \text{for} \quad u \in V_i, 0 \leq i \leq r-1,$$

and $x_u = 0$ for all u such that $\operatorname{dist}(u, v) \geq r$. We claim that

$$\frac{x^T A x}{x^T x} \geq 2\sqrt{d-1}\left(1 - \frac{1}{2r}\right). \tag{4.2}$$

To show this, we compute

$$x^T x = \sum_{i=0}^{r-1} |V_i| w_i^2,$$

and

$$\begin{aligned} x^T A x &= \sum_{u \in V} x_u \sum_{u' \in N(u)} x_{u'} \\ &\geq \sum_{i=0}^{r-1} |V_i| w_i (w_{i-1} + (d-1) w_{i+1}) - (d-1)|V_{r-1}| w_{r-1} w_r \\ &= 2\sqrt{d-1}\left(\sum_{i=0}^{r-1} |V_i| w_i^2 - \frac{1}{2}|V_{r-1}| w_r^2\right). \end{aligned}$$

The inequality comes from the fact that each neighbor of $u \in V_i$ has distance at most $i+1$ from v and at least one neighbor has distance $i-1$ (note that the w_i are decreasing). However, since $x_u = 0$ for $\operatorname{dist}(u, w) \geq r$, so we must subtract off $(d-1)|V_{r-1}| w_{r-1} w_r$. Note that $|V_{i+1}| \leq (d-1)|V_i|$, so the above expression is

$$\geq 2\sqrt{d-1}\left(\sum_{i=1}^{r-1} |V_i| w_i^2\right)\left(1 - \frac{1}{2r}\right).$$

This proves (4.2). But we need $\langle z, 1\rangle = 0$. If $n > 1 + (d-1) + (d-1)^2 + \cdots + (d-1)^{2r-1}$, then there exist vertices $u, v \in V(G)$ at distance at least $2r$ from each other. Let $x \in \mathbb{R}^V$ be the vector obtained from the above construction centered at v. Let $y \in \mathbb{R}^V$ be the vector obtained from the above construction centered at u. Then x and y are supported on disjoint vertex sets with no edges between them. Thus, $x^T A y = 0$.

Choose a constant $c \in \mathbb{R}$ such that $z = x - cy$ has $\langle z, 1\rangle = 0$. Then

$$z^T z = x^T x + c^2 y^T y$$

and

$$z^T Az = x^T Ax + c^2 y^T Ay \geq 2\sqrt{d-1}\left(1 - \frac{1}{2r}\right) z^T z.$$

Taking $r \to \infty$ as $n \to \infty$ gives the theorem. □

We give a second proof of a slightly weaker result, but which is still in the spirit of Theorem 4.15.

Proof 2 (slightly weaker result). We'll show that $\max\{|\lambda_2|, |\lambda_n|\} \geq 2\sqrt{d-1} - o(1)$. This is an illustration of the trace method, also called the moment method. We have

$$\sum_{i=1}^{n} \lambda_i^{2k} = \operatorname{tr}(A^{2k}).$$

The right hand side is the number of closed walks of length $2k$ on G. Now, the number of closed walks of length $2k$ starting at a fixed vertex v in a d-regular graph is at least the number of closed walks of length $2k$ starting at a fixed v in an infinite d-regular tree. To see why this is true, note that given any walk on the infinite d-regular tree, we can walk in the same way on G by assigning an orientation to each vertex. But G may have more walks if it has cycles.

There are at least $C_k(d-1)^k$ closed walks of length $2k$ starting at a fixed v in an infinite d-regular tree, where $C_k = \frac{1}{k+1}\binom{2k}{k}$ is the k-th Catalan number. Thus, the number of walks of length $2k$ on G is at least $\frac{n}{k+1}\binom{2k}{k}(d-1)^k$. On the other hand,

$$d^{2k} + (n-1)\lambda^{2k} \geq \sum_{i=1}^{n} \lambda_i^{2k}.$$

Thus,

$$\lambda^{2k} \geq \frac{1}{k+1}\binom{2k}{k}(d-1)^k - \frac{d^{2k}}{n}.$$

The term $\frac{1}{k+1}\binom{2k}{k}$ is $(2 - o(1))^{2k}$ as $k \to \infty$. Letting $k \to \infty$ and $k = o(\log n)$ as $n \to \infty$ gives us $\lambda \geq 2\sqrt{d-1} - o(1)$. □

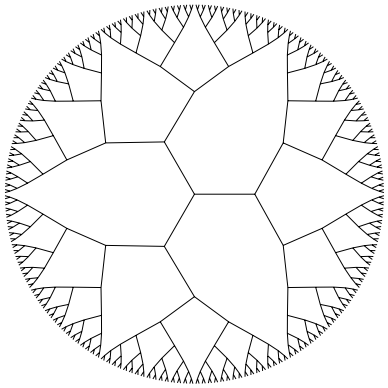


Figure 4.4: Infinite 3-regular tree. Image taken from the excellent survey on expander graphs: Shlomo, Linial, and Wigderson (2006)

Remark 4.16. Note that $2\sqrt{d-1}$ is the spectral radius of the infinite d-regular tree.

4.5 *Ramanujan graphs*

Definition 4.17. A ***Ramanujan graph*** is a d-regular graph whose adjacency matrix has eigenvalues $d = \lambda_1 \geq \cdots \geq \lambda_n$ so that $|\lambda_2|, |\lambda_n| \leq 2\sqrt{d-1}$, i.e. an (n, d, λ)-graph with $\lambda \leq 2\sqrt{d-1}$.

One example of a Ramanujan graph is K_{d+1}, as $\lambda_2 = \cdots = \lambda_n = -1$, but we are more interested in fixing d. For fixed d, do there exist infinitely many d-regular Ramanujan graphs?

Conjecture 4.18. *For all $d \geq 3$, there exist infinitely many d-regular Ramanujan graphs.*

We will discuss some partial results towards this conjecture.

Theorem 4.19 (Lubotzky–Phillips–Sarnak, Margulis)**.** *The above conjecture is true for all d with $d-1$ prime.*

Lubozsky, Phillips, and Sarnak (1988)
Margulis (1988)

Theorem 4.19 is proven by explicitly constructing a Cayley graph on the group $PSL(2,q)$ by invoking deep results from number theory relating to conjectures of Ramanujan, which is where the name comes from. In 1994, Morgenstern strengthened Theorem 4.19 result to all d for which $d-1$ is a prime power. This is essentially all that is known. In particular, Conjecture 4.18 is open for $d = 7$.

Morgenstern (1994)

It is interesting to consider the case of random graphs. What is the distribution of the largest non-λ_1 eigenvalue?

Theorem 4.20 (Friedman)**.** *Fix $d \geq 3$. A random n-vertex d-regular graph is, with prability $1 - o(1)$, a nearly-Ramanujan graph in the sense that*

$$\max\{|\lambda_2|, |\lambda_n|\} \leq 2\sqrt{d-1} + o(1)$$

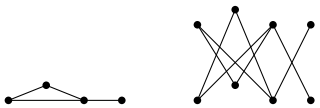
where the $o(1)$ term goes to 0 as $n \to \infty$.

Friedman (2004)

Experimental evidence suggests that, for all fixed d, a fixed proportion (between 0 and 1) of graphs on n vertices should be Ramanujan as $n \to \infty$. However, no rigorous results are known in this vein.

Recently, there has been some important progress on a bipartite analogue of this problem:

Note that for all bipartite graphs, $\lambda_i = -\lambda_{n+1-i}$. To see this, let the parts be A and B and take an eigenvector v with eigenvalue λ. Let v consist of v_A on A and v_B on B. Then negating v_B gives an eigenvector v' with eigenvalue $-\lambda$. So, a bipartite graph is called ***bipartite Ramanujan*** if $\lambda_2 \leq 2\sqrt{d-1}$.



G $\quad$ $G \times K_2$

An example of a graph G and its corresponding graph $G \times K_2$

Every Ramanujan graph G has an associated bipartite Ramanujan graph: we can construct $G \times K_2$; if G has eigenvalues $\{\lambda_i\}$ then $G \times K_2$ has eigenvalues $\{\lambda_i\} \cup \{-\lambda_i\}$, so the d-regular bipartite Ramanujan graph problem is a weakening of the original problem.

Theorem 4.21 (Marcus–Spielman–Srivastava)**.** *For all d, there exist infinitely many d-regular bipartite Ramanujan graphs.*

Marcus, Spielman, and Srivastava (2015)

Theorem 4.21 uses a particularly clever construction of randomized graphs.

4.6 *Sparse graph regularity and the Green–Tao theorem*

We will now combine the concepts of pseudorandom graphs with regularity involving sparse graphs. ***Sparse*** means edge density $o(1)$

— here we always consider a sequence of graphs on n vertices as $n \to \infty$, and $o(1)$ is with respect to n. The naïve analogue of the triangle removal lemma in a sparse setting is not true; we need an additional constraint:

Meta-Theorem 4.22 (Sparse triangle removal lemma)**.** For all $\epsilon > 0$, there exists $\delta > 0$ so that, if Γ is a sufficiently pseudorandom graph on n vertices with edge density p and G is a subgraph of Γ with fewer than $\delta n^3 p^3$ triangles, then G can be made triangle-free by deleting $\epsilon n^2 p$ edges.

We call this a ***meta-theorem*** as the condition "sufficiently pseudorandom" is not made explicit: the result is precisely true for some pseudorandomness conditions on which we will elaborate later. We can consider the traditional triangle removal lemma to be a special case of this where Γ is a complete graph.

Remark 4.23. Meta-Theorem 4.22 is not true without the hypothesis of Γ: take G as in Corollary 3.18 to have n vertices and $n^{2-o(1)}$ edges, where every edge belongs to exactly one triangle.

Remark 4.24. If $\Gamma = G(n,p)$ is an Erdős–Rényi graph with $p \geq \frac{C}{\sqrt{n}}$, then the conclusion of Meta-Theorem 4.22 holds. Conlon and Gowers (2014)

The motivation for the above is the Green–Tao Theorem:

Theorem 4.25 (Green–Tao)**.** *The primes contain arbitrarily long arithmetic progressions.* Green and Tao (2008)

This is in some sense a sparse extension of Szemerédi's Theorem: the density of the primes up to n decays like $\frac{1}{\log n}$ by the Prime Number Theorem.

The strategy for proving the Theorem 4.25 is to start with the primes and embed them (with high relative density) in what we will call *pseudoprimes*: numbers with no small prime divisors. This set is easier to analyze with analytic number theory, specifically using sieve methods. In particular, we can more easily show that the pseudoprimes are sufficiently pseudorandom, allowing the use of sparse hypergraph removal lemmas.

Recall the three main steps of using regularity: partitioning, cleaning, and counting. Naïve attempts to apply this approach to prove the sparse triangle removal lemma result in serious difficulties, and new ideas are needed. We require a sparse notion of regularity separate from the standard notion:

Definition 4.26. Given a graph G, a pair $(A,B) \subset V(G)^2$ is called ***(ϵ,p)-regular*** if, for all $U \subset A, W \subset B$ with $|U| \geq \epsilon|A|, |W| \geq \epsilon|B|$, then

$$|d(U,W) - d(A,B)| < \epsilon p.$$

An equitable partition $V(G) = V_1 \sqcup \cdots \sqcup V_k$ is said to be (ϵ, p)-***regular*** if all but at most ϵ proportion of pairs are (ϵ, p)-regular.

Theorem 4.27 (Sparse regularity lemma)**.** *For all $\epsilon > 0$ there exists some $M \in \mathbb{N}$ for which every graph with edge density at most p has an (ϵ, p)-regular partition into at most M parts.*

Scott (2010)

Sparse objects have in some sense more freedom of structure, which is why statements like the sparse regularity lemma are much more intricate than the dense regularity lemma.

Theorem 4.27 is true but quite misleading: it could be true that most edges are inside irregular pairs. This makes the cleaning step more difficult as it might clean away too many of your edges. One example of this is a clique on $o(n)$ vertices.

In practice, G is often assumed to satisfy some "upper-regularity" hypothesis. For example, a graph is said to have ***no dense spots*** if there exists $\eta = o(1)$ and a constant $C > 0$ such that, for all $X, Y \subseteq V(G)$, if $|X|, |Y| \geq \eta |V|$, then

$$d(X, Y) \leq Cp.$$

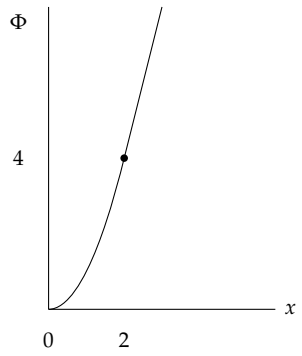
We will now prove Theorem 4.27 with the "no dense spots" hypothesis:

Proof sketch of Theorem 4.27 under the "no dense spots" hypothesis. This is essentially the same proof as in Szemerédi's Regularity Lemma. The key property we used in the energy increment argument was that the energy was bounded above by 1 and increased by ϵ^5. Now the energy increases by $\epsilon^5 p^2$. This depends on p, which could break the proof. However, as there are no dense spots, the final energy is at most $O(C^2 p^2)$, so the number of bad steps is bounded (depending on ϵ). □

Theorem 4.27 is still true without the condition "no dense spots," however:

Proof sketch of Theorem 4.27 in generality. We repeat the proof of Theorem 3.5 and instead of using x^2 as the energy, consider

$$\Phi(x) = \begin{cases} x^2 & \text{if } 0 \leq x \leq 2 \\ 4x - 4 & \text{if } x > 2. \end{cases}$$



Scott's energy function $\Phi(x)$.

This function has the boosting step: for all random variables $X \geq 0$, if $\mathbb{E}[X] \leq 1$,

$$\mathbb{E}\Phi(X) - \Phi(\mathbb{E}X) \geq \frac{1}{4} \operatorname{Var} X.$$

Furthermore, the inequality

$$\mathbb{E}\Phi(X) \leq 4\mathbb{E}X$$

allows us to bound the total energy of a partition by $O(1)$. □

Theorem 4.27 shows that the hard part of Meta-Theorem 4.22 is not the regularity lemma but the counting step. There is no counting lemma for sparse regular graphs. However, given our hypothesis that G is a subgraph of a pseudorandom graph Γ, we can construct a counting lemma which will allow us to prove the sparse triangle removal lemma.

We want something like the following to be true:

If you have three sets V_1, V_2, V_3 so that (V_i, V_j) are (ϵ, p)-regular $\forall i \neq j$ with edge density $d_{i,j}$, the number of triangles with one vertex in each part is

$$(d_{12}d_{23}d_{31} + O(\epsilon^c))\, p^3 |V_1||V_2||V_3|.$$

However, no such statement holds; take $G(n,p)$ with $p \ll \frac{1}{\sqrt{n}}$ and remove an edge from each triangle.

There is another example, due to Alon:

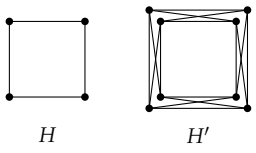
Example 4.28. There exists a triangle-free pseudorandom d-regular graph Γ with $d = \Theta\left(n^{2/3}\right)$ that is a (n, d, λ)-graph with $\lambda = \Theta\left(\sqrt{d}\right)$. Alon (1995)

To fix the issues with the above attempt, we have the following "meta-theorem:"

Meta-Theorem 4.29. Given three sets V_1, V_2, V_3 in G where G is a subgraph of a sufficiently pseudorandom graph with edge density p so that (V_i, V_j) are (ϵ, p)-regular for all $i \neq j$ with edge density $d_{i,j}$, the number of triangles with one vertex in each part is

$$(d_{12}d_{23}d_{31} + O(\epsilon^c))\, p^3 |V_1||V_2||V_3|.$$

We will now create a precise "sufficiently pseudorandom" condition for Meta-Theorem 4.22 and Meta-Theorem 4.29. We say that, given a graph H, a graph Γ is *pseudorandom with respect to H-density* if it has H-density $(1 + o(1))p^{e(H)}$. It turns out that the sparse triangle counting lemma Meta-Theorem 4.22 holds if Γ is pseudorandom with respect to H-density for every subgraph H of $K_{2,2,2}$.

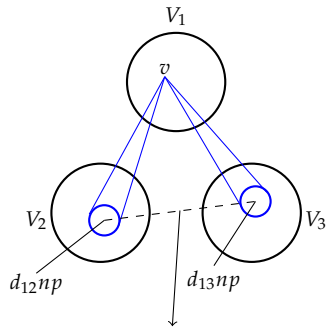


H and its 2-blowup H'.

Remark 4.30. This condition cannot necessarily be replaced by any of the other conditions given in Theorem 4.1 as our implication chain does not hold in a sparse setting.

This plays an analogous role to the C_4 condition in Theorem 4.1; C_4 was the ***2-blowup*** of an edge, while $K_{2,2,2}$ is a 2-blowup of a triangle. This acts somewhat like a graph-theoretic analogue of a second-moment: controlling copies of a graph H's second moment allows us to control copies of H in a subset of $V(G)$.

The proof Theorem 3.13 no longer works in the sparse case. Given three parts V_1, V_2, and V_3, that are pairwise (ϵ, p)-regular, we can no



There are not enough vertices to use (ϵ, p)-regularity.

longer take the neighbors of a vertex in V_1 that are in V_2 and V_3 and say that, as there are enough of them, they have enough overlap. This fails due to the extra factor of p in the sparse case.

Theorem 4.31 (Sparse counting lemma). *There exists a sparse counting lemma for counting H in $G \subset \Gamma$ if Γ is pseudorandom with respect to the density of every subgraph of the 2-blowup of H.*

Conlon, Fox, and Zhao (2015)

With this sparse counting lemma, one can prove Meta-Theorem 4.22 with the same proof structure as that of Theorem 3.15, using this pseudorandom property as our "sufficiently pseudorandom" condition on Γ.

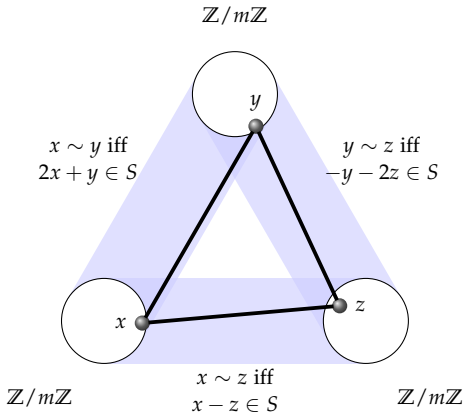
We state a equivalent version of Roth's theorem (Theorem 3.19):

Theorem 4.32 (Density Roth's Theorem). *If $A \subset \mathbb{Z}/n\mathbb{Z}$ with $|A| = \delta n$, then A contains at least $c(\delta)n^2$ 3-APs where $c(\delta) > 0$ is a constant depending only on δ.*

This can be proven by applying the proof structure from the proof of Theorem 3.19 using Theorem 3.15 (alternatively, we can use a supersaturation argument). Similarly to this, we can use Meta-Theorem 4.22 to prove a sparse analogue of Roth's Theorem:

Meta-Theorem 4.33 (Relative Roth's Theorem). If $S \subset \mathbb{Z}/n\mathbb{Z}$ is sufficiently pseudorandom with $|S| = pn$, and $A \subset S$ with $|A| \geq \delta|S|$, then A contains at least $c(\delta)n^2p^3$ 3-APs where $c(\delta) > 0$ is a constant depending only on δ.

What should "pseudorandom" mean here? Recall our proof of Roth's Theorem: creating three copies X, Y, Z of $\mathbb{Z}/n\mathbb{Z}$ and putting edges among $x \in X, y \in Y, z \in Z$ if $2x + y \in S$, $x - z \in S$, $-y - 2z \in S$. From this construction, we can read out the pseudorandom properties we want this graph Γ_S to have from our counting lemma.



Definition 4.34. We say that $S \subset \mathbb{Z}/n\mathbb{Z}$ satisfies a ***3-linear-forms condition*** if, for uniformly randomly chosen $x_0, x_1, y_0, y_1, z_0, z_1 \in \mathbb{Z}/n\mathbb{Z}$, the probability that the twelve numbers formed by the linear forms corresponding to those above:

$$\left\{ \begin{array}{lll} -y_0 - 2z_0, & x_0 - z_0, & 2x_0 + y_0, \\ -y_1 - 2z_0, & x_1 - z_0, & 2x_1 + y_0, \\ -y_0 - 2z_1, & x_0 - z_1, & 2x_0 + y_1, \\ -y_1 - 2z_1, & x_1 - z_1, & 2x_1 + y_1 \end{array} \right\}$$

are all in S is within a $1 + o(1)$ factor of the expectation if $S \subset \mathbb{Z}/n\mathbb{Z}$ were random with density p, and the same holds for any subset of these 12 expressions.

We also have a corresponding theorem, a simplification of the Relative Szemerédi Theorem used by Green–Tao:

Green and Tao (2008)

Theorem 4.35 (Relative Szemerédi Theorem). *Fix $k \geq 3$. If $S \subset \mathbb{Z}/n\mathbb{Z}$ satisfies the k-linear-forms condition then any $A \subset S$ with $|A| \geq \delta|S|$ has a lot of k-APs.* Conlon, Fox, and Zhao (2015)

There are still interesting open problems involving sparse regularity, particularly involving what sorts of pseudorandomness hypotheses are required to get counting lemmas.

Remark 4.36. Theorems like Theorem 4.35 can also be proven without the use of regularity, in particular by using the technique of ***transference***: Szemerédi's Theorem can be treated as a black box, and applied directly to the sparse setting. For more about this, see "Green–Tao theorem: an exposition" by Conlon, Fox, Zhao. Conlon, Fox, and Zhao (2014)

5
Graph limits

5.1 Introduction and statements of main results

Graph limits seeks a generalization of analytic limits to graphs. Consider the following two examples that shows the potential parallel between the set of rational numbers and graphs:

Example 5.1. For $x \in [0,1]$, the minimum of $x^3 - x$ occurs at $x = 1/\sqrt{3}$. But if we restrict ourselves in $\mathbb{Q}$ (pretending that we don't know about real numbers), a way to express this minimum is to find a sequence $x_1, x_2, \ldots$ of rational numbers that converges to $1/\sqrt{3}$.

Example 5.2. Given $p \in (0,1)$, we want to minimize the density of C_4's among all graphs with edge density p. From Theorem 4.1 we see that the minimum is p^4, which is obtained via a sequence of quasirandom graphs. (There is no single finite graph that obtains this minimum.)

We can consider the set of all graphs as a set of discrete objects (analogous to $\mathbb{Q}$), and seek its "completion" (analogously $\mathbb{R}$).

Definition 5.3. A ***graphon*** ("graph function") is a symmetric measurable function $W : [0,1]^2 \to [0,1]$.

Remark 5.4. Definition 5.3 can be generalized to $\Omega \times \Omega \to [0,1]$ where Ω is any measurable probability space, but for simplicity we will usually work with $\Omega = [0,1]$. (In fact, most "nice" measurable probability space can be represented by $[0,1]$.)

The codomain of the function can also be generalized to $\mathbb{R}$, in which case we will refer to the function as a ***kernel***. Note that this naming convention is not always consistent in literature.

Graphons can be seen as a generalized type of graphs. In fact, we can convert any graph into a graphon, which allow us to start imagining what the limits of some sequences of graph should look like.

Example 5.5. Consider a ***half graph*** G_n, which is a bipartite graph where one part is labeled $1, 2, \ldots, n$ and the other part is labeled $n+1, \ldots, 2n$, and vertices i and $n+j$ is connected if and only if $i \leq j$. If we treat the adjacency matrix $\mathrm{Adj}(G_n)$ as a $0/1$ bit image, we can define graphon $W_{G_n} : [0,1]^2 \to [0,1]$ (which consists of $(2n)^2$ "pixels" of size $1/(2n) \times 1/(2n)$ each). When n goes to infinity, the graphon converges (pointwise) to a function that looks like Figure 5.2.

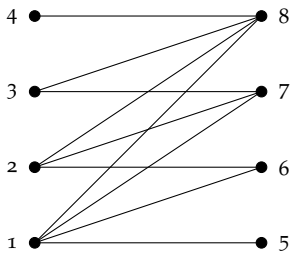
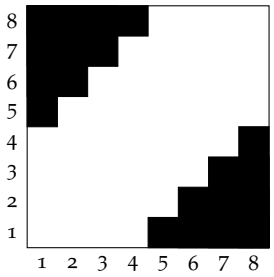


Figure 5.1: The half graph G_n for $n = 4$

This process of converting graphs to graphons can be easily generalized.

Definition 5.6. Given a graph G with n vertices (labeled $1, \ldots, n$), we define its ***associated graphon*** as $W_G : [0,1]^2 \to [0,1]$ obtained by partitioning $[0,1] = I_1 \cup I_2 \cup \cdots I_n$ with $\lambda(I_i) = 1/n$ such that if $(x,y) \in I_i \times I_j$, then $W(x,y) = 1$ if i and j are connected in G and 0 otherwise. (Here $\lambda(I)$ is the Lebesgue measure of I.)

However, as we experiment with more examples, we see that using pointwise limit as in Example 5.5 does not suffice for our purpose in general.



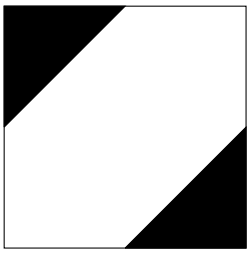


Figure 5.2: The graph of W_{G_n} (for $n = 4$) and the limit as n goes to infinity (black is 1, white is 0)

Example 5.7. Consider any sequence of random (or quasirandom) graphs with edge density $1/2$ (with number of vertices approaching infinity), then the limit (should) approach the constant function $W = 1/2$, though it certainly does not do so pointwise.

Example 5.8. Consider a complete bipartite graph $K_{n,n}$ with the two parts being odd-indexed and even-indexed vertices. Since the adjacency matrix looks like a checkerboard, we may expect limit to look like the $1/2$ constant function as well, but this is not the case: if we instead label the two parts $1, \ldots, n$ and $n+1, \ldots 2n$, then we see that the graphons should in fact converge to a 2×2 checkerboard instead.

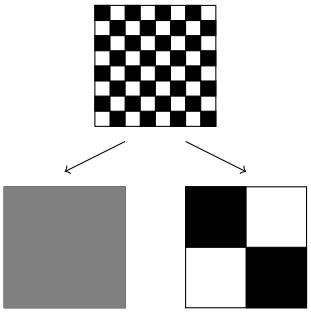


Figure 5.3: A graph of $W_{K_{n,n}}$ and two possible limits of $W_{K_{n,n}}$ as n goes to infinity

The examples above show that we need to (at the very least) take care of relabeling of the vertices in our definition of graph limits.

Definition 5.9. A ***graph homomorphism*** from H to G is a map $\phi : V(H) \to V(G)$ such that if $uv \in E(H)$ then $\phi(u)\phi(v) \in E(G)$. (Maps edges to edges.) Let $\mathrm{Hom}(H,G)$ be the set of all such homomorphisms. and let $\hom(H,G) = |\mathrm{Hom}(H,G)|$. Define ***homomorphism density*** as

$$t(H,G) = \frac{\hom(H,G)}{|V(G)|^{|V(H)|}}.$$

This is also the probability that a uniformly random map is a homomorphism.

Example 5.10. • $\hom(K_1, G) = |V(G)|$,

- $\hom(K_2, G) = 2|E(G)|$,
- $\hom(K_3, G)$ is 6 times the number of triangles in G,
- $\hom(G, K_3)$ is the number of proper 3-colorings of G (where the colors are labeled, say red/green/blue).

Remark 5.11. Note that the homomorphisms from H to G do not quite correspond to copies of subgraphs H inside G, because the homomorphisms can be non-injective. Since the number of non-injective homomorphisms contribute at most $O_H(n^{|V(H)-1|})$ (where $n = |V(G)|$), they form a lower order contribution as $n \to \infty$ when H is fixed.

Definition 5.12. Given a symmetric measurable function $W : [0,1]^2 \to \mathbb{R}$, define

$$t(H, W) = \int_{[0,1]^{|V(H)|}} \prod_{ij \in E(H)} W(x_i, x_j) \prod_{i \in V(H)} dx_i.$$

Note that $t(H, G) = t(H, W_G)$ for every G and H.

Example 5.13. When $H = K_3$, we have

$$t(K_3, W) = \int_{[0,1]^3} W(x,y)W(y,z)W(z,x)\, dxdydz.$$

This can be viewed as the "triangle density" of W.

We may now define what it means for graphs to converge and what the limit is.

Definition 5.14. We say that a sequence of graphs G_n (or graphons W_n) is ***convergent*** if $t(H, G_n)$ (or $t(H, W_n)$) converges as n goes to infinity for every graph H. The sequence ***converges to*** W if $t(H, G_n)$ (or $t(H, W_n)$) converges to $t(H, W)$ for every graph H.

Remark 5.15. Though not necessary for the definition, we can think of $|V(G_n)|$ going to infinity as n goes to infinity.

A natural question is whether a convergent sequence of graphs has a "limit". (Spoiler: yes.) We should also consider whether the "limit" we defined this way is consistent with what we expect. To this end, we need a notion of "distance" between graphs.

One simple way to define the distance between G and G' to be $\sum_k 2^{-k}|t(H_k, G) - t(H_k, G')|$ for some sequence $H_1, H_2, \ldots$ of all the graphs. (Here 2^{-k} is added to make sure the sum converges to a number between 0 and 1.) This is topologically equivalent to the concept of convergence in Definition 5.14, but it is not useful.

Another possibility is to consider the ***edit distance*** between two graphs (number of edge changes needed), normalized by a factor of

$1/|V(G)|^2$. This is also not very useful, since the distance between any two $G(n, 1/2)$ is around $1/4$, but we should expect them to be similar (and hence have $o(1)$ distance).

This does, however, inspires us to look back to our discussion of quasirandom graphs and consider when a graph is close to constant p (i.e. similar to $G(n, p)$). Recall the DISC criterion in Theorem 4.1, where we expect $|e(X, Y) - p|X||Y||$ to be small if the graph is sufficiently random. We can generalize this idea to compare the distance between two graphs: intuitively, two graphs (on the same vertex set, say) are close if $|e_G(X, Y) - e_{G'}(X, Y)|/n^2$ is small for all subsets X and Y. We do, however, need some more definitions to handle (for example) graph isomorphisms (which should not change the distances) and graphs of different sizes.

Definition 5.16. The ***cut norm*** of $W : [0, 1]^2 \to \mathbb{R}$ is defined as

$$\|W\|_\square = \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W \right|,$$

where S and T are measurable sets.

For future reference, we also define some related norms.

Definition 5.17. For $W : [0, 1]^2 \to \mathbb{R}$, define the L^p ***norm*** as $\|W\|_p = \left(\int |W|^p\right)^{1/p}$, and the L^∞ ***norm*** as the infimum of all the ream numbers m such that the set of all the points (x, y) for which $W(x, y) > m$ has measure zero. (This is also called the ***essential supremum*** of W.)

Definition 5.18. We say that $\phi : [0, 1] \to [0, 1]$ is ***measure-preserving*** if $\lambda(A) = \lambda(\phi^{-1}(A))$ for all measurable $A \subseteq [0, 1]$.

Example 5.19. The function $\phi(x) = x + 1/2 \bmod 1$ is clearly measure-preserving. Perhaps less obviously, $\phi(x) = 2x \bmod 1$ is also measure-preserving, since while each interval is dilated by a factor of 2 under ϕ, every point has two pre-images, so the two effects cancel out. This only works because we compare A with $\phi^{-1}(A)$ instead of $\phi(A)$.

Definition 5.20. Write $W^\phi(x, y) = W(\phi(x), \phi(y))$ (intuitively, "relabelling the vertices"). We define the ***cut distance***

$$\delta_\square(U, W) = \inf_\phi \|U - W^\phi\|_\square$$

where ϕ is a measure-preserving bijection.

For graphs G, G', define the ***cut distance*** $\delta_\square(G, G') = \delta_\square(W_G, W_{G'})$. We also define the cut distance between a graph and a graphon as $\delta_\square(G, U) = \delta_\square(W_G, U)$.

Note that ϕ is not quite the same as permuting vertices: it is allowed to also split vertices or overlay different vertices. This allows us to optimize the minimum discrepancy/cut norm better than simply considering graph isomorphisms.

Remark 5.21. The inf in the definition is indeed necessary. Suppose $U(x,y) = xy$ and $W = U^\phi$, where $\phi(x) = 2x \bmod 1$, we cannot attain $\|U - W^{\phi'}\|_\square = 0$ for any ϕ' (although the cut distance is 0) since ϕ is not bijective.

Now we present the main theorems in graph limit theory that we will prove later. First of all, one might suspect that there is an alternative definition of convergence using the cut distance metric, but it turns out that this definition is equivalent to Definition 5.14.

Theorem 5.22 (Equivalence of convergence)**.** *A sequence of graphs or graphons is convergent if and only if it is a Cauchy sequence with respect to the cut (distance) metric.*

Borgs, Chayes, Lovász, Sós, and Vesztergombi (2008)

(A Cauchy sequence with respect to metric d is a sequence $\{x_i\}$ that satisfies $\sup_{m\geq 0} d(x_n, x_{n+m}) \to 0$ as $n \to \infty$.)

Theorem 5.23 (Existence of limit)**.** *Every convergent sequence of graphs or graphons has a limit graphon.*

Lovász and Szegedy (2006)

Denote $\tilde{\mathcal{W}}_0$ as the space of graphons, where graphons with cut distance 0 are identified.

Theorem 5.24 (Compactness of the space of graphons)**.** *The set $\tilde{\mathcal{W}}_0$ is a compact metric space under the cut metric.*

Lovász and Szegedy (2007)

Remark 5.25. Intuitively, this means that the spaces of "essentially different" graphs is not very large. This is similar to the regularity lemma, where every graph has a constant-size description that approximates the graph well. In fact, we can consider this compactness theorem as a qualitative analytic version of the regularity lemma.

5.2 *W-random graphs*

Recall the Erdős-Rényi random graphs $G(n,p)$ we've seen before. We now introduce its graphon generalization. Let's start with a special case, **the stochastic block model**. It is a graph with vertices colored randomly (blue or red), and two red vertices are connected with probability p_{rr}, a red vertex and a blue vertex are connected with probability $p_{rb} = p_{br}$, and two blue vertices are connected with probability p_{bb}.

Definition 5.26. Uniformly pick $x_1, \dots, x_n$ from the interval $[0,1]$. A ***W*-random graph**, denoted $\mathbb{G}(n,W)$, has vertex set $[n]$ and vertices i and j are connected with probability $W(x_i, x_j)$.

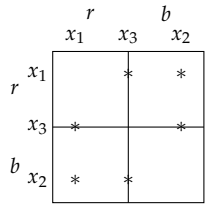


Figure 5.4: 2-block model

An important statistical question is that given a graph, whether there is a good model for where this graph comes from. This gives some motivation to study W-random graphs. We also learnt that the sequence of Erdős-Rényi random graphs converges to the constant graphon, where below is an analogous result.

Theorem 5.27. *Let W be a graphon. Suppose that for all n, G_n are chosen from W-random graphs independently, then $G_n \to W$ almost surely.*

Remark 5.28. In particular, every graphon W is the limit of some sequence of graphs. This gives us some form of graph approximations.

The proof for the above theorem uses Azuma's inequality in order to show that $t(F, G_n) \approx t(F, W)$ with high probability.

5.3 Regularity and counting lemmas

We now develop a series of tools to prove Theorem 5.24.

Theorem 5.29 (Counting Lemma)**.** *For graphons W, U and graph F, we have*

$$|t(F, W) - t(F, U)| \leq |E(F)|\, \delta_\square(W, U).$$

Proof. It suffices to prove $|t(F, W) - t(F, U)| \leq |E(F)| \|W - U\|_\square$. Indeed, by considering the above over U replaced by U^ϕ, and taking the infimum over all measure-preserving bijections ϕ, we obtain the desired result.

Recall that the cut norm $\|W\|_\square = \sup_{S,T \subseteq [0,1]} |\int_{S \times T} W|$. Now we prove its useful reformulation: for measurable functions u and v,

$$\sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W \right| = \sup_{u,v : [0,1] \to [0,1]} \left| \int_{[0,1]^2} W(x,y) u(x) v(y) dx dy \right|.$$

Here's the reason for the equality to hold: we take $u = 1_S$ and $v = 1_T$ so the left hand side is no more than the right hand side, and then the bilinearity of the integrand in u, v yields the other direction (the extrema are attained for u, v taking values at 0 or 1).

We now illustrate the case when $F = K_3$. Observe that

$$\begin{aligned} t(K_3, W) - t(K_3, U) &= \int ((W(x,y)W(x,z)W(y,z) - U(x,y)U(x,z)U(y,z)) dx dy dz \\ &= \int (W - U)(x,y) W(x,z) W(y,z) dx dy dz \\ &\quad + \int U(x,y)(W - U)(x,z) W(y,z) dx dy dz \\ &\quad + \int U(x,y) U(x,z) (W - U)(y,z) dx dy dz. \end{aligned}$$

Take the first term as an example: for a fixed z,

$$\left| \int (W - U)(x,y) W(x,z) W(y,z) dx dy dz \right| \leq \|W - U\|_\square$$

by the above reformulation. Therefore, the whole sum is bounded by $3\|W-U\|_\square$ as we desire.

For a general graph F, by the triangle inequality we have

$$|t(F,W)-t(F,U)| = \left|\int \Big(\prod_{u_iv_i\in E} W(u_i,v_i) - \prod_{u_iv_i\in E} U(u_i,v_i)\Big)\prod_{v\in V} dv\right|$$

$$\le \sum_{i=1}^{|E|}\left|\int\left(\prod_{j=1}^{i-1} U(u_j,v_j)(W(u_i,v_i)-U(u_i,v_i))\prod_{k=i+1}^{|E|} W(u_k,v_k)\right)\prod_{v\in V} dv\right|.$$

Here, each absolute value term in the sum is bounded by $\|W-U\|_\square$ the cut norm if we fix all other irrelavant variables (everything except u_i and v_i for the i-th term), altogether implying that $|t(F,W)-t(F,U)| \le |E(F)|\,\delta_\square(W,U)$.

□

We now introduce an "averaging function" for graphon W.

Definition 5.30. For a partition $\mathcal{P}=\{S_1,\dots,S_k\}$ of $[0,1]$ into measurable subsets, and $W:[0,1]^2\to\mathbb{R}$ a symmetrical measurable function, define the ***stepping operator*** $W_\mathcal{P}:[0,1]^2\to\mathbb{R}$ constant on each $S_i\times S_j$ such that $W_\mathcal{P}(x,y)=\frac{1}{\lambda(S_i)\lambda(S_j)}\int_{S_i\times S_j} W$ if $(x,y)\in S_i\times S_j$.

(We ignore the defined term when the denominator equals 0, because the sets are measure-zero anyway).

This is actually a projection in Hilbert space $L^2([0,1]^2)$, onto the subspace of functions constant on each step $S_i\times S_j$. It can also be viewed as the conditional expectation with respect to the σ-algebra generated by $S_i\times S_j$.

Theorem 5.31 (Weak regularity lemma)**.** *For any $\epsilon>0$ and any graphon $W:[0,1]^2\to\mathbb{R}$, there exists a partition $\mathcal{P}$ of $[0,1]$ into no more than $4^{1/\epsilon^2}$ measurable sets such that $\|W-W_\mathcal{P}\|_\square\le\epsilon$.*

Definition 5.32. Given graph G, a partition $\mathcal{P}=\{V_1,\dots,V_k\}$ of $V(G)$ is called ***weakly ϵ-regular*** if for all $A,B\subset V(G)$,

$$\left|e(A,B)-\sum_{i,j=1}^{k} d(V_i,V_j)|A\cap V_i||B\cap V_j|\right| \le \epsilon|V(G)|^2.$$

These are similar but different notions we have seen when introducing Theorem 3.5.

Theorem 5.33 (Weak Regularity Lemma for Graphs)**.** *For all $\epsilon>0$ and graph G, there exists a weakly ϵ-regular partition of $V(G)$ into up to $4^{1/\epsilon^2}$ parts.* Frieze-Kannan (1999)

Lemma 5.34 (L^2 energy increment)**.** *Let W be a graphon and $\mathcal{P}$ a partition of $[0,1]$, satisfying $\|W-W_\mathcal{P}\|_\square>\epsilon$. There exists a refinement $\mathcal{P}'$ of $\mathcal{P}$ dividing each part of $\mathcal{P}$ into no more than 4 parts, such that $\|W_{\mathcal{P}'}\|_2^2>\|W_\mathcal{P}\|_2^2+\epsilon^2$.*

Proof. Because $\|W - W_{\mathcal{P}}\|_{\square} > \epsilon$, there exist subsets $S, T \subset [0,1]$ such that $|\int_{S\times T}(W - W_{\mathcal{P}})| > \epsilon$. Let $\mathcal{P}'$ be the refinement of $\mathcal{P}$ by introducing S and T (divide $\mathcal{P}$ based on whether it's in $S \setminus T$,$T \setminus S$, $S \cap T$ or $\overline{S} \cap \overline{T}$), and that gives at most 4 sub-parts each.

Define $\langle W, U\rangle$ to be $\int WU$. We know that $\langle W_{\mathcal{P}}, W_{\mathcal{P}}\rangle = \langle W_{\mathcal{P}'}, W_{\mathcal{P}}\rangle$ because $W_{\mathcal{P}}$ is constant on each step of $\mathcal{P}$, and $\mathcal{P}'$ is a refinement of $\mathcal{P}$. Thus, $\langle W_{\mathcal{P}'} - W_{\mathcal{P}}, W_{\mathcal{P}}\rangle = 0$. By Pythagorean Theorem,

$$\|W_{\mathcal{P}'}\|_2^2 = \|W_{\mathcal{P}'} - W_{\mathcal{P}}\|_2^2 + \|W_{\mathcal{P}}\|_2^2 > \|W_{\mathcal{P}}\|_2^2 + \epsilon^2,$$

where the latter inequality comes by the Cauchy–Schwarz inequality,

$$\|1_{S\times T}\|_2 \|W_{\mathcal{P}'} - W_{\mathcal{P}}\|_2 \geq |\langle W_{\mathcal{P}'} - W_{\mathcal{P}}, 1_{S\times T}\rangle| = |\langle W - W_{\mathcal{P}}, 1_{S\times T}\rangle| > \epsilon.$$

□

Proposition 5.35. *For any $\epsilon > 0$, graphon W, and $\mathcal{P}_0$ partition of $[0,1]$, there exists partition $\mathcal{P}$ refining part of $\mathcal{P}_0$ into no more than $4^{1/\epsilon^2}$ parts, such that $\|W - W_{\mathcal{P}}\|_{\square} \leq \epsilon$.*

This proposition specifically tells us that starting with any given partition, the regularity argument still works.

Proof. We repeatedly apply Lemma 5.34 to obtain $\mathcal{P}_0, \mathcal{P}_1, \ldots$ partitions of $[0,1]$. For each step, we either have $\|W - W_{\mathcal{P}}\|_{\square} \leq \epsilon$ and thus stop, or we know $\|W_{\mathcal{P}'}\|_2^2 > \|W_{\mathcal{P}}\|_2^2 + \epsilon^2$.

Because $\|W_{P_i}\|_2^2 \leq 1$, we are guaranteed to stop after fewer than than ϵ^{-2} steps. We also know that each part is subdivided into no more than 4 parts at each step, obtaining $4^{\epsilon^{-2}}$ as we desire. □

We hereby introduce a related result in computer science, the MAXCUT problem: given a graph G, we want to find $\max e(S, \bar{S})$ among all vertex subsets $S \subset V(G)$. Polynomial-time approximation algorithms developed by Goemans and Williamson that finds a cut within around 0.878 fraction of the optimum. conjecture known as the Unique Games Conjecture would imply that the it would not be possible to obtain a better approximation than the Goemans–Williamson algorithm.2306295 states the impossibility of beating this. It is shown that approximating beyond $\frac{16}{17} \approx 0.941$ is NP-hard.

Goemans and Williamson (1995)

Khot, Kindler, Mossel, and O'Donnell (2007)

Håstad (2001)

On the other hand, the MAXCUT problem becomes easy to approximate for dense graphs, i.e., approximating the size of the maximum cut of an n-vertex graph with in to ϵn^2 additive error in time polynomial in n, where $\epsilon > 0$ is a fixed constant. One can apply an algorithmic version of the weak regularity lemma and brute-force search through all possible partition sizes of the parts. This application was one of the original motivations of the weak regularity lemma.

5.4 *Compactness of the space of graphons*

Definition 5.36. A ***martingale*** is a random sequence $X_0, X_1, X_2, \ldots$ such that for all n, $\mathbb{E}[X_n | X_{n-1}, X_{n-2}, \ldots, X_0] = X_{n-1}$.

Example 5.37. Let X_n denotes the time n balance at a fair casino, where the expected value of each round's gain is 0. Then $\{X_n\}_{n \geq 0}$ is a martingale.

Example 5.38. For a fixed random variable X, we define $X_n = \mathbb{E}(X|$ information up to time n$)$, so that this sequence also forms a martingale.

Theorem 5.39 (Martingale Convergence Theorem)**.** *Every bounded martingale converges almost surely.*

Remark 5.40. Actually, instead of bounded, it is enough for the martingales to be L^1-bounded or uniform integrable, both of which gives $\sup \mathbb{E}(X_n^+) < \infty$.

We sketch a idea inspired by a betting strategy. The proof below omits some small technical details that can be easily filled in for those who are familiar with the basic language of probability theory.

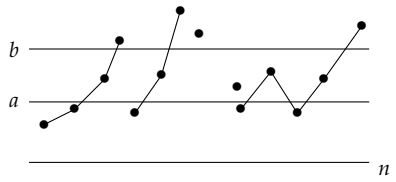


Figure 5.5: examples of "upcrossings"

Proof. An "upcrossing" of $[a, b]$ consists of an interval $[n, n + t]$ such that $X_n < a$, and X_{n+t} is the first instance after X_n such that $X_{n+t} > a$. We refer to the figure on the right instead of giving a more precise definition.

Suppose there is a sequence of bounded martingale $\{X_n\}$ that doesn't converge. Then there exists rational numbers $0 < a < b < 1$ such that $\{X_n\}$ upcrosses the interval $[a, b]$ infinitely many times. We will show that this event occurs with probability 0 (so that after we sum over $a, b \in \mathbb{Q}$, $\{X_n\}$ converges with probability 1).

Denote u_N to be the number of upcrossings (crossings from below to above the interval) up to time N. Consider the following betting strategy: at any time, we hold either 0 or 1 share. If $X_n < a$, then buy 1 share and hold it until the first time that the price (X_n) reads more than b (i.e. we sell at time m such that $X_m > b$ for the first time and $m > n$).

How much profit do we make from this betting strategy? We pocket $b - a$ for each upcrossing. Accounting for difference between our initial and final balance, our profit is at least $(b - a)u_N - 1$. On the other hand, the optional stopping theorem tells us that every "fair" betting strategy on a margingale has zero expected profit. So because the profits of a martingale is zero,

$$0 = \mathbb{E} \text{ profit} \geq (b - a)\mathbb{E}u_N - 1,$$

which implies $\mathbb{E}u_N \leq \frac{1}{b-a}$. Let $u_\infty = \lim_{N\to\infty} u_N$ denotes the total number of upcrossings. By the monotone convergence theorem, we have $\mathbb{E}u_\infty \leq \frac{1}{b-a}$ too, hence $\mathbb{P}(u_\infty = \infty) = 0$, implying our result. □

We now prove the main theorems of graph limits using the tools developed in previous sections, namely the weak regularity lemma (Theorem 5.31) and the martingale convergence theorem (Theorem 5.39). We will start by proving that the space of graphons is compact (Theorem 5.24). In the next section we will apply this result to prove Theorem 5.23 and Theorem 5.22, in that order. We will also see how compactness can be used to prove a graphon-reformulation of the strong regularity lemma.

Recall that $\widetilde{W}_0$ is the space of graphons modulo the equivalence relation $W \sim U$ if $\delta_\square(W, U) = 0$. We can see that $(\widetilde{W}_0, \delta_\square)$ is a metric space.

Theorem 5.41 (Compactness of the space of graphons)**.** *The metric space $(\widetilde{W}_0, \delta_\square)$ is compact.* Lovász and Szegedy (2007)

Proof. As $\widetilde{W}_0$ is a metric space, it suffices to prove sequential compactness. Fix a sequence $W_1, W_2, \ldots$ of graphons. We want to show that there is a subsequence which converges (with respect to $\delta_\square$) to some limit graphon.

For each n, apply the weak regularity lemma (Theorem 5.31) repeatedly, to obtain a sequence of partitions

$$\mathcal{P}_{n,1}, \mathcal{P}_{n,2}, \mathcal{P}_{n,3}, \ldots$$

such that

(a) $\mathcal{P}_{n,k+1}$ refines $\mathcal{P}_{n,k}$ for all n, k,

(b) $|\mathcal{P}_{n,k}| = m_k$ where m_k is a function of only k, and

(c) $\|W_n - W_{n,k}\|_\square \leq 1/k$ where $W_{n,k} = (W_n)_{\mathcal{P}_{n,k}}$.

The weak regularity lemma only guarantees that $|\mathcal{P}_{n,k}| \leq m_k$, but if we allow empty parts then we can achieve equality.

Initially, each partition may be an arbitrary measurable set. However, for each n, we can apply a measure-preserving bijection ϕ to $W_{n,1}$ and $\mathcal{P}_{n,1}$ so that $\mathcal{P}_{n,1}$ is a partition of $[0,1]$ into intervals. For each $k \geq 2$, assuming that $\mathcal{P}_{n,k-1}$ is a partition of $[0,1]$ into intervals, we can apply a measure-preserving bijection to $W_{n,k}$ and $\mathcal{P}_{n,k}$ so that $\mathcal{P}_{n,k}$ is a partition of $[0,1]$ into intervals, and refines $\mathcal{P}_{n,k-1}$. By induction, we therefore have that $\mathcal{P}_{n,k}$ consists of intervals for all n, k. Properties (a) and (b) above still hold. While property (c) may not hold, and it's no longer true that $W_{n,k} = (W_n)_{\mathcal{P}_{n,k}}$, we still know that $\delta_\square(W_n, W_{n,k}) \leq 1/k$ for all n, k. This will suffice for our purposes.

Now, the crux of the proof is a diagonalization argument in countably many steps. Starting with the sequence $W_1, W_2, \ldots$, we will repeatedly pass to a subsequence. In step k, we pick a subsequence $W_{n_1}, W_{n_2}, \ldots$ such that:

1. the endpoints of the parts of $\mathcal{P}_{n_i,k}$ all individually converge as $i \to \infty$, and
2. $W_{n_i,k}$ converges pointwise almost everywhere to some graphon U_k as $i \to \infty$.

There is a subsequence satisfying (1) since each partition $\mathcal{P}_{n,k}$ has exactly m_k parts, and each part has length in $[0,1]$. So consider a subsequence $(W_{a_i})_{i=1}^{\infty}$ satisfying (1). Each $W_{a_i,k}$ can be naturally identified with a function $f_{a_i,k} : [m_k]^2 \to [0,1]$. The space of such functions is bounded, so there is a subsequence $(f_{n_i})_{i=1}^{\infty}$ of $(f_{a_i})_{i=1}^{\infty}$ converging to some $f : [m_k]^2 \to [0,1]$. Now f corresponds to a graphon U_k which is the limit of the subsequence $(W_{n_i})_{i=1}^{\infty}$. Thus, (2) is satisfied as well.

To conclude step k, the subsequence is relabeled as $W_1, W_2, \ldots$ and the discarded terms of the sequence are ignored. The corresponding partitions are also relabeled. Without loss of generality, in step k we pass to a subsequence which contains $W_1, \ldots, W_k$. Thus, the end result of steps $k = 1, 2, \ldots$ is an infinite sequence with the property that $(W_{n,k})_{n=1}^{\infty}$ converges pointwise almost everywhere (a.e.) to U_k for all k:

$$\begin{array}{cccccccc}
 & W_1 & W_2 & W_3 & \ldots & & \\
k=1 & W_{1,1} & W_{2,1} & W_{3,1} & \ldots & \to U_1 & \text{pointwise a.e.} \\
k=2 & W_{1,2} & W_{2,2} & W_{3,2} & \ldots & \to U_2 & \text{pointwise a.e.} \\
k=3 & W_{1,3} & W_{2,3} & W_{3,3} & \ldots & \to U_3 & \text{pointwise a.e.} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots &
\end{array}$$

Similarly, $(\mathcal{P}_{n,k})_{n=1}^{\infty}$ converges to an interval partition $\mathcal{P}_k$ for all k.

By property (a), each partition $\mathcal{P}_{n,k+1}$ refines $\mathcal{P}_{n,k}$, which implies that $W_{n,k} = (W_{n,k+1})_{\mathcal{P}_{n,k}}$. Taking $n \to \infty$, it follows that $U_k = (U_{k+1})_{\mathcal{P}_k}$ (see Figure 5.6 for an example). Now each U_k can be thought of as a random variable on probability space $[0,1]^2$. From this view, the equalities $U_k = (U_{k+1})_{\mathcal{P}_k}$ exactly imply that the sequence $U_1, U_2, \ldots$ is a martingale.

The range of each U_k is contained in $[0,1]$, so the martingale is bounded. By the martingale convergence theorem (Theorem 5.39), there exists a graphon U such that $U_k \to U$ pointwise almost everywhere as $k \to \infty$.

Recall that our goal was to find a convergent subsequence of $W_1, W_2, \ldots$ under $\delta_\square$. We have passed to a subsequence by the above diagonalization argument, and we claim that it converges to U under

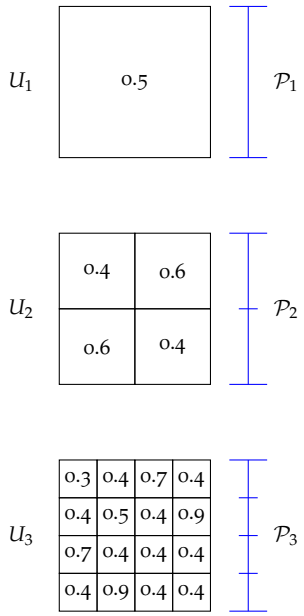


Figure 5.6: An example of possible U_1, U_2, and U_3, each graphon averaging the next.

$\delta_\square$. That is, we want to show that $\delta(W_n, U)_\square \to 0$ as $n \to \infty$. This follows from a standard "3-epsilons argument": let $\epsilon > 0$. Then there exists some $k > 3/\epsilon$ such that $\|U - U_k\|_1 < \epsilon/3$, by pointwise convergence and the dominated convergence theorem. Since $W_{n,k} \to U_k$ pointwise almost everywhere (and by another application of the dominated convergence theorem), there exists some $n_0 \in \mathbb{N}$ such that $\|U_k - W_{n,k}\|_1 < \epsilon/3$ for all $n > n_0$. Finally, since we chose $k > 3/\epsilon$, we already know that $\delta(W_n, W_{n,k})_\square < \epsilon/3$ for all n. We conclude that

$$\begin{aligned}\delta(U, W_n)_\square &\leq \delta(U, U_k)_\square + \delta(U_k, W_{n,k})_\square + \delta(W_{n,k}, W_n)_\square \\ &\leq \|U - U_k\|_1 + \|U_k - W_{n,k}\|_1 + \delta(W_{n,k}, W_n)_\square \\ &\leq \epsilon.\end{aligned}$$

The second inequality uses the general bound that

$$\delta(W_1, W_2)_\square \leq \|W_1 - W_2\|_\square \leq \|W_1 - W_2\|_1$$

for graphons W_1, W_2. □

5.5 *Applications of compactness*

We will now use the compactness of $(\widetilde{W}_0, \delta_\square)$ to prove several results, notably the strong regularity lemma for graphons, the equivalence of the convergence criteria defined by graph homomorphism densities and by the cut norm, and the existence of a graphon limit for every sequence of graphons with convergent homomorphism densities.

As a warm-up, we will prove that graphons can be uniformly approximated by graphs under the cut distance. The following lemma expresses what we could easily prove without compactness:

Lemma 5.42. *For every $\epsilon > 0$ and every graphon W, there exists some graph G such that $\delta_\square(G, W) < \epsilon$.*

Proof. By a well-known fact from measure theory, there is a step function U such that $\|W - U\|_1 < \epsilon/2$. For any constant graphon p there is a graph G such that $\|G - p\|_\square < \epsilon/2$; in fact, a random graph $G(n, p)$ satisfies this bound with high probability, for sufficiently large n. Thus, we can find a graph G such that $\|G - U\|_\square < \epsilon/2$ by piecing together random graphs of various densities. So

$$\delta_\square(G, W) \leq \|W - U\|_1 + \|U - G\|_\square < \epsilon$$

as desired. □

However, in the above lemma, the size of the graph may depend on W. This can be remedied via compactness.

Proposition 5.43. *For every $\epsilon > 0$ there is some $N \in \mathbb{N}$ such that for any graphon W, there is a graph G with N vertices such that $\delta_\square(G, W) < \epsilon$.*

Proof. For a graph G, define the ϵ-ball around G by $B_\epsilon(G) = \{W \in \widetilde{W}_0 : \delta_\square(G, W) < \epsilon\}$.

As G ranges over all graphs, the balls $B_\epsilon(G)$ form an open cover of $\widetilde{W}_0$, by Lemma 5.42. By compactness, this open cover has a finite subcover. So there is a finite set of graphs $G_1, \dots, G_k$ such that $B_\epsilon(G_1), \dots, B_\epsilon(G_k)$ cover $\widetilde{W}_0$. Let N be the least common multiple of the vertex sizes of $G_1, \dots, G_k$. Then for each G_i there is some N-vertex graph G_i' with $\delta_\square(G_i, G_i') = 0$, obtained by replacing each vertex of G_i with $N/|V(G_i)|$ vertices. But now W is contained in an ϵ-ball around some N-vertex graph. □

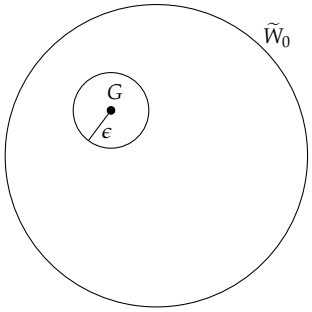


Figure 5.7: Cover of $\widetilde{W}_0$ by open balls

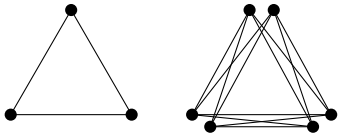


Figure 5.8: A K_3 and its 2-blowup. Note that the graphs define equal graphons.

Remark 5.44. Unfortunately, the above proof gives no information about the dependence of N on ϵ. This is a byproduct of applying compactness. One can use regularity to find an alternate proof which gives a bound.

Intuitively, the compactness theorem has a similar flavor to the regularity lemma; both are statements that the space of graphs is in some sense very small. As a more explicit connection, we used the weak regularity lemma in our proof of compactness, and the strong regularity lemma follows from compactness straightforwardly.

Theorem 5.45 (Strong regularity lemma for graphons)**.** *Let $\epsilon = (\epsilon_1, \epsilon_2, \dots)$ be a sequence of positive real numbers. Then there is some $M = M(\epsilon)$ such that every graph W can be written*

$$W = W_{str} + W_{psr} + W_{sml}$$

where

- *W_{str} is a step function with $k \le M$ parts,*
- *$\|W_{psr}\|_\square \le \epsilon_k$,*
- *$\|W_{sml}\|_1 \le \epsilon_1$.*

Lovász and Szegedy (2007)

If $\epsilon_k = \epsilon/k^2$, then this theorem approximately recovers Szemerédi's Regularity Lemma. If $\epsilon_k = \epsilon$, then it approximately recovers the Weak Regularity Lemma.

Proof. It is a well-known fact from measure theory that any measurable function can be approximated arbitrarily well by a step function. Thus, for every graphon W there is some step function U such that $\|W - U\|_1 \le \epsilon_1$. Unfortunately, the number of steps may depend on W; this is where we will use compactness.

For graphon W, let $k(W)$ be the minimum k such that some k-step graphon U satisfies $\|W - U\|_1 \le \epsilon_1$. Then $\{B_{\epsilon_{k(W)}}\}_{W \in \widetilde{W}_0}$ is clearly an open cover of $\widetilde{W}_0$, and by compactness there is a finite set of graphons $\mathcal{S} \subset \widetilde{W}_0$ such that $\{B_{\epsilon_{k(W)}}(W)\}_{W \in \mathcal{S}}$ covers $\widetilde{W}_0$.

Let $M = \max_{W \in \mathcal{S}} k(W)$. Then for every graphon W, there is some $W' \in \mathcal{S}$ such that $\delta_\square(W, W') \leq \epsilon_{k(W')}$. Furthermore, there is a k-step graphon U with $k = k(W') \leq M$ such that $\|W' - U\|_1 \leq \epsilon_1$. Hence,

$$W = U + (W - W') + (W' - U)$$

is the desired decomposition, with $W_{\text{str}} = U$, $W_{\text{psr}} = W - W'$, and $W_{\text{sml}} = W' - U$. □

Earlier we defined convergence of a sequence of graphons in terms of the sequences of F-densities. However, up until now we did not know that the limiting F-densities of a convergent sequence of graphons are achievable by a single graphon. Without completing the space of graphs to include graphons, this is in fact not true, as we saw in the setting of quasirandom graphs. Nonetheless in the space of graphons, the result is true, and follows swiftly from compactness.

Theorem 5.46 (Existence of limit)**.** *Let $W_1, W_2, \dots$ be a sequence of graphons such that the sequence of F-densities $\{t(F, W_n)\}_n$ converges for every graph F. Then the sequence of graphons converges to some W. That is, there exists a graphon W such that $t(F, W_n) \to t(F, W)$ for every F.*

Lovász and Szegedy (2006)

Proof. By sequential compactness, there is a subsequence $(n_i)_{i=1}^\infty$ and a graphon W such that $\delta_\square(W_{n_i}, W) \to 0$ as $i \to \infty$. Fix a graph F. By Theorem 5.29, it follows that $t(F, W_{n_i}) \to t(F, W)$. But by assumption, the sequence $\{t(F, W_n)\}_n$ converges, so all subsequences have the same limit. Therefore $t(F, W_n) \to t(F, W)$. □

The last main result of graph limits is the equivalence of the two notions of convergence which we had defined previously.

Theorem 5.47 (Equivalence of convergence)**.** *Convergence of F-densities is equivalent to convergence under the cut norm. That is, let $W_1, W_2, \dots$ be a sequence of graphons. Then the following are equivalent:*

Borgs, Chayes, Lovász, Sós, and Vesztergombi (2008)

- *The sequence of F-densities $\{t(F, W_n)\}_n$ converges for all graphs F*
- *The sequence $\{W_n\}_n$ is Cauchy with respect to $\delta_\square$.*

Proof. One direction follows immediately from Theorem 5.29, the counting lemma: if the sequence $\{W_n\}_n$ is Cauchy with respect to $\delta_\square$, then the counting lemma implies that for every graph F, the sequence of F-densities is Cauchy, and therefore convergent.

For the reverse direction, suppose that the sequence of F-densities converges for all graphs F. Let W and U be limit points of $\{W_n\}_n$ (i.e. limits of convergent subsequences). We want to show that $W = U$.

Let $(n_i)_{i=1}^\infty$ be the subsequence such that $W_{n_i} \to W$. By the counting lemma, $t(F, W_{n_i}) \to t(F, W)$ for all graphs F, and by convergence of F-densities, $t(F, W_n) \to t(F, W)$ for all graphs F. Similarly, $t(F, W_n) \to t(F, U)$ for all F. Hence, $t(F, U) = t(F, W)$ for all F.

By the subsequent lemma, this implies that $U = W$. □

Lemma 5.48 (Moment lemma)**.** *Let U and W be graphons such that $t(F, W) = t(F, U)$ for all F. Then $\delta_\square(U, W) = 0$.*

This lemma is named in analogy with the moment lemma from probability, which states that if two random variable have the same moments (and are sufficiently well-behaved) then they are in fact identically distributed.

Proof. We will sketch the proof. Let $\mathbb{G}(k, W)$ denote the W-random graph on k vertices (see Definition 5.26). It can be shown that for any k-vertex graph F,

$$\Pr[\mathbb{G}(k, W) \cong F \text{ as labelled graph}] = \sum_{F' \supseteq F} (-1)^{E(F')-E(F)} t(F', W).$$

In particular, this implies that the distribution of W-random graphs is entirely determined by F-densities. So $\mathbb{G}(k, W)$ and $\mathbb{G}(k, U)$ have the same distributions.

Let $\mathbb{H}(k, W)$ be an edge-weighted W-random graph on vertex set $[k]$, with edge weights sampled as follows. Let $x_1, \ldots, x_k \sim \text{Unif}([0, 1])$ be independent random variables. Set the edge-weight of (i, j) to be $W(x_i, x_j)$.

We claim two facts, whose proofs we omit

- $\delta_\square(\mathbb{H}(k, W), \mathbb{G}(k, W)) \to 0$ as $k \to \infty$ with probability 1, and
- $\delta_1(\mathbb{H}(k, W), W) \to 0$ as $k \to \infty$ with probability 1.

Since $\mathbb{G}(k, W)$ and $\mathbb{G}(k, U)$ have the same distribution, it follows from the above facts and the triangle inequality that $\delta_\square(W, U) = 0$. □

A consequence of compactness and the moment lemma is that the "inverse" of the graphon counting lemma also holds: a bound on F-densities implies a bound on the cut distance. The proof is left as an exercise.

Corollary 5.49 (Inverse counting lemma)**.** *For every $\epsilon > 0$ there is some $\eta > 0$ and integer $k > 0$ such that if U and W are graphons with*

$$|t(F, U) - t(F, W)| \leq \eta$$

for every graph F on at most k vertices, then $\delta_\square(U, W) \leq \epsilon$.

Remark 5.50. The moment lemma implies that a graphon can be recovered by its F-densities. We might ask whether all F-densities are necessary, or whether a graphon can be recovered from, say, finitely many densities. For example, we have seen that if W is the pseudorandom graphon with density p, then $t(K_2, W) = p$ and $t(C_4, W) = p^4$; furthermore, it is uniquely determined by these densities. If the equalities hold then $\delta_\square(W, p) = 0$.

The graphons which can be recovered from finitely many F-densities in this way are called "finitely forcible graphons". Among

the graphons known to be finitely forcible are any step function and the half graphon $W(x,y) = \mathbf{1}_{x+y\geq 1}$. More generally, $W(x,y) = \mathbf{1}_{p(x,y)\geq 0}$ is finitely forcible for any symmetric polynomial $p \in \mathbb{R}[x,y]$ which is monotone decreasing on $[0,1]$.

Lovász and Sós (2008)
Lovász and Szegedy (2011)

5.6 *Inequalities between subgraph densities*

One of the motivations for studying graph limits is that they provide an efficient language with which to think about graph inequalities. For instance, we could be able to answer questions such as the following:

Question 5.51. If $t(K_2, G) = 1/2$, what is the minimum possible value of $t(C_4, G)$?

We know the answer to this question; as discussed previously, by Theorem 4.1 we can consider a sequence of quasirandom graphs; their limit is a graphon W such that $t(K_2, W) = 2^{-4}$.

In this section we work on these kind of problems; specifically, we are interested in homomorphism density inequalities. Two graph inequaities have been discussed previously in this book; Mantel's theorem (Theorem 2.2) and Turán's theorem (Theorem 2.6):

Theorem 5.52 (Mantel's Theorem)**.** *Let $W : [0,1]^2 \to [0,1]$ be a graphon. If $t(K_3, W) = 0$, then $t(K_2, W) \leq 1/2$.*

Theorem 5.53 (Turán's theorem)**.** *Let $W : [0,1]^2 \to [0,1]$ be a graphon. If $t(K_{r+1}, W) = 0$, then $t(K_2, W) \leq 1 - 1/r$.*

Our goal in this section is to determine the set of all feasible ***edge density, triangle density*** pairs for a graphon W, which can be formally written as

$$D_{2,3} = \{(t(K_2, W), t(K_3, W)) \ : \ W \text{ graphon } \} \subseteq [0,1]^2.$$

We know that the limit point of a sequence of graphs is a graphon (Theorem 5.23), hence the region $D_{2,3}$ is closed. Moreover, Mantel's Theorem (Theorem 5.52) tells us that the horizontal section of this region when triangle density is zero extends at most until the point $(1/2, 0) \in [0,1]^2$ (see Figure 5.9).

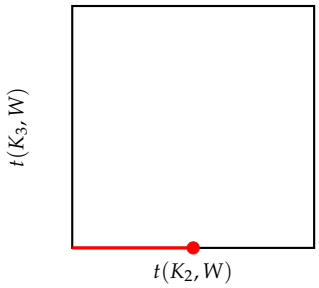


Figure 5.9: Mantel's Theorem implication in the plot of $D_{2,3}$ (red line)

A way in which we can describe $D_{2,3}$ is by its cross sections. A simple argument below shows that each vertical cross section of $D_{2,3}$ is a line segment:

Proposition 5.54. *For every $0 \leq r \leq 1$, the set $D_{2,3} \cap [0,1] \times \{r\}$ is a line segment with no gaps.*

Proof. Consider two graphons W_0, W_1 with the same edge density; then, we can consider

$$W_t = (1-t)W_0 + tW_1,$$

which is a graphon; moreover, its triangle density is mapped continuously as t varies from 0 to 1. Its initial and final values are $t(K_3, W_0)$ and $r(K_3, W_1)$, respectively, so every triangle density between these values can be achieved. □

Then, in order to better understand the shape of $D_{2,3}$, we would like to determine the minimum and maximum subgraph densities that can be achieved given a fixed edge density. We begin by addressing this question:

Question 5.55. What is the maximum number of triangles in an n-vertex m-edge graph?

An intuitive answer would be that the edges should be arranged so as to form a clique. This turns out to be the correct answer: a result known as the Kruskal–Katona theorem implies that a graph with $\binom{k}{2}$ has at most $\binom{k}{3}$ triangles. Here we prove an slightly weaker version of this bound.

The Kruskal–Katona theorem can be proved using a "compression argument": we repeatedly "push" the edges towards the clique and show that number of triangles can never decrease in the process.

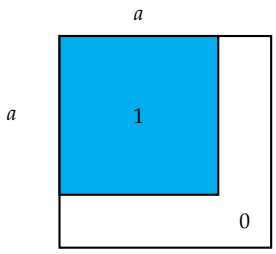


Figure 5.10: Graphon which achieves upper boundary of $D_{2,3}$: $t(K_2, W) = a^2$ and $t(K_3, W) = a^3$

Theorem 5.56. *For every graphon $W : [0,1]^2 \to [0,1]$,*

$$t(K_3, W) \leq t(K_2, W)^{3/2}.$$

Remark 5.57. This upper bound is achieved by a graphon like the one shown in Figure 5.10, which is a limit graphon of a sequence of cliques in G; for each of these graphons, edge and triangle densities are, respectively,

$$t(K_2, W) = a^2, \qquad t(K_3, W) = a^3.$$

Therefore, The upper boundary of the region $D_{3,2}$ is given by the curve $y = x^{3/2}$, as shown by Figure 5.11.

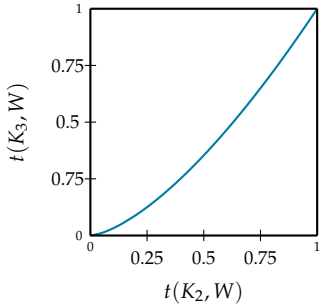


Figure 5.11: Plot of upper boundary of $D_{2,3}$, given by the curve $y = x^{3/2}$ in $[0,1]^2$

Proof of Theorem 5.56. It suffices to prove the following inequality for every graph G:

$$t(K_3, G) \leq t(K_2, G)^{3/2}.$$

Let us look at $\hom(K_3, G)$ and $\hom(K_2, G)$; these count the number of closed walks in the graph of length 3 and 2, respectively. These values correspond to the second and third moments of the spectrum of the graph G:

$$\hom(K_3, G) = \sum_{i=1}^{k} \lambda_i^3 \quad \text{and} \quad \hom(K_2, G) = \sum_{i=1}^{k} \lambda_i^2$$

Where $\{\lambda_i\}_{i=1}^n$ are the eigenvalues of the adjacency matrix A_G. We then have that

$$\hom(K_3, G) = \sum_{i=1}^{n} \lambda_i^3 \leq \left(\sum_{i=1}^{n} \lambda_i^2 \right)^{3/2} = \hom(K_2, G)^{3/2}. \tag{5.1}$$

After dividing by $|V(G)|^3$ on both sides, the result follows. □

Note that in the last proof, we used the following useful inequality, with $a_i = \lambda_i^2$ and $t = 3/2$:

Claim 5.58. Let $t > 1$, and $a_1, \cdots, a_n \geq 0$. Then,

$$a_1^t + \cdots + a_n^t \leq (a_1 + \cdots + a_n)^t$$

Proof. This inequality is homogeneous with respect to the variables a_i, so we can normalize and assume that $\sum a_i = 1$; therefore, each of the $a_i \in [0, 1]$, so that $a_i^t \leq a_i$ for each i. Therefore,

$$LHS = a_1^t + \cdots + a_n^t \leq a_1 + \cdots + a_n = 1 = 1^t = RHS. \qquad \square$$

The reader might wonder whether there is a way to prove this without using eigenvalues of the graph G. We have following result, whose proof does not require spectral graph theory:

Theorem 5.59. *For every $W : [0, 1]^2 \to \mathbb{R}$ which is symmetric,*

$$t(K_3, W) \leq t(K_2, W^2)^{3/2}$$

where W^2 corresponds to the graphon W, squared pointwise.

Note that above, $t(K_2, W)^{3/2}$ falls in between these two terms when W is a graphon because all the terms would be bounded between 0 and 1; therefore, the above result is stronger than that of Theorem 5.56. The proof of this result follows from applying the Cauchy–Schwarz inequality three times; one corresponding to each edge of a triangle K_3.

Proof. We have

$$t(K_3, W) = \int_{[0,1]^3} W(x, y)W(x, z)W(y, z)dxdydz.$$

From now on, we drop the notation for our intervals of integration. We can apply the Cauchy–Schwarz inequality on the following integral; first with respect to the variable dx, and subsequently with respect to the variables dy, dz, each time holding the other two vari-

ables constant:

$$\begin{aligned}
t(K_3, W) &= \int W(x,y)W(x,z)W(y,z)dxdydz \\
&\leq \int \left(\int W(x,y)^2 dx\right)^{1/2} \left(\int W(x,z)^2 dx\right)^{1/2} W(y,z)dydz \\
&\leq \int \left(\int W(x,y)^2 dxdy\right)^{1/2} \left(\int W(x,z)^2 dx\right)^{1/2} \left(\int W(y,z)^2 dy\right)^{1/2} dz \\
&\leq \left(\int W(x,y)^2 dxdy\right)^{1/2} \left(\int W(x,z)^2 dxdz\right)^{1/2} \left(\int W(y,z)^2 dydz\right)^{1/2} \\
&= \|W\|_2^3 \\
&= t(K_2, W)^{3/2},
\end{aligned}$$

completing the proof. □

Remark 5.60. If we did not have the condition that W is symmetric, we could still use Hölder's inequality; however, we would obtain a weaker statement. In this situation, Hölder's inequality would imply that

$$\int_{[0,1]^3} f(x,y)g(x,z)h(y,z)dxdydz \leq \|f\|_3 \|g\|_3 \|h\|_3,$$

and by setting $f = g = h = W$, we could derive a weaker bound than the one obtained in the proof of Theorem 5.59 because, in general, $\|W\|_2 \leq \|W\|_3$.

The next theorem allows us to prove linear inequalities between clique densities.

Theorem 5.61 (Bollobás). *Let $c_1, \cdots, c_n \in \mathbb{R}$. The inequality* Bollobás (1986)

$$\sum_{r=1}^{n} c_r t(K_r, G) \geq 0$$

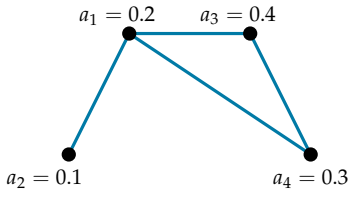
holds for every graph G if and only if it holds for every $G = K_m$ with $m \geq 1$. More explicitly, the inequality holds for all graphs G if and only if

$$\sum_{r=1}^{n} c_r \cdot \frac{m(m-1)\cdots(m-r+1)}{m^r} \geq 0$$

for every $m \geq 1$.

Proof. One direction follows immediately because the set of clique graphs is a subset of the set of all graphs.

We now prove the other direction. The inequality holds for all graphs if and only if it holds for all graphons, again since the set of graphs is dense in $\widetilde{W_0}$ with respect to the cut distance metric. In particular, let us consider the set $\mathcal{S}$ of node-weighted simple graphs, with a normalization $\sum a_i = 1$.



a_4	0	1	1	1
a_3	1	0	0	0
a_2	1	0	0	1
a_1	1	0	1	0
	a_1	a_2	a_3	a_4

Figure 5.12: Example of a node weighted graph on four vertices, whose weights sum to 1, and its corresponding graphon.

As Figure 5.12 shows, each node weighted graph can be represented by a graphon. The set $\mathcal{S}$ is dense in $\widetilde{W_0}$, because this set contains the set of unweighted simple graphs. Then, it suffices to prove this inequality for graphs in $\mathcal{S}$.

Suppose for the sake of contradiction that there exists a node weighted simple graph H such that

$$f(H) := \sum_{r=1}^{n} c_r t(K_r, H) < 0$$

Among all such H, we choose one with smallest possible number m of nodes. We choose node weights $a_1, \cdots, a_m$ with sum equal to 1 such that $f(H)$ is minimized. We can find such H because we have a finite number of parameters, and f is a continuous function over a compact set.

We have that $a_i > 0$ without loss of generality; otherwise we would have a contradiction because we could delete that node and decrease the quantity $|V(H)|$, while $f(H) < 0$ would still hold.

Moreover, H is a complete graph; otherwise there exist i, j such that $ij \notin E(H)$. Note that the clique density is a polynomial in terms of the node weights; this polynomial would not have an a_i^2 term because the set of graphons $\mathcal{S}$ corresponds to simple graphs, and the vertex i would not be adjacent to itself. This polynomial does not have an $a_i a_j$ term either, because i and j are not adjacent. Therefore, $f(H)$ is multilinear in the variables a_i and a_j.

Fixing all of the other node weights and considering a_i, a_j as our variables of the multilinear function $f(H)$, this function would be minimized by setting $a_i = 0$ or $a_j = 0$. If one of these weights were set to zero, this would imply a decrease in the number of nodes, while $a_i + a_j$ would be preserved, hence not increasing $f(H)$. This is a contradiction to the minimality of number of nodes in H such that $f(H) < 0$.

In other words, H must be a complete graph; further, the polynomial $f(H)$ on the variables a_i has to be symmetric:

$$f(H) = \sum_{r=1}^{n} c_r r! s_r,$$

where each s_r is an elementary symmetric polynomial of degree r

$$s_r = \sum_{i_1 < \cdots < i_r} a_{i_1} \cdots a_{i_r}.$$

In particular, by making constant all variables but a_1, a_2, the polynomial $f(H)$ can be written as

$$f(H) = A + B_1 a_1 + B_2 a_2 + C a_1 a_2,$$

where A, B_1, B_2, C are constants; by symmetry, we have $B_1 = B_2$; also, since $\sum a_i = 1$, we have that $a_1 + a_2$ is constant, so that

$$f(H) = A' + Ca_1a_2.$$

If $C > 0$ then f would be minimized when $a_1 = 0$ or $a_2 = 0$; this cannot occur because of the minimality of the number of nodes in H. If $C = 0$ then any value of a_1, a_2 would yield the same minimum value of $f(H)$; in particular we could set $a_1 = 0$, again contradicting minimality on the number of nodes. Therefore, the constant C must be negative,implying that $f(H)$ would be minimized when $a_1 = a_2$. Then, all of the a_i have to be equal, and H can also be regarded as an unweighted graph.

In other words, if the inequality of interest fails for some graph H, then it must fail for some unweighted clique H; this completes the proof. □

Remark 5.62. In the proof above, we only considered clique densities; an inequality over other kinds of graphs would not necessarily hold.

Thanks to the theorem above, it is relatively simple to test linear inequalities between densities, since we just have to verify them for cliques. We have the following corollary:

Corollary 5.63. *For each n, the extremal points of the convex hull of*

$$\{(\ (t(K_2, W), t(K_3, W), \cdots, t(K_n, W))\ :\ W \text{ graphon}\} \subset [0,1]^{n-1}$$

are given by $W = K_m$ for all $m \geq 1$.

Note that the above claim implies Turán's theorem, because by Theorem 5.61, the extrema of the set above are given in terms of clique densities, which can be understood by taking W to be a clique. Thus, if $t(K_{r+1}, W) = 0$, then this cross section on the higher dimensional cube $[0,1]^r$ will be bounded by the value $t(K_2, W) = 1 - \frac{1}{r}$.

In the particular case that we want to find the extremal points in the convex hull of $D_{2,3} \subset [0,1]^2$, they correspond to

$$p_m = \left(\frac{m-1}{m}, \frac{(m-1)(m-2)}{m^2}\right)$$

All of the points of these form in fact fall into the curve given by $y = x(2x-1)$, which is the dotted red curve in Section 5.6.

Because the region $D_{2,3}$ is contained in the convex hull of the red points $\{p_m\}_{m\geq 0}$, it also lies above the curve $y = x(2x-1)$. We can moreover draw line segments between the convex hull points, so as to obtain a polygonal region that bounds $D_{2,3}$.

The region $D_{2,3}$ was determined by Razborov, who developed the theory of ***flag algebras***, which have provided a useful framework in

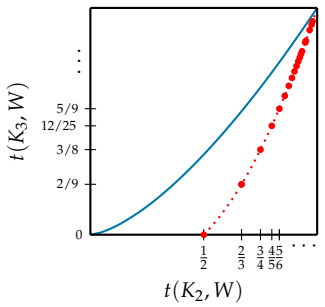


Figure 5.13: Set of lower boundary points of $D_{2,3}$, all found in the curve given by $y = x(2x-1)$
Razborov (2007)

which to set up sum of squares inequalities, e.g., large systematic applications of the Cauchy–Schwarz inequalities, that could be used in order to prove graph density inequalities.

Theorem 5.64 (Razborov). *For a fixed edge density $t(K_2, W)$, which falls into the following interval, for some $k \in \mathbb{N}$*

$$t(K_2, W) \in \left[1 - \frac{1}{k-1}, 1 - \frac{1}{k}\right],$$

the minimum feasible $t(K_3, W)$ is attained by a unique step function graphon corresponding to a k-clique with node weights $a_1, a_2, \cdots, a_k$ with sum equal to 1, and such that $a_1 = \cdots = a_{k-1} \geq a_k$.

Razborov (2008)

The region $D_{2,3}$ is illustrated on the right in Section 5.6. We have exaggerated the drawwings of the concave "scallops" in the lower boundary of the region for better visual effects.

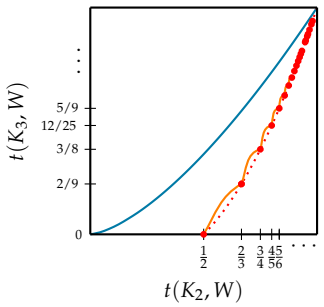


Figure 5.14: Complete description of the region $D_{2,3} \subset [0,1]^2$

Note that in Turán's theorem, the construction for the graphs which correspond to extrema value (Chapter 2, definition 2.5) are unique; however, in all of the intermediate values $t(t_2, W) \neq 1 - 1/k$, this theorem provides us with non-unique constructions.

To illustrate why these constructions are not unique, the graphon in Figure 5.15, which is a minimizer for triangle density when $t(t_2, W) = 2/3$ can be modified by replacing the highlighted region by any graphon with the same edge density.

α_3	0	1	1
α_2	1	0	1
α_1	1	1	0
	α_1	α_2	α_3

Figure 5.15: A non unique optimal graphon in the case $k = 3$.

Nikiforov (2011)

Reiher (2016)

Non-uniqueness of graphons that minimize $t(K_3, W)$ implies that this optimization problem is actually difficult.

The problem of minimizing the K_r-density in a graph of given edge density was solved for $r = 4$ by Nikiforov and all r by Reiher., respectively.

More generally, given some inequality between various subgraph densities, can we decide if the inequality holds for all graphons?

For polynomial inequalities between homomorphism densities, it suffices to only consider linear densities, since $t(H, W)t(H', W) = t(H \sqcup H', W)$.

Let us further motivate with a related, more classical question regarding nonnegativity of polynomials:

Question 5.65. Given a multivariable polynomial $p \in \mathbb{R}[x_1, x_2, \cdots, x_n]$, is $p(x) \geq 0$ for every $x = (x_1, \cdots, x_n)$?

This problem is decidable, due to a classic result of Tarski that every the first-order theory of the reals is decidable. In fact, we have the following characterization of nonnegative real polynomials.

Theorem 5.66 (Artin). *A polynomial $p \in \mathbb{R}[x_1, x_2, \cdots, x_n]$ is nonnegative if and only if it can be written as a sum of squares of rational functions.*

However, when we turn our interest into the set of lattice points, the landscape changes:

Question 5.67. Given a multivariable polynomial $p \in \mathbb{R}[x_1, x_2, \cdots, x_n]$, can it be determined whether $p(x_1, \cdots, x_n) \geq 0$ for all $x \in \mathbb{Z}^n$?

The answer to the above question is no. This is related to the fact that one cannot solve diophantine equations, or even tell whether there is a solution:

Theorem 5.68 (Matiyasevich; Hilbert's 10th problem). *Given a general diophantine equation is an undecidable problem to find its solutions, or even to determine whether integer solutions exist.* Matiyasevich (2011)

Turning back to our original question of interest, we want to know whether the following question is decidable

Question 5.69. For a given set of graphs $\{H_i\}_{i\in[k]}$ and $a_1, \cdots, a_k \in \mathbb{R}$, is $\sum_{i=1}^{k} a_i t(H_i, G) \geq 0$ true for every graph G?

The following theorem provides an answer to this question:

Theorem 5.70 (Hatami - Norine). *Given a set of graphs $\{H_i\}_{i\in[k]}$ and $a_1, \cdots, a_k \in \mathbb{R}$, whether the inequality* Hatami and Norine (2011)

$$\sum_{i=1}^{k} a_i t(H_i, G) \geq 0$$

is true for every graph G is undecidable.

A rough intuition for why the above theorem is true is that we actually have a discrete set of points along the lower boundary of $D_{2,3}$; one could reduce the above problem into proving the same inequalities along the points in the intersection of the red curve and the region. The set of points in this intersection forms a discrete set, and the idea is to encode integer inequalities (which are undecidable) into graph inequalities by using the special points on the lower boundary of $D_{2,3}$.

Another kind of interesting question is to ask whether specific inequalities are true; there are several open problems of that type. Here is an important conjecture in extremal graph theory:

Conjecture 5.71 (Sidorenko's Conjecture). *If H is a bipartite graph then* Sidorenko (1993)

$$t(H, W) \geq t(K_2, W)^{e(H)}.$$

We worked recently with an instance of the above inequality, when $H = C_4$, when we were discussing quasirandomness. However, the above problem is open. Let us consider the Möebius strip graph

- which consists in removing a 10-cycle from a complete bipartite graph $K_{5,5}$ (Section 5.6).

The name of this graph comes from its realization as a face-vertex incidence graph of the usual simplicial complex of the Möebius strip. The graph above is the first one for which this inequality remains an open problem.

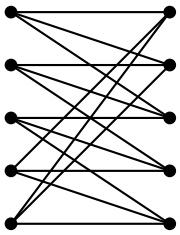


Figure 5.16: The Möebius strip graph.

Even if nonnegativeness of a general linear graph inequalities is undecidable, if one wants to decide whether they are true up to an ε-error, the problem becomes more accessible:

Theorem 5.72. *There exists an algorithm that, for every $\varepsilon > 0$ decides correctly that*

$$\sum_{i=1}^{n} c_i t(H_i, G) \geq -\varepsilon$$

for all graphs G, or outputs a graph G such that

$$\sum_{i=1}^{n} c_i t(H_i, G) < 0.$$

Proof sketch. As a result of weak regularity lemma, we can take a weakly ε-regular partition. All the information regarding edge densities can be represented by this partition; in other words, one would only have to test a bounded number of possibilities on weighted node graphs with $\leq M(\varepsilon)$ parts whose edge weights are multiples of ε. If the estimate for the corresponding weighted sum of graph densities is true for the auxiliary graph one gets from weak regularity lemma, then it is also true for the original graph up to an ε-error; otherwise, we can output a counterexample. □

Part II

Additive combinatorics

6
Roth's theorem

In Chapter 3.3, we proved Roth's theorem using Szemerédi regularity lemma via the triangle removal lemma. In this chapter, we will be instead be studying Roth's original proof of Roth's Theorem using Fourier analysis. First, let us recall the statement of Roth's Theorem. Let $r_3([N])$ denote the maximum size of a 3-AP-free subset of $[N]$. Then Roth's theorem states that $r_3([N]) = o(N)$.

One of the drawbacks of using Szemerédi regularity which shows an upper bound that is something like $\frac{N}{\log^* N}$. Roth's Fourier analytic proof would instead give us an upper bound of something like $\frac{N}{\log \log N}$, which is a much more reasonable bound.

Sanders (2011)
Bloom (2016)

Remark 6.1. The current best upper bound known is $r_3([N]) \leq N(\log N)^{1-o(1)}$ and the best lower bound known is $r_3(N) \geq Ne^{-O(\sqrt{\log N})}$ due to the Behrend construction. There is some evidence that seem to suggest that the lower bound is closer to truth, but closing the gap is still an open problem.

6.1 Roth's theorem in finite fields

We will begin by examining a finite field analogue to Roth's Theorem. Finite field models are a good sandbox for testing methods before applying to general integer cases; in particular, it is a good starting point because a lot of technicalities go away.

Let $r_3(\mathbb{F}_3^n)$ denote the maximum size of 3 AP-free subset of $\mathbb{F}_3^n$. Note that given x, y, z in $\mathbb{F}_3^n$, the following are equivalent:

- x, y, z for a 3 term arithmetic progression
- $x - 2y + z = 0$
- $x + y + z = 0$
- x, y, z form a line

- for all i, the ith coordinate of x, y, z are all distinct or all equal.

This is relevant to the game of SET, which can be thought of as finding 3 APs in $\mathbb{F}_3^4$.

We will state and prove a version of Roth's theorem in the finite field model. The proof is in the same spirit as the general Roth's theorem, but is slightly easier.

Meshulam (1995)

Theorem 6.2.

$$r_3(\mathbb{F}_3^n) = O\left(\frac{3^n}{n}\right)$$

The proof using triangle removal lemma copies verbatim so we can get $r_3(\mathbb{F}_3^n) = o(3^n)$ but the above theorem gives a better dependence.

We comment briefly on the history of this problem. In 2004, Edel found that $r_3(\mathbb{F}_3^n) \geq 2.21^n$. It was open for a long time whether $r_3(\mathbb{F}_3^n) = (3 - o(1))^n$. Recently, a surprising breakthrough showed that $r_3(\mathbb{F}_3^n) \leq 2.76^n$.

Edel (2004)

Croot, Lev, Pach (2016)
Ellenberg and Gijswijt (2016)

We had an energy increment argument during the proof of Szemerédi Regularity lemma. The strategy for Roth's theorem is a variant of energy increment. Instead, we will consider density increment. Given $A \subset \mathbb{F}_3^n$, we employ the follow strategy.

1. If A is pseudorandom (which we will see is equivalent to it being Fourier uniform, which roughly translates to all its Fourier coefficients are small) then there is a counting lemma which will show that A has lots of 3-AP.

2. If A is not pseudorandom, then we will show that A has a large Fourier coefficient. Then we can find a codimension 1 affine subspace (i.e. hyperplane) where density of A will increase. Now we consider A restricted to this hyperplane, and repeat the previous steps.

3. Each time we repeat, we obtain a density increment. Since density is bounded above by 1, this gives us a bounded number of steps.

Next, we recall some Fourier analytic ideas that will be important in our proof. In $\mathbb{F}_3^n$, we consider the Fourier characters $\gamma_r : \mathbb{F}_3^n \to \mathbb{C}$, indexed by $r \in \mathbb{F}_3^n$, which are defined to be $\gamma_r(x) = \omega^{r \cdot x}$ where $\omega = e^{2\pi i/3}$ and $r \cdot x = r_1 x_1 + \cdots + r_n x_n$. We define a ***Fourier transform***. For $f : \mathbb{F}_3^n \to \mathbb{C}$, the Fourier transform is given by $\widehat{f} : \mathbb{F}_3^n \to \mathbb{C}$ where

$$\widehat{f}(r) = \mathop{\mathbb{E}}_{x \in \mathbb{F}_3^n} f(x)\omega^{-r \cdot x} = \langle f, \gamma_r \rangle.$$

Effectively, the fourier transform is the inner product of f and the Fourier characters.

Remark 6.3. We use the following convention on normalization: in a finite group, for a physical space we will use average measure but in frequency we will always use sum measure.

We note some key properties of the Fourier transform.

Proposition 6.4. • $\widehat{f}(0) = \mathbb{E} f$

- *(Plancheral/Parseval)* $\mathbb{E}_{x\in\mathbb{F}_3^n} f(x)\overline{g(x)} = \sum_{r\in\mathbb{F}_3^n} \widehat{f}(r)\overline{\widehat{g}(r)}$.
- *(Inversion)* $f(x) = \sum_{r\in\mathbb{F}_3^n} \widehat{f}(r)\omega^{r\cdot x}$
- *(Convolution) Define* $(f*g)(x) = \mathbb{E}_y f(y)g(x-y)$. *Then we claim that* $\widehat{f*g}(x) = \widehat{f}(x)\widehat{g}(x)$.

To prove these properties notice that Fourier characters form an orthonormal basis. Indeed, we can check

$$\langle \gamma_r, \gamma_s\rangle = \mathop{\mathbb{E}}_x \gamma_r(x)\overline{\gamma_s(x)} = \mathop{\mathbb{E}}_x \omega^{-(r-s)\cdot x} = \begin{cases} 1 \text{ if } r = s, \\ 0 \text{ otherwise.}\end{cases}$$

If we think of Fourier transform as a unitary change of basis, inversion and Parseval's follows immediately. To see the formula for convolution, note that

$$\mathop{\mathbb{E}}_x (f*g)\omega^{r\cdot x} = \mathop{\mathbb{E}}_{x,y} f(y)g(x-y)\omega^{-r(y+(x-y))} = \mathop{\mathbb{E}}_r f(x)\omega^{-r\cdot x}\mathop{\mathbb{E}}_s g(x)\omega^{-s\cdot x}.$$

The following key identity relates Fourier transform with 3-APs.

Proposition 6.5. *If* $f, g, h : \mathbb{F}_3^n \to \mathbb{C}$, *then*

$$\mathop{\mathbb{E}}_{x,y} f(x)g(x+y)h(x+2y) = \sum_r \widehat{f}(r)\widehat{g}(-2r)\widehat{h}(r).$$

We will give two proofs of this proposition, with the second being more conceptual.

First proof. We expand the LHS using the formula for Fourier inversion.

$$\begin{aligned} LHS &= \mathop{\mathbb{E}}_{x,y}\left(\sum_{r_1}\widehat{f}(r_1)\omega^{r_1\cdot x}\right)\left(\sum_{r_2}\widehat{g}(r_2)\omega^{r_2\cdot(x+y)}\right)\left(\sum_{r_3}\widehat{h}(r_3)\omega^{r_3\cdot(x+2y)}\right) \\ &= \sum_{r_1,r_2,r_3}\widehat{f}(r_1)\widehat{g}(r_2)\widehat{h}(r_3)\mathop{\mathbb{E}}_x\omega^{x\cdot(r_1+r_2+r_3)}\mathop{\mathbb{E}}_y\omega^{y\cdot(r_2+2r_3)} \\ &= \sum_r \widehat{f}(r)\widehat{g}(-2r)\widehat{h}(r) \end{aligned}$$

The last equality follows because

$$\mathop{\mathbb{E}}_x \omega^{x\cdot(r_1+r_2+r_3)} = \begin{cases} 1 & \text{if } r_1+r_2+r_3 = 0, \\ 0 & \text{otherwise}\end{cases}$$

and

$$\mathop{\mathbb{E}}_{y} \omega^{y\cdot(r_2+2r_3)} = \begin{cases} 1 & \text{if } r_2 + 2r_3 = 0, \\ 0 & \text{otherwise.} \end{cases}$$ □

Second proof. In this proof, we think of the LHS as a convolution.

$$\begin{aligned} \mathop{\mathbb{E}}_{x,y,z:x+y+z=0} f(x)g(y)h(z) &= (f * g * h)(0) \\ &= \sum_r \widehat{f * g * h}(r) \\ &= \sum_r \widehat{f}(r)\widehat{g}(r)\widehat{h}(r) \square \end{aligned}$$

In particular, note that if we take $f, g, h = 1_A$ where $A \subset \mathbb{F}_3^n$, then

$$3^{-2n}\#\{(x,y,z) \in A^3 : x + y + z = 0\} = \sum_r \widehat{1}_A(r)^3. \tag{6.1}$$

Remark 6.6. If $A = -A$ then this gives the same formula that counts closed walks of length 3 in Cayley graphs. In particular, $\{\widehat{1}_A(r) = r\}$ correspond eigenvalues of Cayley(G, A).

Lemma 6.7 (Counting Lemma)**.** *If $A \subset \mathbb{F}_3^n$ with $|A| = \alpha 3^n$, let $\Lambda_3(A) = \mathbb{E}_{x,y} 1_A(x)1_A(x+y)1_A(x+2y)$. Then,*

$$\left|\Lambda_3(A) - \alpha^3\right| \le \alpha \max_{r\neq 0} \left|\widehat{1_A}(r)\right|.$$

Proof. By Proposition 6.5,

$$\Lambda_3(A) = \sum_r \widehat{1_A}(r)^3 = \alpha^3 + \sum_{r\neq 0} \widehat{1_A}(r)^3.$$

Therefore,

$$\begin{aligned} \left|\Lambda_3(A) - \alpha^3\right| &\le \sum_{r\neq 0} \left|\widehat{1_A}(r)\right|^3 \\ &\le \max_{r\neq 0} \left|\widehat{1_A}(r)\right| \cdot \sum_r \left|\widehat{1_A}(r)\right|^2 \\ &= \max_{r\neq 0} \left|\widehat{1_A}(r)\right| \cdot \mathbb{E} 1_A^2 \qquad \text{(Parseval)} \\ &= \alpha \max_{r\neq 0} \left|\widehat{1_A}(r)\right|. \end{aligned}$$

□

Proof of Theorem 6.2. Let $N = 3^n$, the number of elements in $\mathbb{F}_3^n$.

Step 1. If the set is 3-AP free, then there is a large Fourier coefficient.

Lemma 6.8. *If A is 3-AP-free and $N \ge 2\alpha^{-2}$, then there is $r \neq 0$ such that $\left|\widehat{1_A}(r)\right| \ge \alpha^2/2$.*

Proof. By counting lemma and the fact that $\Lambda_3(A) = \frac{|A|}{N^2} = \frac{\alpha}{N}$,

$$\alpha \max_{r \neq 0} \left|\widehat{1_A}(r)\right| \geq \alpha^3 - \frac{\alpha}{N} \geq \frac{\alpha^3}{2}.$$

□

Step 2. Large Fourier coefficient implies density increment on a hyperplane.

Lemma 6.9. *If $\left|\widehat{1_A}(r)\right| \geq \delta$ for some $r \neq 0$, then A has density at least $\alpha + \frac{\delta}{2}$ when restricted to some hyperplane.*

Proof. We have

$$\begin{aligned}\widehat{1_A}(r) &= \mathop{\mathbb{E}}_{x \in \mathbb{F}_3^n} 1_A(x) w^{-r \cdot x} \\ &= \frac{1}{3}(\alpha_0 + \alpha_1 w + \alpha_2 w^2)\end{aligned}$$

where $\alpha_0, \alpha_1, \alpha_2$ are densities of A on the cosets of $r^\perp$. Notice that $\alpha = \frac{\alpha_0+\alpha_1+\alpha_2}{3}$. By triangle inequality,

$$\begin{aligned}3\delta &\leq \left|\alpha_0 + \alpha_1 w + \alpha_2 w^2\right| \\ &= \left|(\alpha_0 - \alpha) + (\alpha_1 - \alpha) w + (\alpha_2 - \alpha) w^2\right| \\ &\leq \sum_{j=0}^{2} |\alpha_j - \alpha| \\ &\leq \sum_{j=0}^{2} (|\alpha_j - \alpha| + (\alpha_j - \alpha)).\end{aligned}$$

(This final step is a trick that will be useful in the next section.) Note that every term in the last summation is non-negative. Consequently, there exists j such that $\delta \leq |\alpha_j - \alpha| + (\alpha_j - \alpha)$. Then, $\alpha_j \geq \alpha + \frac{\delta}{2}$. □

Step 3 : Iterate density increment.

So far, we have that if A is 3-AP-free and $N \geq 2\alpha^{-2}$, then A has density at least $\alpha + \alpha^2/4$ on some hyperplane. Let our initial density be $\alpha_0 = \alpha$. At the i-th step, we restrict A to some hyperplane, so that the restriction of A inside the smaller space has density

$$\alpha_i \geq \alpha_{i-1} + \alpha_{i-1}^2/4.$$

Let $N_i = 3^{n-i}$. We can continue at step i as long as $N_i \geq 2\alpha_i^{-2}$.

We note that the first index i_1 such that $\alpha_{i_1} \geq 2\alpha_0$ satisfies $i_1 \leq \frac{4}{\alpha} + 1$. This is because $\alpha_{i+1} \geq \alpha + i\frac{\alpha^2}{4}$. Similar calculations shows that if i_ℓ is the first index such that $\alpha_{i_\ell} \geq 2^\ell \alpha_0$ then

$$i_\ell \leq \frac{4}{\alpha} + m\frac{2}{\alpha} + \cdots + \frac{4}{2^{\ell-1}\alpha} + \ell \leq \frac{8}{\alpha} + \log_2 \frac{1}{\alpha}.$$

Suppose the process terminates after m steps with density α_m. Then we find that the size of the subspace in the last step is given by $3^{n-m} < 2\alpha_m^{-2} \leq 2\alpha^{-2}$. So

$$n \leq \frac{8}{\alpha} + \log_3\left(\frac{2}{\alpha^2}\right) = O\left(\frac{1}{\alpha}\right)$$

Thus $\frac{|A|}{N} = \alpha = O(1/n)$. Equivalently, $|A| = \alpha N = O\left(\frac{3^n}{n}\right)$ as desired. □

Remark 6.10. This proof is much more difficult in integers, because there is no subspace to pass down to.

A natural question is whether this technique can be generalized to bound 4-AP counts. In the regularity-based proof of Roth's theorem, we saw that the graph removal lemma was not sufficient, and we actually needed hypergraph regularity and a hypergraph removal lemma to govern 4-AP counts. Similarly, while the counting lemma developed here shows that Fourier coefficients control 3-AP counts, they do not in fact control 4-AP counts. For example, consider the set $A = \{x \in \mathbb{F}_5^n : x \cdot x = 0\}$. One can show that the nonzero Fourier coefficients corresponding to A are all small. However, one can also show that A has the wrong number of 4-APs, thus implying that Fourier coefficients cannot control 4-AP counts. The field of higher-order Fourier analysis, namely quadratic Fourier analysis, was developed by Gowers specifically to extend this proof of Roth's Theorem to prove Szemeredi's Theorem for larger APs. An example of quadratic Fourier analysis is given by the following theorem.

Gowers (1998)

Theorem 6.11 (Inverse theorem for quadratic Fourier analysis)**.** *For all $\delta > 0$, there exists a constant $c(\delta) > 0$ such that if $A \subset \mathbb{F}_5^n$ has density α, and $|\Lambda_4(A) - \alpha^4| > \delta$, then there exists a non-zero quadratic polynomial $f(x_1, \ldots, x_n)$ over $\mathbb{F}_5$ satisfying*

$$|\mathbb{E}_{x \in \mathbb{F}_5^n} 1_A(x) \omega^{f(x)}| \geq c(\delta).$$

6.2 *Roth's proof of Roth's theorem in the integers*

In Section 6.1 we saw the proof of Roth's theorem in the finite field setting, specifically for the set $\mathbb{F}_3^n$. We will now extend this analysis to prove the following bound, which will imply Roth's theorem in the integers:

Theorem 6.12.

Roth (1953)

$$r_3([N]) = O\left(\frac{N}{\log\log N}\right)$$

.

The subsequent proof of this bound is the original one given by Roth himself. Recall that the proof of Roth's theorem in finite fields had the following 3 steps:

1. Show that a 3-AP-free set admits a large Fourier coefficient.
2. Deduce that there must exist a subspace with a density increment.
3. Iterate the density increment to upper bound the size of a 3-AP free set.

The proof of Roth's theorem on the integers will follow the same 3 steps. However, the execution will be quite different. The main difference lies in step 2, where there is no obvious notion of a subspace of $[N]$.

Previously we defined Fourier analysis in terms of the group $\mathbb{F}_3^n$. There is a general theory of Fourier analysis on Abelian groups which relates a group G to its set of characters $\widehat{G}$, also referred to as its dual group. For now, however, we work with the group $\mathbb{Z}$.

The dual group of $\mathbb{Z}$ is $\widehat{\mathbb{Z}} = \mathbb{R}/\mathbb{Z}$. The Fourier Transform of a function $f : \mathbb{Z} \to \mathbb{C}$ is given by the function $\widehat{f} : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ satisfying

$$\widehat{f}(\theta) = \sum_{x \in \mathbb{Z}} f(x)e(-x\theta),$$

where $e(t) = e^{2\pi i t}$. This is commonly referred to as the Fourier series of f.

As they were in $\mathbb{F}_3^n$, the following identities are also true in $\mathbb{Z}$. Their proofs are the same.

- $\widehat{f}(0) = \sum_{x \in \mathbb{Z}} f(x)$
- (Plancherel/Parseval) $\sum_{x \in \mathbb{Z}} f(x)\overline{g(x)} = \int_0^1 \widehat{f}(\theta)\overline{\widehat{g}(\theta)}d\theta$
- (Inversion) $f(x) = \int_0^1 \widehat{f}(\theta)e(x\theta)d\theta$
- Define $\Lambda(f, g, h) = \sum_{x,y \in \mathbb{Z}} f(x)g(x+y)h(x+2y)$. Then

$$\Lambda(f, g, h) = \int_0^1 \widehat{f}(\theta)\widehat{g}(-2\theta)\widehat{h}(\theta)d\theta.$$

In the finite field setting, we defined a counting lemma, which showed that if two functions had similar Fourier transforms, then they had a similar number of 3-APs. We can define an analogue to the counting lemma in $\mathbb{Z}$ as well.

Theorem 6.13 (Counting Lemma)**.** *Let $f, g : \mathbb{Z} \to \mathbb{C}$ such that $\sum_{n \in \mathbb{Z}} |f(n)|^2, \sum_{n \in \mathbb{Z}} |g(n)|^2 \leq M$. Define $\Lambda_3(f) = \Lambda(f, f, f)$. Then*

$$|\Lambda_3(f) - \Lambda_3(g)| \leq 3M \left\| \widehat{f - g} \right\|_\infty.$$

Proof. We can rewrite

$$\Lambda_3(f) - \Lambda_3(g) = \Lambda(f-g, f, f) + \Lambda(g, f-g, f) + \Lambda(g, g, f-g).$$

We want to show that each of these terms is small when $f - g$ has small Fourier coefficients. We know that

$$\begin{aligned}
|\Lambda(f-g,f,f)| &= \left| \int_0^1 \widehat{(f-g)}(\theta)\widehat{f}(-2\theta)\widehat{f}(\theta)d\theta \right| \\
&\leq \left\| \widehat{f-g} \right\|_\infty \left| \int_0^1 \widehat{f}(-2\theta)\widehat{f}(\theta)d\theta \right| \quad \text{(triangle inequality)} \\
&\leq \left\| \widehat{f-g} \right\|_\infty \left(\int_0^1 |\widehat{f}(-2\theta)|^2 d\theta \right)^{1/2} \left(\int_0^1 |\widehat{f}(\theta)|^2 d\theta \right)^{1/2} \\
&\qquad \text{(Cauchy-Schwarz)} \\
&\leq \left\| \widehat{f-g} \right\|_\infty \left(\sum_{x \in \mathbb{Z}} |f(x)|^2 \right) \quad \text{(Plancherel)} \\
&\leq M \left\| \widehat{f-g} \right\|_\infty .
\end{aligned}$$

Bounding the other two terms is identical. □

We can now proceed with proving Roth's Theorem.

Proof of Theorem 6.12. We follow the same 3 steps as in the finite field setting.

Step 1: 3-AP free sets induce a large Fourier coefficient

Lemma 6.14. *Let $A \subset [N]$ be a 3-AP free set, $|A| = \alpha N$, $N \geq 5/\alpha^2$. Then there exists $\theta \in \mathbb{R}$ satisfying*

$$\left| \sum_{n=1}^{N} (1_A - \alpha)(n) e(\theta n) \right| \geq \frac{\alpha^2}{10} N$$

Proof. Since A has no 3-AP, the quantity $1_A(x)1_A(x+y)1_A(x+2y)$ is nonzero only for trivial APs, i.e. when $y = 0$. Thus $\Lambda_3(1_A) = |A| = \alpha N$. Now consider $\Lambda_3(1_{[N]})$. This counts the number of 3-APs in $[N]$. We can form a 3-AP by choosing the first and third elements from $[N]$, assuming they are the same parity. Therefore $\Lambda_3(1_{[N]}) \geq N^2/2$. Now, we apply the counting lemma to $f = 1_A, g = \alpha 1_{[N]}$

Remark 6.15. The spirit of this whole proof is the theme of structure versus pseudorandomness, an idea we also saw in our discussion graph regularity. If A is "pseudorandom", then we wish to show that A has small Fourier coefficients. But that would indicate that f and g have similar Fourier coefficients, implying that A has many 3-AP counts, which is a contradiction. Thus A cannot be pseudorandom, it must have some structure.

Applying Theorem 6.13 yields (where we use the notation $f^\wedge = \widehat{f}$)

$$\frac{\alpha^3 N^2}{2} - \alpha N \leq 3\alpha N \left\| (1_A - \alpha 1_{[N]})^\wedge \right\|_\infty$$

and thus

$$\begin{aligned} \left\| (1_A - \alpha 1_{[N]})^\wedge \right\|_\infty &\geq \frac{\frac{1}{2}\alpha^3 N^2 - \alpha N}{3\alpha N} \\ &= \frac{1}{6}\alpha^2 N - \frac{1}{3} \\ &\geq \frac{1}{10}\alpha^2 N, \end{aligned}$$

where in the last inequality we used the fact that $N \geq 5/\alpha^2$. Therefore there exists some θ with

$$\left| \sum_{n=1}^{N} (1_A - \alpha)(n) e(\theta n) \right| = (1_A - \alpha 1_{[N]})^\wedge(\theta) \geq \frac{1}{10}\alpha^2 N,$$

as desired. □

Step 2: A large Fourier coefficient produces a density increment.

In the finite field setting our Fourier coefficients corresponded to hyperplanes. We were then able to show that there was a coset of a hyperplane with large density. Now, however, θ is a real number. There is no concept of a hyperplane in $[N]$, so how can we chop up $[N]$ in order to use the density increment?

On each coset of the hyperplane each character was exactly constant. This motivates us to partition $[N]$ into sub-progressions such that the character $x \mapsto e(x\theta)$ is roughly constant on each sub-progression.

As a simple example, assume that θ is a rational a/b for some fairly small b. Then $x \mapsto e(x\theta)$ is constant on arithmetic progressions with common difference b. Thus we could partition $[N]$ into arithmetic progressions with common difference b.

Before formalizing this idea, we require the following classical lemma from Dirichlet.

Lemma 6.16. *Let $\theta \in \mathbb{R}$ and $0 < \delta < 1$. Then there exists a positive integer $d \leq 1/\delta$ such that $\|d\theta\|_{\mathbb{R}/\mathbb{Z}} \leq \delta$ (here, $\|\cdot\|_{\mathbb{R}/\mathbb{Z}}$ is defined as the distance to the nearest integer).*

Proof. Pigeonhole principle. Let $m = \left\lfloor \frac{1}{\delta} \right\rfloor$. Consider the $m+1$ numbers $0, \theta, \cdots, m\theta$. By the pigeonhole principle, there exist i, j such that the fractional parts of $i\theta$ and $j\theta$ differ by at most δ. Setting $d = |i - j|$ gives us $\|d\theta\|_{\mathbb{R}/\mathbb{Z}} \leq \delta$, as desired. □

The next lemma formalizes our previous intuition for partitioning $[N]$ into subprogressions such that the map $x \mapsto e(x\theta)$ is roughly constant on each progression.

Lemma 6.17. *Let $0 < \eta < 1$ and $\theta \in \mathbb{R}$. Suppose $N > C\eta^{-6}$ (for some universal constant C). Then one can partition $[N]$ into sub-APs P_i, each with length $N^{1/3} \leq |P_i| \leq 2N^{1/3}$, such that $\sup_{x,y\in P_i} |e(x\theta) - e(y\theta)| < \eta$ for all i.*

Proof. By Lemma 6.16, there exists an integer $d \leq \frac{4\pi N^{1/3}}{\eta}$ such that $\|d\theta\|_{\mathbb{R}/\mathbb{N}} \leq \frac{\eta}{4\pi N^{1/3}}$. Since $N > C\eta^{-6}$, for $C = (4\pi)^6$ we get that $d < \sqrt{N}$. Therefore we can partition $[N]$ into APs with common difference d, each with lengths between $N^{1/3}$ and $2N^{1/3}$. Then inside each sub-AP P, we have that

$$\sup_{x,y\in P} |e(x\theta) - e(y\theta)| \leq |P||e(d\theta) - 1| \leq 2N^{1/3} \cdot 2\pi \|d\theta\|_{\mathbb{R}/\mathbb{Z}} \leq \eta,$$

where we get the inequality $|e(d\theta) - 1| \leq 2\pi \|d\theta\|_{\mathbb{R}/\mathbb{Z}}$ from the fact that the length of a chord is at most the length of the corresponding arc. □

We can now apply this lemma to obtain a density increment.

Lemma 6.18. *Let $A \subset [N]$ be 3-AP-free, with $|A| = \alpha N$ and $N > C\alpha^{-12}$. Then there exists a sub-AP $P \subset [N]$ with $|P| \geq N^{1/3}$ and $|A \cap P| \geq (\alpha + \alpha^2/40)|P|$.*

Proof. By Lemma 6.14, there exists θ satisfying $|\sum_{x=1}^{N} (1_A - \alpha)(x)e(x\theta)| \geq \alpha^2 N/10$. Next, apply Lemma 6.17 with $\eta = \alpha^2/20$ to obtain a partition $P_1, \ldots, P_k$ of $[N]$ satisfying $N^{1/3} \leq |P_i| \leq 2N^{1/3}$. We then get that

$$\frac{\alpha^2}{10} N \leq \left| \sum_{x=1}^{N} (1_A - \alpha)(x)e(x\theta) \right| \leq \sum_{i=1}^{k} \left| \sum_{x\in P_i} (1_A - \alpha)(x)e(x\theta) \right|.$$

For $x, y \in P_i$, $|e(x\theta) - e(y\theta)| \leq \alpha^2/20$. Therefore we have that

$$\left| \sum_{x\in P_i} (1_A - \alpha)(x)e(x\theta) \right| \leq \left| \sum_{x\in P_i} (1_A - \alpha)(x) \right| + \frac{\alpha^2}{20}|P_i|.$$

Altogether,

$$\begin{aligned} \frac{\alpha^2}{10} N &\leq \sum_{i=1}^{k} \left(\left| \sum_{x\in P_i} (1_A - \alpha)(x) \right| + \frac{\alpha^2}{20}|P_i| \right) \\ &= \sum_{i=1}^{k} \left| \sum_{x\in P_i} (1_A - \alpha)(x) \right| + \frac{\alpha^2}{20} N \end{aligned}$$

Thus

$$\frac{\alpha^2}{20}N \leq \sum_{i=1}^{k} \left| \sum_{x \in P_i} (1_A - \alpha)(x) \right|$$

and hence

$$\frac{\alpha^2}{20} \sum_{i=1}^{k} |P_i| \leq \sum_{i=1}^{k} \big| |A \cap P_i| - \alpha |P_i| \big|.$$

We want to show that there exists some P_i such that A has a density increment when restricted to P_i. Naively bounding the RHS of the previous sum does not guarantee a density increment, so we use the following trick

$$\begin{aligned} \frac{\alpha^2}{20} \sum_{i=1}^{k} |P_i| &\leq \sum_{i=1}^{k} \big| |A \cap P_i| - \alpha |P_i| \big| \\ &= \sum_{i=1}^{k} \left(\big| |A \cap P_i| - \alpha |P_i| \big| + |A \cap P_i| - \alpha |P_i| \right). \end{aligned}$$

Thus there exists an i such that

$$\frac{\alpha^2}{20} |P_i| \leq \big| |A \cap P_i| - \alpha |P_i| \big| + |A \cap P_i| - \alpha |P_i|.$$

Since the quantity $|x| + x$ is always strictly greater than 0, this i must satisfy $|A \cap P_i| - \alpha |P_i| \geq 0$, and thus we have

$$\frac{\alpha^2}{20} |P_i| \leq 2(|A \cap P_i| - \alpha |P_i|),$$

which yields

$$|A \cap P_i| \geq (\alpha + \frac{\alpha^2}{40}) |P_i|.$$

Thus we have found a subprogression with a density increment, as desired. □

Step 3: Iterate the density increment.

Step 3 is very similar to the finite field case. Let our initial density be $\alpha_0 = \alpha$, and the density after each iteration be α_i. We have that $\alpha_{i+1} \geq \alpha_i + \alpha_i^2/40$, and that $\alpha_i \leq 1$. We double α (i.e. reach T such that $\alpha_T \geq 2\alpha_0$) after at most $40/\alpha + 1$ steps. We double α again (i.e. go from $2\alpha_0$ to $4\alpha_0$) after at most $20/\alpha + 1$ steps. In general, the kth doubling requires at most $\frac{40}{2^{k-1}\alpha}$ steps. There are at most $\log_2(1/\alpha) + 1$ doublings, as α must remain less than 1. Therefore the total number of iterations must be $O(1/\alpha)$.

Lemma 6.18 shows that we can pass to a sub-AP and increment the density whenever $N_i > C\alpha^{-12}$. Therefore if the process terminates at step i, we must have $N_i \leq C\alpha_i^{-12} \leq C\alpha^{-12}$. Each iteration reduces the size of our set by at most a cube root, so

$$N \leq N_i^{3^i} \leq (C\alpha^{-12})^{3^{O(1/\alpha)}} = e^{e^{O(1/\alpha)}}.$$

Therefore $\alpha = O(1/\log\log N)$ and $|A| = \alpha N = O(N/\log\log N)$, as desired. □

Remark 6.19. This is the same proof in spirit as last time. A theme in additive combinatorics is that the finite field model is a nice playground for most techniques.

Let us compare this proof strategy in both $\mathbb{F}_3^n$ and $[N]$. We saw that $r_3(\mathbb{F}_3^n) = O(N/\log N)$. However, the bound for $[N]$ is $O(N/\log\log N)$, which is weaker by a log factor. Where does this stem from? Well, in the density increment step for $\mathbb{F}_3^n$, we were able to pass down to a subset which had size a constant factor of the original one. However, in $[N]$, each iteration gives us a subprogression which has size equal to the cube root of the previous subspace. This poses a natural question—is it possible to pass down to subsprogressions of $[N]$ which look more like subspaces? It turns out that this is indeed possible.

For a subset $S \subset \mathbb{F}_3^n$, we can write its *orthogonal complement* as

$$U_S = \{x \in \mathbb{F}_3^n : x \cdot s = 0 \text{ for all } s \in S\}.$$

In $[N]$, the analogous concept is known as a ***Bohr set***, an idea developed by Bourgain to transfer the proof in Section 6.1 to $\mathbb{Z}$. This requires us to work in $\mathbb{Z}/N\mathbb{Z}$. For some subset $S \subset \mathbb{Z}/N\mathbb{Z}$, we can define its Bohr set as

Bourgain, 1999

$$\text{Bohr}(S, \epsilon) = \{x \in \mathbb{Z}/N\mathbb{Z} : \left\|\frac{sx}{N}\right\| \leq \epsilon \text{ for all } s \in S\}.$$

This provides a more natural analogy to subspaces, and is the basis for modern improvements on bounds to Roth's Theorem. We will study Bohr sets in relation to Freiman's Theorem in Chapter 7.

6.3 *The polynomial method proof of Roth's theorem in the finite field model*

Currently, the best known bound for Roth's Theorem in $\mathbb{F}_3^n$ is the following:

Theorem 6.20. $r_3(\mathbb{F}_3^n) = O(2.76^n)$. Ellenberg and Gijswijt (2017)

This bound improves upon the $O(3^n/n^{1+\epsilon})$ bound (for some $\epsilon > 0$) proved earlier by Bateman and Katz. Bateman and Katz used Fourier-analytic methods to prove their bound, and until very recently, it was open whether the upper bound could be improved to a power-saving one (one of the form $O(c^n)$ for $c < 3$), closer to the lower bound given by Edel of 2.21^n.

Bateman and Katz (2012)

Edel (2004)

Croot–Lev–Pach gave a similar bound for 3-APs over $(\mathbb{Z}/4\mathbb{Z})^n$, proving that the maximum size of a set in $(\mathbb{Z}/4\mathbb{Z})^n$ with no 4-APs is

$O(3.61^n)$. They used a variant of the polynomial method, and their proof was made easier by the fact that there are elements of order 2. Ellenberg and Gijswijt used the Croot–Lev–Pach method, as it is often referred to in the literature, to prove the bound for $\mathbb{F}_3^n$.

Croot, Lev, and Pach (2017)

We will use a formulation that appears on Tao's blog.

Tao (2016)

Let $A \subseteq \mathbb{F}_3^n$ be 3-AP-free (this is sometimes known as a *cap set* in the literature). Then we have the identity

$$\delta_0(x+y+z) = \sum_{a \in A} \delta_a(x)\delta_a(y)\delta_a(z) \tag{6.2}$$

for $x, y, z \in A$, where δ_a is the Dirac delta function, defined as follows:

$$\delta_a(x) := \begin{cases} 1 & \text{if } x = a, \\ 0 & \text{if } x \neq a. \end{cases}$$

Note that (6.2) holds because $x + y + z = 0$ if and only if $z - y = y - x$ in $\mathbb{F}_3^n$, meaning that x, y, z form an arithmetic progression, which is only possible if $x = y = z = a$ for some $a \in \mathbb{F}_3^n$.

We will show that the left-hand side of (6.2) is "low-rank" and the right-hand side is "high-rank" in a sense we explain below.

Recall from linear algebra the classical notion of rank: given a function $F\colon A \times A \to \mathbb{F}$, for a field $\mathbb{F}$, we say F is rank 1 if it is nonzero and can be written in the form $F(x,y) = f(x)g(y)$ for some functions $f, g\colon A \to \mathbb{F}$. In general, we define rank F to be the minimum number of rank 1 functions required to write F as a linear combination of rank 1 functions. We can view F as a matrix.

How should we define the rank of a function $F\colon A \times A \times A \to \mathbb{F}$? We might try to extend the above notion by defining such a function F to be rank 1 if $F(x,y,z) = f(x)g(y)h(z)$, known as *tensor rank*, but this is not quite what we want. Instead, we say that F has ***slice-rank*** 1 if it is nonzero and it can be written in one of the forms $f(x)g(y,z)$, $f(y)g(x,z)$, or $f(z)g(x,y)$. In general, we say the ***slice-rank*** of F is the minimum number of slice-rank 1 functions required to write F as a linear combination. For higher powers of A, we generalize this definition accordingly.

What is the rank of a diagonal function? Recall from linear algebra that the rank of a diagonal matrix is the number of nonzero entries. A similar result holds true for the slice-rank.

Lemma 6.21. *If* $F\colon A \times A \times A \to \mathbb{F}$ *equals*

$$F(x,y,z) = \sum_{a \in A} c_a \delta_a(x)\delta_a(y)\delta_a(z),$$

then

$$\text{slice-rank } F = |\{a \in A : c_a \neq 0\}|.$$

Here the coefficients c_a correspond to diagonal entries.

Proof. It is clear that slice-rank $F \leq |\{a \in A : c_a \neq 0\}|$, as we can write F as a sum of slice-rank 1 functions by

$$F(x,y,z) = \sum_{\substack{a \in A \\ c_a \neq 0}} c_a \delta_a(x)(\delta_a(y)\delta_a(z)).$$

For the other direction, assume that all diagonal entries are nonzero; if $c_a = 0$ for some a, then we can remove a from A without increasing the slice-rank. Now suppose slice-rank $F < |A|$. So we can write

$$\begin{aligned} F(x,y,z) = {} & f_1(x)g_1(y,z) + \cdots + f_\ell(x)g_\ell(y,z) \\ & + f_{\ell+1}(y)g_{\ell+1}(x,z) + \cdots + f_m(y)g_m(x,z) \\ & + f_{m+1}(z)g_{m+1}(x,y) + \cdots + f_{|A|-1}(z)g_{|A|-1}(x,y). \end{aligned}$$

Claim 6.22. There exists $h\colon A \to \mathbb{F}_3$ with $|\operatorname{supp} h| > m$ such that

$$\sum_{z \in A} h(z) f_i(z) = 0 \tag{6.3}$$

for all $i = m+1, \ldots, |A|-1$.

Here $\operatorname{supp} h$ is the set $\{z \in A : h(z) \neq 0\}$.

Proof. In the vector space of functions $A \to \mathbb{F}_3$, the set of h satisfying (6.3) for all $i = m+1, \ldots, |A|-1$ is a subspace of dimension greater than m. Furthermore, we claim that every subspace of dimension $m+1$ has a vector whose support has size at least $m+1$. For a subspace X of dimension $m+1$, suppose we write $m+1$ vectors forming a basis of X in an $|A| \times (m+1)$ matrix Y. Then, this matrix has rank $m+1$, so there must be some non-vanishing minor of order $m+1$; that is, we can delete some rows of Y to get an $(m+1) \times (m+1)$ matrix with nonzero determinant. If the column of this matrix are the vectors v_1 through v_{m+1}, then these vectors generate all of $\mathbb{F}_3^{m+1}$. In particular, some linear combination of $v_1, v_2, \ldots, v_{m+1}$ is equal to the vector of all ones, which has support $m+1$. So, taking that linear combination of the original vectors (the columns of Y) gives a vector of support at least $m+1$. □

Pick the h from the claim. We find

$$\sum_{z \in A} F(x,y,z)h(z) = \sum_{a \in A} \sum_{z \in A} c_a \delta_a(x)\delta_a(y)\delta_a(z)h(z) = \sum_{a \in A} c_a h(a)\delta_a(x)\delta_a(y),$$

but also

$$\begin{aligned} \sum_{z \in A} F(x,y,z)h(z) = {} & f_1(x)\widetilde{g_1}(y) + \cdots + f_\ell(x)\widetilde{g_\ell(y)} \\ & + f_{\ell+1}(y)\widetilde{g_{\ell+1}}(x) + \cdots + f_m(y)\widetilde{g_m}(x), \end{aligned}$$

where $\widetilde{g_i}(y) = \sum_{z\in A} g_i(y,z)h(z)$ for $1 \le i \le \ell$, and
$\widetilde{g_i}(x) = \sum_{z\in A} g_i(x,z)h(z)$ for $\ell+1 \le i \le m$. Thus

$$\sum_{a\in A} c_a h(a)\delta_a(x)\delta_a(y) = f_1(x)\widetilde{g_1}(y) + \cdots + f_\ell(x)g_\ell(y) + f_{\ell+1}(y)\widetilde{g_{\ell+1}}(x) + \cdots + f_m(y)\widetilde{g_m}(x).$$

Note the left-hand side has more than m diagonal entries (namely the a where $h(a) \neq 0$), but the left-hand side has rank at most m, which is a contradiction as we have reduced to the 2-dimensional case. □

Using induction, we can easily generalize (from 3 variables) to any finite number of variables, the proof of which we omit.

We have thus proved that the slice-rank of the right hand side of (6.2) is $|A|$, and is therefore "high-rank." We now show that the left hand side has "low-rank."

Lemma 6.23. *Define* $F\colon A \times A \times A \to \mathbb{F}_3$ *as follows:*

$$F(x+y+z) := \delta_0(x+y+z).$$

Then slice-rank $F \le 3M$, *where*

$$M := \sum_{\substack{a,b,c\ge 0\\ a+b+c=n\\ b+2c\le 2n/3}} \frac{n!}{a!b!c!}.$$

Proof. In $\mathbb{F}_3$, one has $\delta_0(x) = 1 - x^2$. Applying this coordinate-wise,

$$\delta_0(x+y+z) = \prod_{i=1}^{n}(1-(x_i+y_i+z_i)^2), \tag{6.4}$$

where the x_i are the coordinates of $x \in \mathbb{F}_3^n$, and so on. If we expand the right-hand side, we obtain a polynomial in $3n$ variables with degree $2n$. We find a sum of monomials, each of the form

$$x_1^{i_1}\cdots x_n^{i_n} y_1^{j_1}\cdots y_n^{j_n} z_1^{k_1}\cdots z_n^{k_n},$$

where $i_1, i_2, \ldots, i_n, j_1, \ldots, j_n, k_1, \ldots, k_n \in \{0,1,2\}$. Group these monomials. For each term, by the pigeonhole principle, at least one of $i_1+\cdots+i_n, j_1+\cdots+j_n, k_1+\cdots+k_n$ is at most $2n/3$.

We can write (6.4) as a sum of monomials, which we write explicitly as

$$\prod_{i=1}^{n}(1-(x_i+y_i+z_i)^2) = \sum_{\substack{i_1,i_2,\ldots,i_n\\ j_1,j_2,\ldots,j_n\\ k_1,k_2,\ldots,k_n}} c_{i_1,\ldots,i_n,j_1,\ldots,j_n,k_1,\ldots,k_n} x_1^{i_1}\cdots x_n^{i_n} y_1^{j_1}\cdots y_n^{j_n} z_1^{k_1}\cdots z_n^{k_n} \tag{6.5}$$

where $c_{i_1,\dots,i_n,j_1,\dots,j_n,k_1,\dots,k_n}$ is a coefficient in $\mathbb{F}_3$. Then, we can group terms to write (6.5) as a sum of slice-rank 1 functions in the following way:

$$\begin{aligned}\prod_{i=1}^{n}(1-(x_i+y_i+z_i)^2) = &\sum_{i_1+\cdots+i_n\le\frac{2n}{3}} x_1^{i_1}\cdots x_n^{i_n} f_{i_1,\dots,i_n}(y,z)\\ &+\sum_{j_1+\cdots+j_n\le\frac{2n}{3}} y_1^{j_1}\cdots y_n^{j_n} g_{j_1,\dots,j_n}(x,z)\\ &+\sum_{k_1+\cdots+k_n\le\frac{2n}{3}} z_1^{k_1}\cdots z_n^{k_n} h_{k_1,\dots,k_n}(x,y),\end{aligned}$$

where

$$f_{i_1,\dots,i_n}(y,z)=\sum_{\substack{j_1,j_2,\dots,j_n\\ k_1,k_2,\dots,k_n}} c_{i_1,\dots,i_n,j_1,\dots,j_n,k_1,\dots,k_n} y_1^{j_1}\cdots y_n^{j_n} z_1^{k_1}\cdots z_n^{k_n},$$

and $g_{j_1,\dots,j_n}(x,z)$ and $h_{k_1,\dots,k_n}(x,y)$ are similar except missing some terms to avoid overcounting.

So, each monomial with degree at most $2n/3$ contributes to the slice-rank 3 times, and the number of such monomials is at most M. Thus the slice-rank is at most $3M$. □

We would like to estimate M. If we let $0 \le x \le 1$, we see that $Mx^{2n/3} \le (1+x+x^2)^n$ if we expand the right-hand side. Explicitly,

$$Mx^{2n/3} \le \sum_{\substack{a,b,c\ge 0\\ a+b+c=n\\ b+2c\le 2n/3}} x^{b+2c}\frac{n!}{a!b!c!} \le (1+x+x^2)^n.$$

So

$$M \le \inf_{0<x<1}\frac{(1+x+x^2)^n}{x^{2n/3}} \le (2.76)^n,$$

where we plug in $x = 0.6$.

Alternatively, we could Stirling's formula, which would give the same bound.

When this proof came out, people were shocked; this was basically a four-page paper, and demonstrated the power of algebraic methods. However, these methods seem more fragile compared to the Fourier-analytic methods we used last time. It is an open problem to extend this technique to prove a power-saving upper-bound for the size of a 4-AP-free subset of $\mathbb{F}_5^n$ (in the above arguments, we can replace $\mathbb{F}_3$ with any other finite field, so the choice of field does not really matter). It is also open to extend the polynomial method to corner-free sets in $\mathbb{F}_2^n \times \mathbb{F}_2^n$, where corners are sets of the form $\{(x,y),(x+d,y),(x,y+d)\}$, or to the integers.

6.4 Roth's theorem with popular differences

After giving a new method for 3-APs in $\mathbb{F}_3^n$ that gave a much better bound than Fourier analysis, we will now give a different proof that gives a much worse bound, but has strong consequences.

This theorem involves a "popular common difference."

Theorem 6.24. *For all $\epsilon > 0$, there exists $n_0 = n_0(\epsilon)$ such that for $n \geq n_0$ and every $A \subseteq \mathbb{F}_3^n$ with $|A| = \alpha 3^n$, there exists $y \neq 0$ such that*

Green (2005)

$$|\{x : x, x+y, x+2y \in A\}| \geq (\alpha^3 - \epsilon)3^n.$$

Here y is the popular common difference; this theorem obtains a lower bound on the number of 3-APs with common difference y in A. Note that $\alpha^3 3^n$ is roughly the expected number of 3-APs with common difference y if A is a random subset of $\mathbb{F}_3^n$ with size $\alpha 3^n$. The theorem states we can find some y such that the number of 3-APs with common difference y is close to what we expect in a random set, and suggests that it is not true that the number of 3-APs is at least what we would expect in a random set.

Green showed that the theorem is true with $n_0 = \operatorname{tow}((1/\epsilon)^{O(1)})$. This bound was improved by Fox–Pham to $n_0 = \operatorname{tow}(O(\log \frac{1}{\epsilon}))$, using the regularity method. They showed that this bound is tight; this is an instance in which the regularity method gives the right bounds, which is interesting. This is the bound we will show.

Fox and Pham (2019+)

Lemma 6.25 (Bounded increments)**.** *Let $\alpha, \epsilon > 0$. If $\alpha_0, \alpha_1, \ldots \in [0,1]$ such that $\alpha_0 \geq \alpha$, then there exists $k \leq \lceil \log_2 \frac{1}{\epsilon} \rceil$ such that $2\alpha_k - \alpha_{k+1} \geq \alpha^3 - \epsilon$.*

Proof. Otherwise, $\alpha_1 \geq 2\alpha_0 - \alpha^3 + \epsilon \geq \alpha^3 + \epsilon$. Similarly $\alpha_2 \geq 2\alpha_1 - \alpha^3 + \epsilon \geq \alpha^3 + 2\epsilon$. If we continue this process, we find $\alpha_k \geq \alpha^3 + 2^{k-1}\epsilon$ for all $1 \leq k \leq \lceil \log_2 \frac{1}{\epsilon} \rceil + 1$. Thus $\alpha_k > 1$ if $k = \lceil \log_2 \frac{1}{\epsilon} \rceil + 1$, which is a contradiction. □

Let $f : \mathbb{F}_3^n \to \mathbb{C}$, and let $U \leq \mathbb{F}_3^n$; this notation means that U is a subspace of $\mathbb{F}_3^n$. Let $f_U(x)$ be the average of $f(x)$ on the U-coset that x is in.

The lemma below is related to an arithmetic analog of the regularity lemma.

Lemma 6.26. *For all $\epsilon > 0$, there exists $m = \operatorname{tow}(O(\log \frac{1}{\epsilon}))$ such that for all $f : \mathbb{F}_3^n \to [0,1]$, there exist subspaces $W \leq U \leq \mathbb{F}_3^n$ with $\operatorname{codim} W \leq m$ such that*

$$\|\widehat{f - f_W}\|_\infty \leq \frac{\epsilon}{|U^\perp|}$$

and

$$2\|f_U\|_3^3 - \|f_W\|_3^3 \geq (\mathbb{E} f)^3 - \epsilon.$$

Proof. Let $\epsilon_0 := 1$ and $\epsilon_{k+1} := \epsilon 3^{-1/\epsilon_k^2}$ for integers $k \geq 0$. Using the recursion, we find that the recursion says $\epsilon_{k+1}^{-2} = \epsilon^{-2} 3^{2/\epsilon_k^2}$, so that

$$\epsilon_{k+1}^{-2} \leq 2^{2^{\epsilon_k^{-2}}}$$

for sufficiently large k. Let

$$R_k := \{r \in \mathbb{F}_3^n : |\hat{f}(r)| \geq \epsilon_k\}.$$

Then $|R_k| \leq \epsilon_k^{-2}$, since by Parseval's identity, $\sum_r |\hat{f}(r)|^2 = \mathbb{E}[f^2] \leq 1$. Now define $U_k := R_k^\perp$ and $\alpha_k := \|f_{U_k}\|_3^3$. Note $\alpha_k \geq (\mathbb{E}f)^3$ by convexity. So by the previous lemma, there exists $k = O(\log \frac{1}{\epsilon})$ such that $2\alpha_k - \alpha_{k+1} \geq (\mathbb{E}f)^3 - \epsilon$. For this choice of k, let $m := \epsilon_{k+1}^{-2}$. With some computation we find $m = \operatorname{tow}(O(\log \frac{1}{\epsilon}))$.

It is not too hard to check that

$$\widehat{f_W}(r) = \begin{cases} \hat{f}(r) & \text{if } r \in W^\perp, \\ 0 & \text{if } r \notin W^\perp. \end{cases}$$

So $\|\widehat{f - f_{U_{k+1}}}\|_\infty \leq \max_{r \notin R_{k+1}} |\hat{f}(r)| \leq \epsilon_{k+1} \leq 3^{-|R_k|}\epsilon \leq \epsilon/|U_k^\perp|$. So if we take $W = U_{k+1}$ and $U = U_k$, we are done, as $\operatorname{codim} U_{k+1} \leq |R_{k+1}| \leq m$. □

With a regularity lemma comes a counting lemma, which is left as an exercise (it is fairly easy to prove). Define

$$\Lambda_3(f; U) = \mathbb{E}_{x \in \mathbb{F}_3^n, y \in U} f(x) f(x+y) f(x+2y).$$

Lemma 6.27 (Counting lemma)**.** *Let $f, g \colon \mathbb{F}_3^n \to [0,1]$ and $U \leq \mathbb{F}_3^n$. Then*

$$|\Lambda_3(f; U) - \Lambda_3(g; U)| \leq 3|U^\perp| \cdot \|\widehat{f - g}\|_\infty.$$

Lemma 6.28. *Let $f \colon \mathbb{F}_3^n \to [0,1]$, with subspaces $W \leq U \leq \mathbb{F}_3^n$. Then*

$$\Lambda_3(f_W; U) \geq 2\|f_U\|_3^3 - \|f_W\|_3^3.$$

Proof. We use Schur's inequality: $a^3 + b^3 + c^3 + 3abc \geq a^2(b+c) + b^2(a+c) + c^2(a+b)$ for $a, b, c \geq 0$. We find

$$\begin{aligned} \Lambda(f_W; U) &= \mathbb{E}_{\substack{x,y,z \\ \text{form a 3-AP in} \\ \text{the same } U\text{-coset}}} f_W(x) f_W(y) f_W(z) \\ &\geq 2\mathbb{E}_{x,\, y \text{ in same } U\text{-coset}} f_W(x)^2 f_W(y) - \mathbb{E} f_W^3 \\ &\geq 2\mathbb{E} f_W^2 f_U - \mathbb{E} f_W^3 \\ &\geq 2\mathbb{E} f_U^3 - \mathbb{E} f_W^3, \end{aligned}$$

where the first inequality follows from Schur's inequality and the last follows from convexity. □

Theorem 6.29. *For all $\epsilon > 0$, there exists $m = \mathrm{tow}(O(\log \frac{1}{\epsilon}))$ such that if $f : \mathbb{F}_3^n \to [0,1]$, then there exists $U \leq \mathbb{F}_3^n$ with codimension at most m such that*

$$\Lambda_3(f; U) \geq (\mathbb{E}f)^3 - \epsilon.$$

Note if n is large enough, then $|U|$ is large enough, so there exists a nonzero "common difference" y.

Proof. Choose U, W as in the regularity lemma. Then

$$\Lambda_3(f; U) \geq \Lambda_3(f_W; U) - 3\epsilon \geq 2\|f_U\|_3^3 - \|f_W\|_3^3 - 3\epsilon \geq (\mathbb{E}f)^3 - 4\epsilon.$$

□

The corresponding statement for popular differences is true in $\mathbb{Z}$ as well.

Theorem 6.30. *For all $\epsilon > 0$, there exists $N_0 = N_0(\epsilon)$ such that if $N > N_0$ and $A \subseteq [N]$ with $|A| = \alpha N$, then there exists $y > 0$ such that*

Green (2005)

$$|\{x : x, x+y, x+2y \in A\}| \geq (\alpha^3 - \epsilon)N.$$

A similar statement also holds for 4-APs in $\mathbb{Z}$:

Theorem 6.31. *For all $\epsilon > 0$, there exists $N_0 = N_0(\epsilon)$ such that if $N > N_0$ and $A \subseteq [N]$ with $|A| = \alpha N$, then there exists $y > 0$ such that*

Green and Tao (2010)

$$|\{x : x, x+y, x+2y, x+3y \in A\}| \geq (\alpha^4 - \epsilon)N.$$

Remark 6.32. Surprisingly, the corresponding statement for 5-APs (or longer) in $\mathbb{Z}$ is false.

Bergelson, Host, and Kra (2005) with appendix by Ruzsa

7
Structure of set addition

7.1 Structure of sets with small doubling

One of the main goals of additive combinatorics can be roughly described as understanding the behavior of sets under addition. In order to discuss this more precisely, we will begin with a few definitions.

Definition 7.1. Let A and B be finite subsets of an abelian group. Their ***sumset*** is defined as $A + B = \{a + b | a \in A, b \in B\}$. We can further define $A - B = \{a - b | a \in A, b \in B\}$ and $kA = \underbrace{A + A + \cdots + A}_{k \text{ times}}$ where k is a positive integer. Note that this is different from multiplying every element in A by k, which we denote the ***dilation*** $k \cdot A = \{kA | a \in A\}$.

Given a finite set of integers A, we want to understand how its size changes under these operations, giving rise to the following natural question:

Question 7.2. How large or small can $|A + A|$ be for a given value of $|A|$ where $A \subset \mathbb{Z}$?

It turns out that this is not a hard question. In $\mathbb{Z}$, we have precise bounds on the size of the sumset given the size of the set.

Proposition 7.3. *If A is a finite subset of $\mathbb{Z}$, then*

$$2|A| - 1 \leq |A + A| \leq \binom{|A| + 1}{2}.$$

Proof. The right inequality follows from the fact that there are only $\binom{|A|+1}{2}$ unordered pairs of elements of A.

If the elements of A are $a_1 < a_2 < \cdots < a_{|A|}$, then note that $a_1 + a_1 < a_1 + a_2 < \cdots < a_1 + a_{|A|} < a_2 + a_{|A|} < \cdots < a_{|A|} + a_{|A|}$ is an increasing sequence of $2|A| - 1$ elements of $|A + A|$, so the left inequality follows. □

The upper bound is tight when there are no nontrivial collisions in $A + A$, that is, there are no nontrivial solutions to $a_1 + a_2 = a_1' + a_2'$ for $a_1, a_2, a_1', a_2' \in A$.

Example 7.4. If $A = \{1, a, a^2, \ldots, a^{n-1}\} \subset \mathbb{Z}$ for $a > 1$, then $|A + A| = \binom{n+1}{2}$.

The lower bound is tight when A is an arithmetic progression. Even if we instead consider arbitrary abelian groups, the problem is similarly easy. In a general abelian group G, we only have the trivial inequality $|A + A| \geq |A|$, and equality holds if A is a coset of some finite subgroup of G. The reason we have a stronger bound in $\mathbb{Z}$ is that there are no nontrivial finite subgroups of $\mathbb{Z}$.

A more interesting question that we can ask is what can we say about sets where $|A + A|$ is small. More precisely:

Definition 7.5. The ***doubling constant*** of a finite subset A of an abelian group is the ratio $|A + A|/|A|$.

Question 7.6. What is the structure of a set with bounded doubling constant (e.g. $|A + A| \leq 100|A|$)?

We've already seen an example of such a set in $\mathbb{Z}$, namely arithmetic progressions.

Example 7.7. If $A \subset \mathbb{Z}$ is a finite arithmetic progression, $|A + A| = 2|A| - 1 \leq 2|A|$, so it has doubling constant at most 2.

Moreover if we delete some elements of an arithmetic progression, it should still have small doubling. In fact, if we delete even most of the elements of an arithmetic progression but leave a constant fraction of the progression remaining, we will have small doubling.

Example 7.8. If B is a finite arithmetic progression and $A \subseteq B$ has $|A| \geq C|B|$, then $|A + A| \leq |B + B| \leq 2|B| \leq 2C^{-1}|A|$, so A has doubling constant at most $2/C$.

A more substantial generalization of this is a d-dimensional arithmetic progression.

Definition 7.9. A ***generalized arithmetic progression (GAP)*** of dimension d is a set of the form

$$\{x_0 + \ell_1 x_1 + \cdots + \ell_d x_d \mid 0 \leq \ell_1 < L_d, \ldots, 0 \leq \ell_d < L_d, \ell_1, \ldots, \ell_d \in \mathbb{Z}\}$$

where $x_0, x_1, \ldots, x_d \in \mathbb{Z}$ and $L_1, \ldots, L_d \in \mathbb{N}$. The ***size*** of a GAP is defined as $L_1 L_2 \cdots L_d$. If there are no nontrivial coincidences among the elements of the GAP, it is called ***proper***.

Remark 7.10. Note that if a GAP is not proper, the size is not equal to the number of distinct elements, i.e. its cardinality.

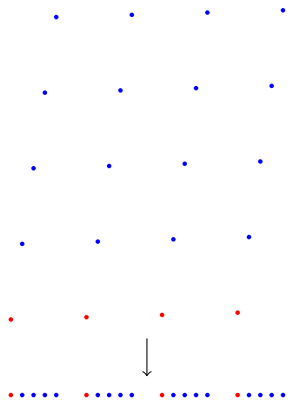


Figure 7.1: Picture of a 2-dimensional arithmetic progression as a projection of a lattice in $\mathbb{Z}^2$ into $\mathbb{Z}$.

It is not too hard to see that a proper GAP of dimension d has doubling constant at most 2^d. Furthermore, we have the same property that deleting a constant fraction of the elements of a GAP will still leave a set of small doubling constant. We have enumerated several examples of sets of small doubling constant, so it is natural to ask whether we can give an exact classification of such sets. We have an "inverse problem" to Question 7.6, asking whether every set with bounded doubling constant must be one of these examples.

This is not such an easy problem. Fortunately, a central result in additive combinatorics gives us a positive answer to this question.

Theorem 7.11 (Freiman's theorem). *If $A \subset \mathbb{Z}$ is a finite set and $|A+A| \leq K|A|$, then A is contained in a GAP of dimension at most $d(K)$ and size at most $f(K)|A|$, where $d(K)$ and $f(K)$ are constants depending only on K.*

Freiman (1973)

Remark 7.12. The conclusion of the theorem can be made to force the GAP to be proper, at the cost of increasing $d(K)$ and $f(K)$, using the fact below, whose proof we omit but can be found as Theorem 3.40 in the textbook by Tao and Vu.

Tao and Vu (2006)

Theorem 7.13. *If P is a GAP of dimension d, then P is contained in a proper GAP Q of dimension at most d and size at most $d^{C_0 d^3}|P|$ for some absolute constant $C_0 > 0$.*

Freiman's theorem gives us significant insight into the structure of sets of small doubling. We will see the proof of Freiman's theorem in the course of this chapter. Its proof combines ideas from Fourier analysis, the geometry of numbers, and classical additive combinatorics.

Freiman's original proof was difficult to read and did not originally get the recognition it deserved. Later on Ruzsa found a simpler proof, whose presentation we will mostly folllow. The theorem is sometimes called the Freiman–Ruzsa theorem. Freiman's theorem was brought into prominence as it and its ideas play central roles in Gowers' new proof of Szemerédi's theorem.

Ruzsa (1994)

If we consider again Example 7.4, then we have $K = \frac{|A|+1}{2} = \Theta(|A|)$. There isn't really a good way to embed this into a GAP. If we let the elements of A be $a_1 < a_2 < \cdots < a_{|A|}$, we can see that it is contained in a GAP of dimension $|A|-1$ and size $2^{|A|-1}$, by simply letting $x_0 = a_1$, $x_i = a_{i+1} - a_1$, and $L_i = 2$ for $1 \leq i \leq |A|-1$. Then this indicates that the best result we can hope for is showing $d(K) = O(K)$ and $f(K) = 2^{O(K)}$. This problem is still open.

Open problem 7.14. Is Theorem 7.11 true with $d(K) = O(K)$ and $f(K) = 2^{O(K)}$?

The best known result is due to Sanders, who also has the best known bound for Roth's Theorem (Theorem 6.12).

Theorem 7.15 (Sanders). *Theorem 7.11 is true with* $d(K) = K(\log K)^{O(1)}$, $f(K) = e^{K(\log K)^{O(1)}}$.

Sanders (2012)

In the asymptotic notation we assume that K is sufficiently large, say $K \geq 3$, so that $\log K$ is not too small.

Similar to how we discussed Roth's theorem, we will begin by analyzing a finite field model of the problem. In $\mathbb{F}_2^n$, if $|A+A| \leq K|A|$, then what would A look like? If A is a subspace, then it has doubling constant 1. A natural analogue of our inverse problem is to ask if all such A are contained in a subspace that is not much larger than A.

Theorem 7.16 ($\mathbb{F}_2^n$-analogue of Freiman). *If $A \subset \mathbb{F}_2^n$ has $|A+A| \leq K|A|$, then A is contained in a subspace of cardinality at most $f(K)|A|$, where $f(K)$ is a constant depending only on K.*

Remark 7.17. If we let A be a linearly independent set (i.e. a basis), then $K = \Theta(|A|)$ and the smallest subspace containing A will have cardinality $2^{|A|}$. Thus $f(K)$ must be exponential in K at least. We'll prove Theorem 7.16 in Section 7.3.

7.2 *Plünnecke–Ruzsa inequality*

Before we can prove Freiman's theorem (Theorem 7.11) or its finite field version (Theorem 7.16), we will need a few tools. We begin with one of many results named after Ruzsa.

Theorem 7.18 (Ruzsa triangle inequality). *If A, B, C are finite subsets of an abelian group, then*

$$|A||B-C| \leq |A-B||A-C|.$$

Proof. We will construct an injection

$$\phi : A \times (B-C) \hookrightarrow (A-B) \times (A-C).$$

For each $d \in B-C$, we can choose $b(d) \in B, c(d) \in C$ such that $d = b(d) - c(d)$. Then define $\phi(a,d) = (a-b(d), a-c(d))$. This is injective because if $\phi(a,d) = (x,y)$, then we can recover (a,d) from (x,y) because $d = y-x$ and $a = x + b(y-x)$. □

Remark 7.19. By replacing B with $-B$ and/or C with $-C$, we can change some of the plus signs into minus signs in this inequality. Unfortunately, this trick cannot be used to prove the similar inequality $|A||B+C| \leq |A+B||A+C|$. Nevertheless, we will soon see that this inequality is still true.

Remark 7.20. Where's the triangle? If we define $\rho(A,B) = \log \frac{|A-B|}{\sqrt{|A||B|}}$, then Theorem 7.18 states that $\rho(B,C) \leq \rho(A,B) + \rho(A,C)$. This looks like the triangle inequality, but unfortunately ρ is not actually a metric because $\rho(A,A) \neq 0$ in general. If we restrict to only looking at subgroups, however, then ρ is a bona fide metric.

The way that we use Theorem 7.18 is to control further doublings of a set of small doubling. Its usefulness is demonstrated by the following example.

Example 7.21. Suppose A is a finite subset of an abelian group with $|2A - 2A| \leq K|A|$. If we set $B = C = 2A - A$ in Theorem 7.18, then we get

$$|3A - 3A| \leq \frac{|2A - 2A|^2}{|A|} \leq K^2|A|.$$

We can repeat this with $B = C = 3A - 2A$ to get

$$|5A - 5A| \leq \frac{|3A - 3A|^2}{|A|} \leq K^4|A|$$

and so on, so for all m we have that $|mA - mA|$ is bounded by a constant multiple of $|A|$.

The condition $|2A - 2A| \leq K|A|$ is stronger than the condition $|A + A| \leq K|A|$. If we want to bound iterated doublings given just the condition $|A + A| \leq K|A|$, we need the following theorem.

Theorem 7.22 (Plünnecke–Ruzsa inequality)**.** *If A is a finite subset of an abelian group and $|A + A| \leq K|A|$, then $|mA - nA| \leq K^{m+n}|A|$.*

Plünnecke (1970)

Ruzsa (1989)

We think of polynomial changes in K as essentially irrelevant, so this theorem just says that if a set has small doubling then any iteration of the set is also small.

Remark 7.23. Plünnecke's original proof of the theorem did not receive much attention. Ruzsa later gave a simpler proof of Plünnecke's theorem. Their proofs involved the study of an object called a commutative layered graph, and involved Menger's theorem for flows and the tensor power trick. Recently Petridis gave a significantly simpler proof which uses some of the earlier ideas, which we will show here.

Petridis (2012)

In proving this theorem, we will generalize to the following theorem.

Set $B = A$ to recover Theorem 7.22

Theorem 7.24. *If A and B are finite subsets of an abelian group and $|A + B| \leq K|A|$, then $|mB - nB| \leq K^{m+n}|A|$.*

Petridis' proof relies on the following key lemma.

Lemma 7.25. *Suppose A and B are finite subsets of an abelian group. If $X \subseteq A$ is a nonempty subset which minimizes $\frac{|X+B|}{|X|}$, and $K' = \frac{|X+B|}{|X|}$, then $|X + B + C| \leq K'|X + C|$ for all finite sets C.*

Remark 7.26. We can think of this lemma in terms of a bipartite graph. If we consider the bipartite graph on vertex set $G_1 \sqcup G_2$, where G_1, G_2 are copies of the ambient abelian group G, with edges from g to $g + b$ for any $g \in G_1$, $g + b \in G_2$ where $b \in B$. Then if $N(S)$ denotes the neighborhood of a set of vertices S, then the lemma is considering

the ***expansion ratio*** $\frac{|N(A)|}{|A|} = \frac{|A+B|}{|A|}$. The lemma states that if X is a set whose expansion ratio K' is less than or equal to the expansion ratio of any of its subsets, then for any set C, $X + C$ also has expansion ratio at most K'.

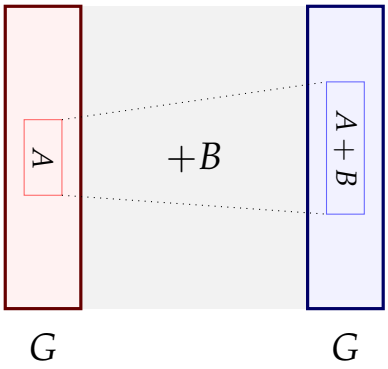


Figure 7.2: Bipartite graph where edges correspond to addition by an element of B.

Proof of Theorem 7.24 assuming Lemma 7.25. Assuming the key lemma, let us prove the theorem. Let X be a nonempty subset of A minimizing $\frac{|X+B|}{|X|}$, and let $K' = \frac{|X+B|}{|X|}$. Note that $K' \leq K$ by minimality. Applying the lemma with $C = rB$ where $r \geq 1$, we have $|X + (r+1)B| \leq K'|X + rB| \leq K|X + rB|$, so by induction $|X + rB| \leq K^r|X|$ for all $r \geq 0$. Applying Theorem 7.18 we have $|mB - nB| \leq \frac{|X+mB||X+nB|}{|X|} \leq K^{m+n}|X| \leq K^{m+n}|A|$. □

Proof of Lemma 7.25. We will proceed by induction on $|C|$. The base case of $|C| = 1$ is clear because for any finite set S, $S + C$ is a translation of S so $|S + C| = |S|$, thus $|X + B + C| = |X + B| = K'|X| = K'|X + C|$.

For the inductive step, assume $|C| > 1$, let $\gamma \in C$ and $C' = C \backslash \{\gamma\}$. Then

$$X + B + C = (X + B + C') \cup \big((X + B + \gamma) \backslash (Z + B + \gamma)\big)$$

where

$$Z = \{x \in X | x + B + \gamma \subseteq X + B + C'\}.$$

$Z \subseteq X$ so by minimality $|Z + B| \geq K'|Z|$. We have

$$\begin{aligned} |X + B + C| &\leq |X + B + C'| + |(X + B + \gamma) \backslash (Z + B + \gamma)| \\ &= |X + B + C'| + |X + B| - |Z + B| \\ &\leq K'|X + C'| + K'|X| - K'|Z| \\ &= K'(|X + C'| + |X| - |Z|). \end{aligned}$$

Now we want to understand the right hand side $X + C$. Note that

$$X + C = (X + C') \sqcup \big((X + \gamma) \backslash (W + \gamma)\big)$$

where

$$W = \{x \in X | x + \gamma \in X + C'\}.$$

In particular this is a disjoint union, so

$$|X + C| = |X + C'| + |X| - |W|.$$

We also have $W \subseteq Z$ because $x + \gamma \in X + C'$ implies $x + B + \gamma \subseteq X + B + C'$. Thus $|W| \leq |Z|$, so

$$|X + C| \geq |X + C'| + |X| - |Z|,$$

which, when combined with the above inequality, completes the induction. □

The key lemma also allows us to replace all the minus signs by pluses in Theorem 7.18 as promised.

Corollary 7.27. *If A, B, C are finite subsets of an abelian group, then $|A||B+C| \leq |A+B||A+C|$.*

Proof. Let $X \subseteq A$ be nonempty such that $\frac{|X+B|}{|X|}$ is minimal. Let $K = \frac{|A+B|}{|A|}$, $K' = \frac{|X+B|}{|X|} \leq K$. Then

$$\begin{aligned} |B+C| &\leq |X+B+C| \\ &\leq K'|X+C| && \text{(Lemma 7.25)} \\ &\leq K'|A+C| \\ &\leq K|A+C| \\ &= \frac{|A+B||A+C|}{|A|} \end{aligned}$$

□

7.3 *Freiman's theorem over finite fields*

We have one final lemma to establish before we can prove the finite field analogue of Frieman's theorem (Theorem 7.16).

Theorem 7.28 (Ruzsa covering lemma)**.** *Let X and B be subsets of an abelian group. If $|X+B| \leq K|B|$, then there exists a subset $T \subset X$ with $|T| \leq K$ such that $X \subset T + B - B$.*

Ruzsa (1999)
In essence, this theorem says that if it looks like $X + B$ is coverable by K translates of the set B (based off only size data), then X is in fact coverable by K translates of the slightly larger set $B - B$.

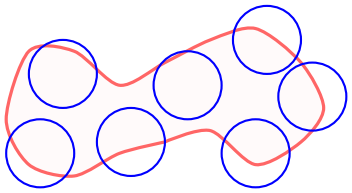


Figure 7.3: A maximal packing of a region with half balls

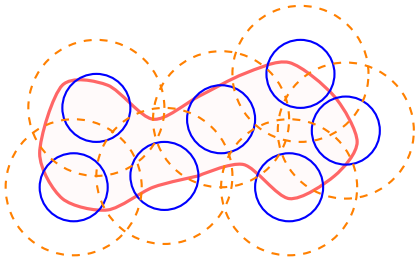


Figure 7.4: The maximal packing leads to a proper covering

The covering analogy provides the intuition for our proof. We treat the covering sets as balls in a metric space. Now, if we have a maximal packing of half-sized balls, expanding each to become a unit ball should produce a covering of the region. Note that maximal here means no more balls can be placed, not that the maximum possible number of balls have been placed. We formalize this to prove the Ruzsa covering lemma.

Proof. Let $T \subset X$ be a maximal subset such that $t + B$ is disjoint for all $t \in T$. Therefore, $|T||B| = |T+B| \leq |X+B| \leq K|B|$. So, $|T| \leq K$.

Now, as T is maximal, for all $x \in X$ there exists some $t \in T$ such that $(t+B) \cap (x+B) \neq \emptyset$. In other words, there exists $b, b' \in B$ such that $t + b = x + b'$. Hence $x \in t + B - B$ for some $t \in T$. Since this applies to all $x \in X$, we have $X \subset T + B - B$. □

The Ruzsa covering lemma is our final tool required for the proof of Freiman's theorem over finite fields (Theorem 7.16). The finite field model is simpler than working over $\mathbb{Z}$, and so it can be done with fewer tools compared to the original Freiman's theorem (Theorem 7.11).

Now, we will prove Freiman's theorem in groups with bounded exponent. This setting is slightly more general than finite fields.

Definition 7.29. The ***exponent*** of an abelian group (written additively) is the smallest positive integer r (if it exists) such that $rx = 0$ for all elements x of the group.

We also use $\langle A \rangle$ to refer to the subgroup of a group G generated by some subset A of G. By this notation, the exponent of a group G is $\max_{x \in G} |\langle x \rangle|$. With that notation, we can finally prove Ruzsa's analogue of Freiman's theorem over finite exponent abelian groups.

Theorem 7.30 (Ruzsa)**.** *Let A be a finite set in an abelian group with exponent $r < \infty$. If $|A + A| \leq K|A|$, then*

$$|\langle A \rangle| \leq K^2 r^{K^4} |A|.$$

Ruzsa (1999)
This theorem is, in a sense, the converse of our earlier observation that if A is a large enough subset of some subgroup H, then A has small doubling

Proof. By the Plünnecke–Ruzsa inequality (Theorem 7.22), we have

$$|A + (2A - A)| = |3A - A| \leq K^4 |A|.$$

Now, from the Ruzsa Covering Lemma (with $X = 2A - A$, $B = A$), there exists some $T \subset 2A - A$ with $|T| \leq K^4$ such that

$$2A - A \subset T + A - A.$$

Adding A to both sides, we have,

$$3A - A \subset T + 2A - A \subset 2T + A - A.$$

Iterating this, we have for any positive integer n,

$$(n+1)A - A \subset nT + A - A \subset \langle T \rangle + A - A.$$

Using the Ruzsa Covering Lemma allowed us to control the expression $nA - A$ nicely. If we had only used the Plünnecke–Ruzsa inequality (Theorem 7.22), the argument would have failed as the exponent of K would've blown up.

For sufficiently large n, we have $nA = \langle A \rangle$. Thus we can say,

$$\langle A \rangle \subset \langle T \rangle + A - A.$$

Due to the bounded exponent, we have,

$$|\langle T \rangle| \leq r^{|T|} \leq r^{K^4}.$$

And by the Plünnecke–Ruzsa inequality (Theorem 7.22),

$$|A - A| \leq K^2 |A|.$$

Thus we have,

$$|\langle A \rangle| \leq r^{K^4} K^2 |A|.$$

□

Example 7.31. In $\mathbb{F}_2^n$, if A is an independent subset (e.g. the basis of some subgroup), then A has doubling constant $K \approx |A|/2$, and $|\langle A \rangle| = 2^{|A|} \approx 2^{2K}|A|$. Thus the bound on $|\langle A \rangle|$ must be at least exponential in K.

It has recently been determined very precisely the maximum possible value of $|\langle A \rangle| / |A|$ over all $A \subset \mathbb{F}_2^\infty$ with $|A+A|/|A| \leq K$. Asymptotically, it is $\Theta\left(2^{2K}/K\right)$. Even-Zohar (2012)

For general r, we expect a similar phenomenon to happen. Ruzsa conjectured that $|\langle A \rangle| \leq r^{CK}|A|$. This result is proven for some r such as the primes. Ruzsa (1999) Even-Zohar and Lovett (2014)

Our proof for Freiman's theorem over abelian groups of finite exponent (Theorem 7.30) does not generalize to the integers. Indeed, in our proof above, $|\langle T \rangle|$ if we were working in $\mathbb{Z}$. The workaround is to model subsets of $\mathbb{Z}$ inside a finite group in a way that partially preserves additive structure.

7.4 *Freiman homomorphisms*

To understand any object, you should understand maps between them and the properties preserved by those maps. This is one of the fundamental principles of mathematics. For example, when studying groups we are not concerned with what the labels of the elements are, but the the relations between them according to the group operation. With manifolds, we do not focus on embeddings in space but instead maps (e.g. diffeomorphisms) which preserve various fundamental properties.

In additive combinatorics, our object of study is set addition. So we must understand maps between sets which preserve, or at least partially preserve, additive structure. Such maps are referred to as ***Freiman homomorphisms***.

Definition 7.32. Let A, B be subsets in (possibly different) abelian groups. We say that $\phi\colon A \to B$ is a ***Freiman s-homomorphism*** (or a ***Freiman homomorphism of order s***), if

Freiman s-homomorphism partially remembers additive structure, up to s-fold sums.

$$\phi(a_1) + \cdots + \phi(a_s) = \phi(a_1') + \cdots + \phi(a_s')$$

whenever $a_1, \ldots, a_s, a_1', \ldots, a_s' \in A$ satisfy

$$a_1 + \cdots + a_s = a_1' + \cdots + a_s'.$$

Definition 7.33. If $\phi\colon A \to B$ is a bijection, and both ϕ and ϕ^{-1} are Freiman s-homomorphisms, then ϕ is said to be a ***Freiman s-isomorphism***.

Let us look at some examples:

Example 7.34. Every group homomorphism is a Freiman homomorphism for any order.

Example 7.35. If ϕ_1 and ϕ_2 are both Freiman s-homomorphisms, then their composition $\phi_1 \circ \phi_2$ is also a Freiman s-homomorphism. And if ϕ_1 and ϕ_2 are both Freiman s-isomorphisms, then their composition $\phi_1 \circ \phi_2$ is a Freiman s-isomorphism.

Example 7.36. Suppose S has no additive structure (e.g. $\{1, 10, 10^2, 10^3\}$). Then an arbitrary map $\phi\colon S \to \mathbb{Z}$ is a Freiman 2-homomorphism.

Example 7.37. Suppose S_1 and S_2 are both sets without additive structure. Then any bijection $\phi\colon S_1 \to S_2$ is a Freiman 2-isomorphism.

Note that Freiman isomorphism and group homomorphisms have subtle differences!

Example 7.38. The natural embedding $\phi\colon \{0,1\}^n \to (\mathbb{Z}/2\mathbb{Z})^n$ is a group homomorphism, so it is a Freiman homomorphism of every order. It is also a bijection. But its inverse map does not preserve some additive relations, thus it is not a Freiman 2-isomorphism!

In general, the mod N map $\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$ is a group homomorphism, but not a Freiman isomorphism. This holds even if we restrict the map to $[N]$ rather than $\mathbb{Z}$. However, we can find Freiman isomorphisms by restricting to subsets of small diameter.

Proposition 7.39. *If $A \subset \mathbb{Z}$ has diameter smaller than N/s, then (mod N) maps A Freiman s-isomorphically to its image.*

If A is restricted to a small interval, then it does not have its additive relations wrap around mod N. Thus it becomes a Freiman isomorphism.

Proof. If $a_1, \dots, a_s, a'_1, \dots, a'_s \in A$ are such that

$$\sum_{i=1}^{s} a_i - \sum_{i=1}^{s} a'_i \equiv 0 \pmod{N},$$

then the left hand side, viewed as an integer, has absolute value less than N (since $|a_i - a'_i| < N/s$ for each i). Thus the left hand side must be 0 in $\mathbb{Z}$. So the inverse of the mod N map is a Freiman s-homomorphism over A, and thus mod N is a Freiman s-isomorphism. □

7.5 *Modeling lemma*

When trying to prove Freiman's theorem over the integers, our main difficulty is that a subset A with small doubling might be spread out over $\mathbb{Z}$. But we can use a Freiman isomorphism to model A inside a smaller space, preserving relative additive stucture. In this smaller space, we have better tools such as Fourier Analysis. To set up this model, we prove a modeling lemma. To warm up, let us prove this in the finite field model.

Theorem 7.40 (Modeling lemma in finite field model)**.** *Let $A \subset \mathbb{F}_2^n$ with $2^m \geq |sA - sA|$ for some positive integer m. Then A is Freiman s-isomorphic to some subset of $\mathbb{F}_2^m$.*

$\mathbb{F}_2^n$ could potentially be very large. But we can model the additive structure of A entirely within $\mathbb{F}_2^m$, which has bounded size.

Remark 7.41. If $|A + A| \leq K|A|$, then by the Plünnecke–Ruzsa inequality (Theorem 7.22) we have $|sA - sA| \leq K^{2s}|A|$, so the hypothesis if the theorem would be satisfied for some $m = O(s \log K + \log |A|)$.

Proof. The following are equivalent for linear maps $\phi\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$:

1. ϕ is Freiman s-isomorphic when restricted to A.
2. ϕ is injective on sA.
3. $\phi(x) \neq 0$ for all nonzero $x \in sA - sA$.

Then let $\phi\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ be the uniform random linear map. Each $x \in sA - sA$ violates condition (3) with probability 2^{-m}. Thus if $2^m \geq |sA - sA|$, then the probability that condition (3) is satisfied is nonzero. This implies the existence of a Freiman s-isomorphism. □

This proof does not work directly in $\mathbb{Z}$ as you cannot just choose a random linear maps. In fact, the model lemma over $\mathbb{Z}$ shows that, in fact, if $A \subset \mathbb{Z}$ has small doubling, then a large fraction of A can be modeled inside a small cyclic group whose size is comparable to $|A|$. It turns out to be enough to model a large subset of A, and we will use the Ruzsa covering lemma later on to recover the structure of the entire set A.

Theorem 7.42 (Ruzsa modeling lemma)**.** *Let $A \subset \mathbb{Z}$, $s \geq 2$, and N be a positive integer such that $N \geq |sA - sA|$. Then there exists $A' \subset A$ with $|A'| \geq |A|/s$ such that A' is Freiman s-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$.*

Ruzsa (1992)

Proof. Let $q > \max(sA - sA)$ be a prime. For every choice of $\lambda \in [q-1]$, we define ϕ as the composition of functions as follows,

We just want to take q large enough to not have to worry about any pesky details. Its actual size does not really matter.

$$\phi\colon \mathbb{Z} \to \mathbb{Z}/q\mathbb{Z} \xrightarrow{\times\lambda} \mathbb{Z}/q\mathbb{Z} \to [q].$$

Any unspecified maps refer to the natural embeddings to and from mod q. The first two maps are group homomorphisms, so they must be Freiman s-homomorphisms. The last map is not a group homomorphism over the whole domain, but it is over small intervals. In fact, by the pigeonhole principle, for all λ there exists an interval $I_\lambda \subset [q]$ of length less than q/s such that $A_\lambda = \{a \in A\colon \phi(a) \in I_\lambda\}$ has more than $|A|/s$ elements. Thus ϕ, when restricted to A_λ, is a Freiman s-homomorphism.

Now, we take this map and send it to a cyclic group, while preserving Freiman s-homomorphism. We define,

$$\psi\colon \mathbb{Z} \xrightarrow{\phi} [q] \to \mathbb{Z}/N\mathbb{Z}.$$

Claim 7.43. If ψ does not map A_λ Freiman s-isomorphically to its image, then there exists some nonzero $d = d_\lambda \in sA - sA$ such that $\phi(d) \equiv 0 \pmod N$.

Proof. Suppose ψ does not map A_λ Freiman isomorphically to its image. Thus, there exists $a_1, \dots, a_s, a'_1, \dots, a'_s \in A_\lambda$ such that

$$a_1 + \cdots + a_s \neq a'_1 + \cdots + a'_s,$$

but

$$\phi(a_1) + \cdots + \phi(a_s) \equiv \phi(a'_1) + \cdots + \phi(a'_s) \pmod N.$$

Since $\phi(A_\lambda) \subset I_\lambda$, which is an interval of length less than q/s, we have,

$$\left|\phi(a_1) + \cdots + \phi(a_s) - \phi(a'_1) - \cdots - \phi(a'_s)\right| \in (-q, q).$$

By swapping $(a_1, \dots, a_s)$ with $(a'_1, \dots, a'_s)$ if necessary, we assume that the LHS above is nonnegative, i.e., lies in the interval $[0, q)$.

We set $d = a_1 + \cdots + a_s - a'_1 - \cdots - a'_s$. Thus $d \in (sA - sA) \setminus \{0\}$. Now, as all the functions composed to form ϕ are group homomorphisms mod q, we have

$$\phi(d) \equiv \phi(a_1) + \cdots + \phi(a_s) - \phi(a'_1) - \cdots - \phi(a'_s) \pmod q,$$

and $\phi(d)$ lies in $[0, q)$ by the definition of ϕ. Thus the two expressions above are equal. As a result,

$$\phi(d) \equiv 0 \pmod N.$$

□

Now, for each $d \in (sA - sA)\setminus\{0\}$, the number of λ such that $\phi(d) \equiv 0 \pmod N$ equals the number of elements of $[q-1]$ divisible by N. This number is at most $(q-1)/N$.

Note that we are fixing d, but ϕ is determined by λ.

Therefore, the total number of λ such that there exists $d \in (sA - sA)\setminus\{0\}$ with $\phi(d) \equiv 0 \pmod N$ is at most $(|sA - sA| - 1)(q - 1)/N < q - 1$. So there exists some λ such that ψ maps A_λ Freiman s-isomorphically onto its image. Taking $A' = A_\lambda$, our proof is complete. □

By summing up everything we know so far, we establish a result that will help us in the proof of Freiman's theorem.

Corollary 7.44. *If $A \subset \mathbb{Z}$ with $|A + A| \leq K|A|$, then there exists a prime $N \leq 2K^{16}|A|$ and some $A' \subset A$ with $|A'| \geq |A|/8$ such that A' is Freiman 8-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$.*

Proof. By the Plünnecke–Ruzsa inequality (Theorem 7.22), $|8A - 8A| \leq K^{16}|A|$. We choose a prime $K^{16} \leq N < 2K^{16}$ by Bertrand's postulate. Then we apply the modeling lemma with $s = 8$ and $N \geq |8A - 8A|$. Thus there exists a subset $A' \subset A$ with $|A'| \geq |A|/8$ which is Freiman 8-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$. □

7.6 Bogolyubov's lemma

In the Ruzsa modeling lemma (Theorem 7.42) we proved that for any set A of integers with small doubling constant, a large fraction of A is Freiman isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$ with N not much larger than the size of A. To prove Freiman's Theorem, we need to prove that we can cover A with GAPs. This leads to the natural question of how to cover large subsets of $\mathbb{Z}/N\mathbb{Z}$ with GAPs. In this section, we first show how to find additive structure within subsets of $\mathbb{Z}/N\mathbb{Z}$. Later on, we will show how to use this additive structure to obtain a covering. It will be easier to first consider the analogous question in the finite field $\mathbb{F}_2^n$. Note a subset of $\mathbb{F}_2^n$ of size $\alpha 2^n$ does not necessarily contain any large structure such as a subspace. However, the key intuition for this section is the following: given a set A, the sumset $A + A$ smooths out the structure of A. With this intuition, we arrive at the following natural question:

Question 7.45. Suppose $A \subset \mathbb{F}_2^n$ and $|A| = \alpha 2^n$ where α is a constant independent of n. Must it be the case that $A + A$ contains a large subspace of codimension $O_\alpha(1)$?

The answer to the above question is no, as evidenced by the following example.

Example 7.46. Let A_n be the set of all points in $\mathbb{F}_2^n$ with hamming weight (number of 1 entries) at most $(n - c\sqrt{n})/2$. Note by the central limit theorem

$$|A_n| \sim k2^n$$

where $k > 0$ is a constant depending only on c. However, $A_n + A_n$ consists of points in the boolean cube whose Hamming weight is at most $n - c\sqrt{n}$ and thus does not contain any subspace of dimension $> n - c\sqrt{n}$. The proof of this claim is left as an exercise to the reader. (The same fact was also used in the proof of (6.3).)

Returning to the key intuition that the sumset $A + A$ smooths out the structure of A, it is natural to consider sums of more copies of A. It turns out that if we replace $A + A$ with $2A - 2A$ in Question 7.45 then the answer is affirmative.

Theorem 7.47 (Bogolyubov's lemma). *If $A \subset \mathbb{F}_2^n$ and $|A| = \alpha 2^n$ where α is a constant independent of n then $2A - 2A$ contains a subspace of codimension at most $1/\alpha^2$.* Bogolyubov (1939)

Proof. Let $f = 1_A * 1_A * 1_{-A} * 1_{-A}$. Note that f is supported on $2A - 2A$. Next, by the convolution property in Proposition 6.4,

$$\widehat{f} = \widehat{1}_A^2 \widehat{1}_{-A}^2 = |\widehat{1}_A|^4.$$

By Fourier inversion, we have

$$f(x) = \sum_{r \in \mathbb{F}_2^n} \widehat{f}(r)(-1)^{r\cdot x} = \sum_{r \in \mathbb{F}_2^n} |\widehat{1}_A(r)|^4 (-1)^{r\cdot x}.$$

Note that it suffices to find a subspace where f is positive since $f(x) > 0$ would imply $x \in 2A - 2A$. We will choose this subspace by looking at the size of the Fourier coefficients. Let

$$R = \{r \in \mathbb{F}_2^n \backslash \{0\} : |\widehat{1}_A(r)| > \alpha^{3/2}\}.$$

By Parseval's identity, $|R| < 1/\alpha^2$. Next note

$$\sum_{r \notin R \cup \{0\}} |\widehat{1}_A(r)|^4 \leq \alpha^3 \sum_{r \notin R \cup \{0\}} |\widehat{1}_A(r)|^2 < \alpha^4.$$

If x is in $R^\perp$, the orthogonal complement of R, then

$$\begin{aligned} f(x) &= \sum_{r \in \mathbb{F}_2^n} |\widehat{1}_A(r)|^4 (-1)^{r\cdot x} \\ &\geq |\widehat{1}_A(0)|^4 + \sum_{r \in R} |\widehat{1}_A(r)|^4 (-1)^{r\cdot x} - \sum_{r \notin R \cup \{0\}} |\widehat{1}_A(r)|^4 \\ &> \alpha^4 + \sum_{r \in R} |\widehat{1}_A(r)|^4 - \alpha^4 \\ &\geq 0. \end{aligned}$$

Thus $R^\perp \subset \operatorname{supp}(f) = 2A - 2A$ and since $|R| < 1/\alpha^2$, we have found a subspace with the desired codimension contained in $2A - 2A$. □

Our goal is now to formulate an analogous result for a cyclic group $\mathbb{Z}/N\mathbb{Z}$. The first step is to formulate an analog of subspaces for the cyclic group $\mathbb{Z}/N\mathbb{Z}$. Note we encountered a similar issue in transferring the proof of Roth's theorem from finite fields to the integers (see Theorem 6.2 and Theorem 6.12). It turns out that the correct analog is given by a Bohr set. Recall the definition of a Bohr set:

Definition 7.48. Suppose $R \subset \mathbb{Z}/N\mathbb{Z}$. Define

$$\operatorname{Bohr}(R, \epsilon) = \{x \in \mathbb{Z}/N\mathbb{Z} : \left\| \frac{rx}{N} \right\| \leq \epsilon, \text{for all } r \in R\}$$

where $\|\cdot\|$ denotes the distance to the nearest integer. We call $|R|$ the dimension of the Bohr set and ϵ the width.

It turns out that Bogolyubov's lemma holds over $\mathbb{Z}/N\mathbb{Z}$ after replacing subspaces by Bohr sets of the appropriate dimension. Note that the dimension of a Bohr set of $\mathbb{Z}/N\mathbb{Z}$ corresponds to the codimension of a subspace of $\mathbb{F}_2^n$.

Theorem 7.49 (Bogolyubov's lemma in $\mathbb{Z}/N\mathbb{Z}$)**.** *If $A \subset \mathbb{Z}/N\mathbb{Z}$ and $|A| = \alpha N$ then $2A - 2A$ contains some Bohr set $\operatorname{Bohr}(R, 1/4)$ with $|R| < 1/\alpha^2$.* Bogolyubov (1939)

Recall the definition of the Fourier Transform over $\mathbb{Z}/N\mathbb{Z}$.

Definition 7.50. Fourier transform of $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ is the function $\hat{f} : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ given by

$$\widehat{f}(r) = \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} f(x)\omega^{-rx}$$

where $\omega = e^{(2\pi i)/N}$.

We leave it as an exercise to the reader to verify the Fourier inversion formula, Parseval's identity, Plancherel's identity and the other basic properties of the Fourier transform. Now we will prove Theorem 7.49. It follows the same outline as the proof of Theorem 7.47 except for a few minor details.

Proof of Theorem 7.49. Let $f = 1_A * 1_A * 1_{-A} * 1_{-A}$. Note that f is supported on $2A - 2A$. Next, by the convolution property in Proposition Proposition 6.4,

$$\widehat{f} = \widehat{1}_A^2 \widehat{1}_{-A}^2 = |\widehat{1}_A|^4.$$

By Fourier inversion, we have

$$f(x) = \sum_{r \in \mathbb{Z}/N\mathbb{Z}} \widehat{f}(r)\omega^{rx} = \sum_{r \in \mathbb{Z}/N\mathbb{Z}} |\widehat{1}_A(r)|^4 \cos\left(\frac{2\pi rx}{N}\right).$$

Let

$$R = \{r \in \mathbb{Z}/N\mathbb{Z} \backslash \{0\} : |\widehat{1}_A(r)| > \alpha^{3/2}\}.$$

By Parseval's identity, $|R| < 1/\alpha^2$. Next note

$$\sum_{r \notin R \cup \{0\}} |\widehat{1}_A(r)|^4 \leq \alpha^3 \sum_{r \notin R \cup \{0\}} |\widehat{1}_A(r)|^2 < \alpha^4.$$

Now note the condition $x \in \mathrm{Bohr}(R, 1/4)$ is precisely equivalent to

$$\cos\left(\frac{2\pi rx}{N}\right) > 0 \text{ for all } r \in R.$$

For $x \in \mathrm{Bohr}(R, 1/4)$, we have

$$\begin{aligned} f(x) &= \sum_{r \in \mathbb{Z}/N\mathbb{Z}} |\widehat{1}_A(r)|^4 \cos\left(\frac{2\pi rx}{N}\right) \\ &\geq |\widehat{1}_A(0)|^4 + \sum_{r \notin R \cup \{0\}} |\widehat{1}_A(r)|^4 \cos\left(\frac{2\pi rx}{N}\right) \\ &> 0. \end{aligned}$$

□

We have now shown that for a set A that contains a large fraction of $\mathbb{Z}/N\mathbb{Z}$, the set $2A - 2A$ must contain a Bohr set of dimension less than $1/\alpha^2$. In the next section we will analyze additive structure within Bohr sets. In particular, we will show that Bohr sets of low dimension contain large GAPs.

7.7 *Geometry of numbers*

Before we can prove the main result of this section, we first introduce some machinery from the geometry of numbers. The geometry of numbers involves the study of lattices and convex bodies and has important applications in number theory.

Definition 7.51. A ***lattice*** in $\mathbb{R}^d$ is a set given by $\Lambda = \mathbb{Z}v_1 \oplus \cdots \oplus \mathbb{Z}v_d$ where $v_1, \ldots, v_d \in \mathbb{R}^d$ are linearly independent vectors.

Definition 7.52. The ***determinant*** $\det(\Lambda)$ of a lattice $\Lambda = \mathbb{Z}v_1 \oplus \cdots \oplus \mathbb{Z}v_d$ is the absolute value of the determinant of a matrix with $v_1, \ldots, v_d$ as columns.

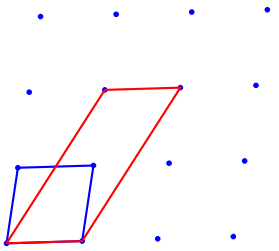


Figure 7.5: A lattice in $\mathbb{R}^2$, the blue shape is a fundamental parallelepiped while the red is not.

Remark 7.53. Note the determinant of a lattice is also equal to the volume of the fundamental parallelepiped.

Example 7.54. $\mathbb{Z} + \mathbb{Z}\omega$ where $\omega = e^{(2\pi i)/3}$ is a lattice. Its determinant is $\sqrt{3}/2$.

Example 7.55. $\mathbb{Z} + \mathbb{Z}\sqrt{2} \subset \mathbb{R}$ is **not** a lattice because 1 and $\sqrt{2}$ are not linearly independent.

We now introduce the important concept of successive minima of a convex body K with respect to a lattice Λ.

Definition 7.56. Given a centrally symmetric convex body $K \subset \mathbb{R}^d$ (by centrally symmetric we mean $x \in K$ if and only if $-x \in K$), define the ***i^{th} successive minimum*** of K with respect to a lattice Λ as

$$\lambda_i = \inf\{\lambda \geq 0 : \dim(\operatorname{span}(\lambda K \cap \Lambda)) \geq i\}$$

for $1 \leq i \leq d$. Equivalently, λ_i is the minimum λ that λK contains i linearly independent lattice vectors from Λ.

A ***directional basis*** of K with respect to Λ is a basis $\mathbf{b}_1, \ldots, \mathbf{b}_d$ of $\mathbb{R}^d$ such that $\mathbf{b}_i \in \lambda_i K$ for each $i = 1, \ldots, d$. (Note that there may be more than one possible directional basis.)

Example 7.57. Let $e_1, \ldots, e_8$ be the standard basis vectors in $\mathbb{R}^8$. Let $v = (e_1 + \cdots + e_8)/2$. Consider the lattice

$$\Lambda = \mathbb{Z}e_1 \oplus \cdots \oplus \mathbb{Z}e_7 \oplus \mathbb{Z}v.$$

Let K be the unit ball in $\mathbb{R}^8$. Note that the directional basis of K with respect to Λ is $e_1, \ldots, e_8$. This example shows that the directional basis of a convex body K is not necessarily a $\mathbb{Z}$-basis of Λ.

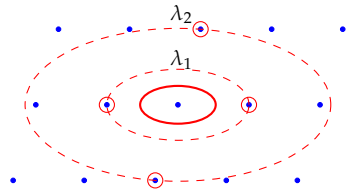


Figure 7.6: A diagram showing the successive minima of the body outlined by the solid red line with respect to the lattice of blue points.

Minkowski's second theorem gives us an inequality to control the product of the successive minima in terms of the volume of K and the determinant of the lattice Λ.

Theorem 7.58 (Minkowski's second theorem)**.** *Let $\Lambda \in \mathbb{R}^d$ be a lattice and K a centrally symmetric body. Let $\lambda_1 \leq \cdots \leq \lambda_d$ be the successive minima of K with respect to Λ. Then*

Minkowski (1896)

$$\lambda_1 \dots \lambda_d \text{vol}(K) \leq 2^d \det(\Lambda).$$

Example 7.59. Note that Minkowski's second theorem is tight when

$$K = \left[-\frac{1}{\lambda_1}, \frac{1}{\lambda_1}\right] \times \cdots \times \left[-\frac{1}{\lambda_d}, \frac{1}{\lambda_d}\right]$$

and Λ is the lattice $\mathbb{Z}^d$.

The proof of Minkowski's second theorem is omitted. We will now use Minkowski's second theorem to prove that a Bohr set of low dimension contains a large GAP.

Theorem 7.60. *Let N be a prime. Every Bohr set of dimension d and width $\epsilon \in (0,1)$ in $\mathbb{Z}/N\mathbb{Z}$ contains a proper GAP with dimension at most d and size at least $(\epsilon/d)^d N$.*

Proof. Let $R = \{r_1, \dots, r_d\}$. Let

$$v = \left(\frac{r_1}{N}, \dots, \frac{r_d}{N}\right).$$

Let $\Lambda \subset \mathbb{R}^d$ be a lattice consisting of all points in $\mathbb{R}^d$ that are congruent mod 1 to some integer multiple of v. Note $\det(\Lambda) = 1/N$ since there are exactly N points of Λ within each translate of the unit cube. We consider the convex body $K = [-\epsilon, \epsilon]^d$. Let $\lambda_1, \dots, \lambda_d$ be the successive minima of K with respect to Λ. Let $\mathbf{b}_1, \dots, \mathbf{b}_d$ be the directional basis. We know

$$\|\mathbf{b}_j\|_\infty \leq \lambda_j \epsilon \text{ for all } j.$$

For each $1 \leq j \leq d$, let $L_j = \lceil 1/(\lambda_j d) \rceil$. If $0 \leq l_j < L_j$ then

$$\|l_j \mathbf{b}_j\|_\infty < \frac{\epsilon}{d}.$$

If we have integers $l_1, \dots, l_d$ with $0 \leq l_i < L_i$ for all i then

$$\|l_1 \mathbf{b}_1 + \cdots + l_d \mathbf{b}_d\|_\infty \leq \epsilon. \tag{7.1}$$

Each $\mathbf{b}_j$ is equal to $x_j v$ plus a vector with integer coordinates for some $0 \leq x_j < N$. The bound for the i^{th} coordinate in (7.1) implies

$$\left\| \frac{(l_1 x_1 + \cdots + l_d x_d) r_i}{N} \right\|_{\mathbb{R}\setminus\mathbb{Z}} \leq \epsilon \text{ for all } i.$$

Thus, the GAP

$$\{l_1 x_1 + \cdots + l_d x_d : 0 \leq l_i < L_i \text{ for all } i\}$$

is contained in $\text{Bohr}(R, \epsilon)$. It remains to show that this GAP is large and that it is proper. First we show that it is large. Using Minkowski's second theorem, its size is

$$\begin{aligned} L_1 \cdots L_k &\geq \frac{1}{\lambda_1 \cdots \lambda_d \cdot d^d} \\ &\geq \frac{\text{vol}(K)}{2^d \det(\Lambda) d^d} \\ &= \frac{(2\epsilon)^d}{2^d \frac{1}{N} d^d} \\ &= \left(\frac{\epsilon}{d}\right)^d N. \end{aligned}$$

Now we check that the GAP is proper. It suffices to show that if

$$l_1 x_1 + \cdots + l_d x_d \equiv l_1' x_1 + \cdots + l_d' x_d \pmod{N},$$

then we must have $l_i = l_i'$ for all i. Setting

$$\mathbf{b} = (l_1 - l_1')\mathbf{b}_1 + \cdots + (l_d - l_d')\mathbf{b}_d,$$

we have $b \in \mathbb{Z}^d$. Furthermore

$$\|\mathbf{b}\|_\infty \leq \sum_{i=1}^{d} \frac{1}{\lambda_i d} \|\mathbf{b}_i\|_\infty \leq \epsilon < 1,$$

so actually $\mathbf{b}$ must be 0. Since $b_1, \ldots, b_d$ is a basis we must have $l_i = l_i'$ for all i, as desired. □

7.8 *Proof of Freiman's theorem*

So far in this chapter, we have demonstrated a number of useful methods and theorems in additive combinatorics on our quest to prove Freiman's theorem (Theorem 7.11). Now, we finally put these tools together to form a complete proof.

The proof method will be as follows. Starting with a set A with small doubling constant, we first map A to a subset, B, of $\mathbb{Z}/N\mathbb{Z}$ using the corollary of the Ruzsa modeling lemma (Theorem 7.42). We then find a large GAP within $2B - 2B$ using Bogolyubov's lemma (Theorem 7.47) and results on the geometry of numbers. This in turn gives us a large GAP in $2A - 2A$. Finally, we apply the Ruzsa covering lemma (Theorem 7.28) to create a GAP that contains A from this GAP contained in $2A - 2A$. Recall the statement of Freiman's theorem (Theorem 7.11):

> If $A \subset \mathbb{Z}$ is a finite set and $|A + A| \leq K|A|$, then A is contained in a GAP of dimension at most $d(K)$ and size at most $f(K)|A|$.

Proof. Because $|A + A| \leq K|A|$, by the corollary to Ruzsa modeling lemma (Corollary 7.44), there exists a prime $N \leq 2K^{16}|A|$ and some $A' \subset A$ with $|A'| \geq |A|/8$ such that A' is Freiman 8-isomorphic to a subset B of $\mathbb{Z}/N\mathbb{Z}$.

Applying Bogolyubov's lemma (Theorem 7.47) on B with

$$\alpha = \frac{|B|}{N} = \frac{|A'|}{N} \geq \frac{|A|}{8N} \geq \frac{1}{16K^{16}}$$

gives that $2B - 2B$ contains some Bohr set, $\text{Bohr}(R, 1/4)$, where $|R| < 256K^{32}$. Thus, by Theorem 7.60, $2B - 2B$ contains a proper GAP with dimension $d < 256K^{32}$ and size at least $(4d)^{-d}N$.

As B is Freiman 8-isomorphic to A', we have $2B - 2B$ is Freiman 2-isomorphic to $2A' - 2A'$. This follows from the definition of Freiman s-isomorphism and by noting that every element in $2B - 2B$ is the sum and difference of four elements in B with a similar statement for $2A' - 2A'$. Note that arithmetic progressions are preserved by Freiman 2-isomorphisms as the difference between any two elements in $2B - 2B$ is preserved. Hence, the proper GAP in $2B - 2B$ is mapped to a proper GAP, Q, in $2A' - 2A'$ with the same dimension and size.

Next we will use the Ruzsa covering lemma to cover the entire set A with translates of Q. Because $Q \subset 2A - 2A$, we have $Q + A \subset 3A - 2A$. By the Plünnecke-Ruzsa inequality (Theorem 7.22), we have

$$|Q + A| \leq |3A - 2A| \leq K^5|A|.$$

As $A' \subset \mathbb{Z}/N\mathbb{Z}$, we have $N \geq |A'| \geq |A|/8$. Because $|Q| \geq (4d)^{-d}N$, we have $K^5|A| \leq K'|Q|$ where $K' = 8(4d)^d K^5 = e^{K^{O(1)}}$. In particular, the above inequality becomes $|Q + A| \leq K'|Q|$. Hence, by the Ruzsa covering lemma (Theorem 7.42), there exists a subset X of A with $|X| \leq K'$ such that $A \subset X + Q - Q$.

All that remains is to show that $X + Q - Q$ is contained in a GAP with the desired bounds on dimension and size. Note that X is trivially contained in a GAP of dimension $|X|$ with length 2 in every direction. Furthermore, because every element in $Q - Q$ lies on some arithmetic progression contained in Q translated to the origin, we have the dimension of $Q - Q$ is d. Hence, by the bounds outlined above, $X + Q - Q$ is contained in a GAP P with dimension

$$\dim(P) \leq |X| + d \leq K' + d = 8(4d)^d K^5 + d = e^{K^{O(1)}}.$$

Because Q is a proper GAP with dimension d and the doubling constant of an arithmetic progression is 2, we have that $Q - Q$ has size at most $2^d|Q|$. The GAP containing X has size $2^{|X|}$. Hence, applying the Plünnecke-Ruzsa inequality, we have that the size of P is

$$\text{size}(P) \leq 2^{|X|}2^d|Q| \leq 2^{K'+d}|2A - 2A| \leq 2^{K'+d}K^4|A| = e^{e^{K^{O(1)}}}|A|.$$

Taking $d(K) = e^{K^{O(1)}}$ and $f(K) = e^{e^{K^{O(1)}}}$ completes the proof of Freiman's theorem. □

Remark 7.61. By considering $A = \{1, 10, 10^2, 10^3, \dots, 10^{|A|-1}\}$ we see that Freiman's theorem is false for $d(K) < \Theta(K)$ and $f(K) < 2^{\Theta(K)}$. It is also conjectured that Freiman's holds for $d(K) = \Theta(K)$ and $f(K) = 2^{\Theta(K)}$.

While the bounds given in the above proof of Freiman's theorems are quite far off this (exponential rather than linear), Chang showed that Ruzsa's arguments can be made to give polynomial bounds ($d(K) = K^{O(1)}$ and $f(k) = \exp(K^{O(1)})$). When we apply Ruzsa's covering lemma, we are somewhat wasteful. Rather than cover A all at once, a better method is to cover A bit by bit. In particular starting with Q we cover parts of A with $Q - Q$. We then repeat the proof on what remains of A to find Q_1 with smaller dimension. We then cover the rest of A with $Q_1 - Q_1$. This method significantly reduces the amount we lose in this step and gives the desired polynomial bounds.

Chang (2002)

As noted before, the best known bound (Theorem 7.15) is given by $d(K) = K(\log K)^{O(1)}$ and $f(K) = e^{K(\log K)^{O(1)}}$, whose proof is substantially more involved.

7.9 *Freiman's theorem for general abelian groups*

We have proved Freiman's theorem for finite fields and for integers, so one might wonder whether Freiman's theorem holds for general abelian groups. This is indeed the case, but first we must understand what such a Freiman's theorem might state.

For $\mathbb{F}_p^n$ for fixed primes p, Freiman's theorem gives that any set with small doubling constant exists in a not too much larger subgroup, while for integers, Freiman's theorem gives the same but for a not too much larger GAP. Because finitely generated abelian groups can always be represented as the direct sum of cyclic groups of prime power orders and copies of $\mathbb{Z}$, to find a generalization of GAPs and subgroups, one might try taking the direct sum of these two types of structures.

Definition 7.62. Define a ***coset progression*** as the direct sum $P + H$ where P is a proper GAP and H is a subgroup. The ***dimension*** of a coset progression is defined as the dimension of P and the ***size*** of a coset progression is defined as the cardinality of the whole set.

By a ***direct sum*** $P + H$ we mean that if $p + h = p' + h'$ for some $p, p' \in P$ and $h, h' \in H$ then $p = p'$ and $h = h'$.

Theorem 7.63 (Freiman's theorem for general abelian groups)**.** *If A is a subset of a arbitrary abelian group and $|A + A| \leq K|A|$, then A is*

Green and Ruzsa (2007)

contained in a coset progression of dimension at most $d(K)$ and size at most $f(k)|A|$, where $d(K)$ and $f(K)$ are constants depending only on K.

Remark 7.64. The proof of this theorem follows a similar method to the given proof of Freiman's theorem but with some modifications to the Ruzsa modeling lemma. The best known bounds for are again given by Sanders and are $d(K) = K(\log K)^{O(1)}$ and $f(K) = e^{K(\log K)^{O(1)}}$. It should be noted that these functions depend only on K, so they remain the same regardless of what abelian group A is a subset of.

Sanders (2013)

7.10 The Freiman problem in nonabelian groups

We may ask a similar question for nonabelian groups: what is the structure of subsets of a nonabelian group that have small doubling? Subgroups still have small doubling just as in the abelian case. Also, we can take a GAP formed by any set of commuting elements. However, it turns out that there are other examples of sets of small doubling, which are not directly derived from either of these examples from abelian groups.

Example 7.65. The ***discrete Heisenberg group*** $H_3(\mathbb{Z})$ is the set of upper triangular matrices with integer entries and only ones on the main diagonal. Multiplication in this group is as follows:

$$\begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a+x & c+z+ay \\ 0 & 1 & b+y \\ 0 & 0 & 1 \end{pmatrix}.$$

Now, let S be the following set of generators of H.

$$S = \left\{ \begin{pmatrix} 1 & \pm 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \pm 1 \\ 0 & 0 & 1 \end{pmatrix} \right\}.$$

Consider the set S^r, which is taken by all products of r sequences of elements from S. By the multiplication rule, the elements of S^r are all of the form

$$\begin{pmatrix} 1 & O(r) & O(r^2) \\ 0 & 1 & O(r) \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus, $|S^r| \leq O(r^4)$, since there are at most $O(r^4)$ possibilities for such a matrix. It can also be shown that $|S^r| = \Omega(r^4)$, and thus $|S^r| = \Theta(r^4)$. Thus, the doubling of S^r is $|S^{2r}|/|S^r| \approx 16$, so S^r has bounded doubling.

It turns out that this is an example of a more general type of construction in a group which is "almost abelian." This is captured by the notion of a nilpotent group.

Definition 7.66. A ***nilpotent group*** G is one whose lower central series terminates. In other words,

$$[\dots[[G,G],G]\dots,G] = \{e\},$$

for some finite number of repetitions. (The commutator subgroup $[H,K]$ is defined as $\{hkh^{-1}k^{-1} : h \in H, k \in K\}$.)

All nilpotent groups have polynomial growth similarly to Example 7.65, defined in general as follows.

Definition 7.67. Let G be a finitely generated group generated by a set S. The group G is said to have ***polynomial growth*** if there are constants $C, d > 0$ such that $|S^r| \leq Cr^d$ for all r. (This definition does not depend on S since for any other set of generators S', there exists r_0 such that $S' \subset S^{r_0}$.)

Gromov's theorem is a deep result in geometric group theory that provides a complete characterization of groups of polynomial growth.

Theorem 7.68 (Gromov's theorem)**.** *A finitely generated group has polynomial growth if and only if it is virtually nilpotent, i.e., has a nilpotent subgroup of finite index.*

Gromov (1981)

The techniques used by Gromov relate to Hilbert's fifth problem, which concerns characterization of Lie groups. A more elementary proof of Gromov's theorem was later given by Kleiner in 2010.

Kleiner (2010)

Now, we have a construction of a set with small doubling in any virtually nilpotent group G: the "nilpotent ball" S^r, where S generates G. It is then natural to ask the following question.

Question 7.69. Must every set of small doubling (or equivalently, sets known as *approximate groups*) behave like some combination of subgroups and nilpotent balls?

Lots of work has been done on this problem. In 2012, Hrushovski, using model theoretic techniques, showed a weak version of Freiman's theorem for nonabelian groups. Later, Breuillard, Green, and Tao, building on Hrushovski's methods, proved a structure theorem for approximate groups, generalizing Freiman's theorem to nonabelian groups. However, these methods provide no explicit bounds due to their use of ultrafilters.

Hrushovski (2012)

Breuillard, Green, and Tao (2012)

7.11 *Polynomial Freiman–Ruzsa conjecture*

In $\mathbb{F}_2^n$, if A is an independent set of size n, its doubling constant is $K = |A+A|/|A| \approx n/2$, and the size of any subgroup that contains A must be at least $2^{\Theta(K)}|A|$.

Another example, extending the previous one, is to let A be a subset of $\mathbb{F}^{m+n}$ defined by $A = \mathbb{F}_2^m \times \{e_1, \dots, e_n\}$ (where $e_1, \dots, e_n$ are generators of $\mathbb{F}_2^n$). This construction has the same bounds as the previous one, but with arbitrarily large $|A|$. This forms an example showing that the bound in the abelian group version of Freiman's theorem cannot be better than exponential.

However, note that in this example, A must contain the very large (affine) subspace $\mathbb{F}_2^m \times \{e_1\}$, which has size comparable to A. We may thus ask whether we could get better bounds in Freiman's theorem if we only needed to cover a large subset of A. In this vein, the Polynomial Freiman–Ruzsa conjecture in $\mathbb{F}_2^n$ asks the following. Green (2004)

Conjecture 7.70 (Polynomial Freiman–Ruzsa conjecture in $\mathbb{F}_2^n$)**.** *If $A \subset \mathbb{F}_2^n$, and $|A+A| \le K|A|$, then there exists an affine subspace $V \subseteq \mathbb{F}_2^n$ with $|V| \le |A|$ such that $|V \cap A| \ge K^{-O(1)}|A|$.*

This conjecture has several equivalent forms. For example, the following three are equivalent to Conjecture 7.70:

Conjecture 7.71. *If $A \subset \mathbb{F}_2^n$, and $|A+A| \le K|A|$, then there exists a subspace $V \subseteq \mathbb{F}_2^n$ with $|V| \le |A|$ such that A can be covered by $K^{O(1)}$ cosets of V.*

Proof of equivalence of Conjecture 7.70 and Conjecture 7.71. Clearly Conjecture 7.71 implies Conjecture 7.70.

Now suppose the statement of Conjecture 7.70 is true, and suppose we have $A \subset \mathbb{F}_2^n$ satisfying $|A+A| \le K|A|$. Then by Conjecture 7.70, there exists some affine subspace V with size at most $|A|$ such that $|V \cap A| \ge K^{-O(1)}|A|$ Applying the Ruzsa covering lemma (Theorem 7.28) with $X = A, B = V \cap A$ gives a set X of size $K^{O(1)}$ such that $A \subseteq V - V + X$. The conclusion of Conjecture 7.71 follows immediately, where the cosets are the shifts of the vector space $V - V$ by each of the elements of X. □

Conjecture 7.72. *If $f : \mathbb{F}_2^n \to \mathbb{F}_2^n$ satisfies*

$$|\{f(x,y) - f(x) - f(y) : x, y \in \mathbb{F}_2^n\}| \le K,$$

then there exists a linear function $g : \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that

$$|\{f(x) - g(x) : x \in \mathbb{F}_2^n\}| \le K^{O(1)}.$$

(In this version, it is straightforward to show a bound of 2^K instead of $K^{O(1)}$, since we can extend f to a linear function based on its values at some basis.)

Conjecture 7.73. *If $f : \mathbb{F}_2^n \to \mathbb{C}$ with $\|f\|_\infty \leq 1$ and $\|f\|_{U_3} \geq \delta$ (where $\|f\|_{U_3}$ is the Gowers U_3 norm, and relates to 4-AP counts), then there exists a quadratic polynomial $q(x_1, \dots, x_n)$ over $\mathbb{F}_2$ such that*

$$|\mathbb{E}_{x \in \mathbb{F}_2^n}[f(x)(-1)^{q(x)}]| \geq \delta^{O(1)}.$$

It turns out that these versions of the conjectures are all equivalent up to polynomial changes in the bounds (or equivalently, linear relations between the $O(1)$ terms). The best bound to date is due to Sanders and achieves a quasipolynomial bound of $e^{(\log K)^{O(1)}}$. The polynomial Freiman–Ruzsa conjecture would be implied by the following strengthening of Bogolyubov's lemma: Sanders (2012)

Conjecture 7.74 (Polynomial Bogolyubov-Ruzsa conjecture in $\mathbb{F}_2^n$)**.** *If $A \subset \mathbb{F}_2^n$ with $|A| = \alpha 2^n$, then $2A - 2A$ contains a subspace of codimension $O(\log(1/\alpha))$.*

The standard form of Bogolyubov's lemma (Theorem 7.47) shows a bound of $O(\alpha^{-2})$. The best result on this conjecture is also due to Sanders, who obtained a quasipolynomial bound of $(\log(1/\alpha))^{O(1)}$. Sanders (2012)

One may similarly make a version of the polynomial Freiman–Ruzsa conjecture in $\mathbb{Z}$ instead of $\mathbb{F}_2^n$. First, we must define a centered convex progression, the analog of a subspace.

Definition 7.75. A ***centered convex progression*** is a set of the form

$$P = \{x_0 + \ell_1 x_1 + \cdots + \ell_d x_d : (\ell_1, \dots, \ell_d) \in \mathbb{Z}^d \cap B\},$$

where B is some convex centrally symmetric body in $\mathbb{R}^d$. In other words, it is a shift of the image of $\mathbb{Z}^d \cap B$ under some homomorphism $\mathbb{Z}^d \to \mathbb{Z}$. Its ***dimension*** is d and its ***size*** is $|\mathbb{Z}^d \cap B|$.

Then, the polynomial Freiman–Ruzsa conjecture in $\mathbb{Z}$ states the following.

Conjecture 7.76 (Polynomial Freiman–Ruzsa conjecture in $\mathbb{Z}$)**.** *If $A \subset \mathbb{Z}$ with $|A + A| \leq K|A|$, then there exists a centered convex progression of dimension $O(\log K)$ and size at most $|A|$ whose intersection with A has size at least $K^{-O(1)}|A|$.*

More generally, the Polynomial Freiman–Ruzsa conjecture in abelian groups uses *centered convex coset progressions*, which are defined as a direct sum $P + H$, where P is the image of some $\mathbb{Z}^d \cap B$ under a homomorphism from $\mathbb{Z}^d$ to the group, and H is some coset of a subgroup.

The best bound on this conjecture (in both the $\mathbb{Z}$ and the abelian group cases) is once again quasipolynomial due to Sanders, who derived it from a quasipolynomial bound for the polynomial Bogolyubov-Ruzsa conjecture: Sanders (2012)

Conjecture 7.77 (Polynomial Bogolyubov-Ruzsa conjecture in $\mathbb{Z}$)**.** *If $A \subset \mathbb{Z}/N\mathbb{Z}$ with N prime, then $2A - 2A$ contains a proper centered convex progression of dimension $O(\log(1/\alpha))$ and size at least $\alpha^{O(1)}N$.*

Again, the version for general abelian groups can be obtained by instead using proper centered convex coset progressions instead.

7.12 *Additive energy and the Balog–Szémeredi–Gowers theorem*

So far, we have measured the amount of additive structure in a set using the doubling constant. Here we introduce ***additive energy***, a new measurement of additive structure in a set; where previously we were interested in sets of high doubling, we are now interested in sets with high additive energy.

Definition 7.78. Let A and B be finite subsets of an abelian group. Their ***additive energy*** is defined to be

$$E(A,B) = |\{(a_1,a_2,b_1,b_2) \in A \times A \times B \times B : a_1 + a_2 = b_1 + b_2\}|.$$

We set the additive energy of a single subset A to be $E(A) := E(A,A)$.

Remark 7.79. We can think of the additive energy as counting 4-cycles in an appropriate Cayley graph. Just as counting 4-cycles turned out to be fundamental in graph theory, we will see that additive energy is fundamental in additive combinatorics.

Definition 7.80. For two finite subsets A and B of an abelian group, define $r_{A,B}(x) := |\{(a,b) \in A \times B : x = a + b\}|$ to count the number of ways x is expressible as a sum in $A + B$.

Remark 7.81. We can compute additive energy as

$$E(A,B) = \sum_x r_{A,B}(x)^2.$$

For additive energy, we have the following analogue of Proposition 7.3.

Proposition 7.82. *If A is a finite subset of $\mathbb{Z}$ then $|A|^2 \leq E(A) \leq |A|^3$.*

Proof. The lower bound comes from the fact that all 4-tuples of the form $(a_1,a_2,a_1,a_2) \in A^4$ are counted by the additive energy $E(A)$. The upper bound is because for any triple $(a_1,a_2,a_3) \in A^3$, we have that $E(A)$ counts at most one 4-tuple with those first three coordinates, with fourth coordinate $a_1 + a_2 - a_3$. □

Remark 7.83. Proposition 7.82 is tight. The lower bound holds when A has no additive structure, while the upper bound holds asymptotically when $A = [n]$.

Thus far, we have likened sets of small doubling and large additive energy. In fact, the former implies the latter.

Proposition 7.84. *If $|A+A| \leq K|A|$ then $E(A) \geq |A|^3/K$.*

Proof. We use Remark 7.81 and the Cauchy-Schwarz inequality to show

$$\begin{aligned} E(A) &= \sum_{x \in A+A} r_{A,A}(x)^2 \geq \frac{1}{|A+A|}\left(\sum_{x \in A+A} r_{A,A}(x)\right)^2 \\ &= \frac{|A|^4}{|A+A|} \geq \frac{|A|^3}{|K|}. \end{aligned}$$

□

It is natural to ask whether the converse of Proposition 7.84 holds. In fact, a set with large additive energy may also have high doubling, as described in Example 7.85 below.

Example 7.85. Consider the set $A = [N/2] \cup \left\{-2, -4, -8, \ldots, -2^{N/2}\right\}$. Note that A is the union of a set of small doubling and a set with no additive structure. The first component forces the additive energy to be $E(A) = \Theta(N^3)$, while the second forces a large doubling $|A+A| = \Theta(N^2)$.

However, Balog and Szemerédi showed that every set with large additive energy must have a highly structured subset with small doubling, even if the set has relatively little additive structure overall. Their proof was later refined by Gowers, who proved polynomial bounds on the constants, and this is the version we will present here.

Theorem 7.86 (Balog–Szemerédi–Gowers theorem)**.** *Let A be a finite subset of an abelian group. If $E(A) \geq |A|^3/K$ then there is a subset $A' \subset A$ with $|A'| \geq K^{-O(1)}|A|$ and $|A'+A'| \leq K^{O(1)}|A'|$.*

Balog and Szemerédi (1994)
Gowers (1998)

We present a stronger version of the theorem, which considers the additive structure between two different sets.

Theorem 7.87. *Let A and B be finite subsets of the same abelian group. If $|A|, |B| \leq n$ and $E(A,B) \geq n^3/K$ then there exist subsets $A' \subset A$ and $B' \subset B$ with $|A'|, |B'| \geq K^{-O(1)}n$ and $|A'+B'| \leq K^{O(1)}n$.*

Proof that Theorem 7.87 implies Theorem 7.86. Suppose $E(A) \geq |A|^3/K$. Apply Theorem 7.87 with $B = A$ to obtain $A', B' \subset A$ with $|A'|, |B'| \geq K^{-O(1)}n$ and $|A'+B'| \leq K^{O(1)}n$. Then by Corollary 7.27, a variant of the Ruzsa triangle inequality, we have

$$|A'+A'| \leq \frac{|A'+B'|^2}{|B'|} \leq K^{O(1)}n.$$

□

To prove Theorem 7.87, we once again reduce from additive combinatorics to graph theory. The proof of Theorem 7.87 relies on the following graph analogue.

Definition 7.88. Let A and B be subsets of an abelian group and let G be a bipartite graph with vertex bipartition $A \cup B$. Then we define the ***restricted sumset*** $A +_G B$ to be the set of sums along edges of G:

$$A +_G B := \{a + b : (a, b) \text{ an edge in } G\}.$$

Theorem 7.89. *Let A and B be finite subsets of an abelian group and let G be a bipartite graph with vertex bipartition $A \cup B$. If $|A|, |B| \leq n$ and G has at least n^2/K edges and $|A +_G B| \leq Kn$ then there exist subsets $A' \subset A$ and $B' \subset B$ with $|A'|, |B'| \geq K^{-O(1)}n$ and $|A' + B'| \leq K^{O(1)}n$.*

Proof that Theorem 7.89 implies Theorem 7.87. Define $r_{A,B}$ as in Definition 7.80. Let $S = \{x \in A + B : r_{A,B}(x) \geq n/2K\}$ be the set of "popular sums." Build a bipartite graph G with bipartition $A \cup B$ such that $(a, b) \in A \times B$ is an edge if and only if $a + b \in S$.

We claim that G has many edges, by showing that "unpopular sums" account for at most half of $E(A, B)$. Note that

$$\frac{n^3}{K} \leq E(A, B) = \sum_{x \in S} r_{A,B}(x)^2 + \sum_{x \notin S} r_{A,B}(x)^2. \tag{7.2}$$

Because $r_{A,B}(x) < n/2K$ when $x \notin S$, we can bound the second term as

$$\sum_{x \notin S} r_{A,B}(x)^2 \leq \frac{n}{2K} \sum_{x \notin S} r_{A,B}(x) \leq \frac{n}{2K}|A||B| \leq \frac{n^3}{2K},$$

and setting back into (7.2) yields

$$\sum_{x \in S} r_{A,B}(x)^2 \geq \frac{n^3}{2K}.$$

Moreover, because $r_{A,B}(x) \leq |A| \leq n$ for all x, it follows that

$$e(G) = \sum_{x \in S} r_{A,B}(x) \geq \sum_{x \in S} \frac{r_{A,B}(x)^2}{n} \geq \frac{n^2}{2K}.$$

Hence, we can apply Theorem 7.89 to find sets $A' \subset A$ and $B' \subset B$ with the desired properties. □

The remainder of this section will focus on proving Theorem 7.89. We begin with a few lemmas.

Lemma 7.90 (Path of length 2 lemma)**.** *Fix $\delta, \epsilon > 0$. Let G be a bipartite graph with bipartition $A \cup B$ and at least $\delta|A||B|$ edges. Then there is some $U \subset A$ with $|U| \geq \delta|A|/2$ such that at least $(1 - \epsilon)$-fraction of the pairs $(x, y) \in U^2$ have at least $\epsilon\delta^2|B|/2$ neighbors common to x and y.*

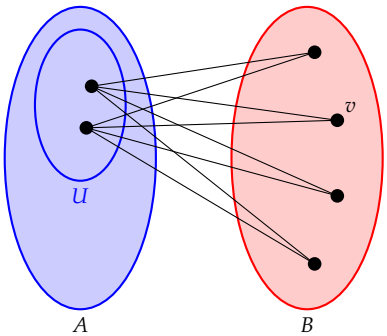


Figure 7.7: Paths of length 2 between two points in U.

Proof. We use the *dependent random choice* method from Section 2.9. Choose $v \in B$ uniformly at random, and let $U = N(v) \subset A$. We have $\mathbb{E}[|U|] \geq \delta|A|$.

We note that pairs with few common neighbors are unlikely to be contained in U. Indeed, if $x, y \in A$ share fewer than $\epsilon\delta^2|B|/2$ common neighbors then $\Pr[\{x, y\} \subset U] < \epsilon\delta^2/2$.

Say two points are *friendly* if they share at least $\epsilon\delta^2|B|/2$ common neighbors. Let X be the number of unfriendly pairs $(x, y) \in U^2$. Then

$$\mathbb{E}[X] = \sum_{\substack{(x,y)\in A^2 \\ \text{unfriendly}}} \Pr[\{x, y\} \subset U] < \frac{\epsilon\delta^2}{2}|A|^2.$$

Hence, we have

$$\mathbb{E}\left[|U|^2 - \frac{X}{\epsilon}\right] \geq (\mathbb{E}[|U|])^2 - \frac{\mathbb{E}[X]}{\epsilon} > \frac{\delta^2}{2}|A|^2,$$

so there is a choice of U with $|U|^2 - X/\epsilon \geq \delta^2|A|^2/2$. For this choice of U, we have $|U|^2 \geq \delta^2|A|^2/2$, so $|U| \geq \delta|A|/2$. Moreover, we have $X \leq \epsilon|U|^2$, so at most ϵ-fraction of pairs $(x, y) \in U^2$ have fewer than $\epsilon\delta^2|B|/2$ common neighbors. □

Lemma 7.91 (Path of length 3 lemma)**.** *There are constants $c, C > 0$ such that the following holds. Fix any $\epsilon, \delta > 0$ and let G be any bipartite graph with bipartition $A \cup B$ and at least $\delta|A||B|$ edges. Then there are subsets $A' \subset A$ and $B' \subset B$ such that every pair $(a, b) \in A' \times B'$ is joined by at least $\eta|A||B|$ paths of length 3, where $\eta = c\delta^C$.*

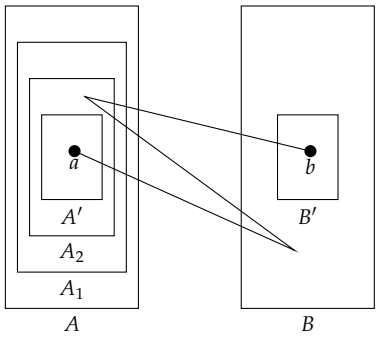


Figure 7.8: The construction for a path of length 3.

Proof. Call vertices a pair of vertices in A *friendly* if they have at least $\frac{\delta^3|B|}{20}$ common neighbors.

Define

$$A_1 := \{a \in A : \deg a \geq \frac{\delta}{2}|B|\}.$$

Restricting A to A_1 maintains an edge density of at least δ between A_1 and B and removes fewer than $\delta|A||B|/2$ edges from G. Because we are left with at least $\delta|A||B|/2$ edges and the max degree of $a \in A_1$ is $|B|$, we have $|A_1| \geq \delta|A|/2$.

Construct $A_2 \subset A_1$ via the path of length 2 lemma (Lemma 7.90) on (A_1, B) with $\epsilon = \delta/10$. Then, $|A_2| \geq \delta|A_1|/2 \geq \delta^2|A|/4$ and at most ϵ-fraction pairs of vertices in A_2 are unfriendly.

Set

$$B' = \{b \in B : \deg(b, A_2) \geq \frac{\delta}{4}|A_2|\}.$$

Restricting from (A_2, B) to (A_2, B') removes at most $\delta|A_2||B|/4$ edges. Because the minimum degree in A_2 is at least $\delta/2$, there are at least $\delta|A_2||B|/2$ edges between A_2 and B. Hence, there are at least

$\delta|A_2||B|/4$ edges between A_2 and B' and because the maximum degree of $b \in B'$ is $|A_2|$, we have $|B'| \geq \delta|B|/4$.

Define

$$A' = \{a \in A_2 : a \text{ is friendly to at least } (1 - \frac{\delta}{5})\text{-fraction of } A_2\}.$$

Then $|A'| \geq |A_2|/2 \geq \delta^2|A|/8$.

We now fix $(a, b) \in A' \times B'$ and lower-bound the number of length-3 paths between them. Because b is adjacent to at least $\delta|A_2|/4$ vertices in A_2 and a is friendly to at least $(1 - \delta/5)|A_2|$ vertices in A_2, there are at least $\delta|A_2|/20$ vertices in A_2 both friendly to a and adjacent to b. For each such $a_1 \in A_2$, there are at least $\delta^3|B|/20$ points $b_1 \in B$ for which ab_1a_1b is a path of length 3, so the number of paths of length 3 from a to b is at least

$$\frac{\delta}{20}|A_2| \cdot \frac{\delta^3}{20}|B| \geq \frac{\delta}{20} \cdot \frac{\delta^2}{4}|A| \cdot \frac{\delta^3}{20}|B| = \frac{\delta^6}{20 \cdot 4 \cdot 80}|A||B|.$$

Taking η equal to the above coefficient, we note that $|A'| \geq \delta^2|A|/8 \geq \eta|A|$ and $|B'| \geq \delta|B|/4 \geq \eta|B|$. □

We can use the path of length 3 lemma to prove the graph-theoretic analogue of the Balog–Szemerédi–Gowers theorem.

Proof of Theorem 7.89. Note that we have $|A|, |B| \geq \frac{n}{K}$. By the path of length 3 lemma (Lemma 7.91), we can find $A' \subset A$ and $B' \subset B$ of sizes $|A'|, |B'| \geq K^{-O(1)}n$ such that for every $(a, b) \in A' \times B'$, there are at least $K^{-O(1)}n^2$ paths ab_1a_1b with $(a_1, b_1) \in A \times B$. Hence, for each $(a, b) \in A' \times B'$, there are at least $K^{-O(1)}n^2$ solutions $x, y, z \in A +_G B$ to the equation $x - y + z = a + b$, as $(x, y, z) = (a + b_1, a_1 + b_1, a_1 + b)$ is a solution along each path ab_1a_1b. It follows that

$$K^{-O(1)}n^2|A' + B'| \leq |A +_G B|^3 = e(G)^3 \leq K^3n^3,$$

so $|A' + B'| \leq K^{O(1)}n$. □

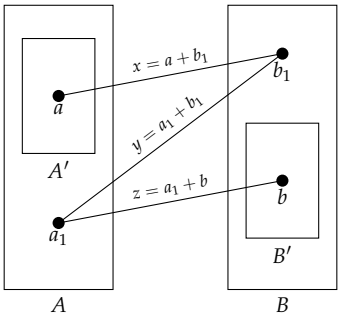


Figure 7.9: Using the path of length 3 lemma to prove the Balog–Szemerédi–Gowers theorem

8
The sum-product problem

In this chapter, we consider how sets behave under both addition and multiplication. The main problem, called ***the sum-product problem***, is the following: can $A + A$ and $A \cdot A = \{ab : a, b \in A\}$ both be small for the same set A?

We take an example $A = [N]$. Then $|A + A| = 2N - 1$, but it turns out that the product set has a large size, $|A \cdot A| = N^{2-o(1)}$. The problem of determining the size of the product set is known as *Erdős multiplication table problem*. One can also see that if A is a geometric progression, then $A \cdot A$ is small, yet $A + A$ is large. The main conjecture concerning the sum-product problem says that either the sum set or the product set has the size very close to the maximum.

Ford (2008)

Conjecture 8.1 (Erdős–Szemerédi's conjecture)**.** *For every finite subset A of $\mathbb{R}$, we have*

Erdős and Szemerédi (1983)

$$\max\{|A + A|, |A \cdot A|\} \geq |A|^{2-o(1)}$$

In this chapter, we will see two proofs of lower bounds on the sum-product problem. To do this, we first develop some tools.

8.1 Crossing number inequality

The ***crossing number*** $\mathrm{cr}(G)$ of a graph G is defined to be the minimum number of crossings in a planar drawing of G with curves. Given a graph with many edges, how big must its crossing number be?

Theorem 8.2 (Crossing number inequality)**.** *If $G = (V, E)$ is a graph satisfying $|E| \geq 4|V|$, then $cr(G) \geq c|E|^3/|V|^2$ for some constant $c > 0$.*

Ajtai, Chvátal, Newborn and Szemerédi (1982)
Leighton (1984)

It follows directly that every n-vertex graph with $\Omega(n^2)$ edges has $\Omega(n^4)$ crossings.

Proof of Theorem 8.2. For any connected planar graph with at least one cycle, we have $3|F| \leq 2|E|$, with $|F|$ denoting the number of

faces. The inequality follows from double-counting of faces using that every face is adjacent to at least three edges and that every edge is adjacent to at most two faces. Applying Euler's formula, we get $|E| \leq 3|V| - 6$. Therefore $|E| \leq 3|V|$ holds for every planar graph G including ones that are not connected or do not have a cycle. Thus we have $cr(G) > 0$ if $|E| > 3|V|$.

Let G be a finite, connected, planar graph and suppose that G is drawn in the plane without any edge intersection. Euler's formula states $|V| - |E| + |F| = 2$.

Suppose G satisfies $|E| > 3|V|$. Since we can get a planar graph by deleting each edge that witnesses a crossing, we have $|E| - \mathrm{cr}(G) \geq 3|V|$. Therefore

$$\mathrm{cr}(G) \geq |E| - 3|V|. \tag{8.1}$$

In order to get the desired inequality, we use a trick from the probabilistic method. Let $p \in [0,1]$ be some real number to be determined and let $G' = (V', E')$ be a graph obtained by randomly keeping each vertex of G with probability p iid. By (8.1), we have $\mathrm{cr}(G') \geq |E'| - 3|V'|$ for every G'. Therefore the same inequality must hold if we take the expected values of both sides:

$$\mathbb{E}\,\mathrm{cr}(G') \geq \mathbb{E}|E'| - 3\mathbb{E}|V'|.$$

One can see that $\mathbb{E}|E'| = p^2|E|$ since an edge remains if and only if both of its endpoints are kept. Similarly $\mathbb{E}|V'| = p|V|$. By keeping the same drawing, we get the inequality $p^4\,\mathrm{cr}(G) \geq \mathbb{E}\,\mathrm{cr}(G')$. Therefore we have

$$\mathrm{cr}(G) \geq p^{-2}|E| - 3p^{-3}|V|.$$

Finally we get the desired inequality by setting $p \in [0,1]$ so that $4p^{-3}|V| = p^{-2}|E|$, which can be done from the condition $|E| \geq 4|V|$. □

8.2 *Incidence geometry*

Another field in mathematics related to the sum-product problem is incidence geometry. The ***incidence*** between the set of points $\mathcal{P}$ and the set of lines $\mathcal{L}$ is defined as

$$I(\mathcal{P}, \mathcal{L}) = |\{(p, \ell) \in \mathcal{P} \times \mathcal{L} : p \in \ell\}|$$

What's the maximum number of incidences between n points and n lines? One trivial upper bound is $|\mathcal{P}||\mathcal{L}|$. We can get a better bound by using the fact that every pair of points is determined by at most one line:

$$\begin{aligned} |\mathcal{P}|^2 &\geq \#\{(p, p', \ell) \in \mathcal{P} \times \mathcal{P} \times \mathcal{L} : pp' \in \ell\,,\ p \neq p'\} \\ &\geq \sum_{\ell \in \mathcal{L}} |\mathcal{P} \cup \ell|(|\mathcal{P} \cup \ell| - 1) \\ &\geq \frac{I(\mathcal{P}, \mathcal{L})^2}{|\mathcal{L}|^2} - I(\mathcal{P}, \mathcal{L}). \end{aligned}$$

The last inequality follows from Cauchy–Schwarz inequality. Therefore, we get $I(\mathcal{P},\mathcal{L}) \leq |\mathcal{P}||\mathcal{L}|^{1/2} + |\mathcal{L}|$. By duality of points and lines, namely by the projection that puts points to lines, we also get $I(\mathcal{P},\mathcal{L}) \leq |\mathcal{L}||\mathcal{P}|^{1/2} + |\mathcal{P}|$. These inequalities give us that n points and n lines have $O(n^{3/2})$ incidences. The order $3/2$ can be found in the first chapter, when we examine $\mathrm{ex}(n, C_4) = \Theta(n^{3/2})$. The proof we will give is basically the same. Recall that the bound was tight and the construction came from finite fields. On the other hand, in the real plane, $n^{3/2}$ is not tight, as we will see in the next theorem.

Theorem 8.3 (Szemerédi–Trotter). *For any set $\mathcal{P}$ of points and $\mathcal{L}$ of lines in $\mathbb{R}^2$,*

$$I(\mathcal{P},\mathcal{L}) = O(|\mathcal{P}|^{2/3}|\mathcal{L}|^{3/2} + |\mathcal{P}| + |\mathcal{L}|).$$

Szemerédi and Trotter (1983)

Corollary 8.4. *For n points and n lines in $\mathbb{R}^2$, the number of incidences is $O(n^{4/3})$.*

Example 8.5. The bounds in both Theorem 8.3 and Corollary 8.4 are best possible up to a constant factor. Here is an exapmle showing that Corollary 8.4 is tight. Let $\mathcal{P} = [k] \times [2k^2]$ and $\mathcal{L} = \{y = mx + b : m \in [k], b \in [k^2]\}$. Then every line in $\mathcal{L}$ contains k points from $\mathcal{P}$, so $I = k^4 = \Theta(n^{4/3})$.

Proof of Theorem 8.3. we first get rid of all lines in $\mathcal{L}$ which contain at most one point in $\mathcal{P}$. One can see that these lines contribute to at most $|\mathcal{L}|$ incidences.

Now we can assume that every line in $\mathcal{L}$ contains at least two points of $\mathcal{P}$. We construct a graph G as the following: first, we assign vertices to all points in $\mathcal{P}$. For every line in $\mathcal{L}$, we assign an edge between consecutive points of $\mathcal{P}$ lying on the line.

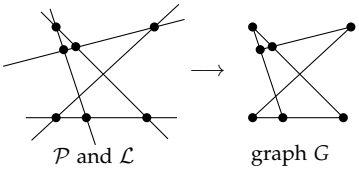


Figure 8.1: Construction of graph G

Since a line with k incidences has $k - 1 \geq k/2$ edges, we have the inequality $|E| \geq I(\mathcal{P},\mathcal{L})/2$. If $I(\mathcal{L},\mathcal{P}) \geq 8|\mathcal{P}|$ holds (otherwise, we get $I(\mathcal{P},\mathcal{L}) \lesssim |\mathcal{P}|$), we can apply Theorem 8.2.

$$\mathrm{cr}(G) \gtrsim \frac{|E|^3}{|V|^2} \gtrsim \frac{I(\mathcal{P},\mathcal{L})^3}{|\mathcal{P}|^2}.$$

Moreover $\mathrm{cr}(G) \leq |\mathcal{L}|^2$ since every pair of lines intersect in at most one point. We rearrange and get $I(\mathcal{P},\mathcal{L}) \lesssim |P|^{2/3}|L|^{2/3}$. Therefore we get that $I(\mathcal{P},\mathcal{L}) \lesssim |\mathcal{P}|^{2/3}|\mathcal{L}|^{3/2} + |\mathcal{P}| + |\mathcal{L}|$. The two linear parts are needed for the cases that we excluded in the proof. □

One can notice that we use the topological property of the real plane when we apply Euler's formula in the proof of Theorem 8.2. Now we will present one example of how the sum-product problem is related to incidence geometry.

Theorem 8.6 (Elekes). *If $A \subset \mathbb{R}$, then $|A + A||A \cdot A| \gtrsim |A|^{5/2}$.*

Elekes (1997)

Corollary 8.7. *If $A \subset \mathbb{R}$, then* $\max\{|A+A|, |A \cdot A|\} \gtrsim |A|^{5/4}$.

Proof of Theorem 8.6. Let $\mathcal{P} = \{(x,y) : x \in A+A, y \in A \cdot A\}$ and $\mathcal{L} = \{y = a(x-a') : a, a' \in A\}$. For a line $y = a(x-a')$ in $\mathcal{L}$, $(a'+b, ab) \in \mathcal{P}$ is on the line for all $b \in A$, so each line in $\mathcal{L}$ contains $|A|$ incidences. By definition of $\mathcal{P}$ and $\mathcal{L}$, we have

$$|\mathcal{P}| = |A+A||A \cdot A| \qquad \text{and} \qquad |\mathcal{L}| = |A|^2.$$

By Theorem 8.3, we obtain

$$\begin{aligned} |A|^3 \leq I(\mathcal{P}, \mathcal{L}) &\leq |\mathcal{P}|^{3/2}|\mathcal{L}|^{3/2} + |\mathcal{P}| + |\mathcal{L}| \\ &\lesssim |A+A|^{3/2}|A \cdot A|^{3/2}|A|^{4/3}. \end{aligned}$$

Rearranging gives the desired result. □

8.3 Sum-product via multiplicative energy

In this chapter, we give a different proof that gives a better lower bound.

Theorem 8.8 (Solymosi). *If $A \subset \mathbb{R}_{>0}$, then* Solymosi (2009)

$$|A \cdot A||A+A|^2 \geq \frac{|A|^4}{4\lceil \log_2 |A| \rceil}$$

Corollary 8.9. *If $A \subset \mathbb{R}$, then*

$$\max\{|A+A|, |A \cdot A|\} \geq \frac{|A|^{4/3}}{2\lceil \log_2 |A| \rceil^{1/3}}$$

We define ***multiplicative energy*** to be

$$E_\times(A) = |\{(a,b,c,d) \in A^4 : \text{there exists some } \lambda \in \mathbb{R} \text{ such that } (a,b) = \lambda(c,d)\}|$$

Note that the multiplicative energy is a multiplicative version of additive energy. We can see that if A has a small product set, then the multiplicative energy is large.

$$\begin{aligned} E_\times(A) &= \sum_{x \in A \cdot A} |\{(a,b) \in A^2 : ab = x\}|^2 \\ &\geq \frac{|A|^4}{|A \cdot A|} \end{aligned}$$

The inequality follows from Cauchy–Schwarz inequality. Therefore it suffices to show

$$\frac{E_\times(A)}{\lceil \log_2 |A| \rceil} \leq 4|A \cdot A|^2.$$

Proof of Theorem 8.8. We use the dyadic decomposition method in this proof. Let A/A be the set $\{a/b : a, b \in A\}$.

$$\begin{aligned} E_\times(A) &= \sum_{s\in A/A} |(s\cdot A)\cap A|^2 \\ &= \sum_{i=0}^{\lceil \log_2 |A|\rceil} \sum_{\substack{s\in A/A \\ 2^i\le ((s\cdot A)\cap A)<2^{i+1}}} |(s\cdot A)\cap A|^2 \end{aligned}$$

By pigeonhole principal, there exists some k such that

$$\frac{E_\times(A)}{\lceil \log_2 |A|\rceil} \le \sum_{\substack{s\in A/A \\ 2^k\le |(s\cdot A)\cap A|<2^{k+1}}} |(s\cdot A)\cap A|^2.$$

We denote $D = \{s : 2^k \le |(s\cdot A)\cap A| < 2^{k+1}\}$ and we sort the elements of D as $s_1 < s_2 < \cdots < s_m$. Then one has

$$\frac{E_\times(A)}{\lceil \log_2 |A|\rceil} \le \sum_{s\in D} |(s\cdot A)\cap A|^2 \le |D| 2^{2k+2}.$$

For each $i \in [m]$ let ℓ_i be a line $y = s_i x$ and let ℓ_{m+1} be the vertical ray $x = \min(A)$ above ℓ_m.

Let $L_j = (A\times A)\cap \ell_j$, then we have $|L_j + L_{j+1}| = |L_j||L_{j+1}|$. Moreover, the sets $L_j + L_{j+1}$ are disjoint for different j, since they span in disjoint regions.

We can get the lower bound of $|A+A|^2$ by summing up $|L_j + L_{j+1}|$ for all j.

$$\begin{aligned} |A+A|^2 &= |A\times A + A\times A| \\ &\ge \sum_{j=1}^{m} |L_j + L_{j+1}| \\ &= \sum_{j=1}^{m} |L_j||L_{j+1}| \\ &\ge m 2^{2k} \ge \frac{E_\times(A)}{4\lceil \log_2 |A|\rceil} \end{aligned}$$

Combining the above inequality with $E_\times(A) \ge |A|^4/|A\cdot A|$, we reach the conclusion. □

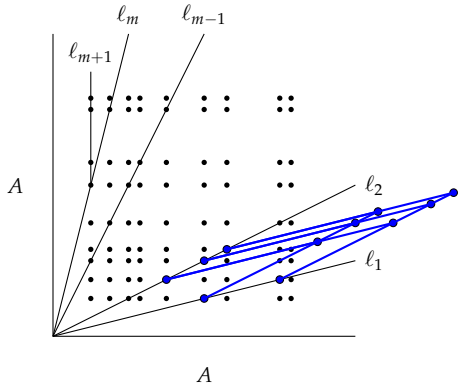


Figure 8.2: Illustration of $L_j + L_{j+1}$

18.217 Graph Theory and Additive Combinatorics

Fall 2019