Graph Theory and Additive Combinatorics

Lecturer: Yufei Zhao

E-mail address: yufeiz@mit.edu
URL: http://yufeizhao.com/gtac/

Contents

About this document

These are the course notes for MIT 18.S997 Special Subject in Mathematics: Graph Theory and Additive Combinatorics, taught by Yufei Zhao in Fall 2017. The notes are collectively written and edited by class participants.[1]

This document is hosted on Overleaf, a collaborative online LaTeX editor. It is viewable by everyone and editable by class participants. Overleaf has a number of features including real-time collaborative editing and LaTeX compilation, version history control, and (optional) Git integration.

Instructions on writing and editing

(For class participants) Summary:

- Sign up to write one (or more, depending on class size) lecture(s).
- Edit these notes on Overleaf for typos/improvements.
- Use Piazza to ask questions and discuss more substantial edits.

Sign up to write lectures notes by putting your name and email in the file writers.txt found in this folder (on Overleaf). Also register for an Overleaf account, so that all your edits will appear under your name (instead of "anonymous user"). If you are registered for credit, sign up to write one (or more, depending on class size) lecture(s), unless you are the editor-in-chief, in which case you are responsible for maintaining the overall state of this document.

You should upload your write-up **within three days of the lecture** (and preferably before the next lecture, for Tuesday lectures). To upload, create a separate file for your notes in the class Overleaf project, and name it "lec#.tex", with "#" replaced by the number of the lecture you're scribing. When writing, follow the formatting and organization conventions used in the notes from the first lecture. Also, try to avoid using custom shortcuts.

[1] The idea of collaboratively created lecture notes was borrowed from Ravi Vakil's undergraduate algebraic geometry course at Stanford.

(You can always write a first draft using your favorite shortcuts, then search and replace them with standard commands before uploading your file to Overleaf.)

Edit the notes on Overleaf for typos/improvements (in any section). More substantial edit requests may be directed towards the writer of the section. Feel free to use the class Piazza forum to discuss anything related to the course or the notes. You are responsible for making sure that your own section(s) are in good shape.

Use `\todo{...}` to add comments like this one.

If you make substantial edits to the project (and in particular, after uploading the notes for a lecture), label the version you've just created, using the History and Revisions button at the top of the Overleaf editor. Then this version of the project can be easily retrieved later if necessary.

Piazza: `https://piazza.com/mit/fall2017/18s997`

Writing assignments

```
Lecturer: Yufei Zhao yufeiz@mit.edu
Editor-in-chief: Gwen McKinley gweneth@mit.edu
Lec 1, 9/7: Yufei Zhao yufeiz@mit.edu
Lec 2, 9/12: Evan Chen evanchen@mit.edu
Lec 3, 9/14: Yibo Gao gaoyibo@mit.edu
Lec 4, 9/19: Ganesh Ajjanagadde gajjanag@mit.edu
Lec 5, 9/21: Morris Ang (Jie Jun) angm@mit.edu
Lec 6, 9/26: Lisa Yang lisayang@mit.edu
Lec 7, 9/28: Albert Soh asoh@mit.edu
Lec 8, 10/3: Ryan Alweiss alweiss@mit.edu
Lec 9, 10/5: Matthew Brennan brennanm@mit.edu
Lec 10, 10/12: Hong Wang hongwang@mit.edu
Lec 11, 10/17: Linus Hamilton luh@mit.edu
Lec 12, 10/19: Yonah Borns-Weil yonahb@mit.edu / Matt Babbitt mwabbitt@MIT.EDU
Lec 13, 10/24: Jake Wellens jwellens@mit.edu
Lec 14, 10/26: Minjae Park minj@mit.edu
Lec 15, 10/31: Diego Roque droque@mit.edu
Lec 16, 11/2: Younhun Kim younhun@mit.edu
Lec 17: 11/7: Jerry Li jerryzli@MIT.EDU / Brynmor Chapman brynmor@mit.edu
Lec 18: 11/9: Jonathan Tidor jtidor@mit.edu / Nicole Wein nwein@mit.edu
Lec 19: 11/14: Kevin Sun sunkevin@mit.edu / Meghal Gupta meghal@mit.edu
Lec 20: 11/16: Pakawut Jiradilok pakawut@mit.edu
```

Lec 21: 11/21 Sarah Tammen setammen@mit.edu

Lec 22: 11/28 Frederic Koehler fkoehler@mit.edu / Christian Gaetz gaetz@mit.edu

Lec 23: 11/30 Saranesh Prembabu saranesh@mit.edu / Benjamin Gunby bgunby@g.harvard.edu

Lec 24: 12/5 Vishesh Jain visheshj@mit.edu

Lec 25: 12/7: Dhroova Aiylam dhroova@mit.edu

Lec 26: 12/12 Paxton Turner pax@mit.edu (last lecture)

[If you are taking the course for credit and have not yet signed up to take notes for a lecture, please double-up with someone for a lecture (preferably for lectures later in the semester). Please make arrangements among yourselves by email and make sure that both people are fine with teaming up.]

CHAPTER 1

Introduction

§1.1 Schur's theorem

Lec1: Yufei Zhao

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. He was unsuccessful. It turns out that, for every n, the equation has nontrivial solutions mod p for all sufficiently large primes p, which Schur established by proving the following classic result.

thm:schur **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 $[N] := \{1, 2, \ldots, N\}$.

thm:schur-finitary **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 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. Such questions tend to difficult even to approximate.

thm:schurthm:schur-finitary

Let us show that the two formulations, Theorem 1.1 and 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 $\phi_N : [N] \to [r]$ 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 $X^n + Y^n \equiv Z^n \pmod{p}$.

thm:schur-fermat **Theorem 1.3 (Schur).** *Let n be a positive integer. For all sufficiently large primes p, there are $X, Y, Z \in \{1, \ldots, p-1\}$ such that $X^n + Y^n \equiv Z^n \pmod{p}$.*

Proof 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 $\{1, 2, \ldots, p-1\}$ into at least 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.

thm:ramsey-triangle **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.*

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, using r colors, the edges of a complete graph on $r(N' - 1) + 2$ Pick an arbitrary vertex v. Of the $r(N' - 1) + 1$ edges incident to v, at least N' edges incident to v have the same color, say red. If there is a red edge between any of these N' vertices, we obtain a red triangle. Otherwise, there are at most $r-1$ colors appearing among these N' vertices, so 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) the color $\phi(|i-j|)$. 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. □

§1.2 Highlights from additive combinatorics

The term *additive combinatorics* describes a rapidly growing body of mathematics motivated by simple-to-state questions about addition and multiplication of integers. Its methods are deep and far-reaching, connecting many different areas of mathematics such as graph theory, harmonic analysis, ergodic theory, and discrete geometry. The rest of this section highlights some important developments in additive combinatorics in the past century.

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

thm:vdw **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.*

Remark 1.6. Having arbitrarily long arithmetic progression 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.)

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 in this course are motivated by these results and the developments around them.

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

thm:szemeredi-intro **Theorem 1.8** (Szemerédi's theorem). *Any subset of the integers with positive upper density contains arbitrarily long arithmetic progressions.*

Szemerédi's theorem is deep, and all known proofs are long and involved. The result 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 the most influential modern perspectives towards Szemerédi's theorem (in historical order):

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

The relationships between these disparate approaches are slowly being understood. A unifying theme underlying all known approaches to Szemerédi's theorem is the *dichotomy between structure and pseudorandomness*, which we will see in several places in this course.

Let us mention 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 $\limsup_{N\to\infty} |A \cap [-N,N]^d|/(2N+1)^d > 0$ (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, where a pattern means a finite set up to dilation and translation. The following multidimensional generalization of Szemerédi's theorem was proved by Furstenberg and Katznelson.

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

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.

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

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.

thm:poly-sz **Theorem 1.11** (Polynomial Szemerédi theorem). *Suppose $A \subset \mathbb{Z}$ has positive upper density. If $P_1, \ldots, P_k \in \mathbb{Z}[X]$ 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 $x + P_1(y), \ldots, x + P_k(y) \in A$.*

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 this course. In fact, currently the only known proof of Theorem 1.11 uses ergodic theory, which is beyond the scope of this course.

Another groundbreaking development is the following famous result of Green and Tao, which settled an old folklore conjecture about prime numbers.

thm:green-tao **Theorem 1.12** (Green–Tao theorem). *The primes contain arbitrarily long arithmetic progressions.*

§1.3 What's next?

An important goal of this course is to understand two different proofs of Roth's theorem, which can be rephrased as:

> 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 course. 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.* He and Ruzsa used the graph regularity lemma to give a new, graph theoretic proof of Roth's theorem. One of the first goals of this course will be to understand this graph theoretic proof.

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 (also for us) in extremal graph theory is the following question:

> 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:

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

We do not know the exact answer, but we will prove, using Szemerédi's regularity lemma, that that any such graph must have $o(n^2)$ edges. Roth's theorem would then follow.

Notes

Schur's results (Theorems 1.1–1.3) were proved in [**44**].

Theorem 1.4 is a special case of Ramsey's theorem [**34**], which tells us that for every k, r, s, there is some N such that if the edges of a k-uniform hypergraph on N are colored with r colors, then there is a monochromatic clique on s vertices.

van der Waerden's theorem (Theorem 1.5) was proved in [**54**]. Erdős and Turán [**15**] conjectured that the real reason for van der Waerden's theorem is that some color class has positive density. Settling the Erdős–Turán conjecture, Roth [**40**] proved it for 3-term arithmetic progressions (Theorem 1.7), and Szemerédi [**50**] proved it for 4-term arithmetic progressions and subsequently [**51**] for k-term arithmetic progressions for any k. There have been several different proofs of Szemerédi's theorem that have been quite influential. Furstenberg [**18**] (see also [**19**, **21**]) proved an ergodic theoretic result that is equivalent to Szemerédi's theorem. Gowers [**22**] generalized Roth's Fourier analytic approach to longer arithmetic progressions and significantly improved early bounds on Szemerédi's theorem for longer progressions. Rödl et al. [**32**, **35**–**39**] and independently Gowers [**23**] generalized the graph regularity lemma to hypergraphs, and gave a new combinatorial proof of Szemerédi's theorem.

The multidimensional Szemerédi theorem (Theorem 1.9) was initially proved by Furstenberg and Katznelson [**20**] using ergodic theory. The first non-ergodic theoretic proof was via the hypergraph regularity lemma mentioned above.

Theorem 1.10 was conjectured by Lovász and proved independently by Furstenberg [**18**] (via ergodic theory) and by Sárkőzy [**42**] (using Fourier analytic methods). The polynomial Szemerédi theorem (Theorem 1.11) was proved by Bergelson and Leibman [**5**] using ergodic theory, and no other proof is currently known (except in some special cases, e.g., [**33**]).

The Green–Tao theorem was proved in [**24**].

Ruzsa and Szemerédi [**41**] proved the triangle removal lemma using the graph regularity lemma [**52**]. This gives a graph theoretic proof of Roth's theorem.

CHAPTER 2

Extremal graph theory

ch:extremal-graph-theory

Lec2: Evan Chen

§2.1 Mantel's theorem: forbidding a triangle

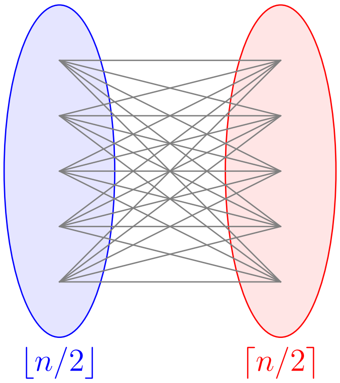
We start with the following question.

> What is the maximum number of edges in an n-vertex triangle-free graph?

These type of questions is the focus of this chapter. Namely we are interested in understanding extremal graphs while forbidden certain substructures.

thm:mantel **Theorem 2.1** (Mantel's theorem)**.** *If a graph G on n vertices contains no triangles, then it has at most $n^2/4$ edges.*

Remark 2.2. The extremal example is to take a complete bipartite graph $K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$. It is easy to see that bipartite graphs are triangle-free. This graph has $\lfloor n^2/4\rfloor$ edges. Thus the bound in Mantel's theorem is tight.



We give two different proofs of this theorem. They illustrate different ideas.

First proof of Theorem 2.1. Suppose G has m edges. We use the notation $d(x)$ for the degree of the vertex x.

Observe that if x is adjacent to y in G, then every other vertex other than x, y cannot be adjacent to both x and y. Consequently,

$$d(x) + d(y) \le n \qquad \text{for every edge } xy \text{ of } G.$$

Sum over all edges, we obtain

$$\sum_{xy \in E(G)} (d(x) + d(y)) \le mn.$$

On the left-hand side, for every vertex $x \in V(G)$, the term $d(x)$ appears once for each edge incident to x, i.e., $d(x)$ times in total. Thus the left-hand side is equal to $\sum_{x \in V(G)} d(x)^2$. It follows that

$$\sum_{x \in V(G)} d(x)^2 \le mn.$$

Now by the Cauchy-Schwarz inequality, we have

$$(2m)^2 = \left(\sum_x d(x) \right)^2 \le n \left(\sum_x d(x)^2 \right) \le mn^2. \qquad \square$$

Second proof of Theorem 2.1. Let A be an independent set of maximum size in G.Since G is triangle-free, the neighborhood of any vertex is an independent set, so $d(x) \le |A|$ for all $x \in V(G)$.

On the other hand, let $B = V(G) \setminus A$. Since A is independent, every edge of G must have at least one endpoint in B. So

$$e(G) \le \sum_{x \in B} d(x) \le \sum_{x \in B} |A| = |A||B| \le \left(\frac{|A| + |B|}{2} \right)^2 = \frac{n^2}{4}. \qquad \square$$

Remark 2.3. By checking the equality conditions in the second proof, we see that $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ is the *unique* triangle-free with the maximum number $\lfloor n^2/4 \rfloor$ number of edges..

Remark 2.4. The proof above works equally well if we replace A by the neighborhood of any maximum degree vertex of G.

§2.2 Turán's Theorem: forbidding a clique

sec:extremal-turan

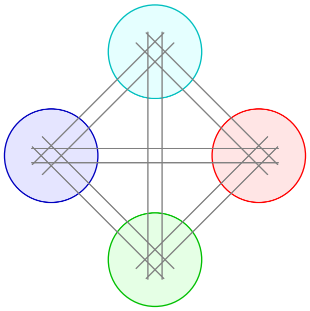
We now consider the following generalization:

> What's the maximum number of edges in an n-vertex graph that does not contain a clique on $r + 1$ vertices?

We begin by constructing the following example.

def:turan-graph *Definition* 2.5. Let the *Turán graph* $T_{n,r}$ be the n-vertex graph obtained by partitioning the vertex set into parts of sizes either $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$ and an edge between any pair of vertices lying in different parts.

Note that there is a unique way to partition n vertices into parts of sizes differing by at most 1.

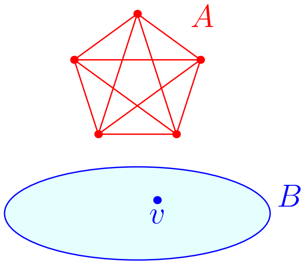


thm:turan **Theorem 2.6** (Turán's Theorem). *If G is an n-vertex graph with no copies of K_{r+1} then it has at most $e(T_{n,r})$ edges.*

We will give three proofs.

First proof of Theorem 2.6. (By induction) We proceed by induction on the number of vertices n. The result is trivial when $n \le r$. So assume $n > r$ from now on.

Let G be a K_{r+1}-free graph on n vertices with the possible maximum number of edges. Then G contains a copy of K_r (otherwise we adding another edge to G would still make it K_{r+1}-free). Let A a set of r vertices of G forming a clique and let $B = V(G) \setminus A$.



Then every vertex in B is adjacent to at most $r-1$ vertices in A, or else this vertex together with B forms a copy of K_{r+1}. [1] There are $\binom{r}{2}$

[1] At this point it's important to remark that we are really "already done" — we have a calculation which is going to be tight at the equality case (we are just deleting a single K_r from $T_{n,r}$). So even before completing the calculation, we know we should get the exact quantity.

edges within A, at most $(r-1)|B|$ edges between A and B by the above observation, and at most $e(T_{n-r}, r)$ in B by induction. Thus

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

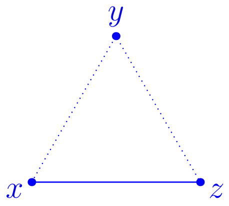
where the final step can be observed by noting that one can obtain $T_{n-r,r}$ from T_n by removing a vertex from each of the r vertex parts. □

thm:turan

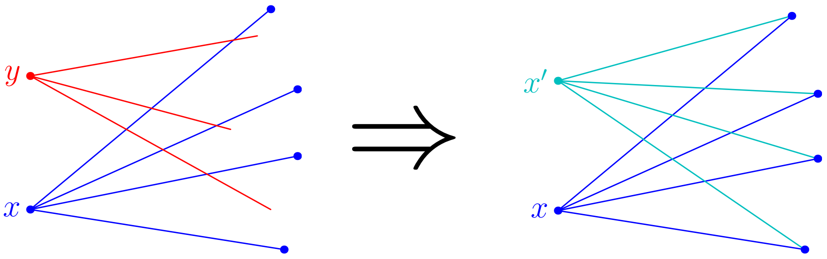
Second proof of Theorem 2.6. (By Zykov symmetrization) Let G be an n-vertex K_{r+1}-free graph with the maximum possible number of edges. We will show that

> For any $x, y, z \in V(G)$, if $xy \notin E(G)$ and $yz \notin E(G)$, then $xz \notin E(G)$.

Assume this is not true, so $xy, yz \notin E(G)$ but $xz \in E(G)$.



Suppose first that $d(y) < d(x)$. Then we replace y by a "clone" x' of x, connected to exactly the neighbors of x.



Then the number of edges increases, and the graph remains still K_{r+1}-free (note that there is no edge between x and x'). Thus, we may assume $d(x) \leq d(y)$. Similarly, we may assume $d(z) \leq d(y)$.

Now let G' be the graph obtained from G by replacing both x and z be clones of y. Then G' is K_{r+1}-free. Since we added $2d(y)$ edges and removed

$d(x) + d(z) - 1$ edges,

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

This contradicts the assumption that G has the maximum number of edges among n-vertex K_{r+1}-free graphs. Therefore, we must have $xz \in E(G)$, thus proving the claim above.

It follows that non-adjacency (i.e., $xy \notin E(G)$) in G is an equivalence relation, so that the complement of G is a disjoint union of cliques. Hence G must be a complete multipartite graph with at most r parts. The parts of G must have sizes differing by at most one, since if two parts differ in more than one vertex in size, then moving a vertex from the bigger part to the smaller part increases the number of edges. It follows that G must be the Turán graph $T_{n,r}$. □

Remark 2.7. Both the first and second proof also show that the Turán graph $T_{n,r}$ is the unique graph achieving the maximum number of edges.

Third proof of Theorem 2.6 (by probabilistic method). Let G be an n-vertex K_{r+1}-free graph with m edges. Take a uniform random ordering of the vertices, and define a random set $X \subset V(G)$ by

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

Then X is a clique. For any $v \in V$, we have $x \in X$ if and only if v appears before all its non-neighbors, so

$$\mathbb{P}(v \in X) = \frac{1}{n - d(v)}.$$

Consequently, by linearity of expectation

$$\mathbb{E}\left[|X|\right] = \sum_{v \in V} \mathbb{P}(v \in X) = \sum_{v \in V} \frac{1}{n - d(v)} \geq n \cdot \frac{1}{n - \frac{2m}{n}} = \frac{1}{1 - \frac{2m}{n^2}},$$

where in the final step follows from Jensen's inequality on the convex function $x \mapsto \frac{1}{n-x}$, noting that $2m/n$ is the average degree in G. On the other hand, since X is a clique and G is K_{r+1}-free, we have $|X| \leq r$. Combining with the above displayed inequality, we obtain $1/r \leq 1 - 2m/n^2$. Hence

$$m \leq \left(1 - \frac{1}{r}\right) \frac{n^2}{2}.$$

This proves the theorem in the case when n/r is an integer. If n is not divisible by r, one can modify the proof in the step where we apply Jensen's inequality, noting that, given $\sum_v d(v)$, the expression $\sum_{v\in V} 1/(n-d(v))$is minimized, if the numbers $d(v)$ are as close to each other as possible (hence differing by at most one from each other). The rest of the analysis is straightforwards, so we omit the details. □

§2.3 Erdős-Stone-Simonovits theorem: forbidding a general graph

In Mantel's theorem and Turán's theorem, we considered the problem of determining the maximum of edges in a graph without a clique of given size. We will generalize this notion by forbidding an arbitrary subgraph.

def:ex *Definition* 2.8. Let $\mathrm{ex}(n,H)$ denote the maximum number of edges in an n-vertex graph that does not contain a copy of H.

Remark 2.9. In the above definition, we do *not* require the copy of H to be an *induced* subgraph. We say that H is a subgraph of G if H is obtained from G by taking a subset of vertices and a subset of edges between those vertices. On the other hand, we say that H is an *induced* subgraph of G if H is obtained from G by taking a subset of the vertices along with all edges of G between those vertices.

For example, C_4, the cycle on four vertices, is a subgraph of K_4, but not as an induced subgraph.

Generally, we only mean induced when we explicitly use the word "induced."

With this notation, Turán's theorem can be restated as

$$\mathrm{ex}\,(n,K_{r+1}) = e(T_{n,r})$$

It is worth noting that

$$\mathrm{ex}\,(n,K_{r+1}) \le \left(1-\frac{1}{r}\right)\frac{n^2}{2}$$

and

$$\mathrm{ex}\,(n,K_{r+1}) = \left(1-\frac{1}{r}+o(1)\right)\frac{n^2}{2} \quad \text{for fixed } r, \text{ as } n\to\infty.$$

Perhaps surprisingly, it is possible to determine the asymptotic behavior of $\mathrm{ex}(n,H)$ for a large class of graphs H simply using the chromatic number of H.

def:chromatic *Definition* 2.10. Let $\chi(H)$ denote the *chromatic number* of a graph H, the minimum number of colors needed to properly color H.

Example 2.11. We have $\chi(K_{r+1}) = r + 1$ and $\chi(T_{n,r}) = r$ (when $n \geq r$).

Remark 2.12. Determining $\chi(H)$ for general graphs is difficult; it is NP-complete even to determine whether a graph is 3-colorable.

Proposition 2.13. *For H a subgraph of G, the inequality $\chi(H) \leq \chi(G)$ holds.*

Proof. Any coloring of G gives a coloring of H. □

Thus, if $\chi(H) = r + 1$ then $T_{n,r}$ is an example of an H-free graph. Consequently,

$$\operatorname{ex}(n, H) \geq e(T_{n,r}) \qquad \text{where } r = \chi(H) - 1.$$

It turns out this example is “asymptotically” optimal in the following sense.

thm:ess **Theorem 2.14** (Erdős-Stone-Simonovits theorem[2]). *Fix a graph H. As $n \to \infty$, we have*

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

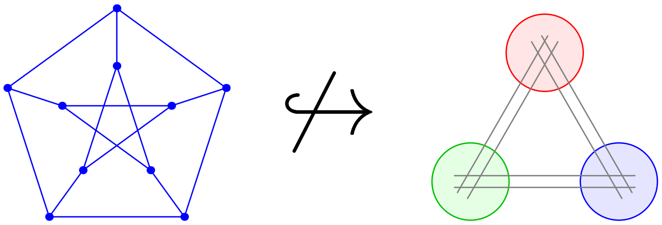
Example 2.15. Here are several examples of this theorem.

- The 5-cycle has chromatic number 3, so
$$\lim_{n\to\infty} \frac{\operatorname{ex}(n, K_3)}{\binom{n}{2}} = \frac{\operatorname{ex}(n, C_5)}{\binom{n}{2}} = \frac{1}{2}.$$
- Let **PG** denote the Petersen graph[3], which has chromatic number 4. Then
$$\lim_{n\to\infty} \frac{\operatorname{ex}(n, K_4)}{\binom{n}{2}} = \frac{\operatorname{ex}(n, \mathbf{PG})}{\binom{n}{2}} = \frac{2}{3}.$$

[2] Cultural remark: the three mathematicians Erdős, Stone, Simonovits never wrote a paper together. Erdős and Stone solved it for H a complete multipartite graph, and Erdős and Simonovits then proved it for general H in this way. Sometimes the theorem is just called the Erdős–Stone theorem.
As a quick digression, let's look at some pronunciations of Hungarian names. The two common mistakes are how these two are pronouced.
's' is pronounced as /sh/. For example, Erdős and Simonovits.
'sz' is pronounced as /s/. For example, Szemerédi and Lovász.
Note that Stone is an English name.

[3] Because every graph theory course needs to feature **PG**. Sometimes superficially.

The equality case is to take $T_{n,3}$ the point is that **PG** can't be embedded into $T_{n,3}$ by virtue of not being 3-colorable.



Remark 2.16. Note that if $\chi(H) = 2$, the theorem only implies that $\text{ex}(n, H) = o(n^2)$, which is much less satisfying. For most bipartite graphs H it is still open what the leading term of $\text{ex}(n, H)$ is.

thm:ess

The proof of Theorem 2.14 begins with the following lemma with a very general and useful idea. A powerful general message is that having "high" average often implies that we can find a "large" subset with "high" minimum. The specific quantifications depend on the applications in mind. In the following lemma, we transform a graph with high average degree into a subgraph with a high minimum degree, and this is done by removing low degree vertices one at a time.

lem:avgtomindeg **Lemma 2.17.** *Whenever $0 < \varepsilon < \delta$, for all n sufficiently large in terms of ε and δ, if G is an n-vertex graph with average degree at least δn then there exists a subgraph G' of G on $n' \geq \frac{\sqrt{\varepsilon}}{2}n$ vertices and with minimum degree at least $(\delta - \varepsilon)n'$.*

Lec3: Yibo Gao *Proof.* Let $G_0 = G$. Construct $G_1, G_2, \ldots$ by successively removing minimal degree vertices one at a time. If $\min\deg(G_i) \geq (\delta - \epsilon)(n - i)$, then we stop the process. We will show that by the time we stop, there are still a lot of vertices remaining. Suppose we stop at $G_{n-n'}$ (after $n - n'$ steps) with n' vertices remaining. Then,

$$\#\text{edges removed} < \sum_{k=n'+1}^{n} (\delta - \epsilon)k = (\delta - \epsilon)\frac{(n + n' + 1)(n - n')}{2}.$$

Since G has average degree at least δn,

$$\begin{aligned}\frac{\delta n^2}{2} \leq e(G) \leq& e(G_{n-n'}) + (\delta - \epsilon)\frac{(n + n' + 1)(n - n')}{2}\\ \leq& \binom{n'}{2} + (\delta - \epsilon)\frac{(n + n' + 1)(n - n')}{2}.\end{aligned}$$

Rearranging, we obtain

$$\frac{\epsilon n^2}{2} - \frac{(\delta-\epsilon)n}{2} \leq \binom{n'}{2} - (\delta-\epsilon)\frac{n'(n'+1)}{2}.$$

Thus, if n is large enough depending on δ and ϵ, the first quadratic term is dominant. We find that $n' \geq \frac{\sqrt{\epsilon}}{2}n$ as a result. The resulting graph $G' = G_{n-n'}$ has minimum degree at least $(\delta-\epsilon)n'$ by our construction. □

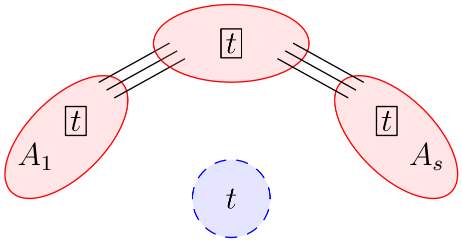
Let's proceed to prove Theorem 2.14, the Erdős-Stone-Simonovits theorem. Suppose that the graph has a lot of edges, then Lemma 2.17 tells us that we can focus on a subgraph with large minimal degree.

Proof of Theorem 2.14. Let $\chi(H) = r+1$ and $\epsilon > 0$. Let G be an n-vertex graph with more than $(1-\frac{1}{r}+2\epsilon)\frac{n^2}{2}$ edges. We will show that if n is large enough, then G contains a copy of H.

By Lemma 2.17, there exists a subgraph on n' vertices where $n' \geq \frac{\sqrt{\epsilon}}{2}n$ and minimum degree at least $(1-\frac{1}{r}+\epsilon)n'$. Note that n' can be made sufficiently large by forcing n to be sufficiently large. To ease notation, replace G by G' and assume that the minimum degree of G is at least $(1-\frac{1}{r}+\epsilon)n$ from now on.

We will show that for all t, if $n \geq n_0(s,t,\epsilon)$ is large enough, then G contains a complete $(r+1)$-partite graph with t vertices in each part. Since $\chi(H) = r+1$, by taking t to be large enough (depending on H), we obtain of copy of H in G.

We will do this by induction on $s = 1,2,\ldots,r+1$ with the induction hypothesis being that for every t, if n is sufficiently large, then G contains a complete s-partite graph with t vertices in each part. The case $s=1$ is trivial since the graph we are trying to embed has no edges.



By induction hypothesis, we can find the sets A_i's of sizes let $|A_i| = T := \lceil \frac{4t}{r\epsilon} \rceil$ for $i = 1,\ldots,s$. The next step is to find t vertices outside all the A_i's such that these t vertices have at least t common neighbors inside each A_i'.

Let's count cliques $(v_1, v_2, \ldots, v_s, w)$ where $v_1 \in A_1, \ldots, v_s \in A_s$, $w \notin A_1 \cup \cdots \cup A_s$. For any $(v_1, \ldots, v_s) \in A_1 \times \cdots \times A_s$, the number of choices for w is at least

$$\begin{aligned} n - s\left(\frac{1}{r} - \epsilon\right)n - |A_1 \cup \cdots \cup A_s| &\geq n - r\left(\frac{1}{r} - \epsilon\right)n - sT \\ &\geq \frac{r\epsilon n}{2}, \end{aligned}$$

for n sufficiently large (depending on r, ϵ, and T). Thus, the number of such cliques is at least $\frac{r\epsilon n}{2}|A_1| \cdots |A_s|$. Let $W \subset V \setminus (A_1 \cup \cdots \cup A_s)$ be the set of w's that appears in at least $\frac{r\epsilon}{4}|A_1| \cdots |A_s|$ such cliques. We then have

$$\frac{r\epsilon n}{2}|A_1| \cdots |A_s| \leq |W| \cdot |A_1| \cdots |A_s| + n \cdot \frac{r\epsilon}{4} \cdot |A_1| \cdots |A_s|.$$

Therefore, $|W| \geq \frac{r\epsilon}{4}n$. Note that every vertex in W has at least $\frac{r\epsilon}{4}|A_i|$ neighbors in each A_i, for $i = 1, \ldots, s$. (Otherwise, w is in less than $\frac{r\epsilon}{4}|A_1| \cdots |A_n|$ cliques.) There are $\binom{T}{t}^s$ ways to select a t-element subset from each A_i. By the pigeonhole principle, if n is large enough, there exist $B_1 \subset A_1, \ldots, B_s \subset A_s$ with $|B_i| = t$ such that there exist t different vertices in W, each adjacent to all vertices in $B_1, \ldots, B_s$. This completes an $(s+1)$-partite graph with t vertices in each part. □

ch:regularity

We will see another proof in Chapter 3 after developing Szemerédi's regularity lemma.

§2.4 Forbidding complete bipartite graphs

The Erdős-Stone-Simonovits theorem tells us that $\mathrm{ex}(n, H) = o(n^2)$ for bipartite H (as $\chi(H) = 2$), but this result does not tell us the precise asymptotic rate of growth of $\mathrm{ex}(n, H)$. For most bipartite graphs, the order of growth of $\mathrm{ex}(n, H)$ is unknown. We will see some upper and lower bounds, and a few cases where we can determine $\mathrm{ex}(n, H)$ up to a constant factor.

The most well-known case is the complete bipartite graph $K_{s,t}$. In particular, every bipartite graph is contained in $K_{s,t}$ for some s, t. So the extremal number of every bipartite graph is bounded by the extremal number of complete bipartite graphs $\mathrm{ex}(n, H) \leq \mathrm{ex}(n, K_{s,t})$.

prob:zar **Problem 2.18 (Zarankiewicz problem).** *Determine* $\mathrm{ex}(n, K_{s,t})$.

The problem remains open for general s, t. Here is an upper bound that has been conjectured to be tight up to a constant factor.

thm:kst **Theorem 2.19** (Kővári-Sós-Turán theorem). *Fix positive integers $s \le t$. There exists constant C such that*

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

For certain values of s and t, we will construct matching lower bound constructions (up to a constant factor). This is perhaps evidence that the upper bound is tight in general, in which case the current difficulty lies in finding good constructions of $K_{s,t}$-free graphs.

Proof. The proof is done via a double counting argument. Suppose G is $K_{s,t}$-free with n vertices and m edges. Let T be the number of copies of $K_{s,1}$ in G. We bound T from above and from below.

Since G is $K_{s,t}$ free so every set of s vertices has at most $t-1$ common neighbors. This gives $T \le (t-1)\binom{n}{s}$.

Letting $d(v)$ denote the degree of v, we have

$$T = \sum_{v \in V(G)} \binom{d(v)}{s} \ge n \binom{\frac{1}{n}\sum_v d(v)}{s} = n\binom{2m/n}{s}$$

by convexity of the function $p(x) = \binom{x}{k} = x(x-1)\cdots(x-k+1)/k!$ for $x \ge k$.[4]

Combining the upper and lower bounds on T, we find,

$$(1+o(1))n\frac{(2m/n)^s}{s!} = n\binom{2m/n}{s} \le (t-1)\binom{n}{s} = (1+o(1))(t-1)\frac{n^s}{s!}$$

where we can view s, t as constants and let $n \to \infty$. This implies

$$m \le (1+o(1))\frac{1}{2}(t-1)^{\frac{1}{s}} n^{2-\frac{1}{s}}. \qquad \square$$

Let us examine a geometric application of the Kővári-Sós-Turán theorem. The following problem, known as the unit distance problem, was originally posed by Erdős [**14**]:

Lec4: Ganesh Ajjanagadde

prob:unit_distance **Problem 2.20** (Unit distance problem). *Determine $u(n)$, the maximum number of times distance 1 occurs among n points placed in $\mathbb{R}^2$.*

For small n, Schade [**43**] obtained the following results by determining the extremal sets up to isomorphism:

[4] Notice that the roots of $p(x)$ are $0, 1, \ldots, k-1$ so $p'(x)$ has exactly one root in each interval $(0,1), (1,2), \ldots, (k-1,k)$ due to the degree of the polynomial. Similarly, each root of $p''(x)$ lies between two adjacent roots of $p'(x)$. It follows that $p''(x) > 0$ for $x \ge k$.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$u(n)$	0	1	3	5	7	9	12	14	18	20	23	27	30	33

The sequence $u(n)$ is also given as OEIS entry A186705 [**1**].

As a quick corollary of the Kövári-Sós-Turán theorem, we prove the following result due to Erdős [**14**, Theorem 2]:

thm:unit_distances **Theorem 2.21.** $u(n) = O(n^{3/2})$

Proof. Define the unit distance graph as a graph with vertex set consisting of the n points in $\mathbb{R}^2$, and an edge between points x and y if they are unit distance apart.

We claim that this graph is $K_{2,3}$-free. Consider two points x and y. Then the locus of points at unit distance from both x and y is the intersection of two circles of unit radius centered at x and y. Two circles in the plane intersect in at most two points, so the unit distance graph is $K_{2,3}$ free.

We may thus apply Theorem 2.19 to conclude that the graph has $O(n^{3/2}$ edges, so that $u(n) = O(n^{3/2})$ as desired. □

The best upper bound known for the unit distance problem is $O(n^{4/3})$, due to [**48**]. An elegant proof of this bound was given by Székely [**49**] based on the "crossing lemma" of Ajtai et al. [**2**] and independently, Leighton [**29**].

The best known lower bound is

$$u(n) \geq n^{1+\frac{c}{\log\log n}} \qquad (1)$$ eq:unit-distance-lower

for some constant $c > 0$. It is believed that in fact the lower bound above is tight, but this remains an open problem. Note that the unit distances problem is connected to the distinct distances problem, which was also raised by Erdős in [**14**]. This connection follows from the trivial inequality $d(n) \geq \frac{\binom{n}{2}}{u(n)}$, where $d(n)$ is the minimum number of distinct distances. By using the square lattice construction below, and the fact that the number of integers $\leq n$ that can be represented as a sum of two squares is $O\left(\frac{n}{\sqrt{\log n}}\right)$ (a standard result of analytic number theory obtained by Landau, see e.g [**31**, Ex. 21, p. 187]), one gets the upper bound $d(n) = O\left(\frac{n}{\sqrt{\log n}}\right)$. On the other hand, the trivial relation between d and u together with the best upper bound on u of $O(n^{4/3})$ yields a lower bound that is polynomially worse. This gap was closed to within a poly-logarithmic factor in work of Guth and Katz [**25**], who obtained $d(n) = \Omega\left(\frac{n}{\log n}\right)$. A recent survey of the unit distance problem is provided in [**53**].

The rest of this section gives a proof of the lower bound (1) to $u(n)$. It uses basic analytic number theory, and should be considered tangential to the rest of the course (it was not covered in lecture).

Before getting into the details, we first sketch the argument of the lower bound, emphasizing the key idea. The proof of the lower bound uses the graph consisting of the integer square lattice of size $\approx \sqrt{n} \times \sqrt{n}$. We focus on a distance $\sqrt{r}$ such that r has a relatively large number of representations as a sum of squares, and such that $r = o(n)$. This will guarantee that for essentially every vertex of the lattice (basically the vertices in the interior which has side length $\approx \sqrt{n} - \sqrt{r}$), there are at least $n^{\frac{c}{\log \log n}}$ neighbors at that distance. Exact details of this argument hinge on two ingredients from number theory:

(1) The number of representations of an integer r as a sum of two squares. This may be encapsulated in the following lemma (due to Lagrange):

lem:count_rep_sum_of_squares **Lemma 2.22.** *Let r have prime factorization $r = 2^e \prod_{i=1}^{k} p_i^{a_i} \prod_{j=1}^{l} q_j^{b_j}$ where the p_i are distinct primes of the form $4t+1$, and the q_j are distinct primes of the form $4t+3$. If one or more of the b_j are odd, then r has no representation as a sum of squares. If all the b_j are even, then the number of representations of r as a sum of squares (**all** representations, including negative integers) is: $N(r) = 4\prod_{i=1}^{k}(a_i + 1)$.*

Proof. The proof of this result is outside the scope of this course, but is elementary and may be found in many references on number theory, e.g [**26**, Theorem 278]. □

(2) From the above Lemma 2.22, it is clear that a “good” r is one which is $o(n)$, is odd and lacks divisors of the form $4t+3$ (since they don’t contribute to the number of representations). Thus, the heart of the matter is getting a handle on how many primes of the form $4t+1$ are there below a given number. What is thus needed is an analog of the celebrated prime number theorem, generalized to arithmetic progressions. Note that Dirichlet’s theorem on the infinitude of primes in arithmetic progressions is insufficient here, as it does not give quantitative estimates. Also note that the Euclidean method, which can be used to deduce that there are infinitely many primes of the form $4t+3$, and which moreover yields (weak) quantitative estimates, does not work for

$4t+1$. As such, one essentially needs the full strength of the prime number theorem for arithmetic progressions:

lem:pnt_ap **Lemma 2.23.** *Let p denote a prime number. Let*

$$\pi_{n,a}(x) = |\{p : p \leq x, p \equiv a \pmod{n}\}|\,.$$

Then if a, n are coprime, we have: $\pi_{n,a}(x) \approx \frac{1}{\phi(n)} \frac{x}{\log x}$, where $\phi(n)$ is Euler's totient function. One has the equivalent form $\psi(x; n, a) \approx \frac{x}{\phi(n)}$, where:

$$\begin{aligned}\psi(x; n, a) &= \sum_{i \leq x, i \equiv a \pmod{n}} \Lambda(i) \\ &= \sum_{p \leq x, p \equiv a \pmod{n}} \lfloor \log_p x \rfloor \log p.\end{aligned}$$

Here ψ is the Chebyshev function, and Λ the von-Mangoldt function.

Proof. This is a classic result of analytic number theory, and may be found in references such as [**10**, Chapter 20]. For a relatively simple, standalone exposition, one may see [**47**]. □

With these remarks in place, we turn to the proof of (1). We first prove the lower bound by elaborating upon the sketch given above. Consider $r = \prod_{p \leq c \log n, p \equiv 1 \pmod{4}} p$ for a $c > 0$. By Lemma 2.23 and 2.22, we see that the number of representations of r as a sum of squares of nonnegative integers is $2^{\Omega\left(\frac{\log n}{\log \log n}\right)} = n^{\Omega\left(\frac{1}{\log \log n}\right)}$. Also by 2.23,

$$\begin{aligned}\log r &\leq \sum_{p \leq c \log n, p \equiv 1 \pmod{4}} \log p \\ &\leq \psi(c \log n; 4, 1) \approx \frac{c}{2} \log n.\end{aligned}$$

Thus, for c small enough, $r = o(n)$. Now consider the set of points $\mathcal{P} = \left\{(x, y) \in \mathbb{Z}^2 : 0 \leq x, y \leq \lfloor \sqrt{n} \rfloor \right\}$, and its "interior"

$$\mathcal{P}^{\mathrm{o}} = \left\{(x, y) \in \mathbb{Z}^2 : 0 \leq x, y \leq \lfloor \sqrt{n} \rfloor - \lceil \sqrt{r} \rceil \right\}.$$

Then, for each point in $\mathcal{P}^{\mathrm{o}}$, we have by the above $n^{\Omega\left(\frac{1}{\log \log n}\right)}$ neighbors at distance $\sqrt{r}$. Now as $|\mathcal{P}^{\mathrm{o}}| \approx |\mathcal{P}|$, we get the desired lower bound $u(n) \geq n^{1+\Omega\left(\frac{1}{\log \log n}\right)}$.

§2.5 Constructing $K_{s,t}$-free graphs

Notice that the upper bound of Theorem 2.21 made use of the incidence geometry of circles in $\mathbb{R}^2$. This technique, applied to algebraic varieties over finite fields $\mathbb{F}_q$ will be the "workhorse" in this section. Our main goal here is to give constructions that solve 2.18 in some special cases by giving graphs with large number of edges that avoid $K_{s,t}$. These algebraic constructions, although richer structurally than Turán's graph, suffer from some defects:

Defer this whole discussion until after the proof of the finite field discussion. It doesn't really make sense until one has seen the construction. -YZ

(1) Finite field geometries only work for prime powers. General n do not have this form; and it is important to have a way to find a prime (or prime power) close to a given n. This is provided by results in analytic number theory that estimate prime gaps. There is a long sequence of such results, beginning with Hoheisel [**27**] (which is sufficient for our purposes), and with the current record in [**4**]:

lem:prime_gaps

Lemma 2.24. *For sufficiently large x, there exists a prime in the interval $[x, x + x^{0.525}]$.*

Basically, what we need is $[x, x + o(x)]$, which is guaranteed by Hoheisel's result itself. As a side note, there are good reasons to believe the $o(x)$ should be $\approx \log^2(x)$, but this is a major open problem on prime gaps. In any case, one could argue that this "defect" is easily patched by the above Lemma.

(2) The general rigidity of algebraic constructions as opposed to random ones. This phenomenon, already alluded to as the "*dichotomy between structure and pseudorandomness*", is at the heart of many things in mathematics. One such illustration, which shows that this is a feature not unique to additive combinatorics, is the problem of capacity achieving error correcting codes in information theory, where "structured" capacity achieving codes were unknown for a very long time, even though random codes with this feature date back to the seminal paper of Shannon [**45**].

We first examine a randomized construction that does not achieve the best possible bounds. This construction is a good illustration of the "alteration method". We have the following Theorem:

thm:rand_subgraph **Theorem 2.25.** *Fix a graph H with ≥ 2 edges. Then* $\mathrm{ex}(n,H) = \Omega\left(n^{2-\frac{v(H)-2}{e(H)-1}}\right)$.

Proof. Consider the Erdős-Rényi random graph $G(n,p)$, i.e a graph with n vertices, and edges placed i.i.d (independent and identically distributed) between pairs of vertices with probability $p = \frac{1}{2}n^{-\frac{v(H)-2}{e(H)-1}}$. Consider the random variable $J = \#(\text{copies of } H \text{ in } G)$. By linearity of expectation,

$$\mathbb{E}\left[J\right] = p^{e(H)}\frac{\prod_{i=0}^{v(H)-1}(n-i)}{|\mathrm{aut}(H)|} \leq p^{e(H)}n^{v(H)}.$$

Also by linearity,

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

For any sample graph G drawn from the random variables, we may remove at most J edges (one for each possible copy of H) to get an H-free graph with at least $e(G) - J$ edges. For n sufficiently large, one has

$$\begin{aligned}\mathbb{E}\left[e(G) - J\right] &\geq p\binom{n}{2} - p^{e(H)}n^{v(H)} \\ &\geq \frac{1}{2}p\binom{n}{2} \\ &\geq \frac{1}{16}n^{2-\frac{v(H)-2}{e(H)-1}}.\end{aligned}$$

Thus there exists a G such that $e(G) - J \geq \frac{1}{16}n^{2-\frac{v(H)-2}{e(H)-1}}$. Take this graph, and remove an edge from each copy of H in G to get an H-free graph with $\Omega\left(n^{2-\frac{v(H)-2}{e(H)-1}}\right)$ edges, as desired. □

The term "alteration" is justified in the above proof by the fact that we first start off with a possibly "bad" random graph, and then do some modifications of that graph to get a graph that has been "patched" to satisfy the given requirements. Note also that the modification procedure can be coupled to the particular random instance, since we only used the linearity of expectation above.

Let us now examine when the bound of Theorem thm:rand_subgraph 2.25 is useful. First off, this is clearly only of use when $\chi(H) = 2$, i.e H is bipartite. Specializing to the complete bipartite case, we get:

$$\mathrm{ex}(n, K_{s,t}) = \Omega\left(n^{2-\frac{s+t-2}{st-1}}\right). \tag{2}$$ eqn:rand_kst

Taking $s = t$ in (2), and combining with Theorem 2.19, we have:

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

Setting $s = t = 2$, we have

$$cn^{2-\frac{2}{3}} \leq \mathrm{ex}(n, K_{2,2}) \leq Cn^{2-\frac{1}{2}}.$$

We now turn to algebraic construction methods based on incidence geometries in finite fields that will allow us to prove the tightness of the upper bound in some cases. First, we give a result of Erdős, Rényi, and Sós [**16**, Corollary 2]:

Given an alternative (informal) discussion of the construction as the incidence graph of points and lines

thm:ers_k22 **Theorem 2.26.** $\mathrm{ex}(n, K_{2,2}) \geq \left(\frac{1}{2} - o(1)\right) n^{\frac{3}{2}}$.

Proof. Suppose $n = p^2 - 1$ for p a prime. Let $V = \mathbb{F}_p^2 \setminus \{(0,0)\}$. Let edges be given by $(x, y) \sim (a, b)$ iff $ax + by = 1$ in $\mathbb{F}_p$, and call the resulting graph G. Then it is clear that $d(v) \geq p-1$ for all $v \in V$, so that $e(G) \geq \left(\frac{1}{2} - o(1)\right) n^{\frac{3}{2}}$. What we now need to show is that G is in fact $K_{2,2}$-free, or equivalently that "two distinct lines can meet in at most one point". Formally, this amounts to showing that the system $ax + by = 1, a'x + b'y = 1$ has at most one solution in $(x, y) \in \mathbb{F}_p^2$ for $(a, b) \neq (a', b')$.

There are two cases. If (a, b) and (a', b') are linearly independent of each other, we are done. If they are linearly dependent, say without loss $(a, b) = \lambda(a', b')$, then we must have $\lambda = 1$ for there to exist a solution (x, y), and this contradicts $(a, b) \neq (a', b')$. Thus G is indeed $K_{2,2}$-free as desired.

For general n, we simply invoke Lemma 2.24. One way to handle it is simply to leave the "residue" $n - (p^2 - 1)$ vertices isolated from the rest of the graph and each other. □

Note that the above proof works for any finite field. Also note that $\frac{1}{2}$ is the correct leading constant, as can be determined by examining the proof of Theorem 2.19. Furthermore, since we solved the case of $s = t = 2$, and since a $K_{2,2}$-free graph is necessarily $K_{2,t}$-free for all $t \geq 2$, we have shown that Theorem 2.19 is tight up to constant factors for $s = 2$.

We now give a result of Brown [**8**] that solves the case of $s = 3$.

thm:bro_k33 **Theorem 2.27.** $\mathrm{ex}(n, K_{3,3}) = \Omega\left(n^{\frac{5}{3}}\right)$.

Proof. The below proof is only a sketch; full details are given in [**8**]. Let the vertex set $V = \mathbb{F}_p^3$. Let edges be given by $(x, y, z) \sim (a, b, c)$ if and only if $(x-a)^2 + (y-b)^2 + (z-c)^2 = \alpha$ in $\mathbb{F}_p$. Let the resulting graph be denoted by G. Here, $\alpha = 1$ if $p \equiv 3 \pmod 4$, and α is taken as a fixed quadratic nonresidue (dependent on p) if $p \equiv 1 \pmod 4$. Then by a theorem of Lebesgue (see e.g [**11**, p. 325]), $d(v) = p^2 - p$ for all $v \in V$. Thus, $e(G) = \Theta(p^5)$. It remains to check that G is $K_{3,3}$-free. Suppose not. Then, we essentially have the situation of three spheres of equal radii intersecting at 3 points. But in "reasonable" geometries, 3 spheres of equal radii can intersect only in 2 points. The above is not rigorous, but gives some idea as to why one would expect a contradiction and thus conclude that G is $K_{3,3}$-free. Invoking Lemma 2.24 then completes the proof. □

For the next construction, it will be helpful to recall some elementary facts about finite fields.

lem:pow_auto **Lemma 2.28.** *Let $\mathbb{F}_q = \mathbb{F}_{p^t}$ be a finite field of characteristic p. Then the map $f : x \to x^p$ is an automorphism.*

Proof. f is obviously multiplicative. Furthermore, $f(x+y) = x^p + y^p$ by the binomial theorem, $\binom{p}{i} \equiv 0 \pmod p$ $\forall 0 < i < p$, and since the characteristic is p. Thus f is an endomorphism. Since the range is finite, it suffices to show that f is injective. Suppose $x^p = y^p$. If $y = 0$, then $x = 0$, and vice versa. Otherwise, we have $a^p = 1$ for $a = xy^{-1}$. Consider the multiplicative group $F_q^\times$. Order of a divides the order of the group, so order of a divides $(p, p^t - 1) = 1$, so $a = 1$ as desired. □

lem:norm_map **Lemma 2.29.** *Let $\mathbb{F}_q = \mathbb{F}_{p^t}$ be a finite field of characteristic p. The* norm map *over $\mathbb{F}_p$ is defined by $N(x) := \prod_{i=0}^{t-1} x^{p^i}$. Then $N(x) \in \mathbb{F}_p$ for all $x \in \mathbb{F}_{p^t}$. Moreover, $|\{x : N(x) = 1\}| = \frac{p^t-1}{p-1}$.*

Proof. First, note that $x^{p^t} = x$ for all $x \in \mathbb{F}_{p^t}$, since if x is not 0, order of x in $\mathbb{F}_q^\times$ divides $p^t - 1$, and if $x = 0$, this is clearly true. Thus, $N(x)^p = \prod_{i=1}^{t} x^{p^i} = \left(\prod_{i=1}^{t-1} x^{p^i}\right) x^{p^0} = N(x)$. The subfield left fixed by the automorphism $x : x \to x^p$ is precisely the base field $\mathbb{F}_p$, so $N(x) \in \mathbb{F}_p$. Now, $N(x) = x^{\frac{p^t-1}{p-1}}$. By letting α be a generator of $\mathbb{F}_q^\times$, we see that $N(\alpha^i)$ over $0 \leq i < p-1$ are distinct. N thus yields a epimorphism of groups $\mathbb{F}_q^\times \twoheadrightarrow \mathbb{F}_p^\times$. By the first isomorphism theorem for groups, $|\{x : N(x) = 1\}| = \frac{p^t-1}{p-1}$ as desired. □

We now give a lemma from commutative algebra, whose proof we omit, and simply refer the reader to the papers [**28**].

lem:comm_alg_sys **Lemma 2.30.** *Let $\mathbb{F}$ be any field. Then the system of equations*

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

has at most $s!$ solutions $(x_1, x_2, \dots, x_s)$ in $\mathbb{F}$, if $a_{ij} \neq a_{i'j}$ for all j, $i \neq i'$, and $a_{ij}, b_k \in \mathbb{F}\ \forall 1 \leq i, j, k \leq s$.

Remark 2.31. If $b_i = 0$ for all $i = 1, \dots, s$, then it is easy to see that there are exactly $s!$ solutions, as one $x_j - a_{ij}$ needs to vanish for each i, and the hypothesis on the a_{ij}'s force no j to be used twice.

With these lemmas in place, we now turn to giving lower bounds on $K_{s,t}$ for $t \geq (s-1)! + 1$. The initial result was due to [**28**], who establish the result for $t \geq s! + 1$. The improvement was due to [**3**]. We present both below.

thm:krs **Theorem 2.32.** *Let $s \geq 2$ and $t > s!$. Then*

$$\mathrm{ex}(n, K_{s,(s-1)!+1}) = \Theta\left(n^{2-\frac{1}{s}}\right).$$

Proof. The upper bound follows from Theorem 2.19.

We define the "norm graph" G to have vertex set $\mathbb{F}_{p^s}$, where $a, b \in \mathbb{F}_{p^s}$ are adjacent if and only if $N(a+b) = 1$. By Lemma 2.29, every vertex has degree $\frac{p^s-1}{p-1} \geq p^{s-1} = n^{1-\frac{1}{s}}$ where $n = p^s$. Thus, G has at least $\Omega\left(n^{2-\frac{1}{s}}\right)$ edges.

We claim that G is $K_{s,t}$-free. Let $y_1, \dots, y_s \in \mathbb{F}_{p^s}$ be distinct. Any common neighbor x must satisfy the system of equations

$$N(x + y_i) = 1 \quad \text{for all } 1 \leq i \leq s.$$

Writing down the system explicitly, we see that we are in the situation of Lemma 2.30 with $b_i = 1$, $a_{ij} = -y_i^{p^{j-1}}$. $a_{ij} \neq a_{i'j}$ since $x \to x^{p^{j-1}}$ is an automorphism by Lemma 2.28. Thus, there are at most $s!$ such x, and thus G is $K_{s,t}$-free for $t > s!$.

Write out this step in more details

We may invoke Lemma 2.24 to conclude the result for all n. □

The above construction was improved in [**3**].

thm:ars **Theorem 2.33.** *Let $s \geq 2$ and $t > s!$. Then*

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

Proof. We define the "projective norm graph" G to have vertex set $V = \mathbb{F}_{p^{s-1}} \times \mathbb{F}_p^\times$, and an edge between (X, x) and (Y, y) if and only if $N(X+Y) = xy$. Similar to the previous proof, we see that G has at least $\Omega\left(n^{2-\frac{1}{s}}\right)$ many edges.

Let $(Y_1, y_1), \ldots, (Y_s, y_s) \in V$ be distinct vertices. Then a common neighbor x satisfies the system of equations

$$N(X + Y_i) = xy_i \quad \text{for all } 1 \leq i \leq s.$$

Note that $Y_i = Y_j$ implies $xy_i = xy_j$ or in other words $y_i = y_j$, which is impossible if $i \neq j$ since the vertices are distinct. Dividing the first $s-1$ equations by the last equation, we have

$$N(\frac{Y_i - Y_s}{X + Y_s} + 1) = \frac{y_i}{y_s} \quad \text{for all} 1 \leq i \leq s-1.$$

Since all Y_i's are distinct, we can divide by $N(Y_i - Y_s)$ and obtain

$$N(\frac{1}{X + Y_s} + \frac{1}{Y_i - Y_s}) = \frac{y_i}{y_s N(Y_i - Y_s)} \quad \text{for all} 1 \leq i \leq s-1.$$

We may now do an invertible change of variables with $x' = \frac{1}{X+Y_s}$, $y_{i'} = \frac{1}{Y_i - Y_s}$, and use $b_i = \frac{y_i}{y_s N(Y_i - Y_s)}$ to enter the same situation as in the weaker result, except that we now have $s-1$ equations as opposed to s (this is where the benefits of using the "projective norm graph" come into play). Once again, invoking Lemma lem:comm_alg_sys 2.30, we conclude that there are at most $(s-1)!$ possible common neighbors. Thus, G' is $K_{s,t}$-free for $t > (s-1)!$.

We may invoke Lemma lem:prime_gaps 2.24 to conclude the result for all n. □

Remark 2.34. It may be possible that the above norm graph constructions are actually $K_{s,t}$-free for even smaller values of t (perhaps exponential in s). We do not know if this is the case.

So far, the cases where the order of $\mathrm{ex}(n, K_{s,t})$, $2 \leq s \leq t$, has been determined are precisely the ones that we have seen as far, namely for $s \in \{2, 3\}$ and all $t \geq (s-1)! + 1$. The order of $\mathrm{ex}(n, K_{4,4})$ remains open. There is some work by [**6**] that shows fundamental limitations of the ability to generalize the $K_{2,2}$ and $K_{3,3}$ constructions.

We have seen two types of constructions so far:

(1) Randomized constructions with alternations, which are quite general and easy to apply, but usually do not give tight bounds. And

(2) Algebraic constructions, which give tight bounds but appears to be somewhat magical and only works in certain special situations.

There is an interesting recent idea of Bukh [**9**] called "random algebraic construciton" that combines these two approches. The basic idea is to construct a graph with vertex set $V = \mathbb{F}_q^s \times \mathbb{F}_q^s$ by choosing a *random* polynomial $f \in \mathbb{F}_q[x_1, \ldots, x_s, y_1, \ldots, y_s]$ (within a certain family, say bounded degree) and letting $(x, y) \in V \times V$ be an edge if and only if $f(x, y) = 0$. The method aims to combine the advantages of both the flexibility of randomized constructions as well as the rigidity of algebraic constructions. It has shown some promising results, but its full potential likely has not been fully realized at this point.

§2.6 Forbidding sparse bipartite graphs

Consider bipartite H, with bipartition $V(H) = A \uplus B$. Clearly $H \subset K_{|A|,|B|}$, and hence Lec5: Morris Ang

$$\mathrm{ex}(n, H) \le \mathrm{ex}(n, K_{|A|,|B|}) \le Cn^{2-\frac{1}{|A|}}.$$

This easy upper bound is often not tight. For most H, we do not even know the leading order asymptotics.

What if H is sparse – more specifically, if vertices in A have bounded degree? In this case, we have a generalization of the Kővári-Sós-Turán Theorem thm:kst 2.19.

thm:ex_for_sparse_bipartite_H **Theorem 2.35.** *Let H be a bipartite graph with bipartition $A \uplus B$, and let Δ be some constant such that $\deg(a) \le \Delta$ for all $a \in A$. Then there exists some constant C such that $\mathrm{ex}(n, H) \le Cn^{2-\frac{1}{\Delta}}$.*

The proof of this theorem uses a technique called *dependent random choice*. The idea is that, in a dense graph, we can always find a large subset U of vertices such that all small subsets of U have many common neighbors.

lem:dependent_random_choice **Lemma 2.36 (Dependent random choice).** *Suppose we have $u, n, r, m, t \in \mathbb{N}$ and $\alpha \in \mathbb{R}$ such that*

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

Then every graph with n vertices and average degree at least αn contains a subset $U \subset V(G)$ with $|U| \geq u$, so that for any subset $S \subset U$ with $|S| = r$, S has at least m common neighbors.

A naïve proof approach would be to choose U uniformly at random, and perhaps alter it to have the desired property. This unfortunately does not work. The idea behind dependent random choice is that, instead of directly randomly specifying U, we instead uniformly select a random subset T, and choose U from the common neighborhood of T.

Proof. Let T be a list of t vertices, chosen uniformly at random from $V(G)$ with replacement. Let A be the common neighborhood of T; for some good choice of T, after a slight alteration of A we will obtain our desired set U. By the linearity of expectation, we have

$$\mathbb{E}[|A|] = \sum_{v \in V(G)} \mathbb{P}(v \in A) = \sum_{v \in V(G)} \left(\frac{\deg(v)}{n}\right)^t,$$

where the last equality follows since $v \in A$ if and only if all vertices in T lie in $n(v)$. The convexity of $x \mapsto x^t$ then allows us to apply Jensen's inequality to obtain

$$\mathbb{E}[|A|] \geq n\alpha^t.$$

For the rest of this proof, we call a vertex set S *bad* if $|S| = r$ and S has fewer than m common neighbors. For any fixed $S \subset V(G)$ of size r, we have $S \subset A$ if and only if the common neighborhood of S contains T. If S is bad, then

$$\mathbb{P}(S \subset A) = \mathbb{P}(T \subset \{\text{common neighbors of } S\}) < \left(\frac{m}{n}\right)^t.$$

Let J denote the number of bad subsets of A. By the linearity of expectation, summing over all bad r-element subsets of $V(G)$, we have

$$\mathbb{E}[J] = \sum_{S \text{ bad}} \mathbb{P}(S \subset A) < \binom{n}{r}\left(\frac{m}{n}\right)^t.$$

Combining these bounds yields

$$\mathbb{E}[|A| - J] > n\alpha^t - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq u.$$

Consequently, there exists some choice of T so that the corresponding A and J satisfy $|A| - J \geq u$. We obtain U by deleting a vertex from each bad subset of A; such U contains no bad sets and has at least $|A| - J \geq u$ elements, as desired. □

For an excellent survey discussing applications of dependent random choice to relevant class topics, see [**17**].

cor:dep_rand_choice_sparse_H **Corollary 2.37.** *For bipartite H with bipartition $A \uplus B$ and $\deg(a) \leq \Delta$ for all $a \in A$, there exists $C > 0$ such that every graph G on n vertices with at least $Cn^{2-\frac{1}{\Delta}}$ edges contains a subset $U \subset V(G)$ with $|U| = |B|$ such that any subset of Δ vertices in U has at least $|A| + |B|$ common neighbors.*

Proof. Apply Lemma 2.36 with $r = t = \Delta$. It suffices to pick $C > 0$ large enough so that

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

Indeed, the first term is $(2C)^{\Delta}$, and all other terms are $O(1)$. Hence any sufficiently large C satisfies the desired inequality. □

We now return to the proof of Theorem 2.35.

Proof of Theorem 2.35. Suppose $e(G) > Cn^{2-\frac{1}{\Delta}}$. By Corollary 2.37, we may pick $U \subset V(G)$ with size $|U| = |B|$ so that any subset of Δ vertices in U has at least $|A| + |B|$ common neighbors.

We embed H into G by first embedding B into U via an arbitrary bijection. For each $v \in A$, the image of $N(v) \subset B$ in G has at least $|A| + |B|$ common neighbors, since $|N(v)| \leq \Delta$. So we can embed v into one of the common neighbors of the image of $N(v)$, and there will always be enough room so that all the vertices of H map to distinct vertices of G. □

CHAPTER 3

Szemerédi's regularity lemma

ch:regularity

Szemerédi's regularity lemma, also known as the *graph regularity lemma* or simply the *regularity lemma*, is one of the central theorems of extremal graph theory. Informally, it states that the vertex set of every large graph can be partitioned into a bounded number of parts so that the graph appears "random-like" between most pairs of vertex parts. It is, in some sense, a rough classification for *all* graphs.

§3.1 Statement and proof

To precisely state the regularity lemma, we introduce some notation. For a graph G and two vertex sets $X, Y \subset V(G)$, we define

$$e_G(X,Y) = \left|\{(x,y) \in X \times Y \,:\, (x,y) \in E(G)\}\right|.$$

Observe that in the case where $X \cap Y = \emptyset$, this is precisely the number of edges between X and Y, but if $X \cap Y \neq \emptyset$, some edges are counted twice; in particular, $e_G(X,X)$ is twice the number of edges within X.

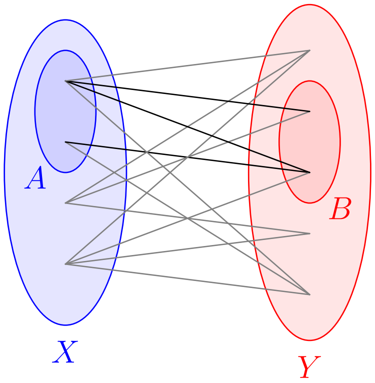
We further define the *edge density*

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

When the ambient graph G is clear from context, we will frequently omit the subscript G and just say $e(X,Y), d(X,Y)$.

Definition 3.1 (ϵ-regular pair). Let G be a graph and $X, Y \subset V(G)$ be vertex sets. We say that (X,Y) is an *ϵ-regular pair* if, for all subsets $A \subset X, B \subset Y$ with $|A| \geq \epsilon|X|$ and $|B| \geq \epsilon|Y|$, we have

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



Roughly speaking, this means that passing to reasonably-sized subsets of X, Y does not significantly change the edge density. Note that for this definition to be interesting, we need the size bounds $|A| \geq \epsilon|X|$ and $|B| \geq \epsilon|Y|$. Otherwise, we could choose A,B to be singletons, and the edge density $d(A, B)$ would be either 0 or 1, and the definition would not be very useful.

Definition 3.2 (ϵ-regular partition). We say that a partition $V(G) = V_1 \uplus \cdots \uplus V_k$ is *ϵ-regular* if

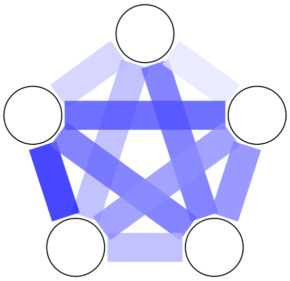
$$\sum \frac{|V_i||V_j|}{n^2} \leq \epsilon,$$

where the sum is taken over all pairs $(i, j) \in [k]^2$ such that (V_i, V_j) is not ϵ-regular.

In other words, a partition is ϵ-regular if at most an ϵ-fraction of pairs of vertices lie between pairs of vertex parts that are not ϵ-regular. In applications, we can bound the contributions from these "bad pairs" as negligible.

The above sum must be allowed to include pairs $i = j$, so as to rule out the trivial partition which has $k = 1$ and $V_1 = V(G)$.

We are finally ready to state the Szemerédi regularity lemma.



thm:szemeredi_reg **Theorem 3.3** (Szemerédi regularity lemma). *For all $\epsilon > 0$, there exists some M such that every graph G has an ϵ-regular partition into at most M parts.*

We state without proof an alternative formulation which is more tedious to prove but easier to use. We say a partition is *equitable* if any two parts of the partition have sizes differing by at most one.

thm:szemeredi_reg_alt **Theorem 3.4** (Szemerédi regularity lemma with equitable partition). *For all $m, \epsilon > 0$, there exists M such that every graph G has an equitable partition $V(G) = V_1 \uplus \cdots \uplus V_k$ such that $m \leq k \leq M$, and the number of non ϵ-regular pairs (V_i, V_j) (with $i < j$) is at most ϵk^2.*

Remark 3.5. In this formulation, we no longer care about the ϵ-regularity of (V_i, V_i). Indeed, if $m \geq 1/\epsilon$, say, then the fraction of pairs of vertices both lying in the same V_i for some i is at most $1/m \leq \epsilon$ (ignoring a tiny error due to rounding).

We now turn to the proof of Theorem thm:szemeredi_reg 3.3. The strategy is to start with the one-part partition $V = V(G)$, and whenever the partition is not ϵ-regular, we further refine the partition into a bounded number of vertex sets. We will define an "energy" associated with each partition, which increases by at least some definite amount with each of the above refinements. As energy is bounded above by 1, this bounds the number of steps the process can take, and hence the number of parts of the partition.[1]

Consider a graph G on n vertices. For $U, W \subset V(G)$, define the *energy* or *mean-squared density*

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

For partitions $\mathcal{P}_U : U = U_1 \uplus \cdots \uplus U_k$ and $\mathcal{P}_W : W = W_1 \uplus \cdots \uplus W_l$, we further define

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

to be the total energy between pairs of parts of $\mathcal{P}_U, \mathcal{P}_W$. Finally, for a partition $\mathcal{P} : V(G) = V_1 \uplus \cdots \uplus V_k$ of the whole vertex set, we say

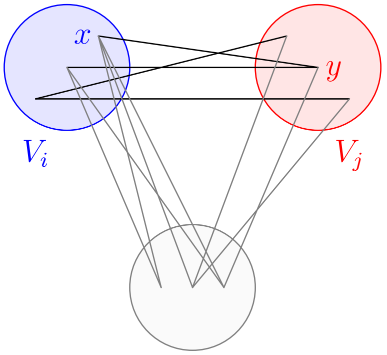
$$q(\mathcal{P}) := q(\mathcal{P}, \mathcal{P}) = \sum_{i,j=1}^{k} q(V_i, V_j).$$

[1] In mathematics, it seems acceptable to refer to any useful L^2 statistic as "energy".

Clearly, since edge densities are bounded above by 1, so is energy:

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

rem:interpretation_energy *Remark* 3.6. Here is one interpretation of energy. Define a random variable Z as follows: Choose x, y uniformly at random from $V(G)$, and suppose $x \in V_i, y \in V_j$. We set $Z = d(V_i, V_j)$.



We have

$$\begin{aligned}
\mathbb{E}[Z] &= \sum_{i,j} \mathbb{P}(x \in V_i, y \in V_j) d(V_i, V_j) \\
&= \sum_{i,j} \frac{|V_i||V_j|}{n^2} d(V_i, V_j) \\
&= \sum_{i,j} \frac{e(V_i, V_j)}{n^2} \\
&= \frac{2e(G)}{n^2},
\end{aligned}$$

$$\mathbb{E}[Z^2] = \sum_{i,j} \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2 = \sum_{i,j} q(V_i, V_j)^2 = q(\mathcal{P}).$$

This explains why we might call $q(\mathcal{P})$ the mean-squared density.

The following lemma shows that energy can never decrease after refinement.

lem:energy_inc **Lemma 3.7.** *If $\mathcal{P}_U$ is a partition of U and $\mathcal{P}_W$ is a partition of W, then*

$$q(\mathcal{P}_U, \mathcal{P}_W) \geq q(U, W).$$

Proof. We first define a random variable Z, which is technically different from that in Remark 3.6, but morally the same. Choose $x \in U$ and $y \in W$ uniformly and independently. If $x \in U_i, y \in W_j$, define $Z = d(U_i, W_j)$. We compute the first two moments of Z:

$$\begin{aligned}\mathbb{E}[Z] &= \sum_{i,j} \frac{|U_i||W_j|}{|U||W|} d(U_i, W_j) \\ &= \sum_{i,j} \frac{|U_i||W_j|}{|U||W|} \frac{e(U_i, W_j)}{|U_i||W_j|} \\ &= \frac{e(U,W)}{|U||W|} \\ &= d(U,W),\end{aligned}$$

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

By convexity, we have $\mathbb{E}[Z^2] \geq \mathbb{E}[Z]^2$, and so

$$\frac{n^2}{|U||W|} q(\mathcal{P}_U, \mathcal{P}_W) \geq d(U,W)^2.$$

Dividing through by $\frac{n^2}{|U||W|}$ yields the desired inequality. □

If a partition is not ϵ-regular, then we can always further partition in a way that significantly increases the energy. It is important to have a definite lower bound on the energy increment so we can show that the refinement process terminates in a bounded number of steps. Lec6: Lisa Yang

lem:energy_inc_irr **Lemma 3.8.** *If (U,W) is not an ϵ-regular pair, then there exist partitions $\mathcal{P}_U \colon U = U_1 \uplus U_2$ and $\mathcal{P}_W \colon W = W_1 \uplus W_2$ such that*

$$q(\mathcal{P}_U, \mathcal{P}_W) > q(U,W) + \epsilon^4 \frac{|U||W|}{n^2}.$$

Proof. Since (U,W) is not an ϵ-regular pair, there exists subsets $U_1 \subseteq U, W_1 \subseteq W$ such that $|U_1| \geq \epsilon|U|, |W_1| \geq \epsilon|W|$ and $|d(U_1, W_1) - d(U,W)| > \epsilon$. Let $U_2 = U \setminus U_1, W_2 = W \setminus W_1$. As in the proof of Lemma 3.7, choose x, y uniformly at random from U, W. Say $x \in U_i, y \in W_j$ and let $Z = d(U_i, W_j)$.

As in the previous proof, the first two moments of Z are

$$\mathbb{E}[Z] = d(U,W) \quad \text{and} \quad \mathbb{E}[Z^2] = \frac{n^2}{|U||W|} q(\mathcal{P}_U, \mathcal{P}_W).$$

So the variance of Z equals to

$$\mathrm{Var}[Z] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = \frac{n^2}{|U||W|}(q(\mathcal{P}_U, \mathcal{P}_W) - q(U,W)). \tag{3}$$ eq:regpf-varZ

On the other hand $\mathrm{Var}[Z] = \mathbb{E}[(Z - \mathbb{E}[Z])^2]$. We have $x \in U_1, y \in W_1$ with probability $\frac{|U_1||W_1|}{|U||W|}$, in which case

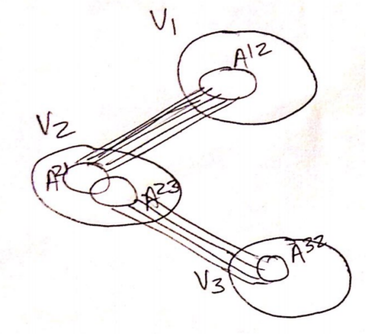
$$|Z - \mathbb{E}[Z]| = |d(U_1, W_1) - d(U,W)| > \epsilon.$$

So $\mathrm{Var}[Z] = \mathbb{E}[(Z - \mathbb{E}[Z])^2] > \frac{|U_1||W_1|}{|U||W|}\epsilon^2 \geq \epsilon^4$. Combining with (3) yields the desired inequality. □

Lemma 3.8 applies to a single pair which is not ϵ-regular. Recall that in an ϵ-regular partition, we are allowed to have a partition where some number of pairs are not ϵ-regular. If too many pairs of vertices are in non-ϵ-regular pairs, then we use the above lemma to refine the partition. This boosts the energy between pairs that are not ϵ-irregular while not decreasing the energy between pairs that are ϵ-regular.

lem:energy_inc_partition **Lemma 3.9.** *If $\mathcal{P} = V_1 \uplus \cdots \uplus V_k = V(G)$ is not an ϵ-regular partition, then there exists a refinement $\mathcal{P}'$ where every V_i is refined into at most 2^{k+1} parts such that $q(\mathcal{P}') \geq q(\mathcal{P}) + \epsilon^5$.*

Proof. For each (i,j) such that (V_i, V_j) is not an ϵ-regular pair, find sets $A \subset V_i$ and $B \subset V_j$ that witness the ϵ-irregularity, i.e. $|A| \geq \epsilon|V_i|, |B| \geq \epsilon|V_j|$ and $|d(A,B) - d(V_i,V_j)| > \epsilon$.



Let $\mathcal{P}'$ be the common refinement of $\mathcal{P}$ and all sets A and B that arise as above for pairs (V_i, V_j) that are not ϵ-regular (including $i = j$). At most $k+1$ sets are introduced in each V_i (at most one set for each (V_i, V_j) with $j \neq i$, and at most two sets when $j = i$), so V_i is divided into at most 2^{k+1} sets in the refinement. Let $\mathcal{P}_{V_i}$ denote the partition of V_i under this refinement.

Apply Lemmas 3.7 and 3.8 to each pair $(\mathcal{P}_{V_i}, \mathcal{P}_{V_j})$. We obtain the following energy increment.

$$\begin{aligned} q(\mathcal{P}') &= \sum_{i,j=1}^{k} q(\mathcal{P}_{V_i}, \mathcal{P}_{V_j}) \\ &\geq \sum_{i,j=1}^{k} q(V_i, V_j) + \epsilon^4 \sum_{\substack{(V_i,V_j) \\ \text{not } \epsilon\text{-regular}}} \frac{|V_i||V_j|}{|V(G)|^2} \\ &> q(\mathcal{P}) + \epsilon^5. \end{aligned}$$

□

Remark 3.10. Note the following subtlety in the proof. We should be careful in analyzing the energy increment while refining the vertex sets V_i. Suppose that (V_1, V_2) and (V_2, V_3) are both not ϵ-regular. After partitioning V_1 and V_2 using the sets that witness their irregularity, the subsets of V_2 and V_3 that witness the irregularity of (V_2, V_3) may no longer witness irregularity in the newly refined partition, so we cannot use Lemma 3.8 to get an energy increment. The proof works around this issue by analyzing the energy increment for each pair (V_i, V_j).

Proof of Theorem 3.3. Starting with the trivial partition (all vertices in one part), repeatedly apply Lemma 3.9 until we get an ϵ-regular partition. Since $0 \leq q(\mathcal{P}) \leq 1$ and the energy increases by at least ϵ^5 after every refinement, the process stops after at most $\leq \epsilon^{-5}$ step. For $\mathcal{P}$ with k parts, we obtain a refinement with $\leq k2^{k+1} \leq 2^{2^k}$ parts. Thus the number of parts in the final ϵ-regular partition is at most $\leq 2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$, an exponential tower of 2's of height at most $2\epsilon^{-5}$. □

Remark 3.11. Note that Szemerédi's regularity lemma is only useful for very large graphs, since for graphs on a small number of vertices, we can trivially obtain an ϵ-regular partition by putting every vertex in its own part.

Remark 3.12. While the bound on the number of parts in Szemerédi's regularity lemma seems ridiculously high, it is actually necessary for the theorem to hold. Gowers constructed graphs whose ϵ-regularity requires at least $2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$ parts, where the tower has height ϵ^{-c}. Intuitively, Gowers' proof reverse-engineers the proof of Szemerédi's regularity lemma by constructing a graph whose regularity partition is forced to be like in the proof of the regularity lemma at each step of the process. This is a good reason to believe that the above proof of the regularity lemma is the "right" proof, and in some non-rigorous sense, the "only" possible way to prove the result.

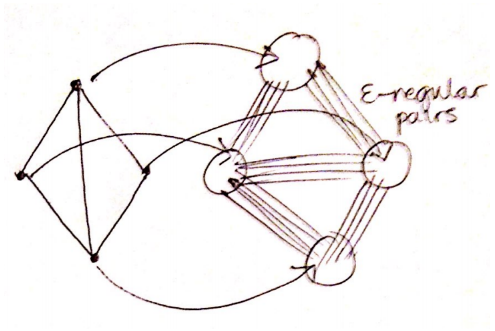
Remark 3.13. Let us say a few words here about how to prove the version of Szemerédi's regularity lemma that gives equitable partitions (Theorem thm:szemeredi_reg_alt 3.4).

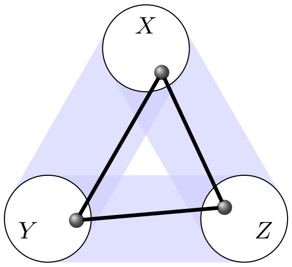
An incorrect way to proceed is to just take an ϵ-regular partition as in Theorem thm:szemeredi_reg 3.3 and then try to make it equitable at the end by further splitting up the vertex sets and rebalancing the parts (moving a small number of vertices around to make the partition equitable). This does not work, since refining a partition does not preserve ϵ-regularity. Indeed, if a pair (X, Y) is ϵ-regular and $A \subset X$ and $B \subset Y$, then we cannot necessarily conclude that (A, B) is ϵ-regular, at least not for the same value of ϵ.

Instead, the way to prove the equitable version of the regularity lemma is to make the partition equitable after every step of the refinement process. As in the above proof, we gain an energy increment at every step. When we make the partition equitable via re-balancing, we may cause a bit of energy loss, but it is dominated by the energy increment from irregularity. So every iteration still gives an energy increment, and as before, the process must end after a bounded number of iterations, at which point the final partition must be both ϵ-regular and equitable.

§3.2 Counting and removal lemmas for triangles

For a partition of G with some pairs of vertex sets that are ϵ-regular and not too sparse, we can embed a subgraph H and show that there are approximately the expected number of such embeddings.

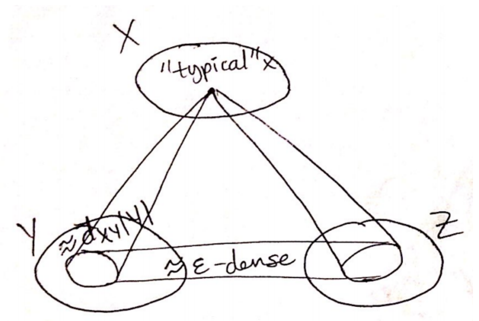




thm:triangle_counting **Theorem 3.14 (Triangle counting lemma).** *Let $X, Y, Z \subset V(G)$. Suppose that the pairs $(X,Y), (X,Z), (Y,Z)$ are all ϵ-regular pairs. Let $d_{XY} = d(X,Y)$, $d_{XZ} = d(X,Z)$, and $d_{YZ} = d(Y,Z)$. If $d_{XY}, d_{XZ}, d_{YZ} \geq 2\epsilon$, then*

$$|\{\text{triangles } (x,y,z) \in X \times Y \times Z\}| \geq (1-2\epsilon)(d_{XY}-\epsilon)(d_{XZ}-\epsilon)(d_{YZ}-\epsilon)|X||Y||Z|.$$

Here is the intuition. A typical $x \in X$ should have about $d_{XY}|Y|$ neighbors in Y. Otherwise, we can use a set of atypical vertices to violate ϵ-regularity. Similarly for the neighborhood of x in Z. Then by the ϵ-regularity of (Y,Z), there will be roughly the expected number of edges between these neighborhoods.



Proof. Let $d_Y(x), d_Z(x)$ be the number of neighbors of x in Y and Z respectively. Note that

$$|\{x \in X : d_Y(x) < (d_{XY} - \epsilon)|Y|\} \leq \epsilon|X|$$

or else this set together with Y violate ϵ-regularity of (X, Y).

Similarly $|\{x \in X : d_Y(x) < (d_{XZ} - \epsilon)|Z|\} \leq \epsilon|X|$.

Let $N_Y(x), N_Z(x)$ be the neighborhoods of x in Y and in Z respectively. This shows there are at least $(1 - 2\epsilon)|X|$ many $x \in X$ where

$$\begin{aligned} d_Y(x) &\geq (d_{XY} - \epsilon)|Y| \qquad\qquad\qquad\qquad \geq \epsilon|Y|, \\ \text{and } d_Z(x) &\geq (d_{XZ} - \epsilon)|Z| \geq \epsilon|Z|. \end{aligned}$$

By ϵ-regularity of (Y, Z) we have

$$e(N_Y(x), N_Z(x)) \geq (d_{YZ} - \epsilon) d_Y(x) d_Z(y) \geq (d_{XY} - \epsilon)(d_{XZ} - \epsilon)(d_{YZ} - \epsilon)|Y||Z|.$$

By summing over $\geq (1 - 2\epsilon)|X|$ many such x, we get the desired lower bound. □

Remark 3.15. We can also get an upper bound on the number of triangles by eliminating vertices that have large neighborhoods in Y and Z. We will also use this method for other graphs beyond triangles though there is more bookkeeping involved.

Next, we will see our first application of Szemerédi's regularity lemma. As with many other standard applications of the regularity lemma, the proof follows the following three-step recipe:

(1) **Partition** the vertex set into an ϵ-regular partition.
(2) **Clean** up the graph by removing edges between pairs of parts that are irregular, low-density, or small. We do not remove many edges in this step.
(3) **Count** or embed the desired graph into a configuration of regular pairs.

thm:triangle_removal **Theorem 3.16** (Triangle removal lemma)**.** *For all $\epsilon > 0$, there exists $\delta > 0$ such that any graph on n vertices with at most δn^3 triangles can be made triangle-free by removing at most ϵn^2 edges.*

thm:szemeredi_reg

Proof. Let M be as in Theorem 3.3. Apply Szemerédi's regularity lemma to find an $\epsilon/4$-regular partition $V_1 \uplus \cdots \uplus V_m$ with $m \leq M$.

Remove edges (x, y) from G that satisfy any of the following conditions:

(1) $(x, y) \in V_i \times V_j$ where (V_i, V_j) is not $\epsilon/4$-regular. At most $\sum_{i \neq j} |V_i||V_j| \leq \frac{\epsilon}{4} n^2$ edges are removed this way.
(2) $(x, y) \in V_i \times V_j$ where $d(V_i, V_j) < \epsilon/2$. At most $\frac{\epsilon}{2} n^2$ edges are removed this way.
(3) $x_i \in V_i$ where $|V_i| \leq \frac{\epsilon n}{4M}$. At most $Mn\frac{\epsilon n}{4M} = \epsilon n^2/4$ edges are removed this way.

Hence we remove at most $\leq \epsilon n^2$ edges from G.

Suppose some triangle $(x, y, z) \in V_i \times V_j \times V_k$ remains (here i, j, k do not have to be distinct). This implies that $(V_i, V_j), (V_i, V_k), (V_j, V_k)$ are $\epsilon/4$-regular with edge density $\geq \epsilon/2$ and $|V_i|, |V_j|, |V_k| > \frac{\epsilon n}{4M}$. By the triangle counting lemma, Theorem 3.16, the number of triangles is at least

thm:triangle_removal

$$(1 - \epsilon/2)(\epsilon/4)^3 |V_i||V_j||V_k| \geq (1 - \epsilon/2)(\epsilon/4)^3 (\epsilon/4M)^3 n^3.$$

So taking $\delta = (1 - \epsilon/2)(\epsilon/4)^3(\epsilon/4M)^3$, any graph with fewer than δn^3 triangles must be triangle free after removing ϵn^2 edges. □

Lec7: Albert Soh

Remark 3.17. Here is a "lazier" (but correct) way to state the triangle removal lemma.

> If a graph on n vertices has $o(n^3)$ triangles, then it can be made triangle-free by removing $o(n^2)$ edges.

This is a lazy version of the theorem because we use o in two different ways. One should read the statement as

> For every function $f(n) = o(n^3)$ there exists a function $g(n) = o(n^2)$ such that if a graph on n vertices has fewer than $f(n)$ triangles, then it can be made triangle-free by removing fewer than $g(n)$ edges.

It is a good exercise to think about why the two version of the triangle removal lemma are equivalent. The "lazy" statement of the triangle removal lemma is pretty standard, though if one ends up chaining too many $o(\cdot)$'s together, it may become too confusing to decode the logical dependence of the parameters.

Remark 3.18. Is there a different approach to proving the triangle removal lemma to tighten the bounds? Fox improved the height of the towers of 2's from $\varepsilon^{-O(1)}$ to $O(\log\left(\frac{1}{\varepsilon}\right))$. The best known lower bound is $\frac{1}{\delta} \geq e^{(\log(\frac{1}{\varepsilon}))^2} = \left(\frac{1}{\epsilon}\right)^{C\log(\frac{1}{\varepsilon})}$, for some constant C. This lower bound growths faster than any polynomial in $1/\varepsilon$, but not by much more. The gap between the two bounds is very large. It is a major open problem to close this gap.

§3.3 Property testing

An example of property testing is the following question. Given a very large graph, is it triangle-free? The goal of property testing is to distinguish graphs that are triangle-free from those that are ε-far from being triangle-free.

def:epsfar *Definition* 3.19 (ε-far). A graph, G is ε-far from some property $\mathcal{P}$ if it is necessary to add or remove at least εn^2 edges to satisfy $\mathcal{P}$.

The goal of property testing is to find an algorithm to decide whether a very large graph, G, is triangle-free or if it is ε-far from being triangle-free. In the case the graph is neither, it doesn't matter what the algorithm decides the graph is.

A very naive algorithm would be to randomly sample k triples of vertices and check for triangles. Then we have two possible cases.

(1) If we find a triangle, then we output that the graph is ε-far from being triangle-free.
(2) If We don't find any triangles, then we output that the triangle-free.

Now we want to see how good our algorithm is. Clearly, if G is triangle-free, the algorithm will always output correctly. However, if G is ε-far from being triangle-free, by the triangle removing lemma, Theorem thm:triangle_removal 3.16, there is some $\delta = \delta(\varepsilon)$ such that G has at least δn^3 triangles. Let s be some constant, and set $k = s/\delta$ as the number of triples that we sample in the algorithm. Then

$$\mathbb{P}(\text{algorithm fails}) \leq \left(1 - \frac{\delta n^3}{\binom{n}{3}}\right)^{\frac{s}{\delta}} \leq (1-6\delta)^{\frac{s}{\delta}} \leq e^{-6s}.$$

So we can make the error probability as small as we wish by choosing s appropriately large.

Definition 3.20 (Hereditary). A graph property is *hereditary* if it is closed under removal of vertices, but not necessarily edges.

Some examples of hereditary properties are: triangle-free, induced-C_4-free, 3-colorable, planar, chordal (no induced C_l, $l \geq 4$), perfect.

Recall that an induced subgraph of G is one formed by taking a subset S of vertices of G and all edges of G in S. For example, C_4 is a subgraph of K_4, but not as an induced one.

A graph property is hereditary if and only if it is equivalent to being induced-$\mathcal{F}$-free, where $\mathcal{F}$ is a possibly infinite set of finite graphs. It is known that there is a property testinng algorith for any hereditary property. In particular, the following result is a strong generalization of the triangle removal lemma.

thm:Alon-Shapira **Theorem 3.21** (Alon-Shapira). *For all family of graphs $\mathcal{F}$, and $\forall \varepsilon > 0$, $\exists \delta > 0, f > 0$ such that if a graph G is ε-far from induced-$\mathcal{F}$-free, then G contains at least $\delta n^{|V(F)|}$ copies of some $F \in \mathcal{F}$ with $|V(F)| \leq f$.*

We won't prove the theorem in this class, as it requires the Strong Regularity Lemma, which is proven by iterating Szemerédi regularity lemma over and over again. The Strong Regularity Lemma has a stronger conclusion, but the bounds are worse. Szemerédi regularity lemma's bounds are tower, but the Strong Regularity Lemma's bounds are wowzer, which comes from iterated application of the tower function.

§3.4 Proof of Roth's theorem

cor:TRL_cor **Corollary 3.22** (Corollary of triangle removal lemma). *If G is an n-vertex graph where every edge is contained in a unique triangle, the maximum number of edges G can have is $o(n^2)$ edges.*

Remark 3.23. There exist graphs as above with $n^{2-o(1)}$ edges, arising from the Behrend construction of 3-AP-free sets shown below.

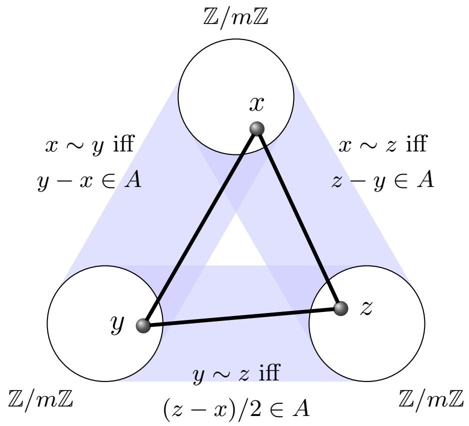
Proof. Since each edge is in a unique triangle, the number of triangles in G must be $\frac{|E(G)|}{3} = O(n^2) = o(n^3)$. By the triangle removal lemma, Theorem thm:triangle_removal 3.16, G can be made triangle free by removing $o(n^2)$ edges, but every edge is in a unique triangle, so we need to remove at least $\frac{|E(G)|}{3}$ edges to make G triangle free. Thus $|E(G)| = o(n^2)$, as desired. □

We now move on to prove Roth's Theorem. Recall that we write 3-AP to mean a 3-term arithmetic progression.

thm:roth **Theorem 3.24** (Roth's theorem). *If $A \subset [n] = \{1, 2, ..., n\}$ contains no 3-APs, then $|A| = o(n)$.*

Proof. Let $m = 4n + 1$. The proof idea is to use a cyclic group to find arithmetic progressions. The only concern with cyclic groups is finding an arithmetic progression that wraps around, so we embed $[n]$ into the cyclic group of m. Also note that we want m to be odd. Now we set up a tripartite graph, $X = Y = Z = \mathbb{Z}/m\mathbb{Z}$. Here we have $A \subset \mathbb{Z}/m\mathbb{Z}$. Then we define the edges in the tripartite graph, G, as follows:

- (x, y) is an edge if and only if $y - x \in A$
- (y, z) is an edge if and only if $z - y \in A$
- (x, z) is an edge if and only if $\frac{z-x}{2} \in A$



Then we have that $|E(G)| = 3|A|m \geq 12|A|n$. If (x, y, z) is a triangle, then $a = y - x$, $b = \frac{z-x}{2}$, $c = z - y \in A$ and a, b, c is a 3-AP. Since A is 3-AP free, the only triangles in the graph correspond to $a = b = c$, i.e. $(x, x + a, x + 2a) \in X \times Y \times Z$. Then we check to see that every edge is in a unique triangle. From Corollary 3.22, the number of edges is $o(n^2)$, so we get $12|A|n = o(n^2)$, resulting in $|A| = o(n)$ as desired. □

If we let $r_3(n) = \max\{|A| : A \subset [n], \text{3-AP-free}\}$, then the above argument gives $r_3(n) \leq \frac{Cn}{\log^* n}$, where $\log^*$ is inverse of tower. The current best bounds are $r_3(n) \leq \frac{Cn(\log \log n)^4}{\log(n)}$, a result by Sanders and Bloom.

As for lower bounds, one way to construct a 3-AP-free set is to take the set of number with only 0's and 1's in its base-3 expansion. This way we include 2^k out of the first 3^k integers, which gives a bound of $r_3(n) \geq n^{\log_3 2}$. For some time, people thought that this was close to the optimal. However, later constructions showed that one can in fact do much better.

thm:behrend **Theorem 3.25** (Behrend's construction). *$\exists c$ such that $\exists A \subset [n]$ that is 3-AP-free with $|A| \geq \frac{n}{e^{c\sqrt{\log(n)}}}$.*

Avoid using $\exists$ and $\forall$ in formal writing. -YZ

Proof. The idea is to look at higher dimensions and take a section of sphere and find a digital projection onto the integers, since there are no 3-APs on a sphere. Let $m, d > 0$ and let $X = [m]^d$ and $X_L = \{(x_1, ..., x_d) \in X : x_1^2 + x_2^2 + ... + x_d^2 = L\}$. Then $X = X_0 \uplus X_1 \uplus ... \uplus X_{dm^2}$. By the pigeonhole principle, one of these is large, i.e. we can find some $L > 0$ such that $|X_L| \geq \frac{m^d}{dm^2}$. Consider $f : X_L \to [n]$, where $n \geq (2m)^d$, defined by setting

$$f(x_1, x_2, ..., x_d) = \sum_{i=1}^{d} x_i (2m)^{i-1}.$$

Note that f is injective, and if $f(\vec{x}) + f(\vec{z}) = 2f(\vec{y})$, then $\vec{x} + \vec{z} = 2 * \vec{y}$, which is impossible since X_L does not contain 3 points on a line. Therefore, the image of f contains no 3-APs.

Thus $|f(X_L)| = |X_L| \geq \frac{m^d}{dm^2} \geq \frac{n}{d2^d m^2}$, which can be made at least $\frac{n}{e^{C\sqrt{\log(n)}}}$ by setting $m = \left\lceil \frac{1}{2} e^{\sqrt{\log(n)}} \right\rceil$ and $d = \left\lceil \sqrt{\log(n)} \right\rceil$. □

We have proved Roth's theorem. We now look at multidimensional Szemeredi's theorem. Instead of avoiding one-dimensional objects like a 3-AP, we can avoid higher dimensional objects. We'll consider here *corners*.

Lec8: Ryan Alweiss

thm:corners **Theorem 3.26** (Corners). *A* corner *is a set of three points (x, y), $(x + d, y)$, $(x, y + d)$ where $d > 0$ is arbitrary.*

If $|A| \subset [n]^2$ contains no corners, then $|A| = o(n^2)$.

The corners theorem is a special case of the multidimensional Szemeredi theorem, Theorem 1.9, where corners are replaced by dilations of any fixed finite set of points.

Proof. We begin with the replace that allows us to replace the $d > 0$ requirement by simply $d \neq 0$. Note $A + A \subset [2n]^2$. By the pigeonhole principle, there is some $z \in [2n]^2$ represented in at least $\frac{|A|^2}{(2n)^2}$ ways by as a sum $a + b$ with $a, b \in A$. Pick this z.

Now, consider $A' = A \cap (z - A)$. We have $|A'| \geq \frac{|A|^2}{(2n)^2}$, because $|A'|$ is the number of ways to write z as the sum of two elements of A. Now, it suffices to show that $|A'| = o(n^2)$. We claim that A' does not contain *any* corners. Suppose A' contains (x, y), $(x + d, y)$, $(x, y + d)$ for $d < 0$, so then $z - A$ does as well and so A contains one with positive d. So we can replace

A by A'. This enables us to get rid of the $d > 0$ condition, and now only to work with $d \neq 0$, through replacing A by A'. So A contains no triple of points (x, y), $(x + d, y)$, $(x, y + d)$ for some any $d \neq 0$.

Consider the tripartite graph with vertex parts X, Y, and Z, where $X = Y = [n]$ and $Z = [2n]$. We will have think of the elements of X, Y, and Z as lines in $[n]^2$. We will make $x' \in X$ correspond to the vertical line $x = x'$, $y' \in Y$ correspond to the horizontal line $y = y'$, and $z' \in Z$ to the slope -1 line $x + y = z'$. Put an edge between two vertices if their corresponding lines intersect at a point in A. So (x', y') is an edge if $(x', y') \in A$, (x', z') is an edge if $(x', z' - x') \in A$, and (y', z') is an edge if $(z' - y', y') \in A$. Note the similarity to the Roth construction.

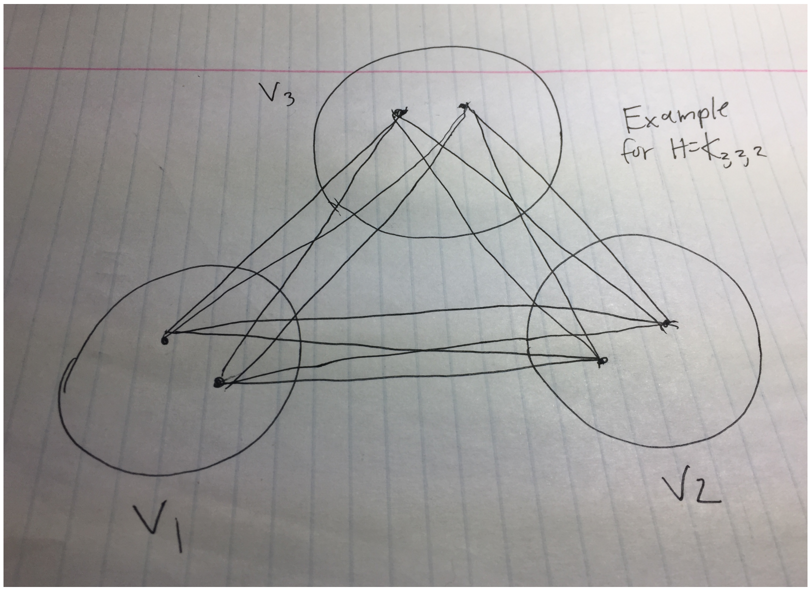
A triangle in the graph corresponds to three lines such that each pair of lines intersects in a point of A. Since A has no corners, the three lines correspond to a triangle in the graph if and only if the three lines all pass through the same point of A (corresponding to a "trivial corner" with $d = 0$). It follows that every edge in the graph lies in a unique triangle, since given two lines intersecting at a point of A, we can take the third line passing through the same point. The number of edges is $3|A|$ and the number of vertices is $4n$, so by Corollary cor:TRL_cor 3.22, we have $|A| = o(n^2)$. □

Note that Roth's theorem follows from the corners theorem. Given $A \subset [n]$, build $B \subset [2n]^2$ by defining B to be the set of pairs $(x, y) \in [2n]^2$ with $x - y \in A$. If B has a corner, then A has a 3-AP. Indeed, if $(x, y), (x + d, y), (x, y + d) \in B$ with $d > 0$ then $x - y - d, x - y, x - y + d \in A$. So B is corner-free, so $n|A| \leq |B| = o(n^2)$, and hence $|A| = o(n)$.

§3.5 Graph embedding and counting lemmas

The triangle counting lemma can be generalized to larger graphs. Here we present two generalizations. In one, vertices are embedded sequentially, and in the other, we take a more global approach. The basic idea is going to use regularity to argue that the number of copies of H in a graph G with some ϵ-regular partition is what we expect, but we must deal carefully with a small number of "bad" vertices which can obstruct the argument.

thm:embed **Theorem 3.27** (Embedding lemma). *Let H be an graph with chromatic number at most r and vertices of degree at most Δ. Let G be a graph and $V_1, \ldots, V_r \subset V(G)$ with $|V_i| \geq |V(H)|/\epsilon$ for all $1 \leq i \leq r$. Furthermore,* *suppose for all $i < j$, (V_i, V_j) is an ϵ-regular in G with $d(V_i, V_j) \ge \lambda$ for some $\lambda > 0$ satisfying $(\Delta + 1)\epsilon \le \lambda^\Delta$. Then G contains a copy of H.*



Note that this theorem is basically a generalization of the triangle removal lemma, but instead of triangles we have H. We could take H to be a complete graph, but the bounds would be worse.

Proof. Let $c\colon V(H) \to [r]$ be a proper r-coloring of H. We construct an injective map $\phi\colon V(H) \to V(G)$, so that $\phi(u) \in V_{c(u)}$ for all $u \in V(H)$. Let $V(H) = [k]$ for notational convenience. We will embed the vertices of H sequentially. Now, for every $0 \le t \le k$ we will construct a partial embedding ϕ_t sending $[t]$ to $V(G)$. For all $t < j \le k$, let X_j^t be the set of compatible images for extending ϕ to some element j. In particular, $X_j^t = \{x \in V_{c(j)} | i \le t, (i,j) \in E(H) \implies (\phi_t(i), x) \in E(G)\}$.

We define $N_H^{\le a}(b) = \{i|(i,b) \in E(H), 1 \le i \le a\}$ and $N_H^{>a}(b) = \{i|(i,b) \in E(H), a < i \le k, \}$.

We will inductively construct X_j^t so that $|X_j^t| \ge (\lambda - \epsilon)^{|N_H^{\le t}(j)|}|V_{c(j)}|$. For $t = 0$, take X_j^0 to be all of $V_{c(j)}$. Suppose we have built ϕ_t for all elements of $[t]$. Now, for all $i \in N_H^{>t}(t+1)$, the number of vertices in X_{t+1}^t with less than $(\lambda - \epsilon)|X_i^t|$ neighbors in X_i^t is smaller than $\epsilon|V_{c(t+1)}|$ or else we witness non-ϵ-regularity, because $|X_i^t| \ge (\lambda - \epsilon)^{N_H^{\le t}(i)}|V_{c(i)}| \ge (\lambda - \epsilon)^\Delta|V_{c(i)}| \ge (\lambda^\Delta - \epsilon\Delta)|V_{c(i)}| \ge \epsilon|V_{c(i)}|$. So those vertices of X_{t+1}^t with less than $(\delta - \epsilon)|X_i^t|$

neighbors in X_i^t on one hand and X_i^t on the other hand would witness the irregularity for $(V_{c(t+1)}, V_{c(i)})$.

Let $m = |N_H^{\le t}(t+1)|$, the neighbors of $t+1$ which are at most t. The inductive hypothesis tells us that $|X_{t+1}^t| \ge (\lambda - \epsilon)^m |V_{c(t+1)|}|$. Eliminate, from X_{t+1}^t, the vertices with fewer than $(\lambda-\epsilon)|X_i^t|$ neighbors in X_i^t for all $i \in N_H^{>t}(t+1)$. Note $|N_H^{>t}(t+1)| \le \Delta - m$. So the remaining number of vertices in X_{t+1}^t is at least

$$((\lambda-\epsilon)^m - (\Delta - m)\epsilon)|V_{c(t+1)}| \ge (\lambda^m - m\epsilon - (\Delta-m)\epsilon)|V_{c(t+1)}|$$
$$\ge (\lambda^\Delta - \Delta\epsilon)|V_{c(t+1)}| \ge \epsilon|V_{c(t+1)}| \ge |V(H)|$$

So we can pick a vertex of $V_{c(t+1)}$ to define $\phi_{t+1}(t+1)$ (and define $\phi_{t+1}(i) = \phi_t(i)$ for $i \le t$), and ϕ is injective.

It suffices to check that this new ϕ_{t+1} has enough compatible images, i.e. that $|X_j^{t+1}| \ge (\lambda-\epsilon)^{|N_H^{\le t+1}(j)|}|V_{c(j)}|$ for any $j > t+1$. If $(t+1, j)$ is not an edge of H, then

$$|X_j^{t+1}| \ge |X_j^t| \ge (\lambda-\epsilon)^{|N_H^{\le t}(j)|}|V_{c(j)}| = (\lambda-\epsilon)^{|N_H^{\le t+1}(j)|}|V_{c(j)}|$$

and we are done. If $(t+1, j) \in E(H)$, then

$$|X_j^{t+1}| \ge (\lambda-\epsilon)|X_j^t| \ge (\lambda-\epsilon)(\lambda-\epsilon)^{|N_H^{\le t}(j)|}|V_{c(j)}| = (\lambda-\epsilon)^{|N_H^{\le t+1}(j)|}|V_{c(j)}|$$

and we are again done.

□

This version is good for graphs with low maximum degree, due to the $(\Delta+1)\epsilon \le \lambda^\Delta$ condition. When H is fairly dense and $\Delta = \Omega(|H|)$, the bound will be closer to the bound of a complete graph.

Recall that we defined a pair (X, Y) to be ϵ-*regular* if $|d(A,B) - d(X,Y)| \le \epsilon$ for all $A \subset X$, $B \subset Y$ with $|A| \ge \epsilon|X|$ and $|B| \ge \epsilon|Y|$. It will be convenient to introduce another notion which is almost equivalent to ϵ-regularity.

def:homog *Definition* 3.28. Let $X, Y \subset V(G)$. We say that the pair (X, Y) is ϵ-*homogeneous* if

$$|e(A,B) - d(X,Y)|A||B|| \le \epsilon|X||Y| \quad \text{for all } A \subset X \text{ and } B \subset Y.$$

It turns out that this definition is equivalent up to a polynomial transformation in ϵ.

lem:reg-homog **Lemma 3.29.** *Let $\epsilon > 0$. In any graph, if a pair of vertex sets (X, Y) is ϵ-regular, then it is ϵ-homogeneous. Conversely, if (X, Y) is ϵ^3-homogeneous, then it is ϵ-regular.*

Proof. Suppose first that (X, Y) is ϵ-regular. For every $A \subset X$ and $B \subset Y$ we have $|e(A, B) - d(X, Y)||A||B|| = |d(A, B) - d(X, Y)||A||B|$. If $|A| \geq \epsilon|X|$ and $|B| \geq \epsilon|Y|$, then by ϵ-regularity we have $|d(A, B) - d(X, Y)||A||B| \leq \epsilon|A||B| \leq \epsilon|X||Y|$. Else, $|d(A, B) - d(X, Y)||A||B| \leq |A||B| \leq \epsilon|X||Y|$. Thus the pair (X, Y) is ϵ-homogeneous.

Now, assume that a pair (X, Y) is ϵ^3-homogeneous. If $A \subset X$ and $B \subset Y$ satisfies $|A| \geq \epsilon|X|$ and $|B| \geq \epsilon|Y|$, then

$$|d(A, B) - d(X, Y)| = \frac{|e(A, B) - d(X, Y)|A||B||}{|A|B|} \leq \frac{\epsilon^3|X||Y|}{\epsilon|X|\epsilon|Y|} \leq \epsilon.$$

Thus (X, Y) is ϵ-regular. □

The notion of homogeneity enables us to more easily formulate a general counting lemma.

thm:counting **Theorem 3.30 (Counting lemma).** *Let H be a graph with $V(H) = [k]$. Let G be a graph with $V_i \subset V(G)$ for $1 \leq i \leq k$ so that (V_i, V_j) is ϵ-homogeneous for all $ij \in E(H)$. Let $N(H)$ denote the number of homomorphisms $H \to G$ where each $i \in V(H)$ is mapped into V_i, i.e.,*

$$N(H) := |\{(v_1, \cdots, v_k) \in V_1 \times \cdots \times V_k : (v_i, v_j) \in E(G) \text{ for all } (i, j) \in E(H)\}|.$$

Then

$$\left| N(H) - \prod_{ij \in E(H)} d(V_i, V_j) \cdot \prod_{i=1}^{k} |V_i| \right| \leq |E(H)|\epsilon \prod_{i=1}^{k} |V_i|$$

Note that the V_i's do not have to be disjoint. The quantity $N(H)$ includes non-injective maps $V(H) \to V(G)$, but since the a negligible fraction of such maps are non-injective (for fixed H and large $|V(G)|$), we can use the theorem to find a genuine copy of H in G.

Proof. The proof proceeds by induction on $|E(H)|$. The result is trivial for empty graphs since $N(H) = \prod_{i=1}^{k} |V_i|$ if H is empty. Without loss of generality, suppose $(1, 2) \in E(H)$. Let H' be H without the edge $(1, 2)$. Suppose we fix $v = (v_3, \cdots, v_k) \in V_3 \times \cdots \times V_k$. Let $X_1(v)$ be the set of $v_1 \in V_1$ such that $(v_1, v_i) \in E(G)$ whenever $(1, i) \in E(H)$. Define $X_2(v)$

similarly. We say that v is *valid* if for any $3 \le i < j \le k$ whenever $(i,j) \in H$, $(v_i, v_j) \in G$.

Then,

$$N(H) = \sum_{\text{valid } v \in V_3 \times \cdots \times V_k} e_G(X_1(v), X_2(v))$$

and

$$N(H') = \sum_v |X_1(v)||X_2(v)|.$$

So $|N(H) - N(H')d(V_1,V_2)| \le \epsilon \prod_{i=1}^k |V_i|$ by ϵ-homogeneity, because the difference is at most $\epsilon|V_1||V_2|$ for each of the at most $\prod_{i=3}^k |V_i|$ choices of v. We have

$$\begin{aligned}
&\left| N(H) - \prod_{ij \in E(H)} d(V_i, V_j) \cdot \prod_{i=1}^k |V_i| \right| \\
&\le \left| N(H) - N(H')d(V_1,V_2) \right| + d(V_1,V_2) \left| N(H') - \prod_{ij \in E(H')} d(V_i,V_j) \cdot \prod_{i=1}^k |V_i| \right| \\
&\le \epsilon \prod_{i=1}^k |V_i| + |E(H')| \epsilon \prod_{i=1}^k |V_i| \\
&= |E(H)| \epsilon \prod_{i=1}^k |V_i|.
\end{aligned}$$

Thus, the induction is complete. □

Lec9: Matthew Brennan

The following removal lemma for general graphs follows from the same proof as the triangle removal lemma.

thm:graph-removal

Theorem 3.31 (Graph removal lemma). *For all H and $\epsilon > 0$, there is a $\delta > 0$ such that if G has n vertices and at most $\delta n^{|V(H)|}$ copies of H, then G can be made H-free by removing at most ϵn^2 edges.*

§3.6 Another proof of Erdős-Stone-Simonovits theorem

Now we give an alternative proof of Erdős-Stone-Simonovits theorem, Theorem 2.14. Recall it says that for a fixed graph H,

$$ex(n,H) = \left(1 - \frac{1}{\chi(H)-1} + o(1)\right)\frac{n^2}{2} \quad \text{as } n \to \infty.$$

We will use Szemerédi's regularity lemma and the counting lemma. The proof will again follow our general recipe for applying the regularity lemma:

(1) (Partition) Find a regular partition
(2) (Clean) Remove edges between irregular, low-density and from small vertex sets
(3) (Count) Find a structure in the in the “reduced graph" and apply the counting lemma

Proof. Let $\chi(H) = r + 1$ and let $\epsilon > 0$. Suppose G has n vertices and at least

$$|E(G)| \geq \left(1 - \frac{1}{\chi(H) - 1} + \epsilon\right) \frac{n^2}{2}$$

edges. We will show that if n is sufficiently large, the graph G contains a copy of H. Applying Szemerédi's regularity lemma, let $V(G) = V_1 \uplus V_2 \uplus \cdots \uplus V_M$ be an ϵ'-regular partition where ϵ' is to be chosen later. Now remove each edge (x, y) of G if

(1) $(x, y) \in V_i \times V_j$ where (V_i, V_j) is not ϵ-regular
(2) $(x, y) \in V_i \times V_j$ where $d(V_i, V_j) < \frac{\epsilon}{8}$
(3) $x \in V_i$ where $|V_i| < \frac{\epsilon n}{4M}$

The number of edges removed is at most $\epsilon n^2/8$ in each step and thus at most $3\epsilon n^2/8$ edges in total are removed. After removing those edges, the resulting graph G' has at least

$$\left(1 - \frac{1}{\chi(H) - 1} + \frac{\epsilon}{4}\right) \frac{n^2}{2}$$

edges. By Turán's theorem, G' contains a copy of K_{r+1}. Let the vertices of this copy of K_{r+1} lie in $V_{i_1}, V_{i_2}, \ldots, V_{i_{r+1}}$, allowing some of the indices to coincide. By the counting lemma,

$$\begin{aligned}\left(\#\text{homomorphisms } H \to G'\right) &\geq \left[\left(\frac{\epsilon}{8}\right)^{|E(H)|} - |E(H)|\epsilon'\right] \prod_{v \in H} |V_{i_{\varphi(v)}}| \\ &\geq \left[\left(\frac{\epsilon}{8}\right)^{|E(H)|} - |E(H)|\epsilon'\right] \prod_{v \in H} \left(\frac{\epsilon n}{4M}\right)^{|V(H)|}\end{aligned}$$

where $\varphi : H \to [r]$ is a coloring. If we set $\epsilon' = \frac{1}{2|E(H)|} \left(\frac{\epsilon}{8}\right)^{|E(H)|}$, then

$$\left(\#\text{homomorphisms } H \to G'\right) \geq cn^{|V(H)|}$$

for some constant $c > 0$ depending on ϵ and H. Since the number of non-injective homomorphisms is $O_H(n^{|V(H)|-1})$, for sufficiently large n, G must contain a genuine copy of H with distinct vertices. □

The intuition in this proof is that the counting lemma yields a method to "boost" a copy of K_{r+1} guaranteed by Turán's theorem to a more complicated graph H. We remark that the phenomenon where above the threshold one copy leads to many copies of a graph H is called "supersaturation".

§3.7 Hypergraph removal lemma and Szemerédi's theorem

Szemerédi's regularity lemma can also be generalized to hold for hypergraphs. However, it is difficult to formalize this lemma and to prove it. One useful application is the hypergraph removal lemma, which we will now state and use to prove Szemerédi's Theorem. To review, an r-uniform hypergraph G has vertex set $V(G)$ and edge set $E(G)$ which consists of a collection of r-element subsets of $V(G)$.

thm:hypergraph-removal **Theorem 3.32** (Hypergraph removal lemma). *For all r-uniform hypergraphs H and $\epsilon > 0$, there is a $\delta > 0$ such that if G is an r-uniform hypergraph with n vertices and at most $\delta n^{|V(H)|}$ copies of H, then G can be made H-free by removing ϵn^r hyperedges.*

cor:hypergraph-cor **Corollary 3.33.** *Fix an r-uniform hypergraph H with $|E(H)| > 1$. If G is an r-uniform hypergraph on n vertices where every hyperedge is contained in a unique copy of H, then G has G has $o(n^r)$ hyperedges.*

This corollary follows from a similar argument to the corollary of the triangle removal lemma. The number of copies of H is $o(n^{|V(H)|})$ and thus removing $e = o(n^r)$ edges renders the hypergraph H-free. However, there are at most $e \cdot |E(H)|$ edges in G since each hyperedge is contained in a unique copy of H. We will use this corollary to deduce Szemerédi's Theorem.

thm:szemeredi **Theorem 3.34** (Szemerédi's theorem). *Fix any $k \geq 3$. Any k-AP-free set $A \subseteq [N]$ has $|A| = o(N)$.*

We will illustrate the proof for $k = 4$, and it will be obvious how to extend this proof to all values of k. Here a tetrahedron consists of edges corresponding to all possible 3-element subsets of a 4-element set.

Proof. Let m be the smallest integer larger than $4n$ and coprime to 6. We remark that this is to avoid wrap-around and divisibility issues. We embed in $\mathbb{Z}/m\mathbb{Z}$. This bound on m ensures that 4-APs in A are the same viewed in $\mathbb{Z}$ as in $\mathbb{Z}/m\mathbb{Z}$. We will build a 3-uniform 4-partite hypergraph G such that the 4-AP's in A correspond to tetrahedra in G. The four parts will be X, Y, Z and W all of which are copies of $\mathbb{Z}/m\mathbb{Z}$. Given $x \in X, y \in Y, z \in Z$ and $w \in W$, we insert an edges as follows:

- $(x, y, z) \in E(G)$ if and only if $3x + 2y + z \in A$;
- $(x, y, w) \in E(G)$ if and only if $2x + y - w \in A$;
- $(x, z, w) \in E(G)$ if and only if $x - z - 2w \in A$;
- $(y, z, w) \in E(G)$ if and only if $-y - 2z - 3w \in A$.

Note that 4-APs correspond to tetrahedra because (x, y, z, w) is a tetrahedron if and only if all four expressions lie in A. But then the expressions form a 4-AP with common difference $-x - y - z - w$. Since A is 4-AP-free, this is possible if and only if $x + y + z + w = 0$. Thus every hyperedge lies in a unique tetrahedron since (x, y, z) can be uniquely completed to a tetrahedron with vertex $w = -x - y - z$. By Corollary 3.33, we have that $|E(G)| = o(n^3)$. Now note that $|E(G)| = 4m^2|A|$. Since $m > 4n$, we have that $|A| = o(n)$. □

We remark that m is coprime to 6 in order to avoid issues with the number of the solutions to equations such as $x + 2y + 3w = a$. We now give some intuition. Loosely, the "complexity" of patterns depends on how many equations they arise from. If they arise from one equation, often graph theory or Fourier analysis is enough. If patterns are described by two or more equations, we often have apply analogous theorems involving hypergraphs. Also, Fourier analysis often stops working and we require quadratic Fourier analysis, which was introduced by Gowers. The ergodic theory approaches also display a parallel hierarchy structure.

We now state some bounds. Each step up in r-uniformity requires a step up in the Ackerman hierarchy. The requirement for the 3-uniform hypergraph regularity lemma is wowzer($\epsilon^{-O(1)}$), which is the tower function iterated $\epsilon^{-O(1)}$ times (similar to how the tower function is iterated exponentials). It was proven by Gowers with higher-order Fourier analysis that the maximum size $r_k(N)$ of a subset of $[N]$ avoiding k-AP's satisfies

$$r_k(N) \leq O\left(\frac{N}{(\log \log N)^{c_k}}\right).$$

The best known lower bound is not so much better than Behrend's bound

$$\frac{N}{e^{(\log N)^{c'_k}}} \leq r_k(N).$$

For 4-APs, Green-Tao showed a slightly better bound of

$$r_4(N) \leq O\left(\frac{N}{(\log N)^c}\right)$$

for some constant $c > 0$. Recall for $r_3(N)$, we know that this bound holds with $c = 1 + o(1)$.

CHAPTER 4

Pseudorandom graphs

ch:pseudorandom-graphs

§4.1 Eigenvalues

In this chapter, we will discuss graphs that "look random" in some sense, and compare the different ways that a graph can be random-like.

We first review some basic facts in spectral graph theory. The adjacency matrix of a graph $G = (V, E)$ with $V = \{v_1, v_2, \dots, v_n\}$ is an $n \times n$ matrix A where

$$a_{ij} = \begin{cases} 1 & \text{if } (v_i, v_j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

It is symmetric and therefore diagonalizable. Therefore:

- There are n real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$
- There is an orthonormal basis of eigenvectors $v_1, v_2, \dots, v_n \in \mathbb{R}^n$

where it holds that

$$Av_i = \lambda_i v_i \quad \text{and} \quad A = \sum_{i=1}^{n} \lambda_i v_i v_i^t.$$

If G is d-regular (i.e. $\deg(v) = d$ for all $v \in V(G)$) then $d = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq -d$ and v_1 is the all ones vector. Some facts:

- If G is connected, then $\lambda_2 < d$.
- G is bipartite if and only if $\lambda_n = -d$ (bipartite in general implies $\lambda_i = -\lambda_{n-i}$, as v_i can be obtained from v_{i-1} by negating the coordinates corresponding to one of the two vertex parts).

We first prove the following lemma about d-regular graphs, which is in a sense a stronger notion of ϵ-regularity since it also holds for small sets X and Y.

thm:eml **Theorem 4.1** (Expander mixing lemma). *Let G be a d-regular graph with n vertices. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of the adjacency matrix* *and let* $\lambda = \max\{|\lambda_2|, |\lambda_n|\}$*. Then for all* $X, Y \subseteq V(G)$*,*

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

Proof. Let A be the adjacency matrix of G and J be the $n \times n$ all 1's matrix. It turns out that $A - \frac{d}{n}J$ has the same eigenvectors as A. We check this. Note that $v_1 = \mathbf{1}$ (the all 1's $n \times 1$ vector) and thus

$$\left(A - \frac{d}{n}J\right)\mathbf{1} = 0.$$

For all $2 \leq i \leq n$, since $\langle v_i, \mathbf{1}\rangle = 0$ due to orthogonality of eigenvectors (and consequently $Jv_i = \mathbf{0}$), we have that

$$\left(A - \frac{d}{n}J\right)v_i = \lambda_i v_i.$$

So $A - \frac{d}{n}J$ has eigenvalues $0, \lambda_2, \lambda_3, \ldots, \lambda_n$. Thus its spectral norm is λ. The spectral norm $\|B\|$ for symmetric matrices B satisfies a Rayleigh quotient or operator norm identity

$$\|B\| = \sup_{u,v\neq 0} \frac{u^t Bv}{|u| \cdot |v|}.$$

Let $\mathbf{1}_X$ be the $n \times 1$ vector with 1's at entries numbered by vertices of subgraph X, and 0's everywhere else. Note that

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

□

Lec10: Hong Wang

Remark 4.2. (a) Expander mixing lemma is stronger than the notion of ϵ–regularity since X and Y could be smaller than linear size.

(b) For d–regular bipartite graphs, $\lambda_i = -\lambda_{n+1-i}$, one should replace λ by λ_2 and restrict X, Y to different parts of the bipartition.

Bilu and Linial showed that the converse to Expander mixing lemma holds as well.

Theorem 4.3 (Converse to expander mixing lemma). *Let* G *be a* d*–regular graph on* n *vertices whose adjacency matrix has eigenvalues* $\lambda_1 \geq \lambda_2 \geq \cdots \geq$

λ_n. If

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

for any $X, Y \subseteq V(G)$, then

$$\max\{|\lambda_2|, |\lambda_n|\} = O(\beta(1 + \log\frac{d}{\beta})).$$

This theorem is tight.

Random graphs mix well. The following result can be proved by bounding eigenvalues. We omit the proof.

Theorem 4.4. *Let $p = p(n) \leq 0.99$, with probability $1 - o(1)$, the random graph $G(n,p)$ satisfies*

$$|e(X,Y) - p|X||Y|| = O(\sqrt{pn|X||Y|})$$

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

We might have different interpretations of d, sometimes we think of d as a constant and let n approach infinity; sometimes we think of d as the degree of a dense graph.

The following theorem is a very difficult result. Its proof is over 100 pages long!

Theorem 4.5 (Friedman's second eigenvalue theorem). *Fix $d \geq 3$, with probability $1 - o(1)$ as $n \to \infty$, a random d–regular graph on n vertices has all but the top eigenvalue bounded in absolute value by $2\sqrt{d-1}+o(1)$, with $o(1) \to 0$ as $n \to \infty$.*

The above bound is nearly best possible, due to the next result (which has a short neat proof, which unfortunately we will not cover).

Theorem 4.6 (Alon-Boppana). *For fixed d, the second largest eigenvalue λ_2 of the adjacency matrix of a d-regular graph satisfies $\lambda_2 \geq 2\sqrt{d-1} - o(1)$, where $o(1)$ goes to zero as the number of vertices goes to infinity.*

Definition 4.7. Ramanujan Graphs are d–regular graphs with $|\lambda_2|, |\lambda_n| \leq 2\sqrt{d-1}$.

The bound $2\sqrt{d-1}$ is natural due to Friedman's second eigenvalue theorem and Alon-Boppana theorem. A more intricate explanation for why

$2\sqrt{d-1}$ is important is that it is the spectral radius of the infinite d-regular tree, which is the universal cover for all d-regular graphs.

It remains a major open problem to find Ramanujan graphs.

Conjecture 4.8. *For any $d \geq 3$, there exists infinitely many d–regular Ramanujan graphs.*

In 1987, Lubotzky, Phillips and Sanark proved the conjecture when $d = p + 1$ for prime $p \equiv 1 \pmod 4$ using deep tools from number theory related to the Ramanujan conjecture (hence the name Ramanujan graphs). Morgenstern generalized this result to $d = q + 1$ where q is a prime power.

In 2013, Marcus, Spielman, and Srivastava devised an elegant probabilistic construction for *bipartite* Ramanujan graphs (where we simply require $\lambda_2 \leq 2\sqrt{d-1}$, noting that $\lambda_i = -\lambda_{n-i}$). They showed that for any $d \geq 3$, there exists infinitely many bipartite d–regular Ramanujan graphs.

§4.2 Fourier analysis on finite abelian groups

A important class of constructions for pseudorandom graphs are Cayley graphs, arising from groups.

Definition 4.9 (Cayley Graph). Let Γ be a finite group and S is a subset of Γ such that $S^{-1} = S$. The *Cayley Graph* $\mathrm{Cay}(\Gamma, S)$ has vertices Γ, edges (g, gs) for any $g \in \Gamma$ and $s \in S$.

We will mostly work with abelian groups, in which case we use additive notation.

Definition 4.10 (Paley graph). A *Paley graph* of order p is defined as $\mathrm{Cay}(\mathbb{Z}/p\mathbb{Z}, S)$ with prime $p \equiv 1 \pmod 4$. Here $S \subset \mathbb{Z}/p\mathbb{Z}$ consists of the non-zero quadratic residues.

We will show that the second largest eigenvalue bound for Paley graph is $O(\sqrt{p})$.

Definition 4.11. Given an abelian group Γ, a *character* is a homomorphism $\chi : \Gamma \to \mathbb{C}^\times = (\{z \in \mathbb{C}, z \neq 0\}, \times)$. In other words, χ satisfies $\chi(a)\chi(b) = \chi(a + b)$ for any $a, b \in \Gamma$.

For finite groups, we have

$$\chi(a)^{|\Gamma|} = \chi(|\Gamma|a) = \chi(0) = 1.$$

Hence $\chi(a)$ is a root of unity.

Let us look at a couple examples. We consider $\Gamma = \mathbb{Z}/N\mathbb{Z}$ (i.e. the integers modulo N) for some integer $N \geq 2$. The characters of Γ are $\chi_r(a) = e^{\frac{2\pi i a r}{N}}$ for $r = 0, \ldots, N-1$. For $\Gamma = (\mathbb{Z}/2\mathbb{Z})^n$, the characters are $\chi_{\mathbf{v}}(\mathbf{x}) = (-1)^{\mathbf{x}\cdot\mathbf{v}}$, with $\mathbf{x}, \mathbf{v} \in (\mathbb{Z}/2\mathbb{Z})^n$.

If χ, ϕ are characters, then so is $\chi\phi$ defined as $(\chi\phi)(a) = \chi(a)\phi(a)$. We can define the inverse character of a character as well: $\chi^{-1}(a) = \chi(-a) = \overline{\chi}(a)$. Through the above multiplication and inverse operation, characters form a group which we denote $\widehat{\Gamma}$.

For finite abelian group Γ, it is a fact that $\Gamma \cong \widehat{\Gamma}$. This group isomorphism is not canonical. One can check that $\widehat{\mathbb{Z}/N\mathbb{Z}} \cong \mathbb{Z}/N\mathbb{Z}$. We have to make a choice in identifying $\widehat{\mathbb{Z}/N\mathbb{Z}}$ with $\mathbb{Z}/N\mathbb{Z}$ by picking a primitive root of unity ω.

We have a canonical isomorphism for $\widehat{G \times H} \cong \widehat{G} \times \widehat{H}$.

For functions $f, g : \Gamma \to \mathbb{C}$, we define inner product

$$\langle f, g\rangle = \frac{1}{|\Gamma|}\sum_{a\in\Gamma} f(a)\overline{g(a)} = \mathbb{E}_{a\in\Gamma} f(a)\overline{g(a)}.$$

Proposition 4.12. *Characters form an orthonormal basis of functions on* Γ.

Proof. For any $\chi \in \widehat{\Gamma}$, $|\chi(a)| = 1$ for any $a \in \Gamma$. So $\langle \chi, \chi\rangle = 1$. If χ is a character, not identically equal to 1, then $\mathbb{E}_{a\in\Gamma}\chi(a) = 0$ since if $\chi(b) \neq 1$,

$$\chi(b)\mathbb{E}_{a\in\Gamma}\chi(a) = \mathbb{E}_{a\in\Gamma}\chi(a+b) = \mathbb{E}_{a\in\Gamma}\chi(a).$$

If χ, ϕ are distinct characters, then $\langle \chi, \phi\rangle = \mathbb{E}_{a\in\Gamma}(\chi\phi^{-1})(a) = 0$. We have showed that characters are orthonormal, the following discrete Fourier transform enable us to show that characters form a basis. □

The completeness of the eigenbasis follows from the explicit construction of characters for finite abelian groups, namely for $\mathbb{Z}/N\mathbb{Z}$ and group products, as above

§4.2.1 Fourier transform.

Let $f : \Gamma \to \mathbb{C}$ be a function on Γ and $\chi \in \widehat{\Gamma}$ be a character. We define the Fourier transform of f as follows:

$$\widehat{f}(\chi) := \langle f, \chi\rangle = \mathbb{E}_{a\in\Gamma} f(a)\overline{\chi(a)}.$$

We have as well the Fourier inverse formula:

$$f = \sum_{\chi\in\widehat{\Gamma}}\langle \Gamma, \chi\rangle\chi = \sum_{\chi\in\widehat{\Gamma}}\widehat{f}(\chi)\chi.$$

Fourier transform is a unitary map by the following Plancherel/Parseval formula.

the unitarity of the Fourier transform is a consequence of the orthogonality of the characters. Parserval is a *con-*

Theorem 4.13 (Plancherel/Parseval's theorem). $\langle f, g\rangle = \sum_{\chi\in\widehat{\Gamma}} \widehat{f}(\chi)\overline{\widehat{g}(\chi)}$.

Proof.

$$\begin{aligned}\langle f, g\rangle &= \langle \sum_{\chi\in\widehat{\Gamma}} \widehat{f}(\chi)\chi, \sum_{\phi\in\widehat{\Gamma}} \widehat{g}(\phi)\phi\rangle \\ &= \sum_{\chi,\phi\in\widehat{\Gamma}} \widehat{f}(\chi)\overline{\widehat{g}(\phi)}\langle\chi,\phi\rangle \\ &= \sum_{\chi\in\widehat{\Gamma}} \widehat{f}(\chi)\overline{\widehat{g}(\chi)}\langle\chi,\chi\rangle\end{aligned}$$

□

We calculate the Fourier transform for a couple examples.

When Γ is the Cyclic group $\mathbb{Z}/N\mathbb{Z}$,

$$\widehat{f}(\chi_r) = \widehat{f}(r) = \frac{1}{N}\sum_{a\in\mathbb{Z}/N\mathbb{Z}} f(a)e^{-\frac{2\pi iar}{N}}.$$

We can write f as $f(a) = \sum_{r=0}^{N-1}\widehat{f}(r)e^{\frac{2\pi iar}{N}}$.

When Γ is the Boolean cube $(\mathbb{Z}/2\mathbb{Z})^n$,

$$\widehat{f}(\chi_{\mathbf{v}}) = \widehat{f}(\mathbf{v}) = \frac{1}{2^n}\sum_{\mathbf{a}\in(\mathbb{Z}/2\mathbb{Z})^n} f(\mathbf{a})(-1)^{\mathbf{v}\cdot\mathbf{a}}.$$

In particular, the characters of Boolean cube are real valued.

Here are some more classic examples with non-finite abelian groups. When Γ is a circle $\Gamma = \mathbb{R}/\mathbb{Z}$, $\widehat{\Gamma} \cong \mathbb{Z}$ (Note that $\Gamma \ncong \widehat{\Gamma}$). We have the Fourier series:

$$\begin{aligned}\widehat{f}(n) &= \int_{\mathbb{R}/\mathbb{Z}} f(x)e^{-2\pi inx}dx, \\ f(x) &= \sum_{n\in\mathbb{Z}}\widehat{f}(n)e^{2\pi inx}.\end{aligned}$$

When Γ is the real line $\Gamma = \mathbb{R}$, $\widehat{\Gamma} \cong \mathbb{R}$. We have the usual Fourier transform:

$$\begin{aligned}\widehat{f}(\xi) &= \int_{-\infty}^{+\infty} f(x)e^{-2\pi ix\xi}dx, \\ f(x) &= \int_{-\infty}^{+\infty}\widehat{f}(\xi)e^{2\pi ix\xi}d\xi.\end{aligned}$$

Proposition 4.14. *Let Γ be a finite abelian group, $S \subset \Gamma\backslash\{0\}$ is a multiplicative subset such that $S = S^{-1}$. Let A be the adjacency matrix of Cay(Γ, S),* *for any $\chi \in \widehat{\Gamma}$, $\big(\chi(a)\big)_{a\in\Gamma}$ is an eigenvector of A with eigenvalue $\sum_{s\in S}\chi(s)$. In particular, this is a complete basis of eigenvectors of A.*

The proposition is an immediate corollary of the following lemma.

Lemma 4.15. *If $f : \Gamma \to \mathbb{C}$ is a function on Γ and let A be a $|\Gamma| \times |\Gamma|$ complex matrix defined by $A_{a,b} = f(b-a)$, then $\chi \in \widehat{\Gamma}$, $\big(\chi(a)\big)_{a\in\Gamma}$ is an eigenvector of A with eigenvalue $|\Gamma|\widehat{f}(\overline{\chi})$.*

Proof. Let $\mathbf{x} = \big(\chi(a)\big)_{a\in\Gamma}$.

$$\begin{aligned}(A\mathbf{x})_a &= \sum_{b\in\Gamma} A_{a,b}\chi(b)\\ &= \sum_b f(b-a)\chi(b) = \sum_b f(b-a)\chi(b-a)\chi(a)\\ &= \widehat{f}(\overline{\chi})\chi(a)|\Gamma|\end{aligned}$$

□

We summarize the connection as: the eigenvalues of the graphs have one to one correspondence with the Fourier transform of 1_S.

§4.2.2 Paley graph.

Proposition 4.16. *Let $p \equiv 1 \pmod 4$ be a prime. The Paley graph Cay$(\mathbb{Z}/p\mathbb{Z}, S)$, where S is the collection of non-zero quadratic residues. In particular, $|S| = \frac{p-1}{2}$ is the degree of the Paley graph. If eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then* $\max\{|\lambda_2|, |\lambda_n|\} = \frac{\sqrt{p}+1}{2}$.

Proof. Let χ be a non-trivial character. $1 + 2\sum_{s\in S}\chi(s) = \sum\limits_{a\in\mathbb{Z}/p\mathbb{Z}} \chi(a^2)$. We calculate the Gauss-Sum:

$$\begin{aligned}|\sum_a \chi(a^2)|^2 &= \sum_{a,b}\chi((a+b)^2)\overline{\chi(a^2)}\\ &= \sum_{a,b}\chi((a+b)^2 - a^2) = \sum_{a,b}\chi(2ab+b^2) = |\Gamma| = p\end{aligned}$$

Thus $\sum_{s\in S}\chi(s) = \frac{\pm\sqrt{p}-1}{2}$. □

§4.3 Quasirandom graphs

The following theorem consolidates many notions of "randomness" we've seen so far.

Lec11: Linus Hamilton (IN PROGRESS)

Proposition 4.17. *Let $p \in (0,1)$ be a constant. Suppose $G_1, G_2, \ldots$ is a sequence of graphs such that G_n has n vertices. Then the following 6 conditions are all equivalent –*

(DISC) *Discrepancy:* $|e(X,Y) - p|X||Y|| = o(n^2)$ *for all* $X, Y \subseteq V(G)$.

(DISC') *Self-discrepancy:* $\left|e(X) - p\binom{|X|}{2}\right| = o(n^2)$ *for all* $X \subseteq V(G)$.

(COUNT) *Subgraph count: For all graphs H, the number of labeled copies of H in G matches what we expect from a random graph is* $\left(p^{|E(H)|} + o(1)\right) n^{|V(H)|}$ *(the rate that the $o(1)$ term goes to zero may depend on H).*

(C4) *4-cycle count:* $|E(G)| \geq (1 + o(1))\, p\binom{n}{2}$ *and the number of labeled 4-cycles in G is at most* $\left(p^4 + o(1)\right) n^4$.

(CODEG) *Codegrees:* $|E(G)| \geq (1 + o(1))\, p\binom{n}{2}$ *and, setting* $\operatorname{codeg}(u,v)$ *as the number of common neighbors of u and v, the codegrees don't vary too much:* $\sum_{u,v} \left|\operatorname{codeg}(u,v) - p^2 n\right| = o(n^3)$.

(EIG) *Eigenvalues:* $|E(G)| \geq (1 + o(1))\, p\binom{n}{2}$*, the largest eigenvalue of G's adjacency matrix is $(p + o(1))\, n$, and the other eigenvalues are all $o(n)$ in absolute value.*

The most surprising condition here is (C4). It seems so weak, compared to e.g. (COUNT). One thing about (C4) that looks strange is that it only enforces an *upper* bound on the number of 4-cycles in G. Let's show that (C4) doesn't need to explicity spell out the matching lower bound, because the lower bound is automatically guaranteed.

lem:C4-count **Lemma 4.18** **(C_4 counting lemma).** *Suppose* $|E(G)| \geq (p + o(1))\binom{n}{2}$. *Then G has at least* $\left(p^4 + o(1)\right) n^4$ *labeled 4-cycles.*

Proof of C_4 counting lemma. Let $\#C_4$ denote the number of labeled 4-cycles in G. Observe that $\#C_4 = \sum_{u,v} \operatorname{codeg}(u,v)^2 + o(n^4)$. This is because two vertices u, v sharing two common neighbors w, z is a manifestiation of the 4-cycle $uwvz$.

Now the proof is a chain of inequalities, mostly Cauchy-Schwarz:

$$\begin{aligned}
\#C_4 &= \sum_{u,v} \operatorname{codeg}(u,v)^2 + o(n^4) \\
&\geq \frac{1}{n^2}\left(\sum_{u,v} \operatorname{codeg}(u,v)\right)^2 + o(n^4) \quad \text{(Cauchy-Schwarz)} \\
&= \frac{1}{n^2}\left(\sum_{u} \deg(u)^2\right)^2 + o(n^4) \\
&\geq \frac{1}{n^2}\left(\frac{1}{n}\left(\sum_{u} \deg(u)\right)^2\right)^2 + o(n^4) \quad \text{(Cauchy-Schwarz)} \\
&= \big(p^4 + o(1)\big)\, n^4
\end{aligned}$$

as desired. □

Now let's start proving that the 6 quasirandomness conditions are equivalent to each other. We'll prove one implication at a time.

(DISC) $\Leftrightarrow$ *(DISC')*

Proof of (DISC) $\Leftrightarrow$ *(DISC').* Since (DISC') is a subcase of (DISC), it suffices to show that (DISC') implies (DISC).

This is true because $e(X,Y) = e(X\cup Y) + e(X\cap Y) - e(X\backslash Y) - e(Y\backslash X)$. If all the quantities on the right-hand side are approximately what we expect from a random graph (plus $o(n^2)$ error), then so is the left-hand side. □

(DISC) $\implies$ *(COUNT)* This is the counting lemma.

(COUNT) $\implies$ *(C4)* Obvious.

(C4) $\implies$ *(CODEG)*

Proof of (C4) $\implies$ *(CODEG).* Suppose (C4) holds. We'll prove (CODEG) using an inequality chain. The first inequality is Cauchy-Schwarz. We'll use $\#C_4$ to denote the number of labeled 4-cycles in G.

$$\begin{aligned}
\sum_{u,v} \left|\operatorname{codeg}(u,v) - p^2 n\right| &\leq n\left(\sum_{u,v}\left(\operatorname{codeg}(u,v) - p^2 n\right)^2\right)^{1/2} \\
&= n\left(\sum_{u,v}\operatorname{codeg}(u,v)^2 - 2p^2 n \sum_{u,v}\operatorname{codeg}(u,v) + p^4 n^4\right)^{1/2} \qquad \text{(Expand)} \\
&\leq n\left(\#C_4 - 2p^2 n \sum_{v}\deg(v)^2 + p^4 n^4 + o(n^4)\right)^{1/2} \\
&\leq n\left(\#C_4 - 2p^4 n^4 + p^4 n^4 + o(n^4)\right)^{1/2} \qquad \text{(Jensen's and } |E(G)| \text{ bound)} \\
&\leq n\left(p^4 n^4 - 2p^4 n^4 + p^4 n^4 + o(n^4)\right)^{1/2} \qquad \text{((C4) condition)} \\
&= o(n^3)
\end{aligned}$$

as desired. □

(CODEG) $\implies$ (DISC)

Proof of (CODEG) $\implies$ (DISC). Again, the proof is a chain of inequalities, using Cauchy-Schwarz. Assume (CODEG) holds. We'll first prove as a lemma that the vertex degrees don't vary too much: to be precise, we'll prove $\sum_u |\deg(u) - pn| = o(n^2)$.

$$\begin{aligned}
\sum_{u} |\deg(u) - pn| &\leq n^{1/2}\left(\sum_{u}(\deg(u) - pn)^2\right)^{1/2} \\
&= n^{1/2}\left(\sum_{u}\deg(u)^2 - pn\sum_{u}\deg(u) + p^2 n^2\right)^{1/2} \qquad \text{(Expand)} \\
&= n^{1/2}\left(\sum_{u,v}\operatorname{codeg}(u,v) - p^2 n^2 + o(n^3)\right)^{1/2} \\
&= o(n^2) \qquad \text{(Use (CODEG))}
\end{aligned}$$

as desired. Now, we can use (CODEG) to prove the inequality (DISC).

$$\begin{aligned}
|e(X,Y)-p|X||Y|| &= \left|\sum_{x\in X}\deg(x,Y)-p|Y|\right| \\
&\leq n^{1/2}\left(\sum_{x\in X}(\deg(x,Y)-p|Y|)^2\right)^{1/2} && \text{(Cauchy-Schwarz)} \\
&\leq n^{1/2}\left(\sum_{x\in V(G)}(\deg(x,Y)-p|Y|)^2\right)^{1/2} \\
&\leq n^{1/2}\left(\sum_{x\in V(G)}\deg(x,Y)^2-p|Y|\sum_{x\in V(G)}\deg(x,Y)+p^2|Y|^2\right)^{1/2} && \text{(Expand)} \\
&\leq n^{1/2}\left(\sum_{u,v\in Y}\operatorname{codeg}(u,v)^2-p|Y|\sum_{u\in Y}\deg(u)+p^2|Y|^2\right)^{1/2} \\
&\leq n^{1/2}\left(p^2|Y|^2-2p^2|Y|^2+p^2|Y|^2+o(n^3)\right)^{1/2} && \text{(CODEG and lemma)} \\
&\leq o(n^2)
\end{aligned}$$

as desired. □

(EIG) ⇔ (C4) So far we've connected everything except (EIG). It's possible to connect (EIG) to (DISC) using a variant of the expander mixing lemma, but we'll see that it's more straightforward to connect (EIG) to (C4). Let A denote the adjacency matrix of G. Let $\lambda_1,\ldots,\lambda_n$ denote the eigenvalues of A, and let $\#C_4$ denote the number of labeled 4-cycles in G. Observe that

$$\begin{aligned}
\sum_{i=1}^{n}\lambda_i^4 &= tr(A^4)=\left[\#C_4+o(n^4)\right] \\
\sum_{i=1}^{n}\lambda_i^2 &= tr(A^2)=2|E(G)|.
\end{aligned}$$

The condition (C4) implies, due to the C_4 counting lemma we proved first 4.18, that $|E(G)|=(p+o(1))\,n^2$ and $\#C_4=\left(p^4+o(1)\right)n^4$. This along with the above trace conditions is enough to prove that (C4) is equivalent to (EIG).

Lemma 4.19. *(EIG) implies (C4).*

Proof of (EIG) implies (C4). A chain of inequalities:

$$
\begin{aligned}
\#C_4 &= \sum_i \lambda_i^4 + o(n^4) \\
\sum_i \lambda_i^4 &\leq \lambda_1^4 + \max\left(|\lambda_2|, |\lambda_n|\right)^2 \sum_{i>1} \lambda_i^2 + o(n^4) \\
&= p^4 n^4 + o(n^2)|E(G)| + o(n^4) \\
&= p^4 n^4 + o(n^4)
\end{aligned}
$$

as desired. □

Lemma 4.20. *(C4) implies (EIG).*

This computation looks wrong since $2|E(G)| = pn^2$? Instead we could use $\lambda_1 \geq \langle u, Au \rangle / \langle u, u \rangle$ with u the ones vector, to get $\lambda_1 = (p + o(1))n$.

Proof of (C4) implies (EIG). The (C4) condition implies

$$
\begin{aligned}
\left(\sum_i \lambda_i^2\right)^2 - \sum_i \lambda_i^4 &= 4|E(G)|^2 - \left(p^4 + o(1)\right) n^4 \\
&= o(n^4).
\end{aligned}
$$

Expand the left-hand side:

$$
\sum_{i \neq j} \lambda_i^2 \lambda_j^2 = o(n^4)
$$

This implies that for every pair of eigenvalues λ_i, λ_j, at least one of λ_i, λ_j is $o(n)$. But not every eigenvalue can be $o(n)$, because otherwise we couldn't have $\sum_i \lambda^4 = \Omega(n^4)$. Therefore, every eigenvalue except for the largest eigenvalue λ_1 must be $o(n)$. From there, the 4-cycle condition $\sum_i \lambda^4 = \left(p^4 + o(1)\right) n^4$ implies that $\lambda_1 = (p + o(1))\, n$. Thus (EIG) holds. □

§4.3.1 Sparse Graphs.

We've proved that many notions of quasirandomness are equivalent for dense graphs. What about sparse graphs, where $p \to 0$ as $n \to \infty$?

Unfortunately, many notions break down.

The counting lemma fails. For $p = o(n^{-1/2})$, take a random graph $G(n, p)$. Then for each triangle, delete one of its edges. This doesn't affect (DISC) much, but now (COUNT) is false.

(DISC) $\implies$ (EIG) fails. Take $G(n,p)$ plus a disjoint clique $K_{\lfloor pn \rfloor}$. If p is small, the new clique doesn't add enough vertices to affect (DISC), but it creates a new large eigenvalue.

Some implications still work, though.

(C4) $\implies$ (CODEG) $\implies$ (DISC) still works, as does **(C4) $\implies$ (EIG) $\implies$ (DISC)**, *as long as (DISC)'s bound is changed from $o(n^2)$ to $o(pn^2)$.*

In sparse settings, we need extra hypotheses for pseudorandomness.

Lec12: Yonah Borns-Weil and Matt Babbitt

Although the properties of pseudorandomness are not true in general for sparse graphs, in the case of Cayley graphs we can still state a meaningful result. Intuitively, this is because Cayley graphs hold some symmetry properties, which prevent a small clique from giving a large eigenvalue.

To make our result explicit, define the following properties of a d-regular graph G:

- $DISC(\epsilon)$ means $\left|e(X,Y) - \frac{d}{n}|X||Y|\right| \leq \epsilon dn$ for all $X, Y \subseteq V(G)$.
- $EIG(\epsilon)$ means than $\max\{|\lambda_2|, |\lambda_n|\} \leq \epsilon d$.

Theorem 4.21 (Conlon-Zhao). *Let Γ be a (not necessarily abelian) group with n elements, and let $S \subseteq \Gamma$ satisfy $S = S^{-1}$. If $G = Cay(\Gamma, S)$ is d-regular, then:*

$$EIG(\epsilon) \implies DISC(\epsilon)$$
$$DISC(\epsilon) \implies EIG(8\epsilon)$$

Proof. The first implication follows directly from the Expander Mixing Lemma, so we only must consider the second.

We will invoke the following well-known result of Grothendieck, which we state without proof below.

grothIneq **Lemma 4.22** (Grothendieck's Inequality). *There is a $K > 0$ such that for all real $n \times n$ matrices $A = (a_{ij})$, the following inequality holds:*

$$\sup_{u_i, v_j \in B_1(H)} \left| \sum_{i,j} a_{ij} \langle u_i, v_j \rangle \right| \leq K \sup_{|x_i| = |y_j| = 1} \left| \sum_{i,j} a_{ij} x_i y_j \right|$$

where $B_1(H)$ is the unit ball in any real Hilbert space H. In particular, taking $K = 2$ works.

The optimal constant K is known as *Grothenieck's constant*, and it is known to be in $(1.67696, 1.78221)$.

Our proof that $DISC(\epsilon)$ implies $EIG(8\epsilon)$ employs this fact. Assume $DISC(\epsilon)$ holds. By Rayleigh's Theorem:

$$\max\{|\lambda_2|,|\lambda_n|\} = \sup_{|u|=|v|=1} u^t\left(A-\frac{d}{n}J\right)v$$

where A is the adjacency matrix of $\mathrm{Cay}(\Gamma,S)$ and J is the all-1's matrix. But then letting $u,v:\Gamma\to\mathbb{R}$ and renormalizing, this in turn is equal to

$$n \sup_{\langle u,u\rangle=\langle v,v\rangle=1} \mathbb{E}_{g,h\in\Gamma}\left(\left(1_S\left(g^{-1}h\right)-\frac{d}{n}\right)u(g)v(h)\right)$$

where 1_S is the indicator function for S.

We now use a clever averaging trick to put the expression into a form to which Grothendieck's Inequality can be applied. Define $u_g, v_g:\Gamma\to\mathbb{R}$ by

$$u_g(a)=u(ag),\; v_g(a)=v(ag).$$

Then by transitivity

$$\begin{aligned}\mathbb{E}_{g,h\in\Gamma}\left(\left(1_S\left(g^{-1}h\right)-\frac{d}{n}\right)u(g)v(h)\right) &= \mathbb{E}_{g,h,a\in\Gamma}\left(\left(1_S\left((ag)^{-1}ah\right)-\frac{d}{n}\right)u(ag)v(ah)\right)\\ &= \mathbb{E}_{g,h\in\Gamma}\left(\left(1_S\left(g^{-1}h\right)-\frac{d}{n}\right)\right)\mathbb{E}_{a\in\Gamma}u(ag)v(ah)\\ &= \mathbb{E}_{g,h\in\Gamma}\left(\left(1_S\left(g^{-1}h\right)-\frac{d}{n}\right)\langle u_g,v_h\rangle\right)\\ &= \frac{1}{n^2}\sum_{g,h\in\Gamma}\left(1_S\left(g^{-1}h\right)-\frac{d}{n}\right)\langle u_g,v_h\rangle\end{aligned}$$

Now let $x_g, y_g\in\{1,-1\}$ for all $g\in\Gamma$.

Then by Grothendieck's Inequality and the Triangle Inequality:

$$\begin{aligned}\max\{|\lambda_2|,|\lambda_n|\} &\le \frac{K}{n}\sup_{x_g,y_g}\left|\sum_{g,h\in\Gamma}\left(1_S\left(g^{-1}h\right)-\frac{d}{n}\right)x_g y_h\right|\\ &\le \frac{K}{n}\sup_{x_g,y_g}\sum_{a,b\in\{1,-1\}}\left|e\left(X^a,Y^b\right)-\frac{d}{n}|X^a|\left|Y^b\right|\right|\end{aligned}$$

where

$$X^a=\{g\in\Gamma|x_g=a\},\; Y^a=\{g\in\Gamma|x_g=a\}$$

Then by the $DISC(\epsilon)$ condition applied to each of the four combnations of X^a, Y^b, this gives:

$$\begin{aligned}\max\{|\lambda_2|, |\lambda_n|\} &\leq \frac{4K}{n} \cdot \epsilon dn \\ &= 4K\epsilon d \\ &\leq 8\epsilon d\end{aligned}$$

which gives precisely $EIG(8\epsilon)$.

□

Intuitively, this theorem should still hold for any graphs with the symmetry properties of Cayley graphs, whether or not they are actually derived from a finite group. Indeed we have the following corollary.

Corollary 4.23. *If G is a vertex-transitive d-regular graph on n vertices, then:*

$$EIG(\epsilon) \implies DISC(\epsilon)$$

$$DISC(\epsilon) \implies EIG(8\epsilon).$$

CHAPTER 5

Graph Limits

ch:graphlimits

§5.1 Introduction

Suppose we only knew about rational numbers (equivalently, we lived in ancient Greece), and somebody tells us to find the value of x that minimizes $x^3 - x$, provided $x \in (0,1)$. The desired value of x is $1/\sqrt{3}$, but we clearly don't know this number. But we could, say, approximate this value by using a sequence that converges to $1/\sqrt{3}$. Now say that we know real numbers again, and fix $0 < p < 1$. How do we minimize 4-cycle density among graphs with edge density at least p?

Last time, we showed that this density is at least p^4. In particular, if the graph is quasirandom, it is equal to $(1+o(1))p^4$. While for any arbitrary real number p we are unable to construct a graph that achieves this minimum value, we are able to create a sequence of quasirandom graphs of edge density p that at least approaches this minimum value.

However, it is inconvenient to solely rely on these sequences of graphs. To the end of developing machinery to talk about our idea of a converging sequence, we create a new object to capture the idea of our graph limit.

def:graphon *Definition* 5.1. A *graphon* (short for "graph function") is a symmetric measureable function $W : [0,1]^2 \to [0,1]$.

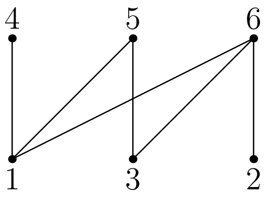
Remark 5.2. Symmetric means $W(x,y) = W(y,x)$ and measurable refers to the Lebesgue measure. In addition, graphons are denoted by W as a matter of convention, and is not necessary.

Remark 5.3. The domain can be replaced by Ω^2 for any probability space Ω. Partly for convention, partly for simplicity, we will generally use $\Omega = [0,1]$. The range can be replaced by $\mathbb{R}$, but we won't call the functions graphons at that point.

Going from a graph G to a graphon W_G can be done (at least here, but we won't do it in general) by splitting $[0,1]^2$ into $|V(G)|^2$ congruent squares, and then shading in each square that corresponds to an entry of 1 in G's adjacency matrix: the value of the graphon at any point is then 1 if it is in a shaded square, and 0 otherwise.

Remark 5.4. Because the only requirements on a graphon are being symmetric and Lebesgue measurable, we disregard any potential confusions about boundary values, in particular at grid-lines of the graphon.

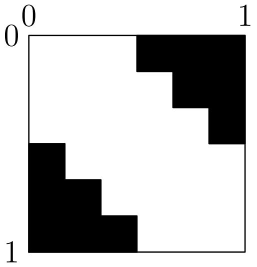
Example 5.5. Consider the half-graph H_3, as defined in Problem Set 2. The graph, its adjacency matrix, and its corresponding graphon are given below.



The half-graph H_3.

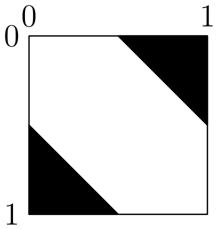
$$\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The adjacency matrix of H_3.



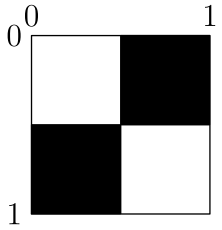
The graphon associated with H_3.

If we then consider the sequence of graphs H_n and take the limit of their graphs as $n \to \infty$, we obtain the following graphon as a result:



Example 5.6. Another example is that of any sequence of random or quasirandom graphs with density 1/2. The graphons associated with these sequence elements converge to the constant function $W(x, y) = 1/2$.

Example 5.7. It is tempting to use these Pretty Picture arguments to gain intuition for graphons, but if not given careful consideration they can lead one astray. For example, consider the graph whose graphon is a $2n \times 2n$ checkerboard configuration. The limit of these pictures seems to be once again the constant function $W(x, y) = 1/2$. However, if we consider the graph associated with each graphon in the sequence, they are in fact the complete bipartite graphs $K_{n,n}$, where vertices labeled in $[2n]$ are partitioned by their parity. With some re-labeling, the graphon associated with each of these graphs, and thus the limit graphon, is given below.



For a graph G, with $V(G) = \{1, 2, \dots, n\}$, let the associated graphon $W_G : [0,1]^2 \to [0,1]$ be constructed first by dividing $[0,1]$ into n equal length intervals $I_1, \dots, I_n$, and set, for $x \in I_i$ and $y \in I_j$,

$$W(x, y) = \begin{cases} 1 \text{ if } (i, j) \in E(G), \\ 0 \text{ otherwise.} \end{cases}$$

Definition 5.8. Given two graphs H and G, a *graph homomorphism* from H to G is a map $\varphi : V(H) \to V(G)$ that preserves edge relationships: $(u, v) \in E(H) \to (\varphi(u), \varphi(v)) \in E(G)$.

Example 5.9. It is possible to make some simple calculations about the number of homomorphisms between a general graph and a given small graph:

$$\begin{aligned} |\mathrm{Hom}(\bullet, G)| &= |V(G)|, \\ |\mathrm{Hom}(\bullet\!-\!\bullet, G)| &= 2 \cdot |E(G)|, \\ |\mathrm{Hom}(\triangle, G)| &= 6 \cdot (\#\text{ triangles in } G), \\ |\mathrm{Hom}(H, \triangle)| &= \#\text{ of proper colorings of } H \text{ w/ labeled colors 1,2,3.} \end{aligned}$$

Use Hom to denote set and hom to denote cardinality

declare a new math operator Hom in the original document.

Remark 5.10. The number of non-injective homomorphisms is $O_H\left(n^{|V(H)|-1}\right)$, where O_H means that the constant in O depends on H.

Definition 5.11. The *homomorphism density* of two graphs G and H is

$$t(H, G) = \frac{|\mathrm{Hom}(H, G)|}{|V(G)|^{|V(H)|}},$$

i.e. it is the probability that a unit random map $V(H) \to V(G)$ is a graph homomorphism.

Remark 5.12. Notice that as $|V(G)| \to \infty$, this is the same as subgraph density.

We can define homomorphism density for graphons as well, which lets us start connecting graphs to graphons.

Definition 5.13. For a graph H and a graphon W,

$$t(H,W) = \int_{[0,1]^{|V(H)|}} \prod_{(i,j)\in E(H)} W(x_i,x_j)dx_1\cdots dx_{|V(H)|}.$$

For example, $t(\triangle, W) = \int_{[0,1]^3} W(x,y)W(y,z)W(z,x)dxdydz$.

These two definitions of density are indeed consistent: for graphs G and H, and the graphon W_G, $t(H,G) = t(H,W_G)$.

Definition 5.14. We say that a sequence of graphs $\{G_n\}$ (or graphons $\{W_n\}$) is *convergent* if all homomorphism densities converge, i.e., $t(F,G_n)$ (respectively $t(F,W_n)$) converges as $n \to \infty$ for all graphs F.

Remark 5.15. While it is not required that $|V(G_n)| = n$, it can be convenient to think of the graphs in this way.

A sequence of quasirandom graphs with edge density $\frac{1}{2} + o(1)$ is convergent (from a previous theorem). But if you have a convergent sequence of graphs, how can we talk about its limit, if it even exists? With the structures we've defined so far, we can successfully answer this question: a convergent sequence of graphs converges to a graphon. In fact, the converse holds: for any graphon, we can create a sequence of graphs that converges to it.

Which theorem?

§5.2 Distances between graphs

Now that we've constructed machinery for convergent sequences of graphs, it is natural to ask whether or not we can impose some distance metric on them as well. There are several ways to do this, the first of which is trivial: the distance between a graph G and G' can be defined as

$$\sum_{k\geq 1} 2^{-k} \cdot |t(F_k,G) - t(F_k,G')|$$

where F_1, F_2, ... is a list of *all* graphs. This gives us a valid distance metric, but it's completely uninformative and we won't use this.

A rather natural notion of distance is the *edit distance*, similar to the Hamming distance. It is defined as the minimum number of edges one needs to change to get from one graph to the other. We can normalize by the number of edges, or by $1/|V(G)|^2$. However, this doesn't capture what we want out of distance. Consider two instances of a random graph $G(n, 1/2)$. We'd like to say that they are close to each other, because they have very similar edge densities, but you need to change around half the edges to get from one to the other.

The most relevant notion of distance of our purposes is based on discrepancy. Recall the DISC criterion from regularity.

Definition 5.16. The *cut-norm* of $W : [0,1]^2 \to \mathbb{R}$ is defined by

$$||W||_\square := \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} W \right|,$$

where S and T are two measurable sets.

This definition of cut-norm is crucial for graphon study.

Remark 5.17. It turns out that the supremum in the above definition is always attained, for some S and T. This is a measure theoretic issue so we won't pay it any heed, but it's nice to know.

Other good norms to know are the L^p norms, defined as $||W||_p = \left(\int |W|^p\right)^{1/p}$ for finite p, and $||W||_\infty$ the essential supremum of W. Need to ignore measure zero contributions. The edit distance corresponds to L^1 for graphons. The norm $\langle W, U\rangle = \int WU$ also comes up occasionally.

These norms satisfy some simple properties. In terms of magnitude, $||W||_\square \leq ||W||_1$ and $||W||_p \leq ||W||_{p+1}$. In addition, a sequence of quasirandom graphs $\{G_n\}$ with edge density $p + o(1)$ has the property that $||W_n - p||_\square \to 0$.

Definition 5.18. We say that $\varphi : [0,1] \to [0,1]$ is *measure-preserving* if $\lambda(\varphi^{-1}(A)) = \lambda(A)$ for all measurable $A \subseteq [0,1]$.

For example, the map $x \hookrightarrow 2x \bmod 1$ is measure-preserving but not bijective (or invertible). Given $W\colon [0,1]^2 \to \mathbb{R}$ and a measure-preserving map $\varphi\colon [0,1] \to [0,1]$, we write $W^\varphi\colon [0,1]^2 \to \mathbb{R}$ to denote

$$W^\varphi(x,y) = W(\varphi(x), \varphi(y)).$$

Definition 5.19. The cut-distance between two graphons W and U is

$$\delta_\square(W,U) = \inf_\varphi ||W - U^\varphi||_\square,$$

where φ ranges over all measure-preserving bijections.

This definition of cut-distance has caveats, as we will see shortly.

Lec13: Jake Wellens

Remark 5.20. We won't belabor the analytic technicalities here, but we remark that this infimum is not always attained, as evidenced by the pair of graphons W and W^ψ, where $W(x,y) = xy$ and

$$\psi(x) = 2x \mod 1.$$

Indeed, ψ is 2-to-1 and hence cannot be undone by any measure preserving bijection φ, so

$$||W - (W^\psi)^\varphi||_\square \neq 0,$$

yet one can construct a sequence of measure preserving bijections φ_n (by chopping up $[0,1]$ into dyadic pieces and permuting them appropriately) which *almost* undoes ψ in the sense that

$$\inf_n ||W - (W^\psi)^{\varphi_n}||_\square \leq \inf_n ||W - (W^\psi)^{\varphi_n}||_1 = 0$$

and hence $\delta_\square(W, W^\psi) = 0$. It is true, however, that for any pair of graphons W and U,

$$\delta_\square(W,U) = \min_{\psi,\varphi} ||W^\varphi - U^\psi||_\square$$

where the minimum is taken over all pairs of measure preserving maps (and is attained).

Definition 5.21. For a graph G, we will use W_G to denote the graphon associated to G, given by scaling G's adjacency matrix to the unit square, and set

$$\delta_\square(G,H) := \delta_\square(W_G, W_H)$$

Definition 5.22. We'll use $\mathcal{W}_0$ to denote[1] the set of all graphons, and $\widetilde{\mathcal{W}_0}$ to denote the space of graphons after identifying graphons with distance zero under the cut metric $\delta_\Box$.

Note, for example, $\delta_\Box(K_3, K_{2,2,2}) = 0$ since the two graphs are represented by the same graphon.

As you may have guessed by now, the metric topology given by $\delta_\Box$ is compatible with our earlier notion of convergence for graphons:

Theorem 5.23 (Equivalence of Convergence). *A sequence of graph(on)s is convergent if and only if it is Cauchy with respect to $\delta_\Box$.*

Missing theorem statement on the existence of limits. -YZ

One half of this equivalence (Cauchy in $\delta_\Box$ implies convergent) comes from the following lemma[2]:

Defer counting lemma until after main theorem statements

Lemma 5.24 (Counting Lemma). *If W and U are graphons and F is a graph, then*

$$|t(F,W) - t(F,U)| \le |E(F)|\delta_\Box(W,U)$$

Proof. It suffices to show that $|t(F,W) - t(F,U)| \le |E(F)| ||W-U||_\Box$, since $t(F,W) = t(F,W^\varphi)$, and hence we can replace W by W^φ for any measure preserving bijection φ. A useful reformulation of the cut norm is

$$||W||_\Box = \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} W \right| = \sup_{U,V:[0,1]\to[0,1]} \left| \int_{[0,1]^2} W(x,y)U(x)V(y)\,dxdy \right|$$

which follows from multilinearity of the final expression inside the $|\cdot|$ bars. With this observation in hand, the proof reduces to a simple trick which we illustrate in the case of $F = K_3$ (the general case is analogous). Write

$$\begin{aligned} t(K_3,W) - t(K_3,U) &= \int (W(x,y)W(x,z)W(y,z) - U(x,y)U(x,z)U(y,z)) \\ &= \int (W-U)(x,y)W(x,z)W(y,z) + \int U(x,y)(W-U)(x,z)W(y,z) \\ &\quad + \int U(x,y)U(x,z)(W-U)(y,z) \end{aligned}$$

[1] The subscript zero is a matter of convention. Some people, e.g. Lovász, use $\mathcal{W}$ to denote the set of all bounded symmetric measurable functions $[0,1]^2 \to \mathbb{R}$

[2] Compare this counting lemma with the counting lemma from the chapter on Szemerédi regularity, in the case that one graphon is constant.

For each fixed z, $W(\cdot, z)$ is a measurable function from $[0,1] \to [0,1]$, so our reformulation of the cut norm implies

$$\left| \int (W-U)(x,y)W(x,z)W(y,z)\,dxdy \right| \leq ||W-U||_\square, \quad \text{for all } z$$

and hence

$$\left| \int (W-U)(x,y)W(x,z)W(y,z)\,dxdydz \right| \leq ||W-U||_\square.$$

Doing the same for the other integrals yields the desired bound of $3||W - U||_\square$. □

The next logical question to ask is whether convergent sequences in $(\widetilde{\mathcal{W}_0}, \delta_\square)$ have limits. We'll answer this question (and more) in the upcoming lecture, when we prove the following theorem:

Theorem 5.25. $(\widetilde{\mathcal{W}_0}, \delta_\square)$ *is compact.*

While this theorem may seem surprising, it turns out to be equivalent to Szemerédi's Regularity lemma. Indeed, for any ϵ, any graphon is ϵ-close to one out of a finite collection $\{W_i\}$ of graphons. Since we can approximate each of these W_i with step graphons and take a common refinement of the associated partitions, we can then approximate any graphon by one with a bounded number of vertex parts which look pseudorandom. This is, of course, only some vague intuition – we'll return to the relationship between regularity and compactness later in more detail.

Remark 5.26. The set of graphs is dense in $(\widetilde{\mathcal{W}_0}, \delta_\square)$. One way to see this is to observe that step graphons are dense in L^1, and that each constant graphon is the limit of a sequence of quasirandom graphs. A more canonical approach, however, is via *W-random graphs* (a graphon generalization of Erdös-Renyi random graphs):

$$\mathbb{G}(n, W) := \begin{cases} 1. \text{ pick } x_1, \ldots, x_n \sim [0,1] \text{ uniformly} \\ 2. \text{ add each edge } (i,j) \text{ independently with probability } W(x_i, x_j) \end{cases}$$

It is known (although we won't prove it) that if $G_n \sim \mathbb{G}(n, W)$, then $G_n \to W$ with probability 1.

Definition 5.27. Given a partition $\mathcal{P} = \{S_i\}$ of $[0,1]$ and $W : [0,1]^2 \to \mathbb{R}$, define

$$W_{\mathcal{P}}(x,y) = \frac{1}{\lambda(S_i)\lambda(S_j)} \int_{S_i \times S_j} W \quad \text{if } (x,y) \in S_i \times S_j$$

This averaging operation is the projection operator in the Hilbert space $L^2([0,1]^2, \mathbb{R})$ onto the subspace of functions which are constant on each rectangle $S_i \times S_j$. Equivalently, it is the conditional expectation operator corresponding to the σ-algebra generated by $\{S_i \times S_j\}$.

Given a graphon W, one may want to find a good approximation by step graphons. From classical analysis we know we can always find a partition $\mathcal{P}$ such that $||W - W_{\mathcal{P}}||_p < \epsilon$, but the number of parts in the partition may be enormous. If we consider the cut norm instead of the L^p norm, then the following theorem says that we don't need too many parts:

Theorem 5.28 (Weak Regularity for Graphons). *For all $\epsilon > 0$ and graphons $W : [0,1]^2 \to [0,1]$, there exists a partition $\mathcal{P}$ of $[0,1]$ into at most $4^{1/\epsilon^2}$ parts such that $||W - W_{\mathcal{P}}||_\square < \epsilon$.*

The proof is an energy increment argument, much like the one used to prove Szemerédi's regularity lemma, and we'll see it in the next lecture. For graphs, closeness in cut-norm to $W_{\mathcal{P}}$ corresponds to the following notion of *weak regularity*:

Definition 5.29. In a graph G, a partition $\mathcal{P} = V_1 \uplus \cdots \uplus V_k$ is called *weakly ϵ-regular* if for any $A, B \subset V(G)$,

$$\left| e(A,B) - \sum_{i,j=1}^{k} d(V_i, V_j)|V_i \cap A||V_j \cap B| \right| \leq \epsilon |V(G)|^2$$

We can now state the weak regularity lemma in this language for graphs, as it was first shown by Frieze and Kannan in 1996:

Theorem 5.30 (Weak Regularity Lemma for Graphs). *For all $\epsilon > 0$ and graphs G, there is a weakly ϵ-regular partition of G into at most $4^{1/\epsilon^2}$ vertex parts.*[3]

The original motivation for the above theorem was algorithmic – indeed, Frieze and Kannan gave an algorithm to construct such partitions in time $O(n/\epsilon^2 + 2^{\tilde{O}(1/\epsilon^2)})$, and used this to give a PTAS for solving Max-Cut on

[3] The partition can be made equitable with only $2^{O(1/\epsilon^2)}$ more parts.

dense graphs.[4] This makes the dense case quite different from the general problem, as the best known polynomial time approximation algorithm for Max-Cut attains an approximation ratio of ≈ .878, and beating this would refute the Unique Games conjecture (and, in any case, beating 16/17 ≈ .941 would show P = NP).

Lec14: Minjae Park

§5.3 Compactness of the graphon space

Many theorems are repeated from the previous lecture. Should I cite to the previous theorems without rewriting? (No need to repeat theorems. Just cite. -YZ)

The space of graphons looks arbitrarily enormous at first, but it is not. Indeed, the graphon space turns out to be compact under the cut-metric.

thm:compactness

Theorem 5.31 (Lovász-Szegedy). *The metric space $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is compact.*

In order to prove the compactness, we need to regularize graphons. The following theorem is an analogue of the weak regularity lemma for graphs.

Definition 5.32. Let W be a graphon and $\mathcal{P} = \{P_1, P_2, \cdots\}$ be a partition of $[0,1]$ into measurable sets. Denote $W_\mathcal{P}$ to be the graphon obtained by taking average of W over each step, i.e. $W_\mathcal{P} = \int_{P_i \times P_j} W$ on $P_i \times P_j$.

thm:weakreg-graphon

Theorem 5.33 (Weak regularity lemma for graphons). *For any $\epsilon > 0$ and any graphon W, there exists a partition $\mathcal{P}$ of $[0,1]$ into at most $4^{1/\epsilon^2}$ measurable sets such that*

$$\|W - W_\mathcal{P}\|_\square \le \epsilon.$$

Recall that small cut-distance captures nothing but a normalized (yet more global) version of ϵ-homogeneity in the weak graph regularity lemma. As in the proof of regularity lemma, we make use of energy increment.

Lemma 5.34 (L^2 energy increment). *Let W be a graphon and $\mathcal{P}$ be a partition of $[0,1]$ into finitely many measurable sets such that $\|W - W_\mathcal{P}\|_\square > \epsilon$. Then there is a refinement $\mathcal{P}'$ of $\mathcal{P}$ dividing each part of $\mathcal{P}$ into at most 4 parts such that*

$$\|W_{\mathcal{P}'}\|_2^2 > \|W_\mathcal{P}\|_2^2 + \epsilon^2.$$

[4] In other words, they provide for each $\epsilon > 0$, an algorithm which runs in time $O(n/\epsilon^2 + 2^{\tilde{O}(1/\epsilon^2)})$ and outputs a cut whose value is at most ϵn^2 less than the maximum cut value.

Proof. Since $\|W - W_{\mathcal{P}}\| > \epsilon$, there exist measurable sets $S, T \subset [0,1]$ with

$$\left| \int_{S\times T} (W - W_{\mathcal{P}}) \right| > \epsilon.$$

Now refine $\mathcal{P}$ into $\mathcal{P}'$ by introducing S and T, that is dividing each part by at most 4 parts depending on whether each element is in either $S \cap T$, $S \backslash T$, $T \backslash S$, or none of them. Note that $\langle W_{\mathcal{P}'}, W_{\mathcal{P}} \rangle = \langle W_{\mathcal{P}}, W_{\mathcal{P}} \rangle$ because $W_{\mathcal{P}}$ is constant over each step with respect to $\mathcal{P}$, and $\mathcal{P}'$ is a refinement of $\mathcal{P}$. Hence we have orthogonality relation $\langle W_{\mathcal{P}'} - W_{\mathcal{P}}, W_{\mathcal{P}} \rangle = 0$ which implies

$$\|W_{\mathcal{P}'}\|_2^2 = \|W_{\mathcal{P}'} - W_{\mathcal{P}}\|_2^2 + \|W_{\mathcal{P}}\|_2^2$$

by the Pythagorean theorem. On the other hand, it follows from the Cauchy-Schwarz inequality and $\langle W, 1_{S\times T} \rangle = \langle W_{\mathcal{P}'}, 1_{S\times T} \rangle$ (by the same reason above) that

$$\begin{aligned} \|W_{\mathcal{P}'} - W_{\mathcal{P}}\|_2 &\geq |\langle W_{\mathcal{P}'} - W_{\mathcal{P}}, 1_{S\times T} \rangle| \\ &= |\langle W - W_{\mathcal{P}}, 1_{S\times T} \rangle| > \epsilon. \end{aligned}$$

Combining this with the last equality gives the desired result. □

thm:weakreg-graphon

Proof of Theorem 5.33. Let an $\epsilon > 0$ and a graphon W be given, and choose any partition of $[0,1]$ into finite measurable sets. Apply the previous lemma repeatedly until we find a regular partition. The number of refinements cannot exceed $1/\epsilon^2$ because the L^2-norm of any graphon is bounded by 1. This proves the weak regularity lemma for graphons. □

Remark 5.35. The regular partition in the proof is obtained by a common refinement of $O(1/\epsilon^2)$ many subsets. In fact, a more efficient formulation of the weak regularity lemma is as follows. For any $\epsilon > 0$ and any graphon W, there exist subsets $\{S_i\}_{i=1}^k, \{T_i\}_{i=1}^k$, and real numbers $\{a_i\}_{i=1}^k$ with $k < 1/\epsilon^2$ such that

$$\left\| W - \sum_{i=1}^k a_i 1_{S_i \times T_i} \right\|_\square \leq \epsilon.$$

This formulation is useful in many applications, for example in computer sciences. But we do not use it for our purpose because the approximation is not obtained by averaging W over steps.

The proof of compactness makes use of a martingale convergence theorem, so we briefly review some facts about martingales.

Definition 5.36. A **martingale** is a random sequence $X_1, X_2, \cdots$ (with $X_0 = \mathbb{E}X_1$) such that for all $n \geq 1$

$$\mathbb{E}\left[X_n \mid X_{n-1}, X_{n-2}, \cdots, X_1\right] = X_{n-1}.$$

Example 5.37. One typical example of martingale is $X_n = X_{n-1} + Y_n$ with $X_0 = 0$ where Y_n's are i.i.d. random variables taking values ± 1 with probability $1/2$. This is so-called coin tossing or the drunkard's walk. Another example is X_n defined as the conditional expectation of X using information revealed upon time n where X is some random variable.

There are two useful properties of martingales. **Martingale concentration**, or Azuma's inequality, tells that a martingale with bounded step cannot drift too far. More precisely, $X_n = O(\sqrt{n})$ with high probability. A second useful property, **martingale convergence**, says that if the martingale itself is bounded, then it converges almost surely.

Theorem 5.38 (Doob's martingale convergence theorem). *Every bounded martingale converges almost surely.*

Proof sketch (by "gambling"). The fundamental principle is that you cannot beat a martingale in expectation. But be careful that this is only valid in finite time.

If $\{X_n\}$ does not converge then there exists $a, b \in \mathbb{R}$ with $a < b$ such that X_n "crosses" $[a, b]$ infinitely often. Thus it is enough to show that this event does not happen almost surely for any $a, b \in \mathbb{Q}$ with $a < b$ (there are only countably many such events, and if we know that each occurs with probability zero, and then with probability 1, none of them occurs).

Let U_N denote the number of "upcrossings" for $[a, b]$, that is the number of events when $X_n < a$ and $X_{n+t} > b$ for the first time $n + t$, up to time N. Now consider a betting strategy in stock market (say, X_n represents some share price) as follows: if $X_n < a$ buy a share and sell it for the first time when $X_{n+t} > b$. Taking advantage of upcrossings, the profit of this strategy at time N is at least $U_n(b - a) - a$. Since the expected profit is 0 for a martingale, we conclude $\mathbb{E}U_n \leq \frac{a}{b-a}$. If we set $U = \lim_{N\to\infty} U_N$, it follows that $\mathbb{E}U \leq \frac{a}{b-a}$ by the monotone convergence theorem. In conclusion, $\mathbb{P}(U = \infty) = 0$. □

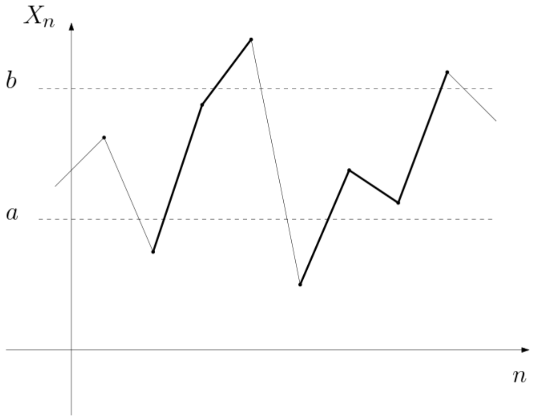


FIGURE 1. Thick segments indicate an individual "upcrossing."

Remark 5.39. The boundedness condition can be relaxed to $\mathbb{E}\,|X_n| < C$ for all n, or more generally to uniform integrability. See graduate probability textbooks (for example, [**12**]) for detailed and precise treatment in this topic.

thm:compactness

Proof of Theorem 5.31. Suffices to show that $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is sequentially compact, i.e. that every sequence $W_1, W_2, \cdots$ has a subsequential limit in the graphon space. Using the weak regularity lemma for each W_n, construct partitions $\mathcal{P}_{n,1}, \mathcal{P}_{n,2}, \cdots$ of $[0,1]$ such that

- $\mathcal{P}_{n,k+1}$ refines $\mathcal{P}_{n,k}$,
- $|\mathcal{P}_{n,k}| = m_k$ which depends only on k,
- $\|W_n - W_{n,k}\| \le 1/k$ where $W_{n,k} = (W_n)_{\mathcal{P}_{n,k}}$.

Replacing W_n by some W_n^ϕ with a measure-preserving transformation ϕ, we may assume that all parts of $\mathcal{P}_{n,k}$ are intervals.

By taking subsequence, assume that the endpoints of the intervals in $\mathcal{P}_{n,1}$ converge as $n \to \infty$ and the corresponding values in each step of $W_{n,1}$ converge. Hence there exists a graphon U_1 such that $W_{n,1} \to U_1$ pointwise almost everywhere as $n \to \infty$. Repeat this for each $k = 2, 3, \cdots$ so that $W_{n,k} \to U_k$ as $n \to \infty$ for fixed k, where U_k is a step graphon associated with $\mathcal{P}_k = \lim_{n\to\infty} \mathcal{P}_{n,k}$.

Recall $W_{n,k} = (W_{n,k+1})_{\mathcal{P}_{n,k}}$ and thus in the limit we also have $U_k = (U_{k+1})_{\mathcal{P}_k}$, for $\mathcal{P}_{k+1}$ refines $\mathcal{P}_k$. Let (X, Y) be a random point chosen in $[0,1]^2$ uniformly. Then $\{U_k(X,Y)\}$ is a martingale because

$$\mathbb{E}\left[U_{k+1}(X,Y) \mid U_k(X,Y)\right] = (U_{k+1})_{\mathcal{P}_k}(X,Y) = U_k(X,Y).$$

(For whom not familiar with conditional expectation, note that $U_k(X,Y)$ is a stepwise constant function with steps associated with $\mathcal{P}_k$, so the conditional expectation is nothing but the integration over each step.) Since it is obvious that $\{U_k\}$ is bounded, by the martingale convergence theorem, there exists a graphon U such that $U_k \to U$ pointwise almost everywhere as $k \to \infty$.

Fix $\epsilon > 0$. There exists some $k > \frac{3}{\epsilon}$, such that $\|U_k - U\| < \frac{\epsilon}{3}$. For sufficiently large n, we have $\|W_{n,k} - U_k\|_1 < \epsilon/3$, and so

$$\begin{aligned}\delta_\square(U, W_n) \le \delta_\square(U, U_k) + \delta_\square(U_k, W_{n,k}) + \delta_\square(W_{n,k}, W_n) \le \|U - U_k\|_1 + \|U_k - W_{n,k}\|_1 + \delta_\square(W_{n,k}, W_n) \\ \le \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.\end{aligned}$$

Since ϵ can be chosen arbitrarily small, we conclude that $W_1, W_2, \ldots$ converges to U in the metric space $(\widetilde{\mathcal{W}}_0, \delta_\square)$. □

Lec15: Diego Roque

Consequences of compactness:

Corollary 5.40. *For all $\epsilon > 0$, there exists N_0 such that for every graphon W, there is some graph G with N_0 vertices such that $\delta_\square(W, G) < \epsilon$.*

Proof. Look at $B_\square(G, \epsilon) = \{\text{graphons } W | \delta_\square(G, W) < \epsilon\}$.

Consider $\{B_\square(G, \epsilon) | G \text{ is a graph}\}$. This is an open cover of the space of graphons $\widetilde{\mathcal{W}}_0$. This because every graphon is a limit of a sequence of graphs.

By compactness, there exists a finite cover using only a finite list of graphs $G_1, \ldots, G_k$. Take $N = \operatorname{lcm}(|V(G_1)|, \ldots, |V(G_k)|)$. All of these G_i's can be identified with a graph on N vertices (with respect to $\delta_\square$), so each graphon must be ϵ close to this graph.

□

How is compactness related to the regularity lemma?

We used the weak regularity lemma to prove compactness. We can use compactness to prove the strong regularity lemma.

Theorem 5.41 (Strong Regularity Lemma). *Let $\vec{\epsilon} = (\epsilon_1, \ldots)$, $\epsilon_k > 0$ for all k. There exist $M = M(\vec{\epsilon})$, such that every graphon W can be written as $W = W_{str} + W_{psr} + W_{sml}$ where*

1. *A structured component W_{str} is a step function graphon with $k \le M$ steps.*
2. *A pseudorandom component W_{psd} such that $\|W_{psr}\|_\square \le \epsilon_k$*
3. *A small component $\|W_{sml}\|_1 \le \epsilon_1$*

Remark 5.42. Weak regularity lemma corresponds $\epsilon_k = \epsilon$ for all k, $(W_{sml} = 0)$.

Szemerendi regularity lemma corresponds to $\epsilon_k = \frac{\epsilon}{k^2}$.

Explain this a bit more

The structured part is breaking into small pieces. The small you completely get rid of. The rest must be pseudorandom between each pair of parts.

Proof. Let's prove the strong regularity lemma using compactness.

Every W has a step function approximation function U such that $\|W - U\|_1 \leq \epsilon$.

Let $k(W) = \min(k : \exists k\text{-step graphon } U \text{ such that } \|W - U\|_1 \leq \epsilon_1)$.

Consider the open balls $B_\square(W, \epsilon_{k(W)} : W \in \widetilde{\mathcal{W}_0})$. This will be an open cover of $\widetilde{\mathcal{W}_0}$. By compactness, there exists a finite subcover $\mathcal{S}$ of graphons such that

$$\bigcup_{W \in \mathcal{S}} B_\square(W, \epsilon_{k(W)}) = \widetilde{\mathcal{W}_0}$$

Set $M = \max_{W \in \mathcal{S}} k(W)$.

For every W, there exists $W' \in \mathcal{S}$, such that $W \in B_\square(W', \epsilon_{k(W')})$. So there exists U such that U has k steps (with $k \leq M$), such that $\|W' - U\|_1 \leq \epsilon_1$. So we have that

- $\|W - W'\|_\square < \epsilon_{k(W')}$
- $\|W' - U\|_1 \leq \epsilon_1$
- U is a step function with $k \leq M$ steps

We can write $W = U + (W - W') + (W' - U)$. If we define $W_{str} = U, W_{psr} = W - W', W_{sml} = W' - U$, we get the required graphons. This finishes the proof.

□

Remark 5.43. We can use the strong regularity lemma to prove an induced removal lemma, where you delete not only copies of H but induced copies of H. There you can add or delete edges.

Theorem 5.44 (Existence of Limit). *If $t(F, W_n)$ converges for all F, then exist W such that $t(F, W_n)$ converges to $t(F, W)$ for all F.*

Proof. Take any limit point W of the convergent subsequence $\{W_{n_i}\}$ of $\{W_n\}$, by compactness of $\tilde{\mathcal{W}}_0$. Note that $\delta_\square(W_{n_i}, W)$ converges to 0 as i

goes to infinity. By counting lemma, $t(F, W_{n_i})$ converges to $t(F, W)$ as i goes to infinity. But we know that $t(F, W_n)$ converges, so it must also converge to $t(F, W)$.

□

Theorem 5.45 (Equivalence of convergence). *$t(F, W_n)$ converges for all F if and only if W_n is Cauchy with respect to $\delta_\square$.*

Proof. The reverse implication ($\Longleftarrow$) follows from the counting lemma.

It remains to prove the forward direction ($\Longrightarrow$). By compactness, not Cauchy implies the existence of at least two distinct limit points, say W and U. We will show this cannot happen; i.e. $\delta_\square(U, W) = 0$.

By assumption, $t(F, W_n)$ converges to $t(F, W)$ for all F, while also converging to $t(F, U)$ for all F. Hence $t(F, U) = t(F, W)$ for all F. The following lemma completes the proof.

lem:moments-graphons **Lemma 5.46** (Moments Result). *If $t(F, W) = t(F, U)$ for all F, then $\delta_\square(U, W) =$ 0.*

Remark 5.47. Why moments? It's analogous to saying that all the moments of two well-behaved random variables agree, then the random variables agree. Taking each F is like a moment.

□

Lec16: Younhun Kim

lem:moments-graphons

Proof sketch of Lemma 5.46. Recall the random graph $\mathbb{G}(k, W)$, which is determined by choosing $x_1, \ldots, x_k$ uniformly at random from $[0, 1]$, then drawing an edge between i and j with probability $W(x_i, x_j)$ for each pair of vertices.

For a k-vertex graph F, we have

$$\begin{aligned}
&\mathbb{P}[\mathbb{G}(k, W) \cong F \text{ as labeled graphs}] \\
&= \int_{[0,1]^k} \prod_{ij \in E(F)} W(x_i, x_j) \prod_{ij \notin E(F)} (1 - W(x_i, x_j)) dx_1 \cdots dx_k \\
&= \sum_{F' \supseteq F} (-1)^{|E(F')| - |E(F)|} t(F', W) \qquad \text{(Inclusion-Exclusion)}
\end{aligned}$$

Since $t(F, W) = t(F, U)$ for all F, $\mathbb{G}(k, W)$ and $\mathbb{G}(k, U)$ are identically distributed – thus, we can couple them so that they produce the same graph. It can be shown that (though we will not provide a proof) $\delta_\square(\mathbb{G}(k, W), W) \to 0$ as $k \to \infty$ with probability 1. Such an event implies $\delta_\square(W, U) = 0$. □

These results yield the following corollary via compactness, which you will prove in the homework.

Corollary 5.48 (Inverse Counting Lemma). *For all $\epsilon > 0$, there exist $k, \eta > 0$ such that if U, W are graphons that satisfy*

$$|t(F, U) - t(F, W)| \le \eta$$

for all k-vertex graphs F, then $\delta_\square(U, W) \le \epsilon$.

Remark 5.49. It is possible to derive the quantitative bound $k = 2^{O(\epsilon^{-2})}, \eta = 2^{-k^2}$ using the regularity lemma.

§5.4 Graph homomorphism inequalities

lem:moments-graphons

§5.4.1 Finitely Forcible Graphons. The moments result (Lemma 5.46) seems like it might be over-constrained for some families of graphons. It begs the question, "can we get away with finitely many, or a small number of moments?"[5] Consider the following examples:

Example 5.50. Think of quasirandom graphs:

(1) $t(K_2, W) = p \;\&\; t(C_4, W) = p^4 \implies \delta_\square(W, p) = 0$

(2) $t(K_2, W) = p \;\&\; t(C_4, W) \le p^4 + \delta \implies \delta_\square(W, p) \le \epsilon$ (Condition (C4))

This motivates us to think about **finitely forcible graphons**: graphons W that can be identified by finitely many $t(F, W)$. Including p, the constant graphon which we considered just now, some other examples include the block graphon, as well as the upper-triangular graphon $W(x, y) = \mathbf{1}_{\{x+y<1\}}$.

Remark 5.51. What about graphs, instead of graphons? It is possible to show, for graphs G_1 and G_2:

- if $\hom(F, G_1) = \hom(F, G_2)$ for all graphs F, then $G_1 \cong G_2$, and
- if $\hom(G_1, F) = \hom(G_2, F)$ for all graphs F, then $G_1 \cong G_2$.

This should be reminiscent of **Yoneda's lemma** from Category Theory. Graphs, much like categories, can be identified by the collection of morphisms they induce.

[5] As a reminder, by "moments" we mean "the homomorphism densities $t(F, W)$ for a collection of graphs F".

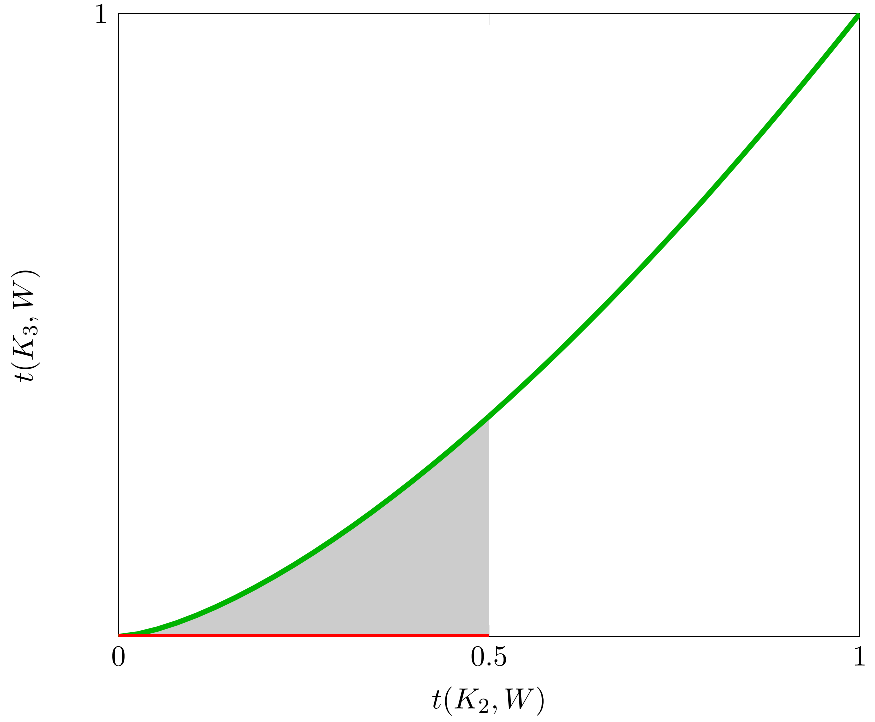


FIGURE 2. A partial illustration of the feasible K_2 and K_3 densities. Red follows from Mantel's Theorem, while green follows from Kruskal and Katona's result. fig:k2k3-density-plot

§5.4.2 Extremal Graph Theory, Revisited. Another question we could ask is, "given the edge density, what are the possible triangle densities?" Stated more rigorously, consider the set of pairs of K_2 and K_3-densities; i.e.

$$\{(t(K_2, W), t(K_3, W)) : \text{graphons } W\} \subset [0, 1]^2.$$

This is a compact set, but what does it look like? We have already covered a partial result! By re-phrasing Mantel's Theorem as

$$t(K_3, W) = 0 \implies t(K_2, W) \leq \frac{1}{2},$$

this now describes one part of the region's boundary, colored in red in Figure fig:k2k3-density-plot 2. The green boundary is the curve $y = x^{3/2}$, which comes from the following result:

prop:kruskal-katona **Proposition 5.52** (Kruskal-Katona). *For all graphons W, $t(K_3, W) \leq t(K_2, W)^{3/2}$.*

This bound is tight. Consider the "worst-case" graphon W that evaluates to 1 in the region $[0,a]^2$, and 0 everywhere else. This gives $t(K_2, W) = a^2$ and $t(K_3, W) = a^3$.

We may also vertically fill in the region in-between the two boundaries. To see why, consider two graphons U, W such that $\int U = \int W$. As t ranges from 0 to 1, $t(K_3, (1-t)W + tW)$ moves continuously from $t(K_3, U)$ to $t(K_3, W)$. (This gives the grey shaded area bounded by the green and red curves.)

We will give **two proofs** of Proposition 5.52.

First proof of Prop. 5.52. It suffices to show $t(K_3, G) \leq t(K_2, G)^{3/2}$ for graphs G, instead of graphons W. In particular, note that this inequality holds if and only if $\hom(K_3, G) \leq \hom(K_2, G)^{3/2}$ since the normalization factors cancel.

Let G be a graph on n vertices, and $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $A = A(G)$. Recall that the number of labelled triangles is equal to $tr(A^3)$, so we may derive

$$\begin{aligned} \hom(K_3, G) &= \sum_{i=1}^{n} \lambda_i^3 \\ &\leq \left(\sum \lambda_i^2\right)^{3/2} \\ &= \hom(K_2, G)^{3/2} \end{aligned}$$

as desired. The inequality in the second line comes from the trivial bound $a^t + b^t \leq (a+b)^t$, which holds for all $t > 1$ and positive a, b.
Note: this proof can also be directly be phrased in terms of the spectrum of graphon W, viewed as a compact operator on $L^2([0,1])$. □

Philosophically, the upcoming second proof should “feel" more natural, because we are not relying on tricks to view properties of functions (graphons) in terms of its spectra. As a result, this latter technique generalizes more easily.

Second proof of Prop. 5.52. We will show the stronger statement

$$t(K_3, W) \leq t(K_2, W^2)^{3/2}$$

where W^2 is the pointwise squaring of W. To this end, we will show a classical result of Loomis and Whitney, from which the desired inequality follows by taking $f = g = h = W$.

Lemma 5.53 (Loomis-Whitney). *Suppose that $f, g, h : \mathbb{R}^2 \to [0,\infty)$ are in $L^2(\mathbb{R}^2)$. Then*

$$\int f(x,y)g(x,z)h(y,z)dxdydz \leq \left(\int f^2\right)^{1/2}\left(\int g^2\right)^{1/2}\left(\int h^2\right)^{1/2}$$

Intuitively, we may think of this lemma as describing the volume of some region $A \subset \mathbb{R}^3$ in terms of its projections A_{xy}, A_{yz}, A_{xz} onto the three coordinate planes: $|A| \leq (|A_{xy}||A_{yz}||A_{xz}|)^{1/2}$.

Proof of Loomis-Whitney. This is a straightforward threefold application of Cauchy-Schwarz, once w.r.t. dx, then dy then finally dz:

$$\begin{aligned}
&\int f(x,y)g(x,z)h(y,z)dxdydz \\
&\leq \int \left(\int f(x,y)^2dx\right)^{1/2}\left(\int g(x,z)^2dx\right)^{1/2} h(y,z)dydz \\
&\leq \int \left(\int f(x,y)^2dxdy\right)^{1/2}\left(\int g(x,z)^2dx\right)^{1/2}\left(\int h(y,z)^2dy\right)^{1/2} dz \\
&\leq \left(\int f(x,y)^2dxdy\right)^{1/2}\left(\int g(x,z)^2dxdz\right)^{1/2}\left(\int h(y,z)^2dydz\right)^{1/2}
\end{aligned}$$

□

□

Theorem 5.54. *Fix $c_1, c_2, \ldots, c_k \in \mathbb{R}$. The inequality*

$$\sum_{r=1}^{k} c_r t(K_r, G) \geq 0 \tag{4}$$

eq:inequality-clique-convexhull

holds for all G if and only if it holds for all cliques $G = K_n$ $(n \geq 1)$.

Remark 5.55. Note that the latter statement is equivalent to

$$\sum_{r=1}^{k} c_r \frac{n(n-1)\cdots(n-r+1)}{n^r} \geq 0 \quad \forall\, n \geq 1.$$

Proof. The forward direction is trivial. For the reverse, we will use Zykov Symmetrization, as in the proof of Turán's theorem (Section sec:extremal-turan 2.2).

Restrict our attention to vertex-weighted simple graphs, using the insight that providing a proof using graphons is sufficient. To be more precise, we will consider graphs whose vertex weights are $\alpha_1, \ldots, \alpha_n$ that sum to 1. Each such graph corresponds to a $\{0,1\}$-valued block graphon with interval sizes $\alpha_1, \ldots, \alpha_n$. Figure 3 provides an example.

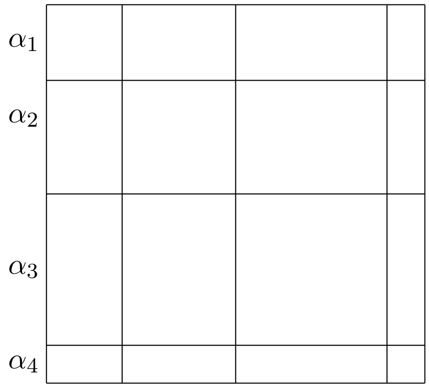


FIGURE 3. Example of a block graphon, interpreted as a vertex-weighted graph.

fig:vtx-weighted-example

Suppose the inequality (Equation 4) fails; our goal is now to show that it fails for some complete graph K_n. Consider the set of all vertex-weighted graphs with the minimum number of vertices that witness the failure; among all such graphs, we may pick the graph H that minimizes

$$f(H) = \sum_{r=1}^{k} c_r t(K_r, H)$$

which is a symmetric polynomial in $\alpha_1, \ldots, \alpha_n$. Observe that $\alpha_i > 0$ for all i, since otherwise we may get rid of the vertex whose weight is zero to obtain a graph with one fewer vertex that still violates the inequality.

We claim that H is a complete graph. To see why, assume otherwise: then there exists $ij \notin E(H)$, which means $f(H)$ is multilinear in α_i and α_j if all other weights are held fixed. Thus, we may shift α_i and α_j, holding their sum constant, so that $f(H)$ never increases and one of α_i, α_j becomes 0, which is a contradiction (by the same observation we made about graphs with $\alpha_i = 0$ for some i).

Now, choose two of the weights, say α_1, α_2, and fix all other values $\alpha_3, \ldots, \alpha_n$. This gives us $f(H) = A + B(\alpha_1 + \alpha_2) + C\alpha_1\alpha_2$ for constants A, B, C. We may additionally simplify this into $f(H) = A' + C\alpha_1\alpha_2$ by noting that $\alpha_1 + \alpha_2$ is also constant. But such a function is minimized by having

- α_1 or $\alpha_2 = 0$, which is not possible, or
- $\alpha_1 = \alpha_2$.

This tells us that all α_i's are equal, which devolves into the symmetric complete graph K_n. □

cor:convex-hull-Kn **Corollary 5.56.** *For each m, the extreme points of the convex hull of*

$$\{(t(K_2, W), \dots, t(K_m, W)) : \textit{graphons } W\}$$

are

$$\{(t(K_2, K_n), \dots, t(K_m, K_n)) : n \geq 1\} \cup \{(1, \dots, 1)\}.$$

(*Note: The extra point* $(1, \dots, 1)$ *comes from the limit as* $n \to \infty$.)

Intuitively, the result tells us that in order to optimize any linear combination of clique densities, subject to linear constraints on other clique densities, one simply has to check the inequalities at the cliques. With a bit of straightforward calculation, one can deduce Turán's theorem, which can be formulated as

$$t(K_{r+1}, W) = 0 \implies t(K_2, W) \leq 1 - \tfrac{1}{r}.$$

It also allows us to begin to draw a more complete picture of set of possible edge and triangle densities (Figure 4). We may plug in $m = 3$ to infer that the convex hull of the region has extreme points

$$\left\{\left(1 - \frac{1}{n}, \frac{(n-1)(n-2)}{n^2}\right) : n \geq 1\right\} \cup \{(1, 1)\}.$$

All of these points lie on the parabola $y = x(2x - 1)$ (the fact that the region lies above this parabola was a problem on your first problem set).

The curves that bound from below in the regime $t(K_2, W) > 0.5$ is a sequence of cubic curves, due to Razborov's theorem (Theorem 5.57), which we state without proof.

th:razborov-hom-density **Theorem 5.57** (Razborov). *For fixed $t(K_2, W) \in [1 - \frac{1}{k-1}, 1 - \frac{1}{k}]$, the min of $t(K_3, W)$ is attained by a complete graph with node weights $\alpha_1, \dots, \alpha_k$, where $\alpha_1 = \alpha_2 = \cdots = \alpha_{k-1} \geq \alpha_k$.*

Remark 5.58. Unless $t(K_2, W) = 1 - \frac{1}{k}$, such a graph is not unique, since we may replace a $0 - 1$ triangle-free graphon with any other triangle-free graphon, making sure it is distinct by scaling it by some $\beta > 0$.

Lec17: Brynmor Chapman and Jerry Li

Razborov's bound is very recent, and quite technical. This result has been generalized to K_2 vs K_4 by Nikiforov. Reiher later extended it to K_2 vs K_r for any r.

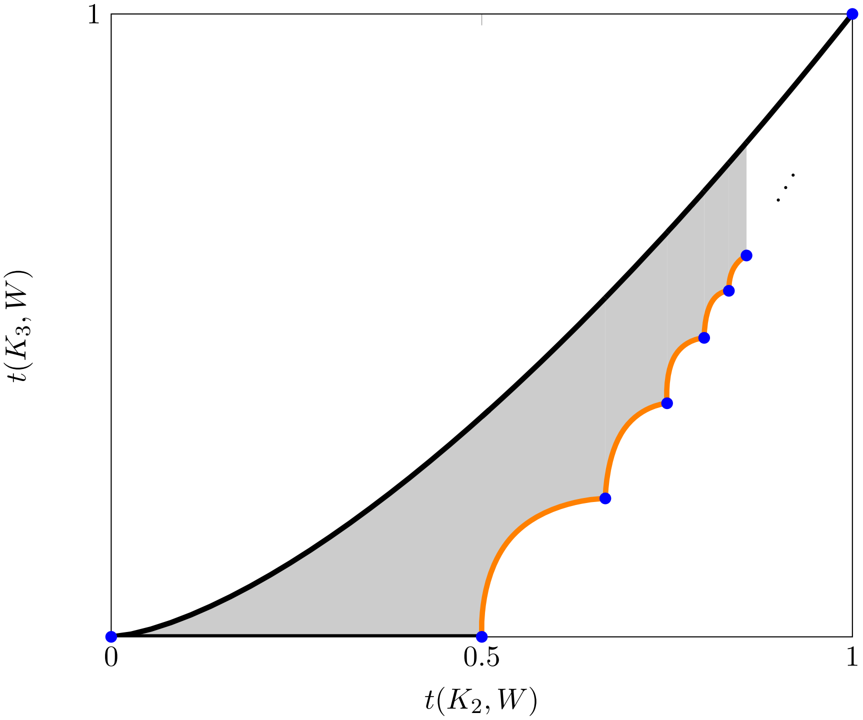


FIGURE 4. A more complete illustration of the feasible K_2 and K_3 densities. Blue shows the corners of the convex hull given by Corollary 5.56, while orange shows the actual boundary given by Razborov's theorem.

cor:convex-hull-Kn

fig:k2k3-density-plot-complete

Consider the following question: How many graphs on n vertices have (approximately) a given (edge, triangle) density pair? For any (σ, τ), define $A(\sigma)$ to be

$$A(\sigma,\tau) = \lim_{\varepsilon_1,\varepsilon_2 \to 0} \lim_{n\to\infty} \# \frac{1}{\binom{n}{2}} \log_2 \{G : |V(G)| = n, |t(K_2, W) - \sigma| < \varepsilon_1, |t(K_3, W) - \tau| < \varepsilon_2\} \,.$$

Recall, the homework problem said that the number of triangle free graphs is roughly $2^{(1+o(1))\frac{n^2}{4}}$. This corresponds calculating this quantity on a specific point on the curve. We wish to have a more general characterization of this quantity. We require the following definition:

Definition 5.59 (Binary entropy). For any $x \in [0, 1]$, let $H(x) = x \log_2 x - (1 - x) \log_2(1 - x)$.

With this in mind, we now state our main bound on $A(\sigma, \tau)$:

Theorem 5.60. *For any σ, τ, we have:*

$$A(\sigma,\tau) = \sup_W \left\{ \int H(W(x,y))dxdy : \text{graphon } W \text{ with } t(K_2,W)=\sigma, t(K_3,W)=\tau \right\} . \tag{5}$$

eq:graphon-density

We pause to make a couple of remarks about this statement. Observe that for a fixed $x \in [0,1]$, we have

$$\binom{N}{xN} = 2^{N(H(x)+o(1))} , \qquad \text{as } N \to \infty ,$$

where we interpret the LHS approximately by rounding xN if necessary. Hence the above statement qualitatively says that the exponent in $A(\sigma,\tau)$ is highly controlled by a single graphon.

By compactness, the supremum is always attained by some graphon.

This is very general. In particular, the equality holds for any graphs, not just K_2, K_3. This is also related to a conjecture of Sidorenko:

Conjecture 5.61 (Sidorenko's conjecture)**.** *If H is bipartite, W graphon, then* $t(H,W) \geq t(K_2,W)^{|E(H)|}$.

We remark we have already seen this for $H = C_4$. In the problem set we will show this for some others. However, in general, we do not know whether this holds for all H. One concrete example we do not know this for is $H = K_{5,5} \setminus C_{10}$. This graph has a name: it is called the Möbius strip. This is because corresponds to the vertex-face incidence graph of the minimal simplicial complex of the Möbius strip.

One reason why these theorems are so hard is because they are in general undecidable:

thm:hatami-norime **Theorem 5.62** (Hatami-Norime)**.** *It is algorithmically undecidable whether an inequality of the form*

$$\sum_{i=1}^{k} c_i t(H_i, G) \geq 0 \tag{6}$$

eq:lin-eq-graphons

holds for all graphs G.

Observe this holds in direct contrast to the case when the H_i's are all cliques, as we saw last time. This is deduced from Matiyasevich's theorem, which proved the undecidablity of Hilbert's 10th problem. The proof of Theorem 5.62 follows by cleverly encoding Diophantine equations in these inequalities at very specific points. The rough idea is to use the vertices of the region in Figure 4 to encode integer solutions to diophantine equations/inequalities.

On the other hand we have:

Theorem 5.63. *There is an algorithm so that, for every $\varepsilon > 0$, decides that either*

$$\sum_{i=1}^{k} c_i t(H_i, G) \geq -\varepsilon \ , \tag{7}$$

eq:lin-eq-graphons-approx

or produces a G for which this inequality fails.

Proof. Apply the weak regularity lemma with some $\varepsilon' \ll \varepsilon$. This implies that we only need to check this for all graphons which are step graphons with side length ε'. Then check the possibilities. If (6) holds for all such graphons, then by the counting lemma, then (7) holds for all graphs, by the counting lemma. Otherwise we have produced a counterexample. □

CHAPTER 6

Roth's theorem

Let $r_3([N])$ denote the maximum size subset of $[N]$ that does not have a 3-AP. Recall Roth's theorem states that $r_3([N]) = o(N)$. We proved this in class using the triangle removal lemma, but this is not how Roth originally proved it.

We now note the history of this theorem. The original proof showed:

Theorem 6.1 (Roth'53). *$r_3([N]) = O(N/\log\log N)$.*

In the 90's, it was shown that:

Theorem 6.2 (Szemeredi'90, Heath-Brown'87). *$r_3([N]) = O(N/\log^c N)$ for some constant $c > 0$.*

In work of Bourgain in '99 and '08 and later Sanders in '12, c was improved to $1/2 + o(1)$ then $2/3 + o(1)$ and $3/4 + o(1)$. Finally, Sanders '11 showed:

Theorem 6.3 (Sanders'11). *$r_3([N]) = O(N(\log\log N)^6/\log N)$.*

Later Bloom '16 improved this to $O(N(\log\log N)^4/\log N)$. It is a major open problem whether we can improve this beyond $O(N/\log N)$. This is related to a famous conjecture of Erdős:

Conjecture 6.4 (Erdős). *Suppose $A \subseteq N$ satisfies $\sum_{a\in A} \frac{1}{a} = \infty$. Then A contains arithmetic progressions of arbitrary size.*

Observe that the hypothesis roughly corresponds to A having density $O(N/\log N)$.

There is a finite field analog of Roth's theorem. Define $r_3(\mathbb{F}_3^n)$ to be the size of the largest subset S of $\mathbb{F}_3^n$ without a solution to $x - 2y + z = 0$ for distinct $x, y, z \in S$. (Note that this is equivalent to $x+y+z = 0$ in characteristic 3.) Brown and Buhler showed in '84 that $r_3(\mathbb{F}_3^n) \in o(3^n)$. Meshulam '95 then showed that $r_3(\mathbb{F}_3^n) \in O(3^n/n)$. This was improved by Bateman-Katz '12, which showed that in fact $r_3(\mathbb{F}_3^n) \in O(3^n/n^{1+c})$ for some $c > 0$. There was then a significant breakthrough of Croot-Lev-Pach '17 which showed that for a slightly different problem, i.e. replacing $\mathbb{F}_3^n$ with $(\mathbb{Z}/4\mathbb{Z})^n$, there was an exponential improvement $r_3((\mathbb{Z}/4\mathbb{Z})^n) \leq 4^{0.926n}$. A week later Ellenberg-Gijswijt showed that this implied that $r_3(\mathbb{F}_3^n) \leq (2.756)^n$.

§6.1 Roth's theorem in $\mathbb{F}_3^n$

thm:meshulam **Theorem 6.5** ([30]). *If $A \subset \mathbb{F}_3^n$ has no 3-APs, then $|A| = O(3^n/n)$.*

The key advantage of this model is that this space has subspaces. The key idea is the following dichotomy: either the associated function is Fourier uniform, which by a counting lemma will imply there are many 3-APs, or there is some large Fourier coefficient γ. Then, looking at $V = \gamma^{\perp}$, the codimension 1 subspace of this coefficient, our density should increment in some coset of V. Since codimension 1 subspaces of $\mathbb{F}_3^n$ are just $\mathbb{F}_3^{n-1}$, we can iterate this argument.

Let us recall some notation on Fourier transforms on finite abelian groups. Let $f : \Gamma \to \mathbb{C}$, and let $\gamma \in \hat{\Gamma}$. Recall $\hat{f}(\gamma) = \langle f, \gamma \rangle = \mathbb{E}_{x \in \Gamma} f(x)\overline{\gamma(x)}$. In particular, when $\Gamma = \mathbb{F}_3^n$, we have $\hat{\Gamma} \cong \mathbb{F}_3^n$, and for any $\gamma \in \mathbb{F}_3^n$, we have

$$\hat{f}(\gamma) = \mathbb{E}_{x \in \mathbb{F}_3^n} f(x)\omega^{-\gamma \cdot x} ,$$

where ω is a primitive 3rd root of unity. The key identity that allows us to relate Fourier analysis to arithmetic progressions is the following:

Theorem 6.6. *Let $f, g, h : \Gamma \to \mathbb{C}$. Let*

$$\Lambda(f, g, h) = \mathbb{E}_{x,y \in \Gamma} f(x)g(x + y)h(x + 2y) .$$

Then

$$\Lambda(f, g, h) = \sum_{\gamma \in \hat{\Gamma}} \hat{f}(\gamma)\hat{g}(\gamma^{-2})\hat{h}(\gamma) .$$

Proof. We will use the Fourier inversion formula, which says that $f(x) = \sum_{\gamma\in\hat{\Gamma}} \hat{f}(\gamma)\gamma(x)$. Expand the LHS using this formula:

$$\begin{aligned}
\Lambda(f,g,h) &= \mathbb{E}_{x,y\in\Gamma} f(x)g(x+y)h(x+2y) \\
&= \mathop{\mathbb{E}}_{x,y}\left(\sum_{\gamma_1}\hat{f}(\gamma_1)\gamma_1(x)\right)\left(\sum_{\gamma_2}\hat{g}(\gamma_2)\gamma_2(x+y)\right)\left(\sum_{\gamma_3}\hat{h}(\gamma_3)\gamma_3(x+2y)\right) \\
&= \sum_{\gamma_1,\gamma_2,\gamma_3\in\hat{\Gamma}} \hat{f}(\gamma_1)\hat{g}(\gamma_2)\hat{h}(\gamma_3)\mathop{\mathbb{E}}_{x,y}\gamma_1(x)\gamma_2(x)\gamma_2(y)\gamma_3(x)\gamma_3(y)^2 \\
&= \sum_{\gamma_1,\gamma_2,\gamma_3\in\hat{\Gamma}} \hat{f}(\gamma_1)\hat{g}(\gamma_2)\hat{h}(\gamma_3)\left(\mathop{\mathbb{E}}_{x}(\gamma_1\gamma_2\gamma_3)(x)\right)\left(\mathop{\mathbb{E}}_{y}(\gamma_2\gamma_3^2)(y)\right)
\end{aligned}$$

Recall that the expectation of a character γ is 1 if $\gamma \equiv 1$ and 0 otherwise. Hence all terms in the final sum vanish except those for which $\gamma_1\gamma_2\gamma_3 \equiv 1$ and $\gamma_2\gamma_3^2 \equiv 1$. This gives:

$$\sum_{\gamma_1,\gamma_2,\gamma_3\in\hat{\Gamma}} \hat{f}(\gamma_1)\hat{g}(\gamma_2)\hat{h}(\gamma_3)\left(\mathop{\mathbb{E}}_{x}(\gamma_1\gamma_2\gamma_3)(x)\right)\left(\mathop{\mathbb{E}}_{y}(\gamma_2\gamma_3^2)(y)\right) = \sum_{\gamma\in\hat{\Gamma}}\hat{f}(\gamma)\hat{g}(\gamma^{-2})\hat{h}(\gamma)$$

This completes the proof.

□

For $\Gamma = \mathbb{F}_3^n$, observe that $x - 2y + z = x + y + z$. So we are just trying to count

$$\frac{1}{3^{2n}}\left|\{x+y+z=0, x,y,z\in A\}\right| = \mathbb{E}_{x,d}1_A(x)1_A(x+d)1_A(x+2d)\ .$$

By applying the identity, we have

$$\mathbb{E}_{x,d}1_A(x)1_A(x+d)1_A(x+2d) = \sum_{\gamma\in\mathbb{F}_3^n}\hat{1}_A(\gamma)^2\hat{1}_A(-2\gamma) = \sum_{\gamma}\hat{1}_A(\gamma)^3\ .$$

If $A \subset \mathbb{F}_3^n$ is 3-AP free, let $|A| = \alpha N$ where $N = 3^n$. Recall that the first step is to show that if there are few 3-APs, then there is a large Fourier coefficient. Recall we need to count trivial 3-APs, so if A is 3-AP free, then the number of 3-APs is exactly $|A| = \alpha N$. Since $\alpha = \hat{1}_A(0)$, by our identity we have

$$\begin{aligned}\frac{1}{N^2}\left|\{x+y+z=0\}\right| &= \sum_{\gamma}\hat{1}_A(\gamma)^3 = \alpha^3 + \sum_{\gamma\neq 0}\hat{1}_A(\gamma)^3\\ &\geq \alpha^3 - \max_{\gamma\neq 0}|\hat{1}_A(\gamma)|\sum_{\gamma}\hat{1}_A(\gamma)^2\\ &= \alpha^3 - \max_{\gamma\neq 0}|\hat{1}_A(\gamma)|\,\mathop{\mathbb{E}}_{x} 1_A(x)^2\\ &= \alpha^3 - \max_{\gamma\neq 0}|\hat{1}_A(\gamma)|\alpha\ .\end{aligned}$$

In other words, if A is 3-AP free, then there exists some $\gamma \neq 0$ so that

$$|\hat{1}_A(\gamma)| \geq \alpha^2 - \frac{1}{N} \geq \frac{\alpha^2}{2}$$

provided that $N \geq \frac{2}{\alpha^2}$.

Now recall the second step is that such a Fourier coefficient should imply a density increment on a codimension 1 subspace. This is formalized in the following lemma:

Lec18: Jonathan Tidor and Nicole Wein

Lemma 6.7. *Let $A \subseteq \mathbb{F}_3^n$ with $|A| = \alpha N$. Let $\gamma \neq 0$ with $|\hat{1}_A(\gamma)| \geq \delta$. Then A has density $\geq \alpha + \frac{\delta}{2}$ on a codimension-1 subspace.*

Proof. By definition, we have

$$\begin{aligned}\hat{1}_A(\gamma) &= \mathbb{E}_{x\in\mathbb{F}_3^N} 1_A(x)\omega^{-\gamma\cdot x}\\ &= \frac{1}{3}\left(\alpha_0 + \alpha_1\omega + \alpha_2\omega^2\right)\end{aligned}$$

where $\alpha_0, \alpha_1, \alpha_2$ are the densities of A on the three cosets of $V = r^{\perp}$. Recall $\frac{1}{3}(\alpha_0 + \alpha_1 + \alpha_2) = \alpha$. Hence

$$\begin{aligned}3\delta &\leq |\alpha_0 + \alpha_1\omega + \alpha_2\omega^2|\\ &= |(\alpha_0 - \alpha) + (\alpha_1 - \alpha)\omega + (\alpha_2 + \alpha)\omega^2|\\ &\leq \sum_{j=0}^{2}|\alpha_j - \alpha|\\ &= \sum_{j=0}^{2}|\alpha_j - \alpha| + (\alpha_j - \alpha)\ .\end{aligned}$$

Since every term in the sum is nonnegative, this implies that there exists j so that $|\alpha_j - \alpha| + (\alpha_j - \alpha) \geq \delta$ which implies $\alpha_j \geq \alpha + \delta/2$. □

thm:meshulam

Proof of Theorem 6.5. Combining the arguments above, we have shown that if $A \subseteq \mathbb{F}_3^n$ with no 3-APs, writing $N = 3^n$ and $|A| = \alpha N$, if $N \geq 2/\alpha^2$ then $\mathbb{F}_3^n$ contains a coset of a codimension 1 subspace, $V + v$, such that $|A \cap (V + v)| \geq (\alpha + \alpha^2/4)|V + v|$.

Write $A' = A \cap (V + v)$. Let $A_1 = A' - v = \{a - v : a \in A'\}$. Note that $A_1 \subseteq V \cong \mathbb{F}_3^{n-1}$ has no 3-APs since if $a, b, c \in A'$ with $(a - v) + (b - v) + (c - v) = 0$, then $a + b + c = 0$ (since we are working over $\mathbb{F}_3$ where $3v = 0$), contradicting the fact that A is 3-AP-free.

Therefore starting with $A = A_0 \subseteq \mathbb{F}_3^n$ with density $\alpha = \alpha_0$, we have produced $A_1 \subseteq \mathbb{F}_3^{n-1}$ with density $\alpha_1 \geq \alpha_0 + \alpha_0^2/4$. Furthermore, A_1 remains 3-AP-free so we can iterate this argument, producing a sequence $A_0, A_1, A_2, \ldots$ with $A_i \subseteq \mathbb{F}_3^{n-i}$ with densities α_i satisfying $\alpha_{i+1} \geq \alpha_i + \alpha_i^2/4$. This process continues as long as $3^{n-i} \geq 2/\alpha_i^2$. In addition, $\alpha_i \leq 1$ since this quantity is a density, so this process cannot continue forever. Careful analysis will give a precise bound on $|A| = \alpha N$.

Since $\alpha_{i+1} \geq \alpha_i(1 + \alpha_i/4)$ then α_i doubles in at most $\lceil 4/\alpha_i \rceil$ steps, i.e., $\alpha_{i+\lceil 4/\alpha_i \rceil} \geq 2\alpha_i$. Then since the α_i's are increasing, they start doubling faster. Let $i_j = \lceil 4/\alpha \rceil + \lceil 2/\alpha \rceil + \cdots + \lceil 2^{3-j}/\alpha \rceil$. Then $\alpha_{i_j} \geq 2^j \alpha$. In particular, this process must halt after at most $i_{\log_2(1/\alpha)}$ steps. Write $k = i_{\log_2(1/\alpha)} \leq 8/\alpha + \log_2(1/\alpha)$.

When the process halts, this means that $3^{n-k} < 2/\alpha_k^2 < 2/\alpha^2$. Combining these facts, we know that

$$n < \log_3(2/\alpha^2) + k \leq 8/\alpha + \log_2(1/\alpha) + \log_3(2/\alpha^2) = O(1/\alpha).$$

This implies that $\alpha = O(1/n)$, so $|A| = \alpha N = O(3^n/n)$, as desired. □

Remark 6.8. If we were being sloppy, we could use the bound $\alpha_i \geq i\alpha^2/4$, which implies that the process must stop after $4/\alpha^2$ steps. This leads to the worse bound $|A| = O(3^n/\sqrt{n})$, so careful analysis of this process is necessary to get the correct bounds. If we were even more careful with the argument and the analysis of it, we could recover Meshulam's original bound of $r_3(\mathbb{F}_3^n) \leq 2 \cdot 3^n/n$.

§6.2 Roth's proof of Roth's theorem

Theorem 6.9 ([40]). *$r_3([N]) = O(N/\log\log N)$.*

The proof will have the same three steps as in $\mathbb{F}_3^n$, but all the steps will be slightly trickier:

(1) few 3-APs implies large Fourier coefficient assuming N is large;
(2) large Fourier coefficient implies density increment on a fairly large arithmetic sub-progression;
(3) iterate the argument on the sub-progression.

The proof uses Fourier analysis on $\mathbb{Z}$. The group of characters $\hat{\mathbb{Z}} \cong \mathbb{R}/\mathbb{Z}$ is not discrete so there will be some integrals but everything works out basically the same as for finite abelian groups.

For $t \in \mathbb{R}/\mathbb{Z}$, define $e(t)\colon \mathbb{Z} \to \mathbb{C}$ by $e(t) = e^{2\pi i t}$. For simplicity we only consider functions $f\colon \mathbb{Z} \to \mathbb{C}$ that are non-zero at finitely-many points. Then $\hat{f}\colon \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ is defined by

$$\hat{f}(\theta) = \sum_{n\in\mathbb{Z}} f(n)e(-n\theta).$$

(Note that by assumption this sum only contains finitely-many non-zero terms. This means we can ignore some annoying details involving integrals of infinite sums.)

We will use the following fact which relates the Fourier coefficients of f to the number of weighted 3-APs that f has. Define

$$T(f,g,h) = \sum_{x,y\in\mathbb{Z}} f(x)g(x+y)h(x+2y).$$

Then

$$T(f,g,h) = \int_{\mathbb{R}/\mathbb{Z}} \hat{f}(\theta)\hat{g}(-2\theta)\hat{h}(\theta)\,d\theta.$$

We write $T(f)$ for $T(f,f,f)$.

Remark 6.10. It is possible to do this proof without using Fourier analysis on $\mathbb{Z}$ by instead embedding $[N]$ into $\mathbb{Z}/(2N+1)\mathbb{Z}$. It does not make very much difference at the end, though doing Fourier analysis in $\mathbb{Z}$ feels more natural for this problem.

§6.2.1 Large Fourier coefficient.

The following proposition should feel similar to the counting lemma. It says that if two functions are similar then they have a similar number of 3-APs.

thm:integer-counting **Lemma 6.11.** *If $f,g\colon \mathbb{Z}\to\mathbb{C}$ satisfy $\sum_{n\in\mathbb{Z}}|f(n)|^2, \sum_{n\in\mathbb{Z}}|g(n)|^2 \le M$, then*

$$|T(f)-T(g)| \le 3M\|\widehat{f-g}\|_\infty.$$

Proof. $T(f) - T(g)$ telescopes as

$$T(f,f,f) - T(g,g,g) = T(f-g,f,f) + T(g,f-g,f) + T(g,g,f-g).$$

Each term is bounded by $M\|\widehat{f-g}\|_\infty$ in absolute value. We prove this for the first term.

$$\begin{aligned} |T(f-g,f,f)| &= \left| \int_0^1 \widehat{f-g}(\theta)\hat{f}(-2\theta)\hat{f}(\theta) \right| \, d\theta \\ &\le \|\widehat{f-g}\|_\infty | \int_0^1 |\hat{f}(-2\theta)\hat{f}(\theta)| \, d\theta \\ &\le \|\widehat{f-g}\|_\infty \left(\int_0^1 |\hat{f}(-2\theta)|^2 \, d\theta \right)^{1/2} \left(\int_0^1 |\hat{f}(\theta)|^2 \, d\theta \right)^{1/2} \\ & \qquad \text{(Cauchy-Schwarz)} \\ &= \|\widehat{f-g}\|_\infty \int_0^1 |\hat{f}(\theta)|^2 \, d\theta \\ &= \|\widehat{f-g}\|_\infty \sum_{n\in\mathbb{Z}} |f(n)|^2 \qquad \text{(Parseval)} \\ &\le M\|\widehat{f-g}\|_\infty. \end{aligned}$$

□

Now if $A \subseteq [N]$ has no 3-APs with $|A| = \alpha N$, then $T(1_A)$ only counts trivial 3-APs, so $T(1_A) = \alpha N$. However $T(\alpha 1_{[N]}) \ge \alpha^3 N^2/2$, i.e., is large. By Lemma 6.11, this implies that $1_A - \alpha 1_{[N]}$ must have a large Fourier coefficient.

In particular, we apply Lemma 6.11 with $M = \alpha N$, implying that

$$\|\widehat{1_A - \alpha 1_{[n]}}\|_\infty \ge \frac{(\alpha^3 N^2/2 - \alpha N)}{3\alpha N} = \alpha^2 N/6 - 1/3.$$

This is at least $\alpha^2 N/10$ if $N \ge 5/\alpha^2$.

Therefore we have shown that if $N \ge 5/\alpha^2$, then there exists $\theta \in \mathbb{R}$ such that

$$\left| \sum_{n=1}^N (1_A - \alpha)(n) e(\theta n) \right| \ge \frac{\alpha^2}{10} N.$$

§6.2.2 Density increment. The following lemma says that a large Fourier coefficient implies a density increment on a large arithmetic sub-progression.

lem:2 **Lemma 6.12.** *If $|\hat{1}_A(r)| \ge \delta$ for some $r \neq 0$, then A has density at least $\alpha + \frac{\delta}{2}$ on a codimension 1 coset.*

The idea of the proof of Lemma 6.12 is as follows. Previously we saw that in the finite field setting, a large Fourier coefficient divides up the space into three cosets where the character is constant on each coset. Here, we don't have the property that the character is constant, but we will be able to find arithmetic sub-progressions such that the character is approximately constant. We will partition $[N]$ into arithmetic progressions of length about $N^{1/3}$ such that $e(n\theta)$ is approximately constant on each progression. To get intuition for why this is possible, consider how $\hat{f}(\theta)$ behaves. It jumps around the unit circle and when θ is close to a fraction with small denominator, it is almost periodic which means that $e(n\theta)$ is approximately constant. Thus, we first approximate θ by a fraction with small denominator using the following lemma.

lem:dir **Lemma 6.13 (Dirichlet).** *Let $\theta \in \mathbb{R}$ and $0 < \delta < 1$. Then there exists a positive integer $d \leq \frac{1}{\delta}$ such that $||d\theta||_{\mathbb{R}/\mathbb{Z}} \leq \delta$ where $||x||_{\mathbb{R}/\mathbb{Z}}$ denotes distance from x to the nearest integer.*

Proof. Let $m = \lfloor \frac{1}{\delta} \rfloor$ and consider $0, \theta, 2\theta, \ldots, m\theta$. By the pigeonhole principle, some pair $i\theta$, $j\theta$ differ by at most δ in their fractional parts. Take $d = |i - j|$. Then $||d\theta||_{\mathbb{R}/\mathbb{Z}} \leq \delta$. □

Now we show how to partition $[N]$ into sub-progressions so that the character is approximately constant on each sub-progression.

lem:pro **Lemma 6.14.** *Let $0 \leq \eta \leq 1$ and $\theta \in \mathbb{R}$. Suppose $N > \frac{C}{\eta^6}$. Then it is possible to partition $[N]$ into sub-progressions P_i of length at least $N^{1/3}$ such that $\sup_{x,x' \in P_i} |e(\theta x) - e(\theta x')| \leq \eta$ for all i.*

Proof. The idea is that if you have an arithmetic progression whose common difference induces a character very close to 1 then the character won't change much on this progression. If P is an arithmetic progression with common different d then by the triangle inequality

$$\begin{aligned}\sup_{x,x' \in P} |e(\theta x) - e(\theta x')| &\leq |P| \cdot |e(\theta d) - 1| \\ &\leq |P| \cdot 2\pi \cdot ||d\theta||_{\mathbb{R}/\mathbb{Z}}\end{aligned}$$

To see the last inequality, notice that $|e(t) - 1|$ is a chord length and $2\pi ||t||_{\mathbb{R}/\mathbb{Z}}$ is the corresponding arc length.

We'll form pretty short arithmetic sub-progressions whose common differences are close to a fraction, that is, $d\theta$ is close to an integer. That is, if $N^{1/3} \leq |P| \leq 2N^{1/3}$ and $||d\theta||_{\mathbb{R}/\mathbb{Z}} \leq \frac{\eta}{2\pi|P_i|}$, then we're done. Applying Lemma 6.13 with $N \geq \frac{C}{\eta^6}$, we can find $d \leq \frac{4\pi N^{1/3}}{\eta} \leq \sqrt{N}$ such that $||d\theta||_{\mathbb{R}/\mathbb{Z}} < \frac{\eta}{4\pi N^{1/3}} \leq \frac{\eta}{2\pi|P|}$. Given such a d, we partition $[N]$ into sub-progressions of common difference d of lengths between $N^{1/3}$ and $2N^{1/3}$. □

Now we'll use the above lemmas to get a density increment, proving Lemma 6.12.

Proof of Lemma 6.12. Recall that there exists $\theta \in \mathbb{R}$ such that $|\sum_{x\in[N]}(1_A - \alpha)(x)e(\theta x)| \geq \frac{\alpha^2}{10}N$ with a partition of $[N]$ into arithmetic sub-progressions P_i as in Lemma 6.14. We will argue that one of these sub-progressions witnesses a density increment. The key point will be that for each of these sub-progressions the character deviates by at most η. We have,

$$\begin{aligned}
\frac{\alpha^2}{10}N &\leq \left|\sum_{x\in[N]}(1_A-\alpha)(x)e(\theta x)\right| \\
&\leq \sum_{i=1}^{k}\left|\sum_{x\in P_i}(1_A-\alpha)(x)e(\theta x)\right| \\
&\leq \sum_{i=1}^{k}\left(\left|\sum_{x\in P_i}(1_A-\alpha)(x)\right| + \frac{\alpha^2}{20}|P_i|\right)
\end{aligned}$$

Note that the error term $\sum_{i=1}^{k}\frac{\alpha^2}{20}|P_i| = \frac{\alpha^2}{20}N$ is only half of our original bound $\frac{\alpha^2}{10}N$. The idea is that we have a lower bound on the average density deviation from α over all P_i and will use this to conclude that one of the sub-progressions must have much bigger density than α. In the finite field setting, we simply used the pigeonhole principle to show existence of such a sub-progression, but here naively applying the pigeonhole principle will not be strong enough. We bound the error term as follows.

$$\begin{aligned}
\sum_{i=1}^{k}\frac{\alpha^2}{20}|P_i| &\leq \sum_{i=1}^{k}\left|\sum_{x\in P_i}(1_A-\alpha)(x)\right| \\
&= \sum_{i=1}^{k}\left(\big||A\cap P_i| - \alpha|P_i|\big| + (|A\cap P_i| - \alpha|P_i|)\right)
\end{aligned}$$

In the last equality, we use the trick of adding the quantity to itself without the absolute value. We can do this because it averages to 0 and it is helpful because now each term is non-negative.

Now by the pigeonhole principle, there exists i such that $||A \cap P_i| - \alpha|P_i|| + (|A \cap P_i| - \alpha|P_i|) \geq \frac{\alpha^2}{20}|P_i|$. So $|A \cap P_i| \geq (\alpha + \frac{\alpha^2}{40})|P_i|$. Thus, if $A \subset [N]$ is 3-AP-free and $N > C\alpha^{-12}$ then $|A \cap P| \geq (\alpha + \frac{\alpha^2}{40})|P|$. for some arithmetic sub-progression P of size at least $N^{1/3}$. □

§6.2.3 **Iteration.** We now iterate the density increment. The recurrence is as follows. $N_0 = N$, $\alpha_0 = \alpha$, $P = [N]$, $N_{i+1} \geq N_1^{1/3}$, $\alpha_{i+1} \geq \alpha_i(1 + \frac{\alpha_i}{40})$, $|P_i| = N_i$. The density doubles after at most $\frac{40}{\alpha}$ steps and doubles again after at most $\frac{20}{\alpha}$ steps etc., resulting in a total of at most $\frac{80}{\alpha}$ steps. When the process stops it must stop because the condition on N is violated, that is, $N_i \leq C\alpha_i^{-12} \leq C\alpha^{-12}$. Thus, $N \leq (C\alpha^{-12})^{3^{80/\alpha}} = e^{e^{O(1/\alpha)}}$ so $\alpha = O(\frac{1}{\log\log N})$. This completes the proof of Roth's Theorem.

§6.2.4 **Comparison between Roth's Theorem for $\mathbb{F}_3^n$ and $[N]$.** In the case of $\mathbb{F}_3^n$, we got a bound on the order of $\frac{N}{\log N}$ where $N = 3^n$ while in the case of $[N]$ we got a bound on the order of $\frac{N}{\log\log N}$. This extra log factor comes from the fact that for $\mathbb{F}_3^n$ we restricted to a subspace of $1/3$ fraction in size, where as for $[N]$ we restricted to a much smaller progression of size $N^{1/3}$. This raises the question of whether sub-progressions are the "correct" analog of subspaces for $[N]$. It turns out they are not and instead the correct analog is something called a *Bohr set*. Bourgain introduced Bohr sets for improving bounds on Roth's theorem in $[N]$.

We define Bohr sets as an analog to subspaces. One way to specify a subspace is as follows. Given characters $r_1, r_2, \ldots, r_k \in \mathbb{F}_3^n$, a subspace is the set $\{x \in \mathbb{F}_3^n | x \cdot r_j = 0 \ \forall j\}$. Similarly, we can specify a Bohr set as follows. Given characters $\gamma_1, \gamma_2, \ldots, \gamma_k \in \hat{\Gamma}$, a Bohr set is the set $\{x \in \Gamma | |\arg \gamma_j(x)| < 2\pi\rho \ \forall j\}$. Unfortunately, Bohr sets are more difficult to work with than subspaces. One reason for this is that subspaces are preserved under addition, but the same is not true for Bohr sets. You should think of a Bohr set kind of like a ball of radius ρ. Addition two elements of such a ball together gives a point in the ball of radius 2ρ, not the original ball. Despite these difficulties Bohr sets can be used to greatly improve the bounds on $r_3([N])$.

Lec19: Kevin Sun and Meghal Gupta

§6.3 Other approaches to Roth's theorem over $\mathbb{F}_3^n$

We now return to the problem of giving bounds on the sizes of subsets $A \subset \mathbb{F}_3^n$ with no three-term arithmetic progression. The following better upper bound was found in 2012.

thm:bateman-katz **Theorem 6.15** (Bateman-Katz). *For $A \subset \mathbb{F}_3^n$ with no 3-AP's, we have $|A| = O(\frac{3^n}{n^{1+c}})$ for some $c > 0$.*

There has also been some progress on the converse problem, finding a large subset of $\mathbb{F}_3^n$ with no 3-AP's.

thm:edel **Theorem 6.16** (Edel). *There exists $A \subset \mathbb{F}_3^n$ with no 3-AP's such that $|A| \geq 2.217^n$ for all sufficiently large n.*

This lower bound was proven by taking a specific finite construction and tensoring it up to constructions for larger n.

However, more recently, a completely different method has improved the upper bound significantly. We show this proof found by Ellenberg and Gijswijt below.

thm:capset **Theorem 6.17.** *For $A \subset \mathbb{F}_3^n$ with no 3-AP's, we have $|A| = O(2.756^n)$.*

To prove this result, we use a general technique which is sometimes referred to as the polynomial method.

Suppose $A \subset \mathbb{F}_3^n$ is 3-AP free (i.e. no $x + y + z = 0$). We can encode this fact in a different way by noting that when A is 3-AP free, then for all $x, y, z \in A$,

$$\delta_0(x + y + z) = \sum_{a \in A} \delta_a(x)\delta_a(y)\delta_a(z)$$

where $\delta_a = 1$ if $x = a$ and 0 otherwise.

Now, we define the notion of "slice rank" for a function, and we will show that the left hand side of the above expression has "low slice rank" while the right side has "high slice rank".

Definition 6.18. Consider functions $F : A \times A \times A \ldots \times A \to \mathbb{F}$ (with k A's) for some field $\mathbb{F}$. We say that a function F has slice-rank 1 if F has the form $F(x_1, x_2, \ldots x_k) = f(x_j)g(x_1, x_2, \ldots x_{j-1}, x_{j+1}, \ldots x_{k-1}, x_k)$ (e.g. $f(x_1)g(x_2, x_3, \ldots x_k)$ or $f(x_2)g(x_1, x_3, x_4, \ldots x_k)$ or $f(x_3)g(x_1, x_2, x_4, x_5, \ldots x_k)$

etc.). The slice rank of a general function $F : A \times A \times A \ldots \times A \to \mathbb{F}$ is the minimum r such that F can be written as the sum of r slice-rank 1 functions.

Remark 6.19. Note that this is analogous to the usual notion of rank of a matrix, which can be defined as follows. We say that $F : A \times A \to \mathbb{F}$ has rank 1 if $F(x, y) = f(x)g(y)$ for some $f, g : A \to \mathbb{F}$. Then define $\mathrm{rank} F = \min\{r : F = F_1 + \cdots + F_r, F_i \text{has rank1}\}$.

Lemma 6.20 (Slice-rank of a diagonal tensor is high). *Consider coefficients $c_a \in \mathbb{F}$, indexed by elements in $A \subseteq \mathbb{F}$. Let $F(x_1, x_2, \ldots x_k) = \sum_{a \in A} c_a \delta_a(x_1) \cdots \delta_a(x_k)$.*

Then the slice-rank of F is exactly the number of c_a that are nonzero.

Proof. One direction is clear, as the slice rank of F must be at most $|\{a \in A : c_a \neq 0\}|$.

We prove the other direction using induction on k.

For $k = 2$, this reduces to standard linear algebra. Now suppose we have $k > 2$ and we have already proven the claim for smaller k. First delete from A all $a \in A$ for which $c_a = 0$. The slice-rank does not increase.

Suppose, for contradiction, that the slice rank is at most $|A| - 1$, so that we can write $\sum_{a \in A} c_a \delta_a(x_1) \cdots \delta_a(x_k)$ as a sum of rank 1 functions.

Then we can group these by x_i and write

$$\sum_{a \in A} c_a \delta_a(x_1) \cdots \delta_a(x_k) = \sum_{i=1}^{k} \sum_{\alpha \in I_i} f_{i,\alpha}(x_i) g_{i,\alpha}(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$$

where $|I_1| + \cdots + |I_k| \leq |A| - 1$.

In order to apply induction, we want to remove the contributions of one such I_k. Consider the orthogonal space to $f_{k,\alpha}$ which is defined by

$$I_k^{\perp} = \{h : A \to \mathbb{F} : \sum_{x \in A} f_{k,\alpha}(x) h(x) = 0 \text{ for all } \alpha \in I_k\}.$$

The dimension of $I_k^{\perp}$ as a vector space must be at least $|A| - |I_k|$. Consider some look at a basis of $I_k^{\perp}$. This gives rise to a matrix with dimensions $(\dim I_k^{\perp}) \times A$, which must contain some full rank minor. In particular, there exists some $h \in I_k^{\perp}$ with $|\operatorname{supp} h| \geq |A| - |I_k|$.

Now multiply our expression from above by $h(x_k)$, and sum over $x_k \in A$. After summing, we are left with

$$\sum_{a \in A} c_a \delta_a(x_1), \cdots \delta_a(x_{k-1}) h(a) = \sum_{i=1}^{k-1} \sum_{\alpha \in I_i} f_{i,\alpha}(x_i) \tilde{g}_{i,\alpha}(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots x_{k-1}),$$

where $\tilde{g} = \sum_{x_k \in A} g_{i,\alpha}(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots x_{k-1}, x_k)h(x_k)$. Each of the $f_{i,\alpha}\tilde{g}_{i,\alpha}$ is a function with slice rank 1.

By our induction hypothesis, the slice rank of the left-hand side is the number of nonzero coefficients, which is the number of a with nonzero $c_a h(a)$, or $|\text{supp } h| \geq |A| - |I_k|$. However, the slice rank of the right-hand side $\leq |I_1| + |I_2| + \cdots + |I_{k-1}| \leq |A| - |I_k| - 1$, which finishes our proof. □

Lemma 6.21. *Let $F(x, y, z) = \delta_0(x + y + z)$ be the function from $A \times A \times A$ to $\mathbb{F}_3$. Then the slice rank of F is at most*

$$3\left(\sum_{\substack{a,b,c \geq 0\\ a+b+c=n\\ b+2c \leq 2n/3}} \frac{n!}{a!b!c!}\right)$$

Proof. Notice that $\delta_0(x) = 1 - x^2$ for $x \in \mathbb{F}_3$. Then we can write $\delta_0(x+y+z)$ as

$$\delta_0(x + y + z) = \prod_{i=1}^{n}(1 - (x_i + y_i + z_i)^2).$$

The right hand side is a polynomial of total degree $2n$ which is a linear combination of monomials

$$x_1^{i_1}x_2^{i_2}\cdots x_n^{i_n}y_1^{j_i}\cdots y_n^{j_n}z_1^{k_1}\cdots z_n^{k_n}$$

By the pigeonhole principle, at least one of the sums $i_1 + \cdots + i_n$, $j_1 + \cdots + j_n$, or $k_1 + \cdots + k_n$ has sum at most $2n/3$.

Now we consider the contributions from all terms with $i_1 + \cdots + i_n \leq \frac{2n}{3}$. We can group these terms by $i_1, \dots, i_n$ and write the sum of those terms in the form $\sum_\alpha f_\alpha(x)g_\alpha(y, z)$, where α ranges over all $(i_1, \dots i_n) \in \{0, 1, 2\}^n$ with $i_1 + \cdots + i_n \leq \frac{2n}{3}$. Each of these is a function with slice-rank 1.

The number of such α is exactly $\sum_{\substack{a,b,c \geq 0\\ a+b+c=n\\ b+2c \leq 2n/3}} \frac{n!}{a!b!c!}$. In this formula, a, b, c represent the number of $0, 1, 2$ among the i_l. Repeating this for $j_1, \dots j_n$ and $k_1, \dots, k_n$ gives the desired bound. □

thm:capset

Proof of Theorem 6.17. Now return to our original equation expressing the fact that A contains no arithmetic progressions of length 3. For all $x, y, z \in$

A,

$$\delta_0(x+y+z) = \sum_{a \in A} \delta_a(x)\delta_a(y)\delta_a(z).$$

If we compare slice ranks on both sides, we have $|A| \leq 3 \left(\sum_{\substack{a,b,c \geq 0 \\ a+b+c=n \\ b+2c \leq 2n/3}} \frac{n!}{a!b!c!} \right)$.

Now fix some $0 \leq x \leq 1$, and compare the sum to $\frac{(1+x+x^2)^n}{x^{2n/3}}$. If we expand this using the multinomial theorem, each (a, b, c) with $a+b+c = n$ gives us a term $\binom{n}{a, b, c} x^{0 \cdot a} x^{1 \cdot b} x^{2 \cdot c} x^{-2n/3}$. Each term in our above sum is bounded above by one of the terms in the multinomial expansion, as $x \leq 1$ and $b + 2c - 2n/3 \leq 0$. Therefore we are able to bound $|A|$ above by $3 \cdot \frac{(1+x+x^2)^n}{x^{2n/3}}$, for any choice of $x < 1$. The optimal value of x to choose is around 0.593, which gives us a bound of $|A| \leq 3 \cdot 2.756^n$.

□

Remark 6.22 (Asymptotics in final step). We can also use Stirling's approximation or Cramer's theorem to finish here.

If $a = (\alpha + o(1))n$, $b = (\beta + o(1))n$, $c = (\gamma + o(1))n$, then $\frac{n!}{a!b!c!} = \left(\alpha^{-\alpha}\beta^{-\beta}\gamma^{-\gamma} + o(1)\right)^n$.

One may recognize this as $e^{H(\alpha,\beta,\gamma)}$, where $H(\alpha, \beta, \gamma) = -\alpha \log \alpha - \beta \log \beta - \gamma \log \gamma$ is an entropy function.

Remark 6.23 (Generalizations). It is unknown if the polynomial method here can be extended to any of the following settings:

- $\mathbb{Z}/N\mathbb{Z}$
- 4-APs in $\mathbb{F}_5^n$.
- Corners $(\{(x, y), (x + d, y), (x, y + d)\})$in $\mathbb{F}_2^N \times \mathbb{F}_2^N$.

CHAPTER 7

Structure of set addition

§7.1 Definitions in additive combinatorics

In this chapter, we will be working in some ambient abelian group (usually $\mathbb{F}_2^n$, $\mathbb{Z}/n\mathbb{Z}$, or $\mathbb{Z}$).

Definition 7.1. $A+B=\{a+b : a\in A, b\in B\}$ and $A-B=\{a-b : a\in A, b\in B\}$. Similarly for an integer k, we have $kA = A+A+\cdots+A$ where there are k terms in the sum, and $k\cdot A=\{k\cdot a : a\in A\}$.

We will primarily be concerned about the sizes of these sets. The most basic question to ask is the following: if we have a set $A\in\mathbb{Z}$ with $|A|=n$, how large or small can $A+A$ be?

We have the following easy inequality: for a finite set $A\subset\mathbb{Z}$,

$$2n-1\le|A+A|\le\binom{n+1}{2}.$$

The left hand side bound follows by considering the terms $a_1+a_1, a_1+a_2, \cdots a_1+a_n$ and $a_2+a_n, \dots, a_n+a_n$, which are all distinct elements of $A+A$. Equality on the left hand side is achieved exactly when A forms an arithmetic progression.

The bound on the right hand side follows by simply considering all $\binom{n+1}{2}$ pairs of elements of A. Equality on the right hand side holds if A is a "Sidon set", when there are no nontrivial solutions to $a+b=c+d$.

More interestingly, we can try to analyze sets A such that $|A+A|\le K|A|$ and determine properties of them. (We usually take K to be a fixed constant.)

We now try to give some examples of sets that satisfy $|A+A|\le K|A|$ for specific constants K.

By our bound above, an arithmetic progression A satisfies $|A+A| = 2|A|-1\le 2|A|$. Therefore arithmetic progressions fall into this category. More generally, we can try to construct sets A with small $A+A$ based on the idea of arithmetic progressions. The most straightforward way to extend this is to take the Cartesian product of arithmetic progressions. Doing this with k arithmetic progressions will give us a set in $\mathbb{Z}^k$ and $|A+A| \le 2^k|A|$. Using this, we can "project" this set A onto $\mathbb{Z}$ using some integer linear transformation, which will give a subset of $\mathbb{Z}$ with the same property.

Definition 7.2. A *generalized arithmetic progression (GAP)* is a set of the form $\{a_0 + a_1d_1 + \cdots + a_kd_k : 0 \le a_i \le N_i - 1, 1 \le i \le k\}$. Note that some of these sums can coincide. We say that the GAP is *proper* if all $\prod N_i$ sums are distinct. We call k the *dimension* of the GAP and $N_1 \cdots N_k$ the *size* of the GAP (note that the size of the set may be smaller if the GAP is not proper)

It is an easy exercise that if A is a GAP of dimension k, then we have $|A+A| \le 2^k|A|$.

gap-subset *Example* 7.3. Subsets of GAP's: Consider A which is a subset of A' where A' is a GAP of dimension k. Let k' be such that A contains $\frac{1}{K'}$ fraction of A's elements, or $K'|A| = |A'|$. Then $|A+A| \le |A'+A'| \le K|A'| = KK'|A|$.

It is natural to ask if we can give some classification of sets that satisfy the property $|A+A| \le K|A|$ for some constant A. In fact, this is possible. It turns out that all such A are contained in some generalized arithmetic progression discussed above, and this result, in some rough sense a classification result for all sets with this property, is known as Freiman's theorem.

Theorem 7.4 (Freiman's theorem)**.** *For every $K > 0$ there exist $d(K), f(K) > 0$ such that if a finite set $A \subset \mathbb{Z}$ satisfies $|A+A| \le K|A|$, then A is contained in a generalized arithmetic progression of dimension at most $d(K)$ and size at most $f(K)|A|$.*

Remark 7.5. We call such problems "inverse problems": we have some examples of sets that satisfy some condition, but our question is whether every set that satisfies this condition belongs to one of our examples.

Lec20: Pakawut Jiradilok

§7.2 Plünnecke–Ruzsa inequality

Freiman's theorem describes the structure of a set with small doubling. One of our goals is to prove the theorem. To do so, we need a few tools.

thm:Ruzsatri **Theorem 7.6** (Ruzsa's triangle inequality). *If A, B, C are finite subsets of an abelian group Γ, then*

$$|A| \cdot |B - C| \le |A - B| \cdot |A - C|.$$

Proof. We will construct an injection from the Cartesian product $A \times (B - C)$ to $(A - B) \times (A - C)$. If d is an element of $B - C$, we pick arbitrarily elements $b(d) \in B$ and $c(d) \in C$ such that $d = b(d) - c(d)$. Define a map $\phi : A \times (B - C) \to (A - B) \times (A - C)$ by $(a, d) \mapsto (a - b(d), a - c(d))$. To see that ϕ is injective, we note that if $\phi(a, d) = (x, y)$ is given, then $d = y - x$ and $a = x + b(d)$. Thus, we can uniquely recover (a, d). □

Remark 7.7. Note that we cannot use this trick to prove the similar inequality $|A| \cdot |B + C| \le |A + B| \cdot |A + C|$, although we will see soon that this is true. The reason this trick would not work is because if we construct a map $\phi' : A \times (B + C) \to (A + B) \times (A + C)$ by $\phi'(a, d) \mapsto (a + b(d), a + c(d))$, where $b(d) \in B$ and $c(d) \in C$ are chosen so that $d = b(d) + c(d)$, then we will not, in general, be able to uniquely recover (a, d) from $(a + b(d), a + c(d))$.

Remark 7.8. On the other hand, by simply switching A to $-A$, B to $-B$, or C to $-C$, we can easily show that $|A| \cdot |B - C| \le |A + B| \cdot |A + C|$, $|A| \cdot |B + C| \le |A + B| \cdot |A - C|$, and $|A| \cdot |B + C| \le |A - B| \cdot |A + C|$.

Remark 7.9. Why is this called Ruzsa's *triangle* inequality? If we write $\rho(A, B) := \log \frac{|A-B|}{\sqrt{|A| \cdot |B|}}$, then the inequality says $\rho(B, C) \le \rho(A, B) + \rho(A, C)$, which looks like the triangle inequality. However, do not take this observation too seriously, since ρ is not a distance. For example, $\rho(A, A) \neq 0$ for a general set A.

In the following example, we will see how Ruzsa's triangle inequality is useful.

Example 7.10. Suppose we set $B = C = 2A - A$. We have $|A| \cdot |3A - 3A| \le |2A - 2A|^2$. In particular, if $|2A - 2A| \le K \cdot |A|$, then $|3A - 3A| \le K^2 \cdot |A|$. Similarly, if we set $B = C = 3A - 2A$, then we get

$$\frac{|5A - 5A|}{|A|} \le \left(\frac{|3A - 3A|}{|A|} \right)^2 .$$

This shows that if we start with a set A for which the size $|2A - 2A|$ is bounded by a constant factor of $|A|$, then we can conclude that a set of the form $kA - kA$ also has its size bounded by a constant factor of $|A|$.

In general, we have an upper bound for the size of a set of the form $kA - \ell A$, when k and ℓ are non-negative integers, provided that A has small doubling, in the following inequality.

thm:PRineq **Theorem 7.11** (Plünnecke-Ruzsa inequality). *Let A be a finite subset of an abelian group, and let K be a positive real number. If $|A+A| \leq K \cdot |A|$, then for integers $k, \ell \geq 0$, $|kA - \ell A| \leq K^{k+\ell}|A|$.*

More generally, it is true that if $|A + B| \leq K \cdot |A|$, where A and B are finite subsets of an abelian group, then $|kB - \ell B| \leq K^{k+\ell}|A|$. We will prove this general version. The following proof, due to Petridis, uses the following key lemma.

Lemma 7.12. *If $X \subseteq A$ is a nonempty subset that minimizes $\frac{|X+B|}{|X|}$, and let $K' = \frac{|X+B|}{|X|}$. Then, $|X + B + C| \leq K' \cdot |X + C|$, for all finite sets C.*

We can think of this lemma via the idea of *expansion ratio*. Let Γ denote the ambient abelian group we are working on. For each finite set $Z \subseteq \Gamma$, the operation "addition by B" sends Z to $Z + B \subseteq \Gamma$. We can consider the expansion ratio $\frac{|Z+B|}{|Z|}$, for each Z. If we fix a finite set $A \subseteq \Gamma$, we can consider a subset $X \subseteq A$ which minimizes this expansion ratio under addition by B. The lemma says that, for any finite set $C \subseteq \Gamma$, the set $X + C$ expands even less than X does under addition by B.

Let's see how the lemma implies the general version of the inequality.

Proof. (assuming the lemma) Take a nonempty subset $X \subseteq A$ that minimizes the ratio $\frac{|X+B|}{|X|}$. Let $K' = \frac{|X+B|}{|X|}$. Note that by the choice of X, we have $K' \leq K$. Therefore, $|X + B| \leq K \cdot |X|$. The lemma says that $X + B$ expands less than X does under addition by B. Therefore, $|X + 2B| \leq K \cdot |X + B| \leq K^2 \cdot |X|$. By iteration, for every non-negative integer $\ell \geq 0$, we have $|X + \ell B| \leq K^{\ell}|X|$.

Now Ruzsa's triangle inequality gives $|X| \cdot |kB - \ell B| \leq |X + kB| \cdot |X + \ell B| \leq K^{k+\ell}|X|^2$. Thus, $|kB - \ell B| \leq K^{k+\ell}|X| \leq K^{k+\ell}|A|$. □

Next, we prove the key lemma.

Proof. (of the key lemma) We proceed by induction on $|C|$.

Base case. If $|C| = 1$, write $C = \{c\}$. Then, for any finite set W, the set $W + C = W + c$ is simply a translation by c, and thus $|W + C| = |W|$. Therefore, $|X + B + C| = |X + B| \leq K'|X| = K'|X + C|$.

Induction step. Assume $|C| > 1$. Write $C = C' \uplus \{c\}$. Note that

$$X + B + C = (X + B + C') \cup [(X + B + c) \setminus (Z + B + c)],$$

where $Z = \{x \in X : x + B + c \subseteq X + B + C'\}$. By minimality of X, we have $|Z + B| \geq K'|Z|$. Therefore,

$$\begin{aligned} |X + B + C| &\leq |X + B + C'| + |(X + B) \setminus (Z + B)| \\ &= |X + B + C'| + |X + B| - |Z + B| \\ &\leq K'|X + C'| + K'|X| - K'|Z|. \qquad \text{(induction hypothesis)} \end{aligned}$$

Observe that $X + C = (X + C') \uplus [(X + c) \setminus (W + c)]$, where $W = \{x \in X : x + c \in X + C'\}$. Note here that unlike the equation for $X + B + C$ above the union here is a disjoint union. We have the equality

$$|X + C| = |X + C'| + |X| - |W|.$$

By definition, $W \subseteq Z$. This gives

$$|X + C| \geq |X + C'| + |X| - |Z|.$$

Using this inequality in the inequality above, we obtain $|X + B + C| \leq K'|X + C|$, as desired. □

With the key lemma, now we can prove "another triangle inequality" we mentioned earlier.

Corollary 7.13. *If A, B, C be finite subsets of an abelian group Γ, then*

$$|A| \cdot |B + C| \leq |A + B| \cdot |A + C|.$$

Proof. The inequality is obvious when A is empty. Assume $A \neq \varnothing$. Consider a nonempty subset $X \subseteq A$ that minimizes $\frac{|X+B|}{|X|}$. Let $\frac{|A+B|}{|A|} = K$ and $\frac{|X+B|}{|X|} = K' \leq K$. The key lemma gives

$$\begin{aligned} |B + C| &\leq |X + B + C| \leq K' \cdot |X + C| \\ &\leq K' \cdot |A + C| \leq K \cdot |A + C| = \frac{|A + B| \cdot |A + C|}{|A|}. \end{aligned}$$

□

§7.3 Freiman's theorem in abelian groups with bounded exponent

The next is an important tool.

thm:Ruzsacover **Theorem 7.14** (Ruzsa's covering lemma). *Let A and S be finite subsets of an abelian group such that $S \neq \varnothing$. Let K be a positive real number. If $|A+S| \leq K \cdot |S|$, then there exists a subset $T \subseteq A$ with $|T| \leq K$ and $A \subseteq T+S-S$.*

Roughly speaking, this theorem says that if the expansion ratio of S by addition of A is at most K, then we can cover A by K translates of $S-S$.

Proof. The idea of this proof is to find a maximal packing. It appears that a maximal packing leads to a good covering, in the sense that we are going to observe here.

Let $T \subseteq A$ be a maximal subset of A such that the sets $t+S$ are disjoint for all $t \in T$. Due to disjointness,

$$|T||S| = |T+S| \leq |A+S| \leq K \cdot |S|,$$

whence $|T| \leq K$. We showed that T has the right size. The rest is to show that $A \subseteq T+S-S$. Since T is maximal, if $a \in A$, then there exists $t \in T$ such that $(t+S) \cap (a+S) \neq \varnothing$. This means there exist $s, s' \in S$ such that $t+s = a+s'$, and hence $a \in t+S-S \subseteq T+S-S$. □

Now we have the tools to prove an analog of Freiman's theorem. Recall that Freiman's theorem roughly says that if A has small doubling, then A is contained in a GAP whose size is not too large with respect to A. Suppose now that we are in the vector space $\mathbb{F}_2^n$. What would an analog of Freiman's theorem be? It turns out that the correct analog of a GAP is a subspace, as subspaces are closed under addition. If $A \subseteq \mathbb{F}_2^n$ satisfies $|A+A| \leq K \cdot |A|$ for some $K > 0$, then there exists a subspace H containing A such that $|H| \leq c(K) \cdot |A|$, for a certain constant $c(K)$ which depends only on K (and, in general, on the abelian group, as we shall see below). Note that this result is equivalent to saying that the subspace $\langle A \rangle$ spanned by A has size at most $c(K)|A|$, because if the upper bound holds for a subspace H, then it must hold for $\langle A \rangle$.

In general, in abelian groups, we will see that the correct analog of a GAP is a subgroup. Below, we will prove a version of Freiman's theorem for abelian groups with bounded exponent.

Definition 7.15. The *exponent* of an abelian group Γ is the smallest positive integer r (if it exists) such that $rx = 0$ for all elements $x \in \Gamma$.

Remark 7.16. The name *exponent* comes from the extension of the definition to general (not necessarily abelian) groups, where the group operation is written multiplicatively, and the corresponding identity is $x^r = 1$.

Theorem 7.17 (Ruzsa). *Let Γ be an abelian group of exponent r, and let A be a finite subset of Γ. If $|A+A| \le K \cdot |A|$, then $|\langle A\rangle| \le K^2 \cdot r^{K^4}|A|$. Here, $\langle A\rangle$ denotes the subgroup generated by A.*

Remark 7.18. Here is an example that an exponent in K is necessary. In $\mathbb{F}_2^n$, let A be a nonempty independent set. Then, we have $K \approx \frac{|A|}{2}$, whereas $|\langle A\rangle| = 2^{|A|} \approx 2^{2K}$.

Proof. The idea is to apply Ruzsa's covering lemma to $2A - A$. Note that, by Plünnecke-Ruzsa (Theorem 7.11), we have

thm:PRineq

$$|A + (2A - A)| = |3A - A| \le K^4|A|.$$

Therefore, the covering lemma implies that there exists a subset $T \subseteq 2A - A$ with $|T| \le K^4$ such that $A + (A-A) \subseteq T + (A-A)$. We then also have that $2A + (A-A) = A + (A + (A-A)) \subseteq A + T + (A-A) = T + (A + (A-A)) \subseteq 2T + (A-A)$. By iterating, we have $nA + (A-A) \subseteq nT + A - A$ for every positive integer n. Let $H = \langle T\rangle$. We therefore have $nA \subseteq nA + (A-A) \subseteq H + (A-A)$, whence $\langle A\rangle \subseteq H + A - A$. Note that we have the bound $|H| \le r^{|T|} \le r^{K^4}$. The inclusion above then gives

$$|\langle A\rangle| \le |H + A - A| \le |H| \cdot |A - A| \le r^{K^4}K^2|A|.$$

We have finished the proof. □

We will need more tools to prove Freiman's theorem in $\mathbb{Z}$.

§7.4 Freiman homomorphisms

Lec21: Sarah Tammen

We are working towards proving Freiman's theorem (Theorem 7.9), which states that, given $K > 0$, there exist $d(K)$ and $f(K)$ so that if $A \subset \mathbb{Z}$ satisfies

$$|A + A| \leq K|A|,$$

then A is contained in a generalized arithmetic progression of dimension at most $d(K)$ and size at most $f(K)|A|$.

Previously, we developed tools such as the Plünnecke-Ruzsa inequality, which enabled us to prove an analogue of Freiman's theorem for subsets of $\mathbb{F}_2^n$ with small doubling. In this setting, one can show that if $A \subset \mathbb{F}_2^n$ has small doubling, then $\langle A \rangle$ is small relative to $|A|$. In fact, the same method of proof can be extended to show that if we replace $\mathbb{F}^2$ by any abelian group Γ with finite exponent, then a finite subset $A \subset \Gamma$ with $|A + A| \leq K|A|$ satisfies $|\langle A \rangle| \leq C_K|A|$ for some constant C_K.

However, to prove Freiman's theorem for $\mathbb{Z}$, even more tools are required. Given a set $A \subset \mathbb{Z}$ with small doubling, we don't really care about the particular set A we are examining; what we care about is what sort of additive relations we have in A. We will consider sets the same if they have the same additive structure. We make this equivalence precise using Freiman isomorphisms.

Definition 7.19. Let Γ, Γ' be abelian groups. Suppose that $A \subset \Gamma$ and $B \subset \Gamma'$. We say that a map

$$\varphi : A \to B$$

is a *Freiman s-homomorphism* if

$$\varphi(a_1) + \cdots + \varphi(a_s) = \varphi(a_1') + \ldots + \varphi(a_s')$$

for any $a_1, \ldots, a_s, a_1', \ldots, a_s' \in A$ with

$$a_1 + \cdots + a_s = a_1' + \cdots + a_s'.$$

If $\varphi : A \to B$ is a bijection so that φ and φ^{-1} are both Freiman s-homomorphisms, then we say that φ is a *Freiman s-isomorphism*.

Remark 7.20. By convention, if the s is dropped from the phrase Freiman s-homorphism (resp. Freiman s-isomorphism), then it should be understood that the map in question is Freiman 2-homomorphism (resp. Freiman 2-isomorphism).

We note that if φ is a Freiman s-homorphism, then φ is a Freiman t-homomorphism for any $t < s$, because if we are given $a_1, \dots a_t, a'_1, \dots a'_t$ with $a_1 + \dots a_t = a'_1 + \dots a'_t$, then, in order to prove that $\varphi(a_1) + \dots \varphi(a_t) = \varphi(a'_1) + \dots \varphi(a'_t)$ we can take $a_{t+1} = a'_{t+1}, \dots, a_s = a'_s$. Since φ is a Freiman s-homorphism, we will have $\varphi(a_1) + \dots \varphi(a_s) = \varphi(a'_1) + \dots \varphi(a'_s)$. Since $\varphi(a_{t+1}) = \varphi(a'_{t+1}), \dots, \varphi(a_s) = \varphi(a'_s)$, this implies that $\varphi(a_1) + \dots \varphi(a_t) = \varphi(a'_1) + \dots \varphi(a'_t)$, and, thus, that φ is a Freiman t-homomorphism. The takeaway from this observation is that if two sets are Freiman s-isomorphic then they behave the same with respect to sums of length up to s.

In order to build intuition about Freiman s-homorphisms and Freiman s-isomorphisms, we consider several examples of maps that may be Freiman s-homorphisms and Freiman s-isomorphisms, or neither.

Example 7.21. If Γ_1 and Γ_2 are abelian groups and $\varphi : \Gamma_1 \to \Gamma_2$ is a group homomorphism, then φ induces a Freiman homomorphism of every order on any finite subset $A \subset \Gamma_1$.

Example 7.22. If $\varphi : \mathbb{Z} \to \mathbb{Z}$ is an affine map (which means that $\varphi(x) = ax + b$ for some $a, b \in \mathbb{Z}$), then φ is a Freiman s-homomorphism for all s.

Example 7.23. An arbitrary map

$$\varphi : \{1, 10, 10^2, 10^3\} \to \{1, 100, 100^2, 100^3\}$$

is a Freiman 2-homomorphism, because there is no additive structure in the set $\{1, 10, 10^2, 10^3\}$. That is, no two elements of $\{1, 10, 10^2, 10^3\}$ sum to another element of $\{1, 10, 10^2, 10^3\}$, which means that the condition of preserving additive structure is vacuously satisfied. If φ is a bijection then φ is a Freiman 2-isomorphism, because $\{1, 100, 100^2, 100^3\}$ also lacks additive structure.

We note that if we replaced the codomain $\{1, 100, 100^2, 100^3\}$ with any finite set B, then an arbitrary map $\varphi : \{1, 10, 10^2, 10^3\} \to B$ would still be a Freiman 2-homomorphism.

Example 7.24. Let $\varphi : \{0, 1\}^n \to \mathbb{F}_2^n$ be the map that sends each element of $\{0, 1\}^n$ to the corresponding vector in $\mathbb{F}_2^n$. (Here, we are defining the sum of two elements of $\{0, 1\}^n$ by addition of integers, *not* addition $\pmod 2$, e.g. $(0, 1) + (1, 1)$ is equal to $(1, 2)$, not $(1, 0)$.) Then φ is a Freiman s-homomorphism for any s. However, φ is not an additive s-isomorphism for any s, because $\mathbb{F}_2^n$ has additive structure that does not pull back to $\{0, 1\}^m$.

mod N map *Example* 7.25. The map

$$\pi_N : \mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$$

defined by sending an integer to its equivalence class $\bmod N$ is a group homomorphism, so it is a Freiman s-homomorphism for all s. However, π_N is not a Freiman s-isomorphism because it is not bijective. However, one might wonder: what happens if we restrict π_N to a finite set $A \subset \mathbb{Z}$? Can we find a criterion for the restriction $\pi_N\Big|_A$ to be a Freiman s-isomorphism? The restriction could fail to be a Freiman s-isomorphism if there is an additive relation in the image of A that does not pull back to an additive relation. We claim that we can prevent this possibility by taking A to be a set that is not very spread out. Specifically, we claim that if $A \subset \mathbb{Z}$ and

$$N > s \cdot \operatorname{diam} A,$$

then the restriction π_N is a Freiman s-isomorphism onto its image. (Here, we are using $\operatorname{diam} A$ to denote the diameter of A. The diameter of a finite subset of $\mathbb{Z}$ is equal to the maximum element minus the minimum element.)

To prove the claim, suppose that $N > s{\cdot}\operatorname{diam} A$ and that $a_1, \dots, a_s, a_1', \dots a_s' \in A$ satisfy

$$a_1 + \cdots + a_s - a_1' - \cdots - a_s' \equiv 0 \, (\bmod N).$$

We must show that

$$a_1 + \cdots + a_s = a_1' + \cdots + a_s'.$$

To do so, we look at differences of the form $a_i - a_i'$. Since each of a_i and a_i' lies in an interval of length less than N/s, we have that

$$|a_i - a_i'| < \frac{N}{s}$$

for $i = 1, \dots, s$. Thus,

$$|a_1 + \cdots + a_s - a_1' - \cdots - a_s'| \leq |a_1 - a_1'| + \cdots + |a_s - a_s'| < s\left(\frac{N}{s}\right) = N.$$

Since $a_1 + \cdots + a_s - a_1' - \cdots - a_s' \equiv 0 \, (\bmod N)$, this implies that $a_1 + \cdots + a_s - a_1' - \cdots - a_s' = 0$.

representative map *Example* 7.26. Finally, we consider the map

$$\psi_N : \mathbb{Z}/N\mathbb{Z} \to \{1, \dots, N\} \subset \mathbb{Z}$$

defined by sending each member of $\mathbb{Z}/N\mathbb{Z}$ to its representative in $\{1, \ldots, N\}$. This map is not a Freiman s-homomorphism, because we can have additive relations in $\mathbb{Z}/N\mathbb{Z}$ that wrap around but are not preserved under ψ_N.

However, we claim that if we choose an appropriate subset $A \subset \mathbb{Z}/N\mathbb{Z}$, then the restriction $\psi_N\Big|_A$ will be a Freiman s-isomorphism onto its image. Specifically, if $A \subset \mathbb{Z}/N\mathbb{Z}$ is such that $\psi_N(A)$ is contained in an interval of length $< N/s$, then $\psi_N\Big|_A$ is a Freiman s-isomorphism onto its image, by the same argument we used in Example 7.25.

The first step in the proof of Freiman's theorem in $\mathbb{Z}$ is to model a set of small doubling by a set in a smaller space, where we have better tools, such as Fourier analysis. A set $A \subset \mathbb{Z}$ may be quite spread out, but we can model a large proportion of A by a dense subset of $\mathbb{Z}/m\mathbb{Z}$ for an m that is comparable to $|A|$. To this end, we use the following result, due to Rusza.

Lemma 7.27. *(Ruzsa's Model Lemma) Suppose that $A \subset \mathbb{Z}$ is finite, that $s \geq 2$, and that $m \geq |sA - sA|$. Then there exists $A' \subset A$ with $|A'| \geq |A|/s$ so that A' is Freiman s-isomorphic to a subset of $\mathbb{Z}/m\mathbb{Z}$.*

Remark 7.28. If we had merely wanted to embed A into $\mathbb{Z}/m\mathbb{Z}$ for some m, then this would be easy - just choose m larger than the maximum element of A. However, this could result in choosing a m that is much larger than $|A|$. The advantage of Ruzsa's Model Lemma is that it holds for any $m \geq |sA - sA|$, so m cannot be too large relative to $|A|$.

Proof. The idea of the proof is as follows: given $s \geq 2$ and a subset $A \subset \mathbb{Z}$ along with an integer m so that $m \geq |sA - sA|$, choose a large prime q - sufficiently large that the reduction $\bmod q$ map π_q is a Freiman s-isomorphism from A onto its image. Multiplying by any nonzero λ in $\mathbb{Z}/q\mathbb{Z}$ defines a Freiman isomorphism $\mathbb{Z}/q\mathbb{Z} \to \mathbb{Z}/q\mathbb{Z}$. Let B denote the image of A under the composition $\cdot\lambda \circ \pi_q$. Using the map ψ_q we lift back to $\mathbb{Z}$. As per our discussion in 7.26, ψ_q may not be a Freiman s-isomorphism, but we can make it a Freiman s-isomorphism by restricting to an interval of appropriate length in $\mathbb{Z}/q\mathbb{Z}$. We represent all of these maps in the following diagram. In the third line, B_j is a subset of B contained within a short interval, and A' is the pre-image of B_j under $\cdot\lambda \circ \pi_q$. All arrows in the third line represent Freiman

s-isomorphisms.

$$\begin{tikzcd}
\mathbb{Z} \arrow[r, "\pi_q"] & \mathbb{Z}/q\mathbb{Z} \arrow[r, "\cdot\lambda"] & \mathbb{Z}/q\mathbb{Z} \arrow[r, "\psi_q"] & \mathbb{Z} \\
A \arrow[r] & \pi_q(A) \arrow[r] & B \arrow[r] & \psi_q(B) \\
A' \arrow[r] & \pi_q(A') \arrow[r] & B_j \arrow[r] & \psi_q(B_j).
\end{tikzcd}$$

Finally, we apply π_m to send $\mathbb{Z}$ to $\mathbb{Z}/m\mathbb{Z}$. This last step is sure to be a Freiman s-homomorphism, but may not be a Freiman s-ismorphism. However, we will show that for *some* choice of λ at the second step, the final step is a Freiman s-isomorphism.

Having proposed a plan, let's elaborate on each of the steps. In the first step of our outline, we had claimed that if we choose a prime q that is sufficiently large, then

$$\pi_q : A \to \mathbb{Z}/q\mathbb{Z}$$
$$a \mapsto a(\operatorname{mod} q)$$

is an isomorphism onto its image. Specifically, taking $q > \operatorname{diam}(A) \cdot s$ is sufficient for $\pi_q\Big|_A$ to be a Freiman s-isomorphism onto its image. After multiplying by a nonzero $\lambda \in \mathbb{Z}/q\mathbb{Z}$, consider the map

$$\psi_q : \mathbb{Z}/q\mathbb{Z} \to \{1, 2, \ldots, q\}.$$

For $j = 1, \ldots, n$, we define a set $B_j \subset B$ by

$$B_j \;=\; \left\{ b \in B : \psi_q(b) \in \left((j-1)\frac{q}{s}, j\frac{q}{s} \right) \right\}.$$

for $j = 1, \ldots, n$. One can check that for each j, the restriction

$$\psi_q\Big|_{B_j}$$

is a Freiman s-isomorphism onto its image. Since $B = B_1 \uplus \cdots \uplus B_s$, we can choose a $j = j(\lambda)$ so that

$$|B_j| \geq |B|/s.$$

Having chosen such a j, let

$$A_\lambda = \{a \in A : \lambda\pi_q(a) \in B_{j(\lambda)}\},$$

and let

$$C_\lambda = \psi_q(B_{j(\lambda)})$$

Then each arrow in the sequence

$$A_\lambda \longrightarrow B_{j(\lambda)} \longrightarrow C_\lambda$$

is a Freiman s-ismorphism. It remains to show that there exists some λ so that the restriction

$$\pi_m\Big|_{C_\lambda}$$

is a Freiman s-isomorphism for some λ. To this end, let

$$\Lambda := \left\{ \lambda \in (\mathbb{Z}/q\mathbb{Z})^x : \begin{array}{c} \pi_m : C_\lambda \to \mathbb{Z}/m\mathbb{Z} \text{ is } \textit{not} \\ \text{a Freiman } s\text{-isomorphism} \end{array} \right\}.$$

We will show that the size of Λ is strictly less than $q-1$.

We note that $\pi_m : C_\lambda \to \mathbb{Z}/m\mathbb{Z}$ is a Freiman s-homomorphism for any λ, because it is induced by a group homomorphism. The only way that π_m can fail to be a Freiman s-isomorphism from C_λ onto its image is if there is an additive relation in the target that does not pull back to an additive relation in the domain, i.e. π_m can fail to be a Freiman s-isomorphism from C_λ onto its image only if there exist $c_1, \ldots, c_s, c_1', \ldots, c_s' \in C_\lambda$ so that

$$c_1 + \cdots + c_s > c_1' + \cdots + c_s'$$

but

$$c_1 + \ldots c_s \equiv c_1' + \ldots c_s' \pmod{m}.$$

Suppose that there exist $c_1, \ldots, c_s, c_1', \ldots, c_s'$ as above, and let

$$b := c_1 + \ldots c_s - c_1' - \cdots - c_s'.$$

Then b is a positive integer divisible by m. Since $\operatorname{diam}(C_\lambda) < q/s$, we have that $b < q$. Since C_λ is Freiman s-isomorphic to A_λ, there exist $a_1, \ldots, a_s, a_1', \ldots, a_s'$ so that $\psi_q(\pi_q(a_i)\lambda) = c_i$ for all i and $\psi_q(\pi_q(a_i')\lambda) = c_i'$ for all i. Let

$$d := a_1 + \cdots + a_s - a_1' - \cdots - a_s'. \tag{8}$$

d definition

Then

$$d \in (sA - sA)\backslash 0.$$

(We know that $d \neq 0$, because $c_1 + \ldots c_s - c_1' - \cdots - c_s' \neq 0$.) If we multiply (refd definition) through by q, we see that $\lambda d \equiv c_1 + \ldots c_s - c_1' - \cdots - c_s' \pmod{q}$, i.e. $\lambda d \equiv b \pmod{q}$. We recall that b is divisible $\pmod{q}$. Recognizing that b is the least positive residue of $\lambda d \pmod{q}$ (i.e. the least positive integer that is congruent to $\lambda d \pmod{q}$), we conclude that if λ is in the "bad set" Λ, then the least positive residue of $\lambda d \pmod{q}$ is a nonzero integer that is divisible by m. There are $\left\lfloor \frac{q-1}{m} \right\rfloor$ positive integers that are less than q and divisible by m. Thus, given a d the total number of λ for which

d represents a failure in the sense of (8) is at most $\left\lfloor \frac{q-1}{m} \right\rfloor$. To find a bound for the size of Λ, take a union bound over all possible failures (i.e. take a union bound over all d that represent a failure). This gives

$$\begin{aligned} |\Lambda| &\leq \sum_{d \in (sA-sA)\setminus\{0\}} \left\lfloor \frac{q-1}{m} \right\rfloor < |sA-sA| \left\lfloor \frac{q-1}{m} \right\rfloor \\ &\leq m \left(\frac{q-1}{m} \right) = q-1, \end{aligned}$$

as claimed. Therefore, there exists a nonzero $\lambda \pmod q$ so that the last step proposed in our outline is a Freiman s-isomorphism. □

Having proved Ruzsa's model lemma, we can deduce the following corollary, which will be useful in our proof of Freiman's theorem.

8isomorphism

Corollary 7.29. *Suppose that $A \subset \mathbb{Z}$ is finite. If $|A+A| \leq K|A|$, then there exists a prime $N \leq 2K^{16}|A|$ and some subset $A' \subset A$ with $|A'| \geq |A|/8$ so that A' is Freiman 8-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$.*

Proof. By the Plünnecke-Ruzsa inequality, $|8A - 8A| \leq K^{16}|A|$. Choose a prime N in the interval $(K^{16}|A|, 2K^{16}|A|]$. Then apply the model lemma. □

§7.5 Bogolyubov's lemma

If A is a set with small doubling, then Corollary 7.29 allows us to view a large subset of A as a subset of a prime cyclic group, where we can do Fourier analysis. As we have seen in the proof of Roth's theorem, Fourier analysis is well-suited to a finite field setting. We will prove a result called Bogolyubov's lemma (Lemma 7.32), which concerns the structure of sets of the form $2A-2A$ for $A \in \mathbb{F}_2^n$. We will later extend this result to the integers. As motivation for our ensuing work, we consider the following question:

Question 7.30. *Suppose that $A \subset \mathbb{F}_2^n$ satisfies $|A| = \alpha 2^n$ for some $\alpha > 0$. Must $|A+A|$ contain a large subspace?*

The analysis varies depending on how big α is. If $\alpha > \frac{1}{2}$, then $A+A = \mathbb{F}_2^n$. This can be deduced by a pigeonhole argument: for any $x \in \mathbb{F}_2^n$, both A and $x - A$ consist of more than half of the additive group $\mathbb{F}_2^n$. Thus, they overlap; that is, there exists $a, a' \in A$, so that $a = x - a'$, which means that $x = a + a'$.

However, for any constant $\alpha < 1/2$, one can choose a sequence $A_n \subset \mathbb{F}_2^n$ so that $|A_n|$ is at least $\alpha 2^n$, but A_n does not contain large subspaces (i.e.

subspaces that consist of at least a constant proportion of $\mathbb{F}_2^n$ or, equivalently, subspaces of bounded codimension). We consider the following example.

Example 7.31. (Niveau set) Let

$$A_n = \left\{ x \in \mathbb{F}_2^n : wt(x) \leq \frac{n - c\sqrt{n}}{2} \right\},$$

where the weight function $wt(x)$ counts the number of 1s in a vector $x \in \mathbb{F}_2^n$. By the central limit theorem, we can choose $c > 0$ so that

$$|A_n| = (\alpha + o(1))2^n.$$

We have that

$$A_n + A_n = \{x \in \mathbb{F}_2^n : wt(x) \leq n - c\sqrt{n}\}.$$

We claim that this set does not contain any subspaces of codimension less than $c\sqrt{n}$. To see this, we use a linear algebra argument, similar to the argument we employed in our polynomial method proof of Roth's theorem. We note that if $A_n + A_n$ *did* contain a subspace of dimension $m > n - c\sqrt{n}$ we could write down a $m \times n$ generator matrix whose columns contained a basis for the subspace. Such a matrix would have an $m \times n$ minor of full rank. By doing row operations, one could arrive at a matrix so that a row of this minor consisted of only 1s, violating our hypothesis that the weight of any vector in $A_n + A_n$ is at most $n - c\sqrt{n}$.

This example shows that even if A consists of a constant proportion of $\mathbb{F}_2^n$, we can not necessarily expect $A + A$ to contain a large subspace. We note that adding A to itself corresponds to "smoothing" A. Even though $A + A$ may not contain a large subspace, one might conjecture that a large subspace can be found if more copies of A are added or subtracted. Indeed, we have the following result, due to Bogolyubov.

Bogolyubov **Lemma 7.32.** *(Bogolyubov's lemma) Suppose that* $A \subset \mathbb{F}_2^n$ *with* $|A| = \alpha 2^n$ *for some constant* α. *Then* $2A - 2A$ *contains a subspace of codimension* $1/\alpha^2$.

To prove Bogolyubov's lemma, it is helpful to use convolutions.

Definition 7.33. Let Γ be an abelian group. Given $f, g : \Gamma \to \mathbb{C}$, we define the convolution $f * g : \Gamma \to \mathbb{C}$ by

$$\begin{aligned}(f * g)(x) &:= \mathop{\mathbb{E}}_{y \in \Gamma} f(y)g(x-y) \\ &= \mathop{\mathbb{E}}_{\substack{y, z \in \Gamma \\ y+z=x}} f(y)g(z).\end{aligned}$$

The second formulation of the definition of $f * g$ suggests that convolutions should be useful for studying sumsets. Indeed, we note that if $A, B \subset \Gamma$, then the convolution $1_A * 1_B$ is supported on $A + B$. As another example of how convolutions correspond to sums, suppose that X is a random variable that takes values in Γ and that $X \sim \mu_X$. This means that for any $a \in \Gamma$, we have that

$$\mathbb{P}(X = a) = \frac{\mu_X(a)}{|\Gamma|}.$$

Similarly, suppose that $Y \sim \mu_Y$. Then $X + Y \sim \mu_X * \mu_Y$.

Convolutions interact nicely with the Fourier transform. In particular, we have the following proposition, which, roughly stated, says that the Fourier transform takes convolutions to multiplication.

Proposition 7.34. *If $f, g : \Gamma \to \mathbb{C}$, then*

$$\widehat{f * g} = \hat{f}\hat{g}. \tag{9}$$

convolution to multiplication

convolution to multiplication

Proof. To prove (9), we must show that for any character γ,

$$\widehat{f * g}(\gamma) = \hat{f}(\gamma)\hat{g}(\gamma).$$

This is a fairly routine computation, which we nevertheless write out: for any character γ, we have

$$\begin{aligned}\widehat{f*g}(\gamma) &= \mathop{\mathbb{E}}_x (f*g)(x)\overline{\gamma(x)}\\ &= \mathop{\mathbb{E}}_x \left(\mathop{\mathbb{E}}_{\substack{y,z\in\Gamma\\ y+z=x}} f(y)g(z)\overline{\gamma(y+z)} \right)\\ &= \mathop{\mathbb{E}}_{y,z\in\Gamma} f(y)g(z)\overline{\gamma(y+z)}\\ &= \mathop{\mathbb{E}}_{y,z\in\Gamma} f(y)g(z)\overline{\gamma(y)}\,\overline{\gamma(z)}\\ &= \left(\mathop{\mathbb{E}}_y f(y)\overline{\gamma(y)}\right)\left(\mathop{\mathbb{E}}_z g(z)\overline{\gamma(z)}\right)\\ &= \hat{f}(\gamma)\hat{g}(\gamma).\end{aligned}$$

□

We now have the technology to prove Bogolyubov's lemma. First we give a proof in the finite field setting. Recall that $A \subset \mathbb{F}_2^n$, $|A| = \alpha 2^n$ and we are trying to show that $2A - 2A$ contains a subspace of codimension at most $1/\alpha^2$.

lec22, part 1: Frederic Koehler

Remark 7.35. Here we are making an immaterial distinction between A and $-A$, i.e. choosing to write $2A - 2A$ instead of $4A$, which will matter later when we move to $\mathbb{Z}/N\mathbb{Z}$. The reason we look at $4A$ is because we previously saw an example where $2A$ does not contain any large subspaces (the Niveau set); but by going to $4A$ we "smooth" the set even more and it turns out we can now find large subspaces.

Proof of Lemma 7.32. Consider $f = 1_A * 1_A * 1_{-A} * 1_{-A}$. Observe that $\hat{f} = \hat{1}_A^2\hat{1}_{-A}^2 = |\hat{1}_A|^4$. By Fourier inversion

$$f(x) = \sum_{r\in\mathbb{F}_2^n} \hat{f}(r)(-1)^{r\cdot x}.$$

Observe that if $f(x) > 0$ then $x \in 2A - 2A$.

How can we make $f(x)$ large?

- $\hat{f}(0) = \alpha^4$ is large, but this term alone is not enough, the other terms can annihilate this positive contribution.

- We want to avoid x s.t. $r \cdot x = 1$ for some r with $|\hat{1}_A(r)|$ large, since these give large negative contributions.

Let $R = \{r \in F_2^n \setminus \{0\} : |\hat{1}_A(r)| > \alpha^{3/2}\}$. By Parseval, there cannot be too many large coefficients:

$$|R|\alpha^3 < \sum_{r \in \mathbb{F}_2^n} |\hat{1}_A(r)|^2 = E|1_A|^2 = \alpha$$

Hence $|R| < \alpha^{-2}$.

Now observe that if $x \in R^\perp$, then

$$\begin{aligned} f(x) &= \sum_{r \in \mathbb{F}_2^n} |\hat{1}_A(r)|^4 (-1)^{r \cdot x} \\ &= |\widehat{1_A}(0)|^4 + \sum_{r \in R} |\hat{1}_A(r)|^4 + \sum_{r \notin (R \cup \{0\})} |\hat{1}(r)|^4 (-1)^{r \cdot x} \\ &\geq \alpha^4 + 0 - \alpha^3 \sum_{r \notin (R \cup \{0\})} |\hat{1}_A(r)|^2 \\ &> \alpha^4 - \alpha^3 \alpha = 0 \end{aligned}$$

since for $r \notin (R \cup \{0\})$ we know $|\hat{1}(r)| \leq \alpha^{3/2}$, and in the last step we used Parseval (the inequality is strict since we added $r = 0$ to the sum).

So $x \in 2A - 2A$ whenever $x \in R^\perp$, which gives the desired subspace of codimension $|R| < \alpha^{-2}$. □

We want to do the same proof in $\mathbb{Z}/n\mathbb{Z}$ but we no longer have subspaces! Instead we will use *Bohr sets*.

Definition 7.36 (Bohr Sets). Suppose $R \subset \mathbb{Z}/N\mathbb{Z}$. We call $|R|$ the dimension and ϵ the width of the Bohr set,

$$Bohr(R, \epsilon) = \left\{ x \in \mathbb{Z}/N\mathbb{Z} : \left\| \frac{rx}{N} \right\|_{\mathbb{R}/\mathbb{Z}} \leq \epsilon \text{ for all } r \in R \right\}$$

where the norm denotes the distance to $\mathbb{Z}$.

To make the analogy to subspaces in the $\mathbb{F}_2^n$ case, we should think of ϵ as close to 0 and then we can think of this as asking that the inner product $r \cdot x \approx 0$, similar to defining a subspace by its orthogonal vectors.

However, Bohr sets are less nice than subspaces because they are not closed under addition; all we know is that $Bohr(R, \epsilon) + Bohr(R, \epsilon) \subset Bohr(R, 2\epsilon)$. (And note that growing the radius of the ball by a factor of 2 typically makes the size exponentially larger in the dimension $|R|$.)

bogolyubov-cyclic **Lemma 7.37** (Bogolyubov's lemma in $\mathbb{Z}/N\mathbb{Z}$). *Suppose $A \subset \mathbb{Z}/N\mathbb{Z}$ and $|A| = \alpha N$. Then $2A - 2A$ contains a Bohr set of dimension at most α^{-2} and width* $1/4$.

Remark 7.38. Note that *dimension* of a Bohr set corresponds to the *codimension* of a subspace under our analogy, so this lemma is indeed analogous to its $\mathbb{F}_2^n$ version.

Proof. This is basically the same proof as the $\mathbb{F}_2^n$ version. Let

$$f = 1_A * 1_A * 1_{-A} * 1_{-A}$$

so

$$\widehat{f} = |\widehat{1}_A|^4.$$

Let

$$R = \{r \in \mathbb{Z}/N\mathbb{Z} : |\widehat{1_A}(r)| > \alpha^{3/2}\}.$$

Observe that if $x \in Bohr(R, 1/4)$ then $\cos(\frac{2\pi rx}{N}) \geq 0$ since $\|rx/N\|_{\mathbb{R}/\mathbb{Z}} \leq 1/4$. By Parseval

$$|R|\alpha^3 < \sum_{r \in \mathbb{Z}/N\mathbb{Z}} |\widehat{1}_A(r)|^2 = E|1_A|^2 = \alpha.$$

By Fourier inversion and the fact that $f(x)$ is real-valued

$$\begin{aligned} f(x) &= \sum_{r \in \mathbb{Z}/N\mathbb{Z}} |\widehat{1}_A(r)|^4 e^{2\pi i (rx/N)} \\ &= \sum_{r \in \mathbb{Z}/N\mathbb{Z}} |\widehat{1}_A(r)|^4 \cos(\frac{2\pi rx}{N}) \\ &= |\widehat{1}_A(0)|^4 + \sum_{r \in R} |\widehat{1}_A(r)|^4 \cos(\frac{2\pi rx}{N}) + \sum_{r \notin (R \cup \{0\})} |\widehat{1}_A(r)|^4 \cos(\frac{2\pi rx}{N}) \\ &\geq \alpha^4 + 0 - \alpha^3 \sum_{r \notin R \cup \{0\}} |\widehat{1}_A(r)|^2 \\ &> \alpha^4 - \alpha^4 = 0 \end{aligned}$$

hence $x \in 2A - 2A$. □

§7.6 Geometry of numbers

Note that a Bohr set is not the same as an arithmetic progression, so the next step towards proving Freiman's theorem is to show that a large Bohr set must contain a large arithmetic progression.

We remind ourselves of the big picture: if a set has small doubling, then its iterated sums are small by Plunecke-Rusza, Ruzsa's model lemma shows we can embed a large subset A' of A into a cyclic group of comparable size to A', Bogolyubov's lemma tells us that $2A' - 2A'$ contains a large Bohr set, and next we show that the large Bohr set contains a large arithmetic progression.

Proposition 7.39. *A Bohr set of dimension d and width ϵ in $\mathbb{Z}/N\mathbb{Z}$ must contain a proper generalized arithmetic progression (recall proper means that all elements that should be distinct are distinct, like a proper parallelipied) of dimension at most d and size at least $(\epsilon/d)^d N$.*

To prove this statement we need tools from *geometry of numbers.* Geometry of numbers basically concerns the study of lattices and convex bodies with application to number theory. Before we start to prove Minkowski's fundamental theorems, we need to briefly review some basic definitions and properties of lattices. Proofs of facts here can be found in e.g. any introductory text on algebraic number theory (such as Neukirch's book).

Definition 7.40. A lattice Λ is a discrete subgroup of $\mathbb{R}^d$.

Obviously integer combinations of any linear independent vectors $v_1, \dots, v_d$ generate a lattice that spans $\mathbb{R}^d$. The next proposition asserts that all lattices are of this form.

Proposition 7.41. *If a lattice spans $\mathbb{R}^d$, then it has an* integral basis $v_1, v_2, \dots, v_d$ *such that*

$$\Lambda = \mathbb{Z}v_1 \oplus \cdots \oplus \mathbb{Z}v_d.$$

Example 7.42. Identify $\mathbb{C}$ with $\mathbb{R}^2$ and let ω be a primitive third root of unity; consider the lattice $\Lambda = \mathbb{Z}1 + \mathbb{Z}\omega$; then we also have $\Lambda = \mathbb{Z}(1+\omega) + \mathbb{Z}(2+\omega)$. This shows that the integral basis is not unique.

Example 7.43. $\mathbb{Z}1 + \mathbb{Z}\sqrt{2} \subset \mathbb{R}$ is a *not* an example of a lattice in $\mathbb{R}$ because it is not discrete.

Definition 7.44. The *determinant of a lattice* is given by

$$\det(\Lambda) = |\det(v_1 \cdots v_d)|$$

where $v_1, \dots, v_d$ is any integral basis, which we treat as the columns of a square matrix.

Definition 7.45. The *fundamental parallelepiped* of a lattice Λ is a set F consisting of the interior of a parallelepiped and part of its boundary, such that $\bigcup_{v \in \Lambda}(v+F) = \mathbb{R}^d$ and these translates of F are all disjoint.

Proposition 7.46. *The determinant of a lattice equals the volume of its fundamental parallelepiped.*

Definition 7.47. A lattice Λ is *nondegenerate* if $\det(\Lambda) \neq 0$, i.e. Λ spans $\mathbb{R}^d$.

We will now prove several results relating the volumes of open subsets of $\mathbb{R}^d$ (which we sometimes also assume are convex and centrally symmetric) to determinants of nondegenerate lattices. lec 22, part 2: Christian Gaetz

blichfeldt **Lemma 7.48** (Blichfeldt's Lemma). *Let $\Lambda \subset \mathbb{R}^d$ be a nondegenerate lattice, and let $K \subset \mathbb{R}^d$ be an open set with $\mathrm{vol}(K) > \det(\Lambda)$. Then there are distinct points $x, y \in K$ such that $x - y \in \Lambda$.*

Proof. Informally, we can use a volume packing argument: if the conclusion were false then the translates $K + v$, $v \in \Lambda$ would all be disjoint (otherwise there are some distinct $x, y \in K$ such that $x + v = y + v'$ for some $v, v' \in \Lambda$, but then $x - y = v' - v \in \Lambda$). Since $\mathrm{vol}(K) > \det(\Lambda)$ this would give rise to a packing of copies of K in $\mathbb{R}^d$ of density greater than one. This argument could be made precise with a suitable definition of density, but we instead give a slightly different formal proof.

Fix a fundamental parallelpiped P of Λ. We know

$$\mathbb{R}^d = \biguplus_{v \in \Lambda} (P + v)$$

and so

$$K = \biguplus_{v \in \Lambda} (K \cap (P + v))$$

Taking volumes, we have

$$\begin{aligned} \mathrm{vol}(K) &= \sum_v \mathrm{vol}(K \cap (P + v)) \\ &= \sum_v \mathrm{vol}((K - v) \cap P) \end{aligned}$$

But $\mathrm{vol}(K) > \det(\Lambda) = \mathrm{vol}(P)$, so some two translates must overlap in P, giving $x - v = y - v'$ for some $x, y \in K$ and distinct $v, v' \in \Lambda$, thus $x - y = v - v' \neq 0 \in \Lambda$. □

As an easy corollary of Blichfeldt's Lemma, we obtain:

Minkowski1 **Theorem 7.49** (Minkowski's First Theorem). *Let Λ be a nondegenerate lattice and K a centrally symmetric convex set, both in $\mathbb{R}^d$. Suppose* $\mathrm{vol}(K) > 2^d \det(\Lambda)$, *then K contains a nonzero point of Λ.*

Proof. By hypothesis, $\mathrm{vol}(\frac{1}{2}K) = 2^{-d}\,\mathrm{vol}(K) > \det(\Lambda)$, so by Lemma 7.48 there exist $x, y \in K$ such that $x - y \in \Lambda \setminus 0$. Since K is centrally symmetric, we see that $-2y \in K$, and therefore $\frac{1}{2}(2x - 2y) = x - y \in K$ by convexity. □

open cube *Remark* 7.50. The constant 2^d in Theorem 7.49 is tight, as can be seen by taking K to be the interior of a d-cube of side length 2, centered at the origin, and Λ to be the standard coordinate lattice.

We will not actually need Theorem 7.49 for our purposes, but rather a generalization known as Minkowski's Second Theorem. Although Remark 7.50 shows that the constant 2^d cannot be uniformly improved, that example is very special: as we scale K we hit lattice points in all directions simultaneously. In order to state Minkowski's Second Theorem, we first need to make some definitions which account more precisely for the shape of K with respect to Λ.

Given a centrally symmetric convex body $K \subset \mathbb{R}^d$, define the *i-th successive minimum* λ_i by

$$\lambda_i := \inf\{\lambda \mid \lambda K \text{ contains } i \text{ linearly independent elements of } \Lambda\}$$

The *i-th directional basis vector* b_i is the unique vector (up to sign and coincidences of the λ_i) such that $b_i \in \lambda_i \bar{K} \cap \Lambda$ and $\dim \mathrm{span}\{b_1, ..., b_i\} = i$.

Intuitively we can think of scaling K by a very small factor λ which we then let grow slowly. Then λ_1 is the first value of λ at which λK hits a nonzero point x of Λ, and b_1 points in the direction of x. We then continue scaling λ until λK hits more points of Λ, except that we require these points to be in genuinely new directions.

Minkowski2 **Theorem 7.51** (Minkowski's Second Theorem). *Let K be a centrally symmetric convex body and Λ a nondegenerate lattice, both in $\mathbb{R}^d$, and let $\lambda_1, ..., \lambda_d$*

be the successive minima. Then

$$\lambda_1\lambda_2\cdots\lambda_d\cdot\mathrm{vol}(K)\le 2^d\cdot\det(\Lambda)$$

Remark 7.52. This is a generalization of Theorem 7.49. We have $(\lambda_1 K^\circ)\cap\Lambda=\{0\}$, and so by Theorem 7.49, we must have $\mathrm{vol}(\lambda_1 K)=\lambda_1^d\,\mathrm{vol}(K)\le 2^d\,\mathrm{vol}(\Lambda)$. Theorem 7.51 allows us to replace the constant λ_1^d with the possibly bigger constant $\lambda_1\lambda_2\cdots\lambda_d$.

Proof. This proof is somewhat unintuitive. The idea is to grow K until we hit a point of Λ, and then continue growing, but only in the complementary direction. However rigorously carrying out this procedure is very tricky.

It suffices to prove the theorem in the case where K is open. Fix a directional basis $b_1,...,b_d$; the fact that K is open implies that $(\lambda_i K)\cap\Lambda\subset\mathrm{span}\{b_1,...,b_{i-1}\}$. For $j=1,...,d$, we define maps $\phi_j: K\to K$ by sending a point $x\in K$ to the center of mass of the $(j-1)$-dimensional slice of K which contains x and is parallel to $b_1,...,b_{j-1}$. (We will see later that the exact "center of mass" property is not essential to the proof; we simply want all points in a hyperplane slice of K to be sent to a unique point also within K that respects K's symmetry). In particular, ϕ_1 is the identity function. We also define a function $\tilde{\phi}: K\to\mathbb{R}^d$ by

$$\tilde{\phi}(x)=\sum_{j=1}^{d}(\lambda_j-\lambda_{j-1})\phi_j(x)/2$$

where by convention we let $\lambda_0=0$. We will see that $\tilde{\phi}(K)$ has volume $\lambda_1\lambda_2\cdots\lambda_d/2^d$ and that $\tilde{\phi}(K)$ contains no lattice points, but this will take a bit more work. □

lec23: Saranesh Prembabu/Benjamin Gunby

$\tilde{\phi}$ is intended to encapsulate this process of growing first in one direction (given by the first term of the sum), then in the second direction (given by the second term of the sum), et cetera. Naively trying to grow in all directions in an ‘affine’ way doesn't work properly.

For $\underline{x}\in\mathbb{R}^d$, let x_i be its entries in the basis (b_i), so that $\underline{x}=\sum_{i=1}^{d}x_i\underline{b}_i$. Note that

$$\phi_j(x)=\sum_{i<j}c_{j,i}(x_j,\ldots,x_d)\underline{b}_i+\sum_{i\ge j}x_i\underline{b}_i$$

for some continuous functions $c_{j,i}$.

Therefore,

$$\tilde{\phi}(\underline{x}) = \sum_{i=1}^{d} (\lambda_i \underline{x}_i/2 + \psi_i(x_{i+1}, \ldots, x_d))\underline{b}_i$$

for some continuous functions ψ_i. So the Jacobian $J(\phi) = \frac{\partial \tilde{\phi}(\underline{x})}{\partial \underline{x}_j}$ is upper triangular with diagonal entries $\lambda_1/2, \ldots, \lambda_d/2$. This implies that

$$\mathrm{vol}(\tilde{\phi}(K)) = \lambda_1 \cdots \lambda_d \mathrm{vol}(K)/2^d.$$

which is conveniently the left hand side of the inequality of Minkowski's Second Theorem $\lambda_1 \cdots \lambda_d \mathrm{vol}(K)/2^d < \det \Lambda$ which we seek to prove..

We return to the proof of Minkowski's Second Theorem.

For any distinct points $\underline{x} = \sum x_i \underline{b}_i$, $\underline{y} = \sum y_i \underline{b}_i$ in K, let k be the largest index such that $x_k \neq y_k$. Then $\phi_i(\underline{x})$ agrees with $\phi_i(\underline{y})$ for all $i > k$. Therefore,

$$\begin{aligned}
\tilde{\phi}(\underline{x}) - \tilde{\phi}(\underline{y}) &= \sum_{j=1}^{d} (\lambda_j - \lambda_{j-1}) \left(\frac{\phi_j(\underline{x}) - \phi_j(\underline{y})}{2} \right) \\
&= \sum_{j=1}^{k} (\lambda_j - \lambda_{j-1}) \left(\frac{\phi_j(\underline{x}) - \phi_j(\underline{y})}{2} \right) \\
&\in \sum_{j=1}^{k} (\lambda_j - \lambda_{j-1}) K \\
&= \lambda_k K.
\end{aligned}$$

The third line is due to convexity and symmetry of K.

The coefficient of $\underline{b}_k$ in $(\tilde{\phi}(\underline{x}) - \tilde{\phi}(\underline{y}))$ is $\lambda_k \left(\frac{x_k - y_k}{2}\right) \neq 0$. In particular, this means $(\tilde{\phi}(\underline{x}) - \tilde{\phi}(\underline{y}))$ lies outside the span of $\underline{b}_1, \underline{b}_2, \ldots \underline{b}_{k-1}$ but inside the interior of $\lambda_k K$ (which is an open set that does not include $\underline{b}_k$); thus it is not a lattice vector.

Since $\tilde{\phi}(K)$ contains no two points separated by a nonzero lattice vector, it must have volume less than or equal to $\det(\Lambda)$ (by Blichfeldt's Lemma). Combining this with the Jacobian volume relation, we have$\lambda_1 \cdots \lambda_d \mathrm{vol}(K)/2^d <$ $\det \Lambda$ and the theorem is proved.

Remark 7.53. We have a reverse inequality $\frac{2^d}{d!} \det \Lambda \leq \lambda_1 \cdots \lambda_d \mathrm{vol} K$.

Proof. The cross polytope with vertices $\pm\frac{b_1}{\lambda_1}, \ldots, \pm\frac{b_d}{\lambda_d} \in K$ has volume $\frac{2^d}{d!} \frac{\det \Lambda}{\lambda_1 \cdots \lambda_d}$ and is contained in the closure of K. □

We now return to the proof that every Bohr set contains a large generalized arithmetic progression of low dimension. Since Bogolyubov's theorem already guarantees the existence of a suitable Bohr set, the proposition we will prove will ultimately make it possible to find a large GAP as needed for Freiman's theorem.

BohrGAPProp **Proposition 7.54.** *Let $R \subseteq \mathbb{Z}/n\mathbb{Z}$ with N prime, $0 < \epsilon < 1$, $|R| = d$. Then Bohr(R, ϵ) contains a proper generalized arithmetic progression of dimension at most d and size $\left(\frac{\epsilon}{d}\right)^d N$.*

Proof. Let $R = \{r_1, \ldots, r_d\}$, and let $\underline{r} = (r_1, \ldots, r_d) \in \mathbb{Z}^d$. $\Lambda := N \cdot \mathbb{Z}^d + \underline{r}\mathbb{Z} = N \cdot \mathbb{Z}^d \oplus \{0, 1, \ldots, (n-1)\} \cdot \underline{r}$ is a lattice, and is a direct sum as N is prime. Λ is the lattice formed by adding N multiples of vector $\underline{r}$ to each lattice point of the rectangular lattice $\Lambda_0 = N \cdot \mathbb{Z}^d$. Because of this, $\det \Lambda = N^{d-1}$ (each unit cell of Λ_0 contains N points and has volume N^d, so the volume of a unit cell of Λ must be N times less).

Let $K = \{\underline{x} \in \mathbb{R}^d : ||\underline{x}||_\infty := \max\{|x_1|, \ldots, |x_d|\} < 1\}$ be the interior of a hypercube centred at the origin with side lengths 2 . Further let $\lambda_1 \leq \cdots \leq \lambda_d$ be the successive minima of K with respect to Λ, and $\underline{b_1}, \ldots, \underline{b_d}$ be a corresponding directional basis (as defined in our proof of Minkowski's second theorem). Note that by the definition of K, $||\underline{b_i}||_\infty = \lambda_i$ (where $||\underline{b_i}||_\infty$ refers to $\max_j((\underline{b_i})_j)$ when $\underline{b_i}$ is written as a vector $((\underline{b_i})_1, (\underline{b_i})_2, \ldots (\underline{b_i})_d)$ in the original basis). Since $\underline{b_i} \in \Lambda$, $\exists x_i \in \mathbb{Z}/N\mathbb{Z}$ such that $\underline{b_i} \equiv x_i \underline{r} \pmod{N}$.

We claim that the generalized arithmetic progression

$$P := \left\{ \ell_1 x_1 + \cdots + \ell_d x_d : 0 \leq \ell_i < \frac{\epsilon N}{\lambda_i d} \right\}$$

is proper and contained in Bohr(R, ϵ).

Indeed,

$$\left|\left| \frac{x_i r_j}{N} \right|\right|_{\mathbb{R}/\mathbb{Z}} = \left|\left| \frac{(\underline{b_i})_j}{N} \right|\right|_{\mathbb{R}/\mathbb{Z}} \leq \left| \frac{(\underline{b_i})_j}{N} \right| \leq \frac{\lambda_i}{N}.$$

Thus $\forall p := \ell_1 x_1 + \cdots + \ell_d x_d \in P$,

$$\left|\left| \frac{p r_j}{N} \right|\right|_{\mathbb{R}/\mathbb{Z}} \leq \sum_{i=1}^{d} \ell_i \left|\left| \frac{x_i r_j}{N} \right|\right|_{\mathbb{R}/\mathbb{Z}} \leq \sum_{i=1}^{d} \left(\frac{\epsilon N}{\lambda_i d} \right) \left(\frac{\lambda_i}{N} \right) = \epsilon.$$

Hence $p \in$ Bohr(R, ϵ).

We now prove P is proper. Suppose $\ell = \sum \ell_i x_i$ and $\ell' = \sum \ell_i' x_i$ such that $\ell \equiv \ell' \pmod{N}$. We know that $0 \leq \ell_i, \ell_i' < \frac{\epsilon N}{\lambda_i d}$.

Let $\underline{b} := \sum_{i=1}^{d} (\ell_i - \ell_i')\underline{b_i} \equiv (\ell - \ell')\underline{r} \equiv 0 \pmod N$ (To clarify notation, note that $\underline{b_i}$ is not the ith component of $\underline{b}$ in the original basis, but rather is the ith basis vector). Then

$$\begin{aligned} ||\underline{b}||_\infty &\le \sum_{i=1}^{d} |\ell_i - \ell_i'|||b_i||_\infty \\ &\le \sum_{i=1}^{d} \frac{\epsilon N}{\lambda_i d}\lambda_i \\ &\le \epsilon N \\ &< N \end{aligned}$$

Thus $\underline{b} = \underline{0}$. Since b_i is a basis and $\sum_{i=1}^{d}(\ell_i - \ell_i')\underline{b_i} = \underline{b} = 0$, $\ell_i = \ell_i'$. Thus P is proper.

It is clear that $\dim(P) \le d$. Its size is

$$\begin{aligned} \prod_{i=1}^{d}\left(\frac{\epsilon N}{\lambda_i d}\right) &= \left(\frac{\epsilon}{d}\right)^d \frac{N^d}{\lambda_1 \cdots \lambda_d} \\ &\ge \left(\frac{\epsilon}{d}\right)^d \frac{N^d \mathrm{vol}(K)}{2^d \det \Lambda} \\ &= \left(\frac{\epsilon}{d}\right)^d N, \end{aligned}$$

where the inequality is by Minkowski's Second Theorem and we use the fact that $\mathrm{vol}(K) = 2^d$ and $\det\Lambda = N^{d-1}$. This completes the proof. □

§7.7 Proof of Freiman's Theorem

Putting the results of the previous two sections together, we have proven the following.

2aminus2a **Proposition 7.55.** *Let $A \subseteq \mathbb{Z}/N\mathbb{Z}$ with N prime and $|A| = \alpha N$. Then $2A - 2A$ contains a proper generalized arithmetic progression of dimension $d \le \frac{1}{\alpha^2}$ and size at least $\left(\frac{1}{4d}\right)^d N$.*

Proof. By Bogolyubov's Lemma, $2A - 2A$ contains a Bohr set of dimension $d \le \frac{1}{\alpha^2}$ and width $\frac{1}{4}$. By Proposition 7.54 with $\epsilon = \frac{1}{4}$, we get a generalized arithmetic progression of the desired size. □

Proof of Freiman's Theorem. The proof outline is as follows.

(1) By the Ruzsa's model lemma, a large subset of A is Freiman-8-isomorphic to a large subset $B \subset \mathbb{Z}/N\mathbb{Z}$, with N prime.
(2) $2B - 2B$ contains a large generalized arithmetic progression by the previous proposition, so $2A - 2A$ does also.
(3) By the Rusza Covering Lemma, A is covered by a small number of translates of this generalized arithmetic progression.

We now give the details.

Cite theorem numbers–done.

(1) By Corollary 7.29, there exists a prime $N \leq 2K^{16}|A|$ and $A' \subset A$, $|A'| \leq \frac{|A|}{8}$ such that A' is Freiman-8-isomorphic to some $B \subset \mathbb{Z}/N\mathbb{Z}$.
(2) By Proposition 7.55 with $\alpha = \frac{|B|}{N} \geq \frac{|A|}{8N} \geq \frac{1}{16K^{16}}$, we see that $2B - 2B$ contains a proper generalized arithmetic progression of dimension $d \leq 256K^{32}$ and size at least $\left(\frac{1}{4d}\right)^d N$.

Because B is Freiman-8-isomorphic to A', $2B - 2B$ is Freiman-2-isomorphic to $2A' - 2A'$ (an easy exercise, as a map that preserves information about length-8 sums for A' necessarily preserves length-2 sums of $2A' - 2A'$). Thus the generalized arithmetic progression in $2B - 2B$ is sent by our Freiman-2-isomorphism to a subset of $2A - 2A$, which must be a proper generalized arithmetic progression, call it P.
(3) Since $P \subset 2A - 2A$, $P + A \subset 3A - 2A$. Therefore,

$$|P + A| \leq |3A - 2A| \leq K^5|A|$$

by Plünnecke-Rusza. Since $N \geq |A'| \geq \frac{|A|}{8}$ and $|P| > \left(\frac{1}{4d}\right)^d N$, we find that $|P + A| \leq K'|P|$, where $K' = 8(4d)^d K^5$.

By the Ruzsa covering lemma, $A \subseteq X + P - P$ for some $X \subset A$ with $|X| \leq K'$. $X + P - P$ is contained in a generalized arithmetic progression by considering

$$X \subset \{\sum_{x \in X} a_x x : a_x \in \{0, 1\}\}.$$

The dimension of this arithmetic progression is at most $d + K'$, and its size is at most $2^{d+K'}|P| \leq 2^{d+K'}|2A - 2A| \leq 2^{d+K'}K^4|A|$, finishing the proof.

□

Recall that Freiman's theorem states that if $A \subset \mathbb{Z}$ and $|A + A| \leq K|A|$, then A is contained in a generalized arithmetic progression (GAP) of dimension at most $d(K)$ and size at most $f(K)|A|$. What can we say about

lec24 (need to add references): Vishesh Jain the growth of $d(K)$ and $f(K)$ as functions of K? An upper bound of $d(K) = 2^{(2K)^{O(1)}}$ and $f(K) = 2^{2^{(2K)^{O(1)}}}$ can be deduced by keeping track of constants in the proof of Freiman's theorem presented earlier. On the other hand, if A is a subset of the integers of size m with no additive relations whatsoever, then $K = \frac{m+1}{2}$, any GAP containing A has dimension at least $m-1$ and the size of the GAP is at least 2^{m-1}. This gives lower bounds on $d(K)$ and $f(K)$ that are exponentially better than the upper bounds mentioned above. It is conjectured that the optimal upper bounds for Freiman's theorem should match these lower bounds (up to constant factors). In the sequel, for ease of notation, we will use $K^{O(1)}$ to refer to $(2K)^{O(1)}$.

OptimalFreiman **Conjecture 7.56** (Optimal bounds for Freiman's Theorem). *In the statement of Freiman's theorem, it is possible to take $d(K) = O(K)$ and $f(K) = 2^{O(K)}$.*

Let us briefly summarize the progress on this conjecture. In step (3) of the proof of Freiman's theorem provided above, if we replace the application of Ruzsa's Covering Lemma by Chang's Covering Lemma (TO DO: Add reference), one can instead obtain the exponentially better bounds $d(K) = K^{O(1)}$ and $f(K) = 2^{K^{O(1)}}$. The best known bounds on Freiman's theorem are due to Sanders (TO DO: Add reference), who improved the constants in Bogulyubov's Lemma to show that one can take $d(K) = K(\log K)^{O(1)}$ and $f(K) = 2^{K(\log K)^{O(1)}}$.

For $\mathbb{F}_2^n$ (and more generally, for $\mathbb{F}_p^n$ for all primes p), the analog of Conjecture 7.56 has already been settled (TO DO: Add reference). More precisely, if $A \subseteq \mathbb{F}_2^n$ and $|A+A| \leq K|A|$, then $|\langle A\rangle| \leq f(K)|A|$ where the *exact* optimal $f(K)$ is known. Asymptotically, $f(K) = K^{O(1)}4^K$. However, one can ask if there always exists a "small" (size $K^{O(1)}|A|$) subspace of $\mathbb{F}_2^n$ containing a "large" fraction ($K^{-O(1)}$) of A. This is the content of one of the most important open problems in additive combinatorics.

PFR-conjecture **Conjecture 7.57** (Polynomial Freiman-Ruzsa Conjecture for $\mathbb{F}_2^n$). *Let $A \subseteq \mathbb{F}_2^n$ such that $|A+A| \leq K|A|$. Then, there exists a subspace $V \leq \mathbb{F}_2^n$ such that $|V| \leq K^{O(1)}|A|$ and $|V \cap A| \geq K^{-O(1)}|A|$.*

Remark 7.58. By an application of Ruzsa's covering lemma, which we leave as an exercise to the reader, this is equivalent to saying that A can be covered by $K^{O(1)}$ many translates of V.

The best known result towards Conjecture 7.57 is due to Sanders (TO DO: Add reference) who proved it with quasipolynomial bounds instead of polynomial bounds i.e. i.e. bounds of the form $e^{(\log K)^{O(1)}}$ instead of $e^{O(\log K)}$.

Part of the reason that the Polynomial Freiman-Ruzsa Conjecture is so attractive is that it admits several different equivalent formulations. Here is one such statement:

equiv-PFR-conjecture **Conjecture 7.59** (Equivalent formulation of Conjecture 7.57). *If $f\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ is such that $|\{f(x+y)-f(x)-f(y)\colon x,y \in \mathbb{F}_2^n\}| \leq K$, then there exists a linear map $g\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that $S = im(f-g) = \{f(x)-g(x)\colon x \in \mathbb{F}_2^n\}$ satisfies $|S| = K^{O(1)}$.*

Remark 7.60. Proving the above conjecture with $|S| \leq 2^K$ is very easy. Simply pick g to agree with f on an arbitrarily chosen basis. Then S is contained in the linear span (over $\mathbb{F}_2$) of at most K vectors $\{f(x+y)-f(x)-f(y)\colon x,y \in \mathbb{F}_2^n\}$ and hence, $|S| \leq 2^K$.

The analog of Conjecture 7.57 for the integers has also attracted a lot of interest.

PFR-conj-integers **Conjecture 7.61** (Polynomial Freiman-Ruzsa Conjecture for $\mathbb{Z}$). *If $A \subset \mathbb{Z}$ satisfies $|A+A| \leq K|A|$, then there exists a convex progression $P := \{a_0+n_1a_1+\cdots+n_da_d : (n_1,\ldots,n_d) \in \mathbb{Z}^d \cap B\}$, where $B \subset \mathbb{R}^d$ is some convex body in $\mathbb{R}^d$ and $a_0,\ldots a_d \in \mathbb{Z}$, such that $d = O(1+\log K)$, $|B \cap \mathbb{Z}^d| \leq K^{O(1)}|A|$ and$|P \cap A| \geq K^{-O(1)}|A|$.*

As before, this is equivalent to saying that A can be covered by $K^{O(1)}$ many translates of P.

Recall that in the proof of Freiman's theorem, the application of Bogulyubov's lemma formed a key step. We also mentioned above that a result of Sanders improving the constants in Bogulyubov's lemma leads to the best known bounds for Freiman's theorem. The situation for the Polynomial Freiman-Ruzsa conjectures is similar in the sense that the best known results have been obtained via improvements for Bogulyubov's lemma. In fact, the following conjectured strengthenings of Bogolyubov's lemma would imply the corresponding Polynomial Freiman-Ruzsa conjecture.

PBR-conjecture-F2 **Conjecture 7.62** (Polynomial Bogulyubov-Ruzsa Conjecture for $\mathbb{F}_2^n$). *If $A \subseteq \mathbb{F}_2^n$ with $|A| = \alpha 2^n$, then $2A-2A$ contains a subspace of codimension $O(\log \frac{1}{\alpha})$.*

The best known result here is also due to Sanders, who proved the conjecture with $(\log \frac{1}{\alpha})^{O(1)}$ instead of $O(\log \frac{1}{\alpha})$.

Remark 7.63. Proving the above conjecture with $2A-2A$ replaced by $\ell A - \ell A$ for any ℓ would still imply Conjecture PFR-conjecture 7.57.

PBR-conjecture-Zmodn **Conjecture 7.64** (Polynomial Bogulyubov-Ruzsa Conjecture for $\mathbb{Z}/N\mathbb{Z}$). *Let N be a prime number. If $A \subset \mathbb{Z}/N\mathbb{Z}$ with $|A| = \alpha N$, then $2A - 2A$ contains a proper convex progression of dimension $O(\log \frac{1}{\alpha})$ and size at least $\alpha^{O(1)} N$*

While the Polynomial Bogulybov-Ruzsa conjectures imply the corresponding Polynomial Freiman-Ruzsa conjectures, it is unknown whether the converse is true.

§7.8 Additive Energy and the Balog-Szemerédi-Gowers Theorem

The theme of Freiman's theorem is to express in multiple ways the fact that a set has a lot of additive structure. We now look at yet another way of expressing (the presence of) additive structure in a set.

AdditiveEnergyDefn *Definition* 7.65. Let Γ be an abelian group. For finite subsets $A, B \subseteq \Gamma$, the *additive energy* between A and B is defined to be $E(A,B) := |\{(a_1,a_2,b_1,b_2) \in A \times A \times B \times B \colon a_1 + b_1 = a_2 + b_2\}|$.

A very useful equivalent way of viewing the additive energy is the following: let $r_{A,B}(n) := |\{(a,b) \in A \times B \,:\, n = a+b\}|$ denote the number of ways the element n can be represented as a sum $a+b$ with $a \in A, b \in B$. Then, it is immediately seen that $E(A,B) = \sum_{n \in \Gamma} r_{A,B}(n)^2$.

We will use $E(A)$ to denote $E(A,A)$. Note that $E(A) \geq |A|^2$ since we have $|A|^2$ "trivial" quadruples of the form (a_1, a_2, a_2, a_1) with $a_1, a_2 \in A$. We also have the trivial upper bound $E(A) \leq |A|^3$ since specifying three elements of any valid quadruple uniquely determines the fourth element. Hence, the additive energy $E(A)$ always ranges between $|A|^2$ and $|A|^3$. As examples, note that if $\Gamma = \mathbb{Z}$ and $A = [N]$, then $E(A) = \Theta(N^3)$ whereas if A has no additive structure, then $E(A) = |A|^2$. In general, we should think of additive energy "close to cubic" as a sign of additive structure. Since we have also seen that "small doubling" is a sign of additive structure, it is natural to ask how, if at all, the conditions $|A+A| \leq K|A|$ and $E(A) \geq c|A|^3$ are related.

SmlDoublingLargeAddEnergy **Proposition 7.66** (Small doubling implies large additive energy). *Let A be a finite subset of an abelian group Γ. If $|A+A| \leq K|A|$, then $E(A) \geq \frac{|A|^3}{K}$.*

Proof. Let $r(x) = |\{(a,a') \in A \times A : \ a + a' = x\}|$. Then, as observed earlier, $E(A) = \sum_x r(x)^2$. Note that $\sum_x r(x) = |A|^2$ since every pair (a,a') is accounted for exactly once (in the summand $r(a+a')$) on the left hand side. Note also that r is supported on $A+A$. Now, an application of Cauchy-Schwarz shows that

$$\begin{aligned} |A+A|.E(A) &= |A+A|\left(\sum_x r(x)^2\right) \\ &\geq \left(\sum_x r(x)\right)^2 \\ &= |A|^4 \end{aligned}$$

which shows that $E(A) \geq \frac{|A|^4}{|A+A|} \geq \frac{|A|^3}{K}$ – the last inequality following from the assumption that $|A+A| \leq K|A|$. □

However, the reverse implication is not true. Indeed, consider $\Gamma = \mathbb{Z}$ and $A = [m] \cup \{m \text{ additional elements introducing no new additive relations}\}$. Then, $E(A) \geq E([m]) = \Theta(|A|^3)$ whereas $|A+A| \geq m^2 \geq \frac{m}{2}|A|$. Essentially, a set can have large additive energy if a "large" fraction of it has additive structure whereas the small doubling condition really takes into account the set as a whole. Therefore, for some sort of converse to Proposition 7.66, it is natural to look for statements about dense subsets of sets with large additive energy. The Balog-Szemerédi-Gowers theorem provides such a converse by showing that any set with large additive energy contains a dense subset with small doubling.

BSG-Thm **Theorem 7.67** (Balog-Szemerédi-Gowers (BSG)). *Let Γ be an abelian group and let $A, B \subseteq \Gamma$ with $|A| = |B| = n$. If $E(A,B) \geq \frac{n^3}{K}$, then there exist $A' \subseteq A, B' \subseteq B$ with $|A'|, |B'| \geq K^{-O(1)}n$ such that $|A'+B'| \leq K^{O(1)}n$.*

Remark 7.68. Setting $A = B$ in the above theorem gives us subsets A', A'' of A such that $|A'|, |A''| \geq K^{-O(1)}n$ and $|A'+A''| \leq K^{O(1)}n$. Then, Ruzsa's triangle inequality shows that

$$\begin{aligned} K^{-O(1)}n.|A'+A'| &\leq |A''|.|A'+A'| \\ &\leq |A'+A''|.|A'+A''| \\ &\leq K^{O(1)}n^2 \end{aligned}$$

which shows that $|A'+A'| \leq K^{O(1)}n$ i.e. A' is a dense subset of A with small doubling.

We will deduce the BSG theorem from the following graph analog, which will be proved via the technique of dependent random choice. Given a bipartite graph G between vertex sets A and B, defined $A+_G B := \{a+b\colon (a,b) \in A \times B \text{ is an edge of } G\}$.

GraphBSG-Thm **Theorem 7.69** (Graph BSG). *Let Γ be an abelian group, and let $A, B \subseteq \Gamma$ with $|A| = |B| = n$. Let G be a bipartite graph between the vertex sets A and B such that G has at least $\frac{n^2}{K}$ edges. If $|A +_G B| \leq Kn$, then there exist $A' \subseteq A, B' \subseteq B$ with $|A'|, |B'| \geq K^{-O(1)}n$ such that $|A'+B'| \leq K^{O(1)}n$.*

Before giving the proof of this theorem, let us show how it implies the BSG theorem.

GraphBSGimpliesBSG **Lemma 7.70.** *The Graph BSG theorem (Theorem 7.69) implies the BSG theorem (Theorem 7.67).*

Proof. Suppose $|A| = |B| = n$ with $E(A,B) \geq \frac{n^3}{K}$. As before, let $r(x) = |\{(a,b) \in A \times B\colon a+b = x\}|$ and let $S = \{x \in A+B\colon r(x) \geq \frac{n}{2K}\}$ denote the set of "popular sums". Note that by definition, $\sum_{x \notin S} r(x)^2 \leq \max_{x\notin S} r(x) \times \sum_{x\notin S} r(x) \leq \frac{n}{2K} \times n^2$, where the last inequality uses the trivial bound $\sum_{x\notin S} r(x) \leq \sum_x r(x) = |A|.|B| = n^2$. Then:

$$\begin{aligned} \frac{n^3}{K} &\leq E(A,B) \\ &= \sum_x r(x)^2 \\ &= \sum_{x\in S} r(x)^2 + \sum_{x\notin S} r(x)^2 \\ &\leq \sum_{x\in S} r(x)^2 + \frac{n^3}{2K} \end{aligned}$$

which shows that

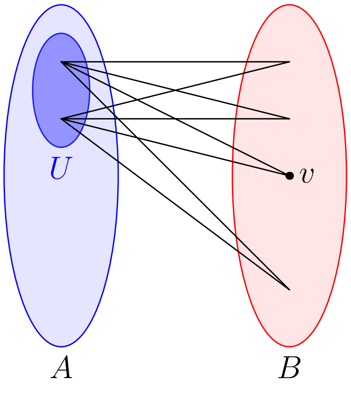
$$\frac{n^3}{2K} \leq \sum_{x\in S} r(x)^2 \leq n \sum_{x \in S} r(x)$$

and hence, $\sum_{x\in S} r(x) \geq \frac{n^2}{2K}$. Let G be the bipartite graph between vertex sets A and B where $(a, b) \in A \times B$ is an edge in G if and only if $a+b \in S$. By the previous paragraph, the number of edges in G equals $\sum_{x\in S} r(x) \geq \frac{n^2}{2K}$. Moreover, since $|A+_G B|$ denotes the number of "popular sums", each popular sum is associated with at least $\frac{n}{2K}$ many distinct pairs (a, b) and since there are only n^2 many such pairs (a, b), it follows that $\frac{n}{2K}.|A +_G B| \leq n^2$ i.e. $|A +_G B| \leq 2Kn$. Since G satisfies the hypotheses of Theorem GraphBSG-Thm 7.69 with K replaced by $2K$, we can apply the Graph BSG theorem to finish the proof. □

We now give a proof of the Graph BSG theorem via a series of lemmas. Lec25: Dhroova Aiylam

PathsOfLength2 **Lemma 7.71** (Paths of Length 2)**.** *Suppose G is a bipartite graph between vertex sets A and B which has at least $\delta|A||B|$ edges, and let $\epsilon > 0$ be given. Then there exists $U \subset A$ with $|U| \geq \delta|A|/2$ such that between at least $(1-\epsilon)|U|^2$ ordered pairs $(u, u') \in U \times U$, there are at least $\epsilon\delta^2|B|/2$ many paths of length 2 between u and u'.*

The proof uses dependent random choice to pick U; in particular, we will pick a random vertex in B random and take U to be its neighborhood. The idea is that if $u, u' \in U$ have few common neighbors in B, then it is unlikely we chose them.



Proof. Choose $v \sim \mathrm{Unif}(B)$ and let $U = N(v)$. By convexity,

$$\mathbb{E}[|U|^2] \geq (\mathbb{E}[|U|])^2 = \delta^2|A|^2.$$

We say a pair $(a, a') \in A^2$ is *bad* if $\operatorname{codeg}(a, , a') < \epsilon\delta^2|B|/2$. For (a, a') a fixed bad pair, we have

$$\mathbb{P}(a, a' \in U) = \mathbb{P}(v \in N(a) \cap N(a')) < \frac{1}{2}\epsilon\delta^2.$$

Let Z be the random variable which counts the number of bad pairs in $U \times U$; from the above and linearity of expectation, $\mathbb{E}[Z] \leq \epsilon\delta^2|A|^2/2$. It follows that

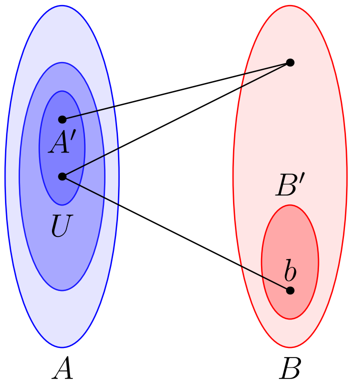
$$\mathbb{E}\left[|U|^2 - \frac{Z}{\epsilon}\right] \geq \delta^2|A|^2 - \frac{1}{2}\delta^2|A|^2 \geq \frac{1}{2}\delta^2|A|^2.$$

Thus for some choice of $v \in B$ we have $|U|^2 - Z/\epsilon \geq \delta^2|A|^2/2$. In this case, $|U|^2 \geq \delta^2|A|^2/2$ so that $|U| \geq \delta|A|/2$; moreover, $Z \leq \epsilon|U|^2$ so that at most ϵ-fraction of pairs in U are *bad*. Equivalently, there are at $(1-\epsilon)|U|^2$ ordered pairs $(u, u') \times U \times U$ which have at least $\epsilon\delta^2|B|/2$ paths of length 2 between them. □

PathsOfLength3 **Lemma 7.72** (Paths of Length 3). *For $0 < \delta < 1$, let G be a bipartite graph between vertex sets A and B with at least $\delta|A||B|$ edges. Then there exist $A' \subset A$ and $B' \subset B$, where $|A'| \geq c\delta|A|$ and $|B'| \geq c\delta|B|$ for some fixed small constant c, such that* every *pair $(a, b) \in A' \times B'$ is connected by at least $\delta^{O(1)}|A||B|$ paths of length 3.*

Proof Sketch. Recall a useful trick from Chapter 2: in a graph with large average degree, there is a large subset of vertices with large minimum degree.

(1) Remove vertices in A which have degree less thant $\delta|B|/2$; by Lemma 2.17 we incur at most constant factor shrinkage.
(2) Apply Lemma 7.71 to find $U \subset A$, which has $|U| \geq \delta|A|$ and almost all pairs in U have codegree at least $\delta^{-O(1)}|B|$.
(3) Remove vertices in U which are in many bad pairs, yielding A' with $|A'| \geq |U|/2$.
(4) Now every vertex in A' has degree at least $\delta|B|/2$ by (1), so there exists $B' \subset B$ with size at least $\delta|B|/4$ such that every vertex in B' has at least $\delta|A'|/4$ neighbors in A'.



To see why this choice of A', B' should work, observe that for all $a \in A'$ and $b \in B'$, b is adjacent to a large fraction of the vertices $a' \in A$, and almost all such a' have large codegree with a. □

GraphBSG-Thm

We are now ready to prove the Graph BSG theorem (Theorem 7.69).

PathsOfLength3

Proof. By Lemma 7.72, there exist $A' \subset A$ and $B' \subset B$, $|A'|, |B'| \geq K^{-O(1)}n$, such that for all $a \in A'$ and $b \in B'$, there are at least $K^{-O(1)}n^2$ many paths of length 3 between a and b. We claim that $|A' + B'| \leq K^{O(1)}n$. Indeed, for all $a' \in A'$ and $b' \in B'$, there exist $K^{-O(1)}n^2$ many $(u, v) \in A \times B$ such that $a + v, v + u, u + b \in A +_G B$. Since

$$a + b = (a + v) - (v + u) + (u + b)$$

every element of $A' + B'$ can be written as $x - y + z$, $x, y, z \in A +_G B$, in at least $K^{-O(1)}n^2$ many ways. Thus

$$K^{-O(1)}n^2|A' + B'| \leq |A +_G B|^3 \leq K^3n^3$$

so that $|A' + B'| \leq K^{O(1)}n$ as desired. □

CHAPTER 8

The Sum-Product Problem

Just as we defined the sumset $A+B=\{a+b : a\in A, b\in B\}$, we can define $A\cdot B=\{ab : a\in A, b\in B\}$.

ESConjecture **Conjecture 8.1** (Erdös-Szemerédi). *For all* $A\subset\mathbb{R}$, $\max\{|A+A|, |A\cdot A|\}\geq |A|^{2-o(1)}$.

The following example shows that the $o(1)$ term is necessary in the exponent of the RHS.

Example 8.2. Let $A=[N]$; clearly $|A+A|=2N-1$. On the other hand, $|A\cdot A|$ (the so-called Erdös multiplication table problem) is harder to bound. Using techniques from analytic number theory, Ford was able to give the precise estimate

$$|A\cdot A|=\Theta\left(\frac{N^2}{(\log N)^{\delta}(\log\log N)^{3/2}}\right)$$

where $\delta=1-(1+\log\log 2)/\log 2\approx 0.086$. However, there is a more elementary way to see that $|A\cdot A|=o(N^2)$.

Indeed, by the Hardy-Ramanujan theorem, all but $o(N)$ of the integers less than or equal to N have $(1+o(1))\log\log N$ prime factors (counting multiplicity). Thus all but at most $o(N^2)$ of the products ab, $a,b\leq N$, have $(2+o(1))\log\log N$ prime factors. Yet by Hardy-Ramanjuan once more, all but $o(N^2)$ of the integers less than or equal to N^2 have $(1+o(1))\log\log N^2=(1+o(1))\log\log N$ prime factors and so cannot appear in the multiplication table.

To furnish a lower bound, we consider integers $n\leq N^2$ of the form pm, where p is a prime in $(N^{2/3}, N]$ and $m\leq N$. By Prime Number Theorem, the number of choices for (p,m) is at least $(1+o(1))N^2/\log N$. On the other hand, every such n has at most 2 such representations since $n\leq N^2$

can have at most two prime factors greater than $N^{2/3}$. It follows that there are at least $(1/2 + o(1))N^2/\log N$ distinct such integers n, whence $|A \cdot A|$ is at least this large.

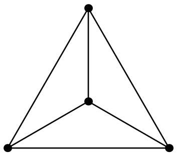
ESConjecture

Conjecture 8.1 is still open, and in fact quite far from being solved. Erdös and Szemerédi first proved a lower bound of $|A|^{1+c}$ for some small absolute constant $c > 0$. Later, Elekes was able to show that $\max\{|A + A|, |A \cdot A|\} \gtrsim |A|^{5/4}$. In 2009, Solymosi proved a lower bound of $|A|^{4/3-o(1)}$; this was recently improved to $|A|^{4/3+c}$ for some small $c > 0$.

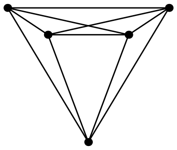
§8.1 Crossing number inequality

Definition 8.3 (Crossing Number). Let the crossing number $\mathrm{cr}(G)$ be the minimum number of edge crossings in a planar drawing of G (edges are allowed to be curves).

For example, the graph K_4 is planar so $\mathrm{cr}(K_4) = 0$.



On the other hand, K_5 is not planar. However, the following drawing shows that it is possible to draw K_5 with a single edge crossing, whence $\mathrm{cr}(K_5) = 1$.



A natural question to ask is: if a graph has lots of edges, must there be lots of crossings? The crossing-number inequality answers this question in the affirmative, but only asymptotically. For instance, it is known that $\mathrm{cr}(K_n) = \Theta(n^4)$ and $\mathrm{cr}(K_{n,n}) = \Theta(n^4)$ but the constants are not known. The problem of determining $\mathrm{cr}(K_{n,n})$ is sometimes referred to as the Turán brick factory problem.

CrossingNumberInequality **Theorem 8.4** (Crossing Number Inequality). *If $G = (V, E)$ is a graph with $|E| \geq 4|V|$, then $cr(G) \geq |E|^3/(64|V|^2)$.*

If G is planar, then $|E| \leq 3|V|$ by Euler's formula. A graph G can clearly be made planar by removing $\mathrm{cr}(G)$ edges, so that $\mathrm{cr}(G) \geq |E| - 3|V|$.

This bound is not quite good enough – for dense graphs the crossing number inequality gives a lower bound which is quadratic in $|E|$ – but it is useful nonetheless.

Proof. Fix G with $|E| \geq 4|V|$, and let $0 < p \leq 1$ be a value which we choose in the sequel. Pick $V' \subset V$ by including each vertex of V independently with probability p, and let $G' = (V', E')$ denote the induced subgraph. As before we have $\mathrm{cr}(G') \geq |E'| - 3|V'|$; we note that $\mathbb{E}[|V'|] = p|V|$, $\mathbb{E}[|E'|] = p^2|E|$. Moreover, $\mathbb{E}[\mathrm{cr}(G')] \leq p^4\mathrm{cr}(G)$ by reusing the optimal drawing of G; a crossing occurs in G' if and only if each of the four vertices of its edges is included in V', which has probability p^4, whence the claim follows from linearity. We conclude that

$$p^4\mathrm{cr}(G) \geq \mathbb{E}[\mathrm{cr}(G')] \geq \mathbb{E}[|E'|] - 3\mathbb{E}[|V'|] = p^2|E| - 3p|V|$$

i.e. $\mathrm{cr}(G) \geq p^{-2}|E| - 3p^{-3}|V|$. The result now follows from putting $p = 4|V|/|E| \leq 1$. □

Lec26: Paxton Turner

§8.2 Szemerédi-Trotter and Incidence Geometry

Methods from geometry, first used by Elekes [**13**], have obtained results on the sum-product problem. First we present a motivating example that uses related techniques on the unit distance problem.

Theorem 8.5. *A set of n points in the real plane $\mathbb{R}^2$ make at most $O(n^{4/3})$ unit length distances.*

This is an old problem of Erdős [**14**], and the conjectured exponent is $1 + o(1)$. Somewhat surprisingly, the exponent $\frac{4}{3}$ is the best known and was obtained by Spencer, Szemerédi, and Trotter [**48**].

Proof. First remove all points that are involved in at most 2 unit distances, and call the remaining set S. Define a graph G whose vertices are the points of S. For all $p \in S$, draw a circle C_p of radius 1 about p. The circle C_p is divided into arcs by the points of $C_p \cap S$, so let the edges of G be arcs connecting consecutive points on the circle. Each pair of circles can cross at most twice, so the number of crossings in G is at most n^2. By the crossing

CrossingNumberInequality

number inequality, Theorem 8.4, we get $n^2 \gtrsim \frac{m^3}{n^2}$. Since each edge represents a unit distance, this proves the theorem. □

Definition 8.6 (Incidences). Let $P \subset \mathbb{R}$ be a set of points and $L \subset \mathbb{R}$ be a set of lines. Then the set of incidences is defined to be

$$I(P, L) = \{(p, \ell) \in P \times L : p \in \ell\}.$$

The Szemerédi-Trotter Theorem tells us how large the set of incidences can be.

Sze-Trot **Theorem 8.7** (Szemerédi-Trotter Theorem). *If $P \subset \mathbb{R}$ is a set of points, and $L \subset RR$ is a set of lines, then*

$$I(P, L) = O(|P|^{2/3}|L|^{2/3} + |P| + |L|)$$

Note that P and L play symmetric roles in this inequality. Moreover this theorem is tight if we let P be a $2n^{2/3} \times n^{1/3}$ grid and let L be the set of lines $y = mx + b$ where $1 < m < n^{1/3}$ and $1 \leq b \leq n^{1/3}$. Here $|P| \asymp |L| \asymp n$ and $|I(P, L)| \asymp n^{4/3}$. The proof shows that the exponent $\frac{4}{3}$ is morally the same as that obtained for the unit distance problem.

Proof. Remove the lines L that have no points in P. Define a graph G whose points are P and whose edges are consecutive points on the lines L (here we remove any trailing rays). The number of edges is $|I(P, L)| - |L|$. Trivially, the number of crossings is at most $\binom{|L|}{2}$. So if $|I(P, L)| - |L| \geq 4|P|$, by the crossing number inequality,

$$|L|^2 \gtrsim \frac{(|I(P, L)| - |L|)^3}{|P|^2}$$

which gives the statement on rearranging. On the other hand if $|I(P, L)| - |L| < 4|P|$, then the statement is trivial. □

Now we come to the first application toward the sum-product problem from incidence geometry.

Theorem 8.8 (Elekes). *If $A \subset \mathbb{R}$, then $|A + A||AA| \gtrsim |A|^{5/2}$.*

Corollary 8.9. *If $A \subset \mathbb{R}$, then* $\max(|A + A|, |AA|) \gtrsim |A|^{5/4}$.

Proof. Let $P = (A + A) \times (AA) \subset \mathbb{R}^2$. The lines L are defined to be all $y = a(x - b)$ where $a, b \in A$. Hence $|L| = |A|^2$ and it's easy to see that each line contains A points from $|P|$. Therefore $|I(P,L)| \geq |A|^3$. By Szemerédi-Trotter,

$$\begin{aligned}|A|^3 \leq |I(P,L)| &\lesssim |P|^{2/3}|L|^{2/3} + |P| + |L| \\ &\lesssim (|A+A||AA|)^{2/3}|A|^{4/3} + |A+A||AA| + |A|^2 \\ &= (|A+A||AA|)^{2/3}(|A|^{4/3} + |A+A|^{1/3}|AA|^{1/3}) + o(|A|^3) \\ &\lesssim (|A+A||AA|)^{2/3}|A|^{4/3}.\end{aligned}$$

Rearranging yields Elekes' result.

□

Remark 8.10. This bound is not true over finite field, for example $\mathbb{F}_p^2$. This is because we're implicitly using topology: the crossing number inequality relies on the Euler characteristic. However, there is a version of the above theorem in $\mathbb{C}^2$, thought we don't mention it here.

The following result of Solymosi is close to the best known.

Theorem 8.11 (Solymosi [46]). *Let* $A \subset \mathbb{R}_{>0}$. *Then* $|AA||A+A|^2 \gtrsim \frac{|A|^4}{4\lceil \log |A| \rceil}$.

Corollary 8.12. *If* $A \subset \mathbb{R}_{>0}$, *then* $\max |A+A|, |AA| \gtrsim \frac{|A|^{4/3}}{\log |A|}$.

The proof relies on the *mutliplicative energy* which is a direct analog of the additive energy.

Definition 8.13 (Multiplicative energy). If A is a subset of an Abelian group, define the *multiplicative energy* $E_x(A)$ to be

$$E_x(A) = |\{(a,b,c,d) \in A^4 : \text{there exists}\, \lambda \in \mathbb{R} \text{ s.t.} (a,b) = \lambda(c,d)\}|.$$

By Cauchy-Schwarz, high multiplicative energy implies that the product set is small:

$$E_x(A)|AA| = \left(\sum_{x \in AA} |\{(a,b) \in A^2 : ab = x\}|\right)|AA| \geq |A|^4.$$

Proof. Observe that Solymosi's result follows from the statement:

$$E_x(A) \leq 4|A+A|^2 \lceil \log_2 |A| \rceil \qquad (10)$$ eqn:soly

Let $A/A := \{a/b : a, b \in A\}$ denote the quotient set. Then

$$E_x(A) = \sum_{s \in A/A} |(s \cdot A) \cap A|^2 = \sum_{i=0}^{\log_2 |A|} \sum_{\substack{s \in A/A \\ 2^i \leq |(s \cdot A) \cap A| < 2^{i+1}}} |(s \cdot A) \cap A|^2.$$

By pigeonhole exists i such that setting $D = \{s : 2^i \leq |(s \cdot A) \cap A| < 2^{i+1}\}$ gives

$$\frac{E_x(A)}{\lceil \log_2 |A| \rceil} \leq \sum_{s \in D} |(s \cdot A) \cap A|^2 \leq |D| 2^{2i+2}$$

Enumerate the elements of D in increasing order to be $\{s_1, \ldots, s_m\}$ with $m := |D|$. Now we define a set L of $m+1$ lines $\ell_1, \ldots, \ell_{m+1}$ as follows. For $1 \leq i \leq m$, define ℓ_i to be the lines $y = s_i x$. The half-line ℓ_{m+1} is defined to be $x = \min(A)$ and we only include the part that lies above the line ℓ_m. Define $L_j = \ell_j \cap (A \times A)$. Then $|L_j + L_{j+1}| = |L_j||L_{j+1}|$ because the sum-set

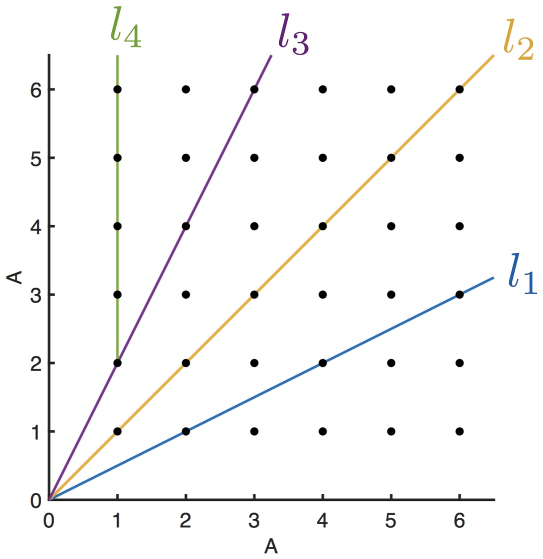


Figure 1. An illustration of the point-line configuration in Solymosi's proof when $m = 3$ and $A = \{1, \ldots, 6\}$.

is the skew-grid between lines ℓ_j and ℓ_{j+1}. Moreover $L_j + L_{j+1}$ and $L_i + L_{i+1}$

are disjoint if $i \neq j$. Thus

$$\begin{aligned} |A+A|^2 &= |(A \times A) + (A \times A)| \\ &\geq \sum_{j=1}^{m} |L_j + L_{j+1}| \\ &= \sum_{j=1}^{m} |L_j||L_{j+1}| \geq m2^{2i} \end{aligned}$$

But above we showed that $\frac{E_x(A)}{\lceil \log_2 |A| \rceil} \leq m2^{2i+2}$. Combining the last two equations implies Equation 10 and the proof of Solymosi's result.

□

Remark 8.14. One can slightly improve the exponent by analyzing what happens between two lines that are non-consecutive in Solymosi's proof.

It is natural to investigate the sum-product problem over the finite fields.

Theorem 8.15 (Bourgain-Katz-Tao [7]). *For all $\delta > 0$, there exists $c > 0$ such that if p is prime, $A \subset \mathbb{F}_p$, and $|A| \leq p^{1-\delta}$, then*

$$\max\{|A+A|, |AA|\} \geq |A|^{1+c}.$$

One can interpret this theorem as saying that $\mathbb{F}_p$ does not have any approximate sub-rings, for in any bona-fide subring A, we know that $A+A = AA = A$.

References

Oei17 1. *The on-line encyclopedia of integer sequences*, `https://oeis.org/A186705`, Accessed: 2017-09-19.

ACNS82 2. Miklós Ajtai, Vašek Chvátal, Monroe M Newborn, and Endre Szemerédi, *Crossing-free subgraphs*, North-Holland Mathematics Studies **60** (1982), 9–12.

ARS99 3. Noga Alon, Lajos Rónyai, and Tibor Szabó, *Norm-graphs: variations and applications*, Journal of Combinatorial Theory, Series B **76** (1999), no. 2, 280–290.

BHP01 4. Roger C Baker, Glyn Harman, and János Pintz, *The difference between consecutive primes, ii*, Proceedings of the London Mathematical Society **83** (2001), no. 3, 532–562.

BL96 5. V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc. **9** (1996), no. 3, 725–753. MR 1325795 (96j:11013)

BBK13 6. Pavle VM Blagojević, Boris Bukh, and Roman Karasev, *Turán numbers for ks, t-free graphs: topological obstructions and algebraic constructions*, Israel Journal of Mathematics **197** (2013), no. 1, 199–214.

BKT04 7. J. Bourgain, N. Katz, and T. Tao, *A sum-product estimate in finite fields, and applications*, Geometric and Functional Analysis **14** (2004), no. 1, 27–57.

Bro66 8. William G Brown, *On graphs that do not contain a thomsen graph*, Canad. Math. Bull **9** (1966), no. 2, 1–2.

Buk15 9. Boris Bukh, *Random algebraic construction of extremal graphs*, Bulletin of the London Mathematical Society **47** (2015), no. 6, 939–945.

Dav13 10. Harold Davenport, *Multiplicative number theory*, vol. 74, Springer Science & Business Media, 2013.

Dic52 11. LE Dickson, *History of the theory of numbers, ii. chelsea*, New York (1952).

durrett2010probability 12. Rick Durrett, *Probability: theory and examples*, Cambridge university press, 2010.

Ele97 13. G. Elekes, *On the number of sums and products*, Acta Arithmetica **LXXX1.4** (1997), 365–367.

Erd46 14. p. Erdős, *On the number of sums and products*, American Mathematical Monthly **53** (1946), 248–250.

ET36 15. P. Erdős and Paul Turán, *On Some Sequences of Integers*, J. London Math. Soc. **11** (1936), 261–264. MR 1574918

ERS66 16. Paul Erdős, A Rényi, and VT Sós, *On a problem of graph theory*, Studia Sci. Math. Hungar **1** (1966), no. 215, C235.

FS11DRC 17. Jacob Fox and Benny Sudakov, *Dependent random choice*, Random Structures Algorithms **38** (2011), no. 1-2, 68–99. MR 2768884

Fur77 18. H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256. MR 0498471 (58 #16583)

Fur81 19. ———, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, Princeton, N.J., 1981, M. B. Porter Lectures. MR 603625

FK78 20. H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291. MR 531279 (82c:28032)

FKO82 21. H. Furstenberg, Y. Katznelson, and D. Ornstein, *The ergodic theoretical proof of Szemerédi's theorem*, Bull. Amer. Math. Soc. **7** (1982), no. 3, 527–552. MR 670131 (84b:28016)

Gow01 22. W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588. MR 1844079 (2002k:11014)

Gow07 23. ———, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. **166** (2007), no. 3, 897–946. MR 2373376 (2009d:05250)

GT08 24. B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. **167** (2008), no. 2, 481–547.

GK10 25. Larry Guth and Nets Hawk Katz, *On the erdos distinct distance problem in the plane*, arXiv preprint arXiv:1011.4105 (2010).

HW79 26. Godfrey Harold Hardy and Edward Maitland Wright, *An introduction to the theory of numbers*, Oxford University Press, 1979.

Hoh30 27. G. Hoheisel, *Primzahlprobleme in der Analysis.*, Sitzungsber. Preuß. Akad. Wiss., Phys.-Math. Kl. **1930** (1930), 580–588 (German).

KRS96 28. János Kollár, Lajos Rónyai, and Tibor Szabó, *Norm-graphs and bipartite turán numbers*, Combinatorica **16** (1996), no. 3, 399–406.

Lei83 29. Frank Thomson Leighton, *Complexity issues in vlsi: optimal layouts for the shuffle-exchange graph and other networks*, MIT press, 1983.

mesh95 30. Roy Meshulam, *On subsets of finite abelian groups with no 3-term arithmetic progressions*, J. Comb. Theory, Ser. A **71** (1995), no. 1, 168–172.

MV06 31. Hugh L Montgomery and Robert C Vaughan, *Multiplicative number theory i: Classical theory*, vol. 97, Cambridge University Press, 2006.

NRS06 32. B. Nagle, V. Rödl, and M. Schacht, *The counting lemma for regular k-uniform hypergraphs*, Random Structures Algorithms **28** (2006), no. 2, 113–179. MR 2198495 (2007d:05084)

Pre17 33. S. Prendiville, *Quantitative bounds in the polynomial Szemerédi theorem: the homogeneous case*, Discrete Anal. (2017), Paper No. 5, 34. MR 3631611

Ram30 34. F. P. Ramsey, *On a problem of formal logic*, Proc. London Math. Soc. Ser. 2 **30** (1930), 264–286.

RNSSK05 35. V. Rödl, B. Nagle, J. Skokan, M. Schacht, and Y. Kohayakawa, *The hypergraph regularity method and its applications*, Proc. Natl. Acad. Sci. USA **102** (2005), no. 23, 8109–8113. MR 2167756

RS07b 36. V. Rödl and M. Schacht, *Regular partitions of hypergraphs: counting lemmas*, Combin. Probab. Comput. **16** (2007), no. 6, 887–901. MR 2351689 (2008j:05238)

RS07a 37. ———, *Regular partitions of hypergraphs: regularity lemmas*, Combin. Probab. Comput. **16** (2007), no. 6, 833–885. MR 2351688 (2008h:05083)

RS04 38. V. Rödl and J. Skokan, *Regularity lemma for k-uniform hypergraphs*, Random Structures Algorithms **25** (2004), no. 1, 1–42.

RS06 39. ———, *Applications of the regularity lemma for uniform hypergraphs*, Random Structures Algorithms **28** (2006), no. 2, 180–194.

Roth53 40. K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109. MR 0051853 (14,536g)

RS78 41. I. Z. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, Colloq. Math. Soc. János Bolyai, vol. 18, North-Holland, Amsterdam, 1978, pp. 939–945. MR 519318 (80c:05116)

Sar78 42. A. Sárkőzy, *On difference sets of sequences of integers. I*, Acta Math. Acad. Sci. Hungar. **31** (1978), no. 1–2, 125–149. MR 0466059 (57 #5942)

Sch93 43. C Schade, *Exakte maximalzahlen gleicher abst ande*, Ph.D. thesis, Diplomarbeit bei H. Harborth, Universit at Braunschweig, 1993.

Sch16 44. I. Schur, *Über die Kongruenz $x^m + y^m = z^m$ (mod. p)*, Jahresber. Deutsche Math.-Verein. **25** (1916), 114–116.

Sha48 45. Claude E Shannon, *A mathematical theory of communication, part i, part ii*, Bell Syst. Tech. J. **27** (1948), 623–656.

Sol05 46. J. Solymosi, *On the number of sums and products*, Bulletin of the London Mathematical Society **37** (2005), 491–494.

Sop10 47. Ivan Soprounov, *A short proof of the prime number theorem for arithmetic progressions*, preprint (2010).

SST84 48. Joel Spencer, Endre Szemerédi, and William T Trotter, *Unit distances in the euclidean plane*, Graph theory and combinatorics (1984), 293–303.

Sze97 49. László A Székely, *Crossing numbers and hard erdős problems in discrete geometry*, Combinatorics, Probability and Computing **6** (1997), no. 3, 353–358.

Sze69 50. E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104. MR 0245555

Sze75 51. ———, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

Sze78 52. ———, *Regular partitions of graphs*, Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), Colloq. Internat. CNRS, vol. 260, CNRS, Paris, 1978, pp. 399–401. MR 540024 (81i:05095)

Sze16 53. Endre Szemerédi, *Erdős's unit distance problem*, pp. 459–477, Springer International Publishing, Cham, 2016.

vdW27 54. B.L. van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw Arch. Wisk. **15** (1927), 212—216.