1 Arithmetic Progressions

Our story begins with the following very famous result in Ramsey Theory:

Theorem 1.1 (van der Waerden 1927)**.** *For any $r, k \in \mathbb{N}$, and any coloring of $\mathbb{N}$ with r colors (namely, for any function $f : \mathbb{N} \to \{1, \ldots, r\}$), there is a monochromatic k-term arithmetic progression, namely a sequence $a, a+d, a+2d, \ldots, a+(k-1)d$ such that*

$$f(a) = f(a+d) = \cdots = f(a+(k-1)d).$$

We won't prove this theorem, but the basic idea is that you apply a very clever induction argument on both r and k. A few decades after van der Waerden, Hales and Jewett realized you could "abstract away" the inductive argument, and prove an abstract statement now called the Hales–Jewett theorem, which implies van der Waerden's theorem as an easy consequence.

This theorem guarantees that whenever we partition $\mathbb{N}$ into subsets $S_1, \ldots, S_r$ (these are just $S_i = f^{-1}(i)$), then for any k, some S_i will contain a k-term arithmetic progression. A natural question to ask is: which one? Additionally, a natural guess is that the "biggest" one will be the one that contains a k-term arithmetic progression. To formalize this, we make the following definition.

Definition 1.2. Given a set $S \subseteq \mathbb{N}$, its *density* is defined as

$$d(S) = \lim_{N \to \infty} \frac{|S \cap [N]|}{N},$$

where $[N] = \{1, 2, \ldots, N\}$, assuming that this limit exists.

Example 1.3. The set of even numbers has density 1/2, as does the set of odd numbers. The set of squares has density 0, since if $\mathbb{S}$ is the set of squares, then

$$|\mathbb{S} \cap [N]| \approx \sqrt{N},$$

so

$$d(\mathbb{S}) \approx \lim_{N \to \infty} \frac{\sqrt{N}}{N} = \lim_{N \to \infty} \frac{1}{\sqrt{N}} = 0.$$

Similarly, the set $\mathbb{P}$ of primes also has density 0; this is because the Prime number theorem says that

$$|\mathbb{P} \cap [N]| \approx \frac{N}{\log N},$$

and thus

$$d(\mathbb{P}) \approx \lim_{N \to \infty} \frac{N/\log N}{N} = \lim_{N \to \infty} \frac{1}{\log N} = 0.$$

Finally, $\mathbb{N}$ itself has density 1.

Lemma 1.4. *If $S, T \subseteq \mathbb{N}$ are disjoint sets, then*

$$d(S \cup T) = d(S) + d(T).$$

Proof.

$$d(S \cup T) = \lim_{N \to \infty} \frac{|(S \cup T) \cap [N]|}{N} = \lim_{N \to \infty} \frac{|S \cap [N]| + |T \cap [N]|}{N} = d(S) + d(T). \qquad \square$$

One consequence of this lemma is that when we color $\mathbb{N}$ with r colors, then one of the color classes S_i must have strictly positive density. So one way of phrasing our "biggest" conjecture above is the following:

Conjecture 1.5 (Erdős–Turán 1936)**.** *If $S \subseteq \mathbb{N}$ has positive density, then it contains a k-term arithmetic progression for any $k \in \mathbb{N}$.*

The first progress towards this theorem was made almost 20 years later:

Theorem 1.6 (Roth 1953)**.** *If $S \subseteq \mathbb{N}$ has positive density, then it contains a 3-term arithmetic progression.*

Finally, the full Erdős–Turán Conjecture was resolved by Szemerédi:

Theorem 1.7 (Szemerédi 1969 ($k = 4$), Szemerédi 1975 (all k))**.** *If $S \subseteq \mathbb{N}$ has positive density, then it contains a k-term arithmetic progression for any $k \in \mathbb{N}$.*

A key component of Szemerédi's proof is now called Szemerédi's regularity lemma, which is actually a statement about graphs. In addition to being an extremely deep and important statement in and of itself, it also demonstrates a remarkable and surprising connection between number theory and graph theory. We won't state or prove the regularity lemma yet, but will begin with one of its most important consequences.

Before that, it is worthwhile to mention two more major ideas related to Szemerédi's theorem. The first is the following result, which is considered one of the most important advances in number theory of recent years:

Theorem 1.8 (Green–Tao, 2004)**.** *For every k, there is a k-term arithmetic progression in the primes, namely some $a, d \in \mathbb{N}$ such that $a, a + d, a + 2d, \ldots, a + (k - 1)d$ are all prime.*

This was a real breakthrough, and took many years and several hundred pages to prove. Note that this is not at all implied from Szemerédi's Theorem, since the primes have density 0, as discussed above. However, many of the ideas that go into the proof of the Green–Tao theorem are closely related to ideas we will discuss in this class; I hope to return to the Green–Tao theorem by the end of the class.

Moreover, both Szemerédi's Theorem and the Green–Tao Theorem are implied by the following conjecture, which is perhaps the biggest open problem in this entire field.

Conjecture 1.9 (Erdős)**.** *If* $S = \{s_1, s_2, \ldots\} \subseteq \mathbb{N}$*, and*

$$\sum_{i=1}^{\infty} \frac{1}{s_i} = \infty$$

then for every k*,* S *contains a* k*-term arithmetic progression.*

This implies the Green–Tao Theorem because Euler proved that the sum of the reciprocals of the primes diverges. Additionally, it implies Szemerédi's Theorem: intuitively, a set of density δ "should" be just a set of numbers that are each roughly $1/\delta$ apart, so we expect that

$$\sum_{i=1}^{\infty} \frac{1}{s_i} \approx \sum_{n=1}^{\infty} \frac{1}{n/\delta} = \delta \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

Thus, if Erdős's Conjecture were proved, then it would imply both Szemerédi's Theorem and the Green–Tao Theorem.

2 Triangle Removal

In order to both prove Roth's theorem and demonstrate the powerful and surprising connection between graphs and arithmetic progressions, we will begin with a very important consequence of Szemerédi's regularity lemma, known as the triangle removal lemma. Recall that a *triangle* in a graph G is a collection of three vertices that are all connected by edges.

Lemma 2.1 (Ruzsa–Szemerédi 1978)**.** *For every* $\varepsilon > 0$*, there exists a* $\delta > 0$ *such that the following holds. If* G *is an* n*-vertex graph with at most* $\delta\binom{n}{3}$ *triangles, we may remove at most* $\varepsilon\binom{n}{2}$ *edges from* G *in order to make it triangle-free.*

Note that G may have up to $\binom{n}{2}$ edges, and up to $\binom{n}{3}$ triangles. Thus, what the triangle removal lemma says is that if our graph has some small constant fraction of all possible triangles, then we can remove some small constant fraction of the edges to make it triangle-free. Said differently, the triangle removal lemma says that the *only* way to construct a graph with *few* triangles is to start with a graph with *no* triangles and then to sprinkle in a few extra edges.

A key thing to note about the triangle removal lemma is the growth rates with n. It is easy to see that if G has at most t triangles, then we can make G triangle-free by removing at most t edges: simply pick an arbitrary edge from each triangle and remove it. But if we apply this simple argument, then it only tells us that if G has at most $\delta\binom{n}{3}$ triangles, then it can be made triangle-free by removing at most $\delta\binom{n}{3}$ edges. But if n is large enough, then $\delta\binom{n}{3} > \binom{n}{2}$, which is the most number of edges G could possibly have. So this is not a very interesting statement: for large n, removing at most $\delta\binom{n}{3}$ edges amounts to removing all edges from G, and obviously this will make G triangle-free!

The triangle removal lemma may seem innocuous, but it is remarkably powerful. Indeed, we will now derive Roth's theorem (namely the $k = 3$ case of Szemerédi's theorem) from it. First, we give a different (but equivalent) formulation of the theorem:

Theorem 2.2 (Roth 1953)**.** *For every $\varepsilon > 0$, there is some integer $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$ and any subset $S \subseteq [N]$ with $|S| \geq \varepsilon N$, S contains a 3-term arithmetic progression.*

This is equivalent to the previous formulation for the following reason: if $T \subseteq \mathbb{N}$ is a set of positive density, then its density is greater than some ε, so by setting $S = T \cap [N_0]$, we conclude that it has a 3-term arithmetic progression. Conversely, by gluing together translated copies of a set $S \subseteq [N]$ with $|S| \geq \varepsilon N$, we get a set of positive density in $\mathbb{N}$, which must contain a 3-term arithmetic progression, and we can conclude that S contains one as well. On the homework, you'll formalize this proof sketch.

One simple and useful remark is that integers a, b, c form a 3-term arithmetic progression if and only if $a + c = 2b$. Indeed, this is equivalent to the condition $b - a = c - b$, which is precisely the condition of being an arithmetic progression.

Proof of Theorem 2.2. We will pick N_0 later. From such a set $S \subseteq [N]$, we construct a graph G as follows. Let X, Y, Z be three copies of the set $[3N]$, and then the vertices of G will be $X \cup Y \cup Z$. We put no edges inside X or Y or Z. Additionally, we connect $x \in X$ to $y \in Y$ if and only if $y - x \in S$, and we connect $y \in Y$ and $z \in Z$ if and only if $z - y \in S$. Finally, we connect $x \in X$ and $z \in Z$ if and only if $z - x \in 2S$, namely $z - x = 2S$ for some $s \in S$.

Now, for every $x \in [N]$ and $s \in S$, we automatically get a triangle in G, namely the triangle $x \in X, x + s \in Y, x + 2s \in Z$; indeed, by definition, all three of these vertices are pairwise adjacent. Therefore, each $s \in S$ yields at least N triangles in G, so we have at least $N|S| \geq \varepsilon N^2$ triangles in G. Moreover, all of these triangles are edge-disjoint, so in order to eliminate all of them, we'd need to delete at least $\varepsilon N^2 \geq \frac{\varepsilon}{100}\binom{9N}{2}$ edges. Since G is a graph on $9N$ vertices, we can apply the contrapositive of the triangle removal lemma (with parameter $\varepsilon/100$) to conclude that there is some $\delta > 0$ such that G has at least $\delta\binom{9N}{3}$ triangles. Now, we choose N_0 large enough that $\delta\binom{9N_0}{3} > \frac{\varepsilon}{100}\binom{9N_0}{2}$. Then we conclude that if $N \geq N_0$, there must be some triangle in G that we haven't yet accounted for.

Since there is no edge within X, Y, or Z, this additional triangle must consist of some $x \in X, y \in Y, z \in Z$. Additionally, we necessarily have that $y - x \neq z - y$, for if these were both equal to some s, then this triangle would just be one of the "simple" triangles we've already considered.

Therefore, we can define $a = y - x, b = \frac{z-x}{2}, c = z - y$. Then by the definition of the edges of G, we know that $a, b, c \in S$. On the other hand, we have that

$$a + c = (y - x) + (z - y) = z - x = 2b,$$

which implies that a, b, c form a 3-term arithmetic progression contained entirely within S, as claimed. □

3 Regularity

The key notion in Szemerédi's regularity lemma is, shockingly, called *regularity*. It is defined for pairs of vertex sets in a graph, and roughly speaking, a pair of vertex sets is called *regular* if the set of edges between them "looks random". We will formalize this shortly.

Definition 3.1. Let G be a graph, and let $A, B \subseteq V(G)$ be sets of vertices. By $e(A, B)$, we denote the number of edges with one endpoint in A and the other in B, namely

$$e(A, B) := |\{(u, v) \in A \times B : uv \in E(G)\}|.$$

Note that in case $A \cap B \neq \varnothing$, then every edge in $A \cap B$ is actually counted twice in this definition. If this bothers you, you can pretend that A and B are disjoint.

Additionally, the *edge density* between A and B is defined as

$$d(A, B) = \frac{e(A, B)}{|A||B|}.$$

Thus, $d(A, B)$ measures what fraction of all possible edges between A and B are actually present.

Definition 3.2. Given a graph G, some parameter $\varepsilon > 0$, and two sets of vertices $A, B \subseteq V(G)$, we say that the pair (A, B) is *ε-regular* if for every $A' \subseteq A, B' \subseteq B$ with $|A'| \geq \varepsilon|A|, |B'| \geq \varepsilon|B|$, we have

$$|d(A, B) - d(A', B')| < \varepsilon.$$

In other words, (A, B) is ε-regular if all the edges between A and B are "well-distributed" throughout A and B; no matter where we look in A and B, we see roughly the same density of edges.

Example 3.3 (Basically the only example)**.** Suppose we fix some parameter $p \in (0, 1)$, and we put edges between A and B by picking them randomly: for every $a \in A, b \in B$, we connect a to b by flipping a p-biased coin and connecting them if and only if it comes up heads. Then one can check that if ε isn't too small (namely $\varepsilon \gtrsim 1/\sqrt{|A| + |B|}$), then the pair (A, B) will be ε-regular in this graph we've defined (with very high probability). Intuitively, this is because the edges are indeed "well-distributed"—they were placed randomly, so how could they not be?

Indeed, this is more or less the only example: if a pair (A, B) is ε-regular, then we can pretty much pretend that it was obtained by putting edges in randomly with probability $p = d(A, B)$.

3.1 Counting

As it turns out, ε-regularity is a very strong and useful condition. For instance, it allows us to compute a huge number of quantities associated to the graph, by pretending that our graph is random and counting the associated quantities there. As an example, suppose we construct a random graph as follows. We start with disjoint sets A, B, C of vertices, and we place edges randomly between A and B with some probability $r = d(A, B)$, between B and C with some probability $s = d(B, C)$, and between A and C with probability $t = d(A, C)$. Then if we fix some vertices $a \in A, b \in B, c \in C$, the probability that they form a triangle is exactly rst. Because of this, one can show that in this random graph, with high probability the number of triangles is very close to

$$\sum_{a \in A} \sum_{b \in B} \sum_{c \in C} rst = rst|A||B||C|.$$

Our next result, known as the *triangle counting lemma*, says that in *any* graph with appropriate regularity conditions, the number of triangles is very close to what it would be in a random graph with the same densities. For vertex sets A, B, C, let $T(A, B, C)$ denote the number of triangles with one vertex in each of A, B, C.

Lemma 3.4 (Triangle counting lemma)**.** *Let $A, B, C \subseteq V(G)$ be disjoint sets of vertices in some graph G, and suppose that each of the three pairs $(A, B), (B, C), (A, C)$ is ε-regular. Let*

$$r = d(A, B), \qquad s = d(B, C), \qquad t = d(A, C).$$

If $r, s, t \geq 2\varepsilon$, then

$$T(A, B, C) \geq (1 - 2\varepsilon)(r - \varepsilon)(s - \varepsilon)(t - \varepsilon)|A||B||C|.$$

Similarly,

$$T(A, B, C) \leq (1 + 8\varepsilon)(r + \varepsilon)(s + \varepsilon)(t + \varepsilon)|A||B||C|.$$

Before proving this, we will state and prove a simple and useful property about the notion of ε-regularity. For a vertex a and a vertex set B, let $\deg_B(a)$ denote the number of neighbors of a in B.

Lemma 3.5. *Suppose (A, B) is an ε-regular pair of vertex sets in some graph. Then fewer than $\varepsilon|A|$ vertices in A have fewer than $(d(A, B) - \varepsilon)|B|$ neighbors in B, and fewer than $\varepsilon|A|$ vertices in A have more than $(d(A, B) + \varepsilon)|B|$ neighbors in B.*

Proof. Let $A' \subseteq A$ be the set of vertices with fewer than $(d(A, B) - \varepsilon)|B|$ neighbors in B. Note that

$$e(A', B) = \sum_{a \in A'} \deg_B(a) < \sum_{a \in A'} (d(A, B) - \varepsilon)|B| = (d(A, B) - \varepsilon)|A'||B|,$$

which implies that

$$d(A', B) = \frac{e(A', B)}{|A'||B|} < d(A, B) - \varepsilon.$$

So if $|A'| \geq \varepsilon|A|$, we get a contradiction to the definition of ε-regularity, which implies that $|A'| < \varepsilon|A|$, as claimed.

The second statement is proved in exactly the same way, simply replacing all minus signs by plus signs. □

With this in hand, we can pretty quickly prove the triangle counting lemma.

Proof of Lemma 3.4. We only prove the first inequality in Lemma 3.4 (i.e. the lower bound on $T(A, B, C)$); the upper bound is proved in essentially the same way. First, let

$$A' = \{a \in A : \deg_B(a) \geq (r - \varepsilon)|B| \text{ and } \deg_C(a) \geq (t - \varepsilon)|C|\}.$$

By Lemma 3.5, the number of $a \in A$ with $\deg_B(a) < (r-\varepsilon)|B|$ is at most $\varepsilon|A|$, and similarly for the number with $\deg_C(a) < (t - \varepsilon)|C|$. So we have that $|A'| \geq (1 - 2\varepsilon)|A|$.

Now, fix some $a \in A'$. Then let $B_a \subseteq B$ be the set of neighbors of a in B, and define $C_a \subseteq C$ similarly. Then since we assumed that $r, t \geq 2\varepsilon$, we get that $|B_a| \geq \varepsilon|B|, |C_a| \geq \varepsilon|C|$, so by regularity of the pair (B, C), we know that

$$|d(B_a, C_a) - s| < \varepsilon,$$

and thus

$$d(B_a, C_a) \geq s - \varepsilon.$$

This, in turn, implies that

$$e(B_a, C_a) \geq (s - \varepsilon)|B_a||C_a| \geq (s - \varepsilon)(r - \varepsilon)(t - \varepsilon)|B||C|.$$

However, every edge between B_a and C_a yields a triangle containing a, since a is adjacent to all vertices in B_a, C_a.

Finally, we sum this result over all $a \in A'$ to conclude that

$$\begin{aligned} T(A, B, C) &\geq \sum_{a \in A'} e(B_a, C_a) \geq |A'| \cdot (s - \varepsilon)(r - \varepsilon)(t - \varepsilon)|B||C| \\ &\geq (1 - 2\varepsilon)(r - \varepsilon)(s - \varepsilon)(t - \varepsilon)|A||B||C|. \end{aligned}$$

□

3.2 Szemerédi's regularity lemma

Hopefully you have been convinced that ε-regularity is a very strong and very useful notion: if we know that some pair of vertex sets in our graph is ε-regular, then we can more or less pretend that that part of the graph is random. This makes the following result, Szemerédi's regularity lemma, so surprising. It says that *every* graph can be cut up into parts so that almost every pair of parts is ε-regular. In other words, we can approximate more or less the whole graph as a random graph. To formalize this, we make the following definition.

Definition 3.6. Let G be an n-vertex graph, and let $\mathcal{P} : V = V_1 \sqcup \cdots \sqcup V_m$ be a partition of its vertex set into m parts. We say that $\mathcal{P}$ is *ε-regular* if

$$\sum_{\substack{(i,j) \in [m]^2 \\ (V_i, V_j) \text{ is not } \varepsilon\text{-regular}}} |V_i||V_j| \leq \varepsilon n^2.$$

Said differently, at most an ε-fraction of the pairs of vertices lie in irregular pairs of parts. Note that if $|V_i| = n/m$ for all i, then this simply means that at most εm^2 pairs (V_i, V_j) are not ε-regular.

Theorem 3.7 (Szemerédi's regularity lemma). *For every $\varepsilon > 0$, there exists some $M \in \mathbb{N}$ so that every graph has an ε-regular partition of its vertex set into at most M parts.*

In other words, we can always partition any graph into a collection of "clusters", in such a way that almost all of the pairs of clusters are ε-regular. Crucially, the number m of clusters is bounded by M, which depends *only* on ε; thus, the "complexity" of the partition depends only on how regular we require our partition to be. In particular, once $n = v(G)$ is much larger than M, then all graphs on n vertices are basically the same: they are composed of at most M clusters, and look like they were randomly generated from these clusters. In other words, all big graphs are basically the same.

Before talking about the proof of the regularity lemma, let's see how it implies the triangle removal lemma. Recall that that lemma said that for any $\varepsilon > 0$, there is some $\delta > 0$ such that if an n-vertex graph G has at most $\delta\binom{n}{3}$ triangles, then we can remove at most $\varepsilon\binom{n}{2}$ edges and make it triangle-free.

Proof of the triangle removal lemma. First, using the regularity lemma, we can find some $(\varepsilon/12)$-regular partition $\mathcal{P} : V(G) = V_1 \sqcup \cdots \sqcup V_m$ of G. Now, we're going to remove a bunch of edges from G:

1. For every pair (V_i, V_j) which is not $(\varepsilon/12)$-regular, we remove all edges between V_i and V_j. The number of edges removed at this step is, by the definition of an $(\varepsilon/12)$-regular partition, at most
$$\sum_{\substack{(i,j)\in[m]^2 \\ (V_i,V_j) \text{ is not } (\varepsilon/12)\text{-regular}}} |V_i||V_j| \le \frac{\varepsilon}{12}n^2.$$

2. Between every pair of clusters (V_i, V_j) with $d(V_i, V_j) < \varepsilon/6$, we remove all edges. The number of edges removed at this step is
$$\sum_{\substack{(i,j)\in[m]^2 \\ d(V_i,V_j)<\varepsilon/6}} e(V_i, V_j) = \sum_{\substack{(i,j)\in[m]^2 \\ d(V_i,V_j)<\varepsilon/6}} d(V_i, V_j)|V_i||V_j| < \sum_{(i,j)\in[m]^2} \frac{\varepsilon}{6}|V_i||V_j| = \frac{\varepsilon}{6}n^2.$$

3. Finally, we remove all edges between V_i and V_j if $\min\{|V_i|, |V_j|\} \le \varepsilon n/(12m)$. The number of edges removed at this step is at most
$$\sum_{\substack{(i,j)\in[m]^2 \\ \min\{|V_i|,|V_j|\}\le\varepsilon n/(12m)}} |V_i||V_j| \le \sum_{\substack{i\in[m] \\ |V_i|\le\varepsilon n/(12m)}} \sum_{j=1}^{m} |V_i||V_j| = n \sum_{\substack{i\in[m] \\ |V_i|\le\varepsilon n/(12m)}} |V_i| \le n\cdot m\cdot \frac{\varepsilon n}{12m} = \frac{\varepsilon}{12}n^2.$$

Thus, the total number of edges we removed is at most
$$\frac{\varepsilon}{12}n^2 + \frac{\varepsilon}{6}n^2 + \frac{\varepsilon}{12}n^2 = \frac{1}{3}\varepsilon n^2 \le \varepsilon\binom{n}{2},$$

since $\frac{1}{3}n^2 \leq \binom{n}{2}$ for all $n \geq 3$.

After removing all these edges, we end up with a subgraph of G, which we'll call G'. If G' is triangle-free, then we're done. If not, then we want to prove that G must have started with many triangles, namely at least δn^3 of them.

So suppose that G' has a triangle, with vertices a, b, c. Each of these vertices lies in some part, say $a \in V_i, b \in V_j, c \in V_k$ (where some of these indices are potentially equal). Since a, b, c form a triangle, we must have not deleted any of the edges between $(V_i, V_j), (V_j, V_k)$, or (V_i, V_k). By the way we deleted edges, we conclude that $|V_i|, |V_j|, |V_k| \geq \varepsilon n/(12m)$, as well as that $(V_i, V_j), (V_j, V_k)$, and (V_i, V_k) are all $(\varepsilon/12)$-regular and have density at least $\varepsilon/6$. So by the triangle counting lemma, Lemma 3.4, we see that

$$\begin{aligned} T(V_i, V_j, V_k) &\geq \left(1 - \frac{\varepsilon}{6}\right)\left(d(V_i, V_j) - \frac{\varepsilon}{12}\right)\left(d(V_j, V_k) - \frac{\varepsilon}{12}\right)\left(d(V_i, V_k) - \frac{\varepsilon}{12}\right)|V_i||V_j||V_k| \\ &\geq \frac{1}{2}\left(\frac{\varepsilon}{6}\right)^3\left(\frac{\varepsilon n}{12m}\right)^3 \\ &\geq \frac{\varepsilon^6}{10000000m^3}\binom{n}{3} \\ &\geq \frac{\varepsilon^6}{10000000M^3}\binom{n}{3}. \end{aligned}$$

So G has at least $\delta\binom{n}{3}$ triangles, where $\delta = \varepsilon^6/(10000000M^3)$. Note that, as M depends only on ε, so does δ. This concludes the proof. □

4 The general counting and removal lemmas

We proved the triangle counting and triangle removal lemmas, but similar results hold for any graph H, rather than just a triangle. Here is the general graph counting lemma.

Theorem 4.1 (Counting lemma)**.** *Fix a graph H with vertices $v_1, \dots, v_h$. For every $\rho > 0$, there exists some $\varepsilon > 0$ so that the following holds. Let $V_1, \dots, V_h$ be vertex sets in a graph H, with the property that (V_i, V_j) is ε-regular if $v_i v_j \in E(H)$. Then the number of copies of H with the ith vertex of H lying in V_i is at least*

$$\left(\left(\prod_{\substack{i<j \\ v_i v_j \in E(H)}} d(V_i, V_j)\right) - \rho\right) \prod_{i=1}^{h} |V_i|$$

and at most

$$\left(\left(\prod_{\substack{i<j \\ v_i v_j \in E(H)}} d(V_i, V_j)\right) + \rho\right) \prod_{i=1}^{h} |V_i|.$$

In other words, the number of copies of H is well-approximated by what it would be in a random graph, which is $\prod_{i<j : v_i v_j \in E(H)} d(V_i, V_j) \prod_{i=1}^{h} |V_i|$. Namely, we can guarantee that this prediction is arbitrarily close to correct, by ensuring that our pairs are sufficiently regular.

As a consequence of the counting lemma, we can also prove a general removal lemma.

Theorem 4.2 (Removal lemma)**.** *For every h-vertex graph H and every $\varepsilon > 0$, there exists some $\delta > 0$ so that the following holds for every n and every n-vertex graph G. If G has at most $\delta\binom{n}{h}$ copies of H, then G can be made H-free by deleting at most $\varepsilon\binom{n}{2}$ edges.*

The proof of the general removal lemma is more or less identical to the proof of the triangle removal lemma. We first apply the regularity lemma, then delete all edges in irregular pairs, in pairs with low edge density, and in pairs where one of the parts is small. By doing so we delete few edges. If there is still a copy of H remaining, it must lie between large, dense, regular pairs, so we can apply the counting lemma and find that we must have started with very many copies of H. The proof of the general counting lemma is also generally the same as the proof of the triangle removal lemma, except that we also need to use induction on h.

5 Proof of the regularity lemma

We are now going to prove Szemerédi's regularity lemma. From now on, we will always have G being an n-vertex graph, without specifying this every time. The proof idea is, at its core, very simple. We start with a trivial partition of the vertex set, $\mathcal{P}_0 = \{V(G)\}$. We will then repeatedly *refine* a partition $\mathcal{P}_\ell$ into a new partition $\mathcal{P}_{\ell+1}$, with the goal of "fixing" irregular pairs. Namely, suppose that $\mathcal{P}_\ell$ has some pair of parts (V_i, V_j) that is not ε-regular. This implies that there exist some $X \subseteq V_i, Y \subseteq V_j$ with $|X| \geq \varepsilon|V_i|, |Y| \geq \varepsilon|V_j|$ and

$$|d(V_i, V_j) - d(X, Y)| \geq \varepsilon.$$

Then in the partition $\mathcal{P}_{\ell+1}$, we essentially want to replace the two parts V_i, V_j with the *four* parts $X, Y, V_i \setminus X, V_j \setminus Y$. The idea is that by refining our partition according to the "witnesses of irregularity", we should be able to eventually get rid of most of the irregular pairs.

The key thing we need to make this proof idea work is some sort of *progress measure*. Namely, we want to somehow say that we are making progress towards an ε-regular partition, in order to bound the number of steps (and thus the number of parts in the final partition). It turns out that many different choices of progress measure work, but probably the simplest one is the following.

Definition 5.1. Let $U, W \subseteq V(G)$ be sets of vertices. The *mean square density* of the pair (U, W) is defined as

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

Let $\mathcal{P}_U : U = U_1 \sqcup \cdots \sqcup U_\ell$ and $\mathcal{P}_W : W = W_1 \sqcup \cdots \sqcup W_m$ be partitions of U, W, respectively. The *mean square density* of $(\mathcal{P}_U, \mathcal{P}_W)$ is defined as

$$q(\mathcal{P}_U, \mathcal{P}_W) \coloneqq \sum_{i=1}^{\ell} \sum_{j=1}^{m} q(U_i, W_j) = \sum_{(i,j) \in [\ell] \times [m]} \frac{|U_i||W_j|}{n^2} d(U_i, W_j)^2.$$

Finally, for a single partition $\mathcal{P} : V(G) = V_1 \sqcup \cdots \sqcup V_m$ of the vertices of G, the *mean square density* of $\mathcal{P}$ is defined as

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

Note that for any partition $\mathcal{P}$ of $V(G)$, we have that

$$0 \leq q(\mathcal{P}) \leq 1. \tag{1}$$

Indeed, $q(\mathcal{P})$ is a sum of non-negative terms, so it's certainly non-negative. On the other hand, $d(V_i, V_j) \leq 1$ for all i, j, so $q(\mathcal{P}) \leq \sum |V_i||V_j|/n^2 = 1$.

The mean square density q will be our progress measure in the proof of Szemerédi's regularity lemma. We will need to collect a few important facts about the mean square density. The first says that when we refine a partition, the mean square density can't go down. Recall that one partition $\mathcal{P}'$ is a *refinement* of $\mathcal{P}$ if every part of $\mathcal{P}'$ is a subset of some part of $\mathcal{P}$.

Lemma 5.2. *Let $U, W \subseteq V(G)$, and let $\mathcal{P}_U : U = U_1 \sqcup \cdots \sqcup U_\ell$ and $\mathcal{P}_W : W = W_1 \sqcup \cdots \sqcup W_m$ be partitions of U, W, respectively. Then we have that*

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

Moreover, suppose that $\mathcal{P}, \mathcal{P}'$ are partitions of $V(G)$ such that $\mathcal{P}'$ is a refinement of $\mathcal{P}$. Then

$$q(\mathcal{P}') \geq q(\mathcal{P}).$$

This proof, as well as all the others we'll see in this section, works by writing some quantity as the sum of squares of some other quantities, and thus concluding that the first thing is non-negative.

Proof. We first observe that

$$\sum_{i=1}^{\ell}\sum_{j=1}^{m}|U_i||W_j| = \left(\sum_{i=1}^{\ell}|U_i|\right)\left(\sum_{j=1}^{m}|W_j|\right) = |U||W|$$

and that

$$\sum_{i=1}^{\ell}\sum_{j=1}^{m}|U_i||W_j|d(U_i,W_j) = \sum_{i=1}^{\ell}\sum_{j=1}^{m}|U_i||W_j|\frac{e(U_i,W_j)}{|U_i||W_j|} = e(U,W) = |U||W|d(U,W).$$

Therefore, we have that

$$\begin{aligned}
&\sum_{i,j}|U_i||W_j|[d(U_i,W_j) - d(U,W)]^2 = \\
&\qquad = \sum_{i,j}|U_i||W_j|\left[d(U_i,W_j)^2 + d(U,W)^2 - 2d(U_i,W_j)d(U,W)\right] \\
&\qquad = \left(\sum_{i,j}|U_i||W_j|d(U_i,W_j)^2\right) + |U||W|d(U,W)^2 - 2d(U,W)\sum_{i,j}|U_i||W_j|d(U_i,W_j) \\
&\qquad = \left(\sum_{i,j}|U_i||W_j|d(U_i,W_j)^2\right) + |U||W|d(U,W)^2 - 2|U||W|d(U,W)^2 \\
&\qquad = n^2\left[q(\mathcal{P}_U,\mathcal{P}_W) - q(U,W)\right].
\end{aligned}$$

Since the left-hand side is a sum of squares, the right-hand side must be non-negative, which proves the first claim.

For the second, suppose that $\mathcal{P}$ partitions $V(G)$ into $V_1 \sqcup \cdots \sqcup V_m$, and $\mathcal{P}'$ refines this partition by splitting each V_i into $\mathcal{P}'_i : V_{i,1} \sqcup \cdots \sqcup V_{i,k_i}$. Then for every i, j, the previous claim shows that

$$q(\mathcal{P}'_i, \mathcal{P}'_j) \geq q(V_i, V_j).$$

By adding this up over all i, j, we find that

$$q(\mathcal{P}) = q(\mathcal{P},\mathcal{P}) = \sum_{i,j} q(V_i,V_j) \leq \sum_{i,j} q(\mathcal{P}'_i,\mathcal{P}'_j) = q(\mathcal{P}'). \qquad \square$$

Our next claim shows that in case a pair (U, W) is not ε-regular, then we can boost the inequality in Lemma 5.2: we can find a partition $\mathcal{P}_U$ of U and a partition $\mathcal{P}_W$ of W so that $q(\mathcal{P}_U, \mathcal{P}_W)$ is greater by some fixed positive amount than $q(U, W)$.

Lemma 5.3. *Suppose that U, W are vertex sets such that (U, W) is not ε-regular, and let $X \subseteq U, Y \subseteq W$ be subsets with $|X| \geq \varepsilon|U|, |Y| \geq \varepsilon|W|$ and*

$$|d(U, W) - d(X, Y)| \geq \varepsilon.$$

Let $\mathcal{P}_U$ be the partition $U = X \sqcup (U \setminus X)$ and $\mathcal{P}_W$ the partition $W = Y \sqcup (W \setminus Y)$. Then

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

Proof. By the proof of Lemma 5.2, we know that we can write $n^2[q(\mathcal{P}_U, \mathcal{P}_W) - q(U, W)]$ as a sum of four non-negative terms, corresponding to the four choices of a part of $\mathcal{P}_U$ and a part of $\mathcal{P}_W$. By discarding three of these terms, we see that

$$n^2[q(\mathcal{P}_U, \mathcal{P}_W) - q(U, W)] \geq |X||Y|[d(X, Y) - d(U, W)]^2 \geq \varepsilon^2|U||W| \cdot \varepsilon^2 = \varepsilon^4|U||W|.$$

Dividing by n^2, we get the desired result. □

Using this, we can refine any non-ε-regular partition into a new partition whose mean square density goes up by some fixed positive amount.

Lemma 5.4. *Let $\mathcal{P} : V(G) = V_1 \sqcup \cdots \sqcup V_m$ be a partition of $V(G)$, and suppose that $\mathcal{P}$ is not ε-regular. Then there exists a refinement $\mathcal{Q}$ of $\mathcal{P}$ with*

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

Moreover, $\mathcal{Q}$ has at most $m2^{2m}$ parts.

Proof. For every pair (V_i, V_j) which is not ε-regular, we can find some $X_{i,j} \subseteq V_i, Y_{i,j} \subseteq V_j$ with $|X_{i,j}| \geq \varepsilon|V_i|, |Y_{i,j}| \geq \varepsilon|V_j|$, and

$$|d(X_{i,j}, Y_{i,j}) - d(V_i, V_j)| \geq \varepsilon.$$

Fix i. For each j such that (V_i, V_j) is not ε-regular, we get a partition of V_i as $V_i = X_{i,j} \sqcup (V_i \setminus X_{i,j})$. Similarly, we also get a partition of V_i as $Y_{j,i} \sqcup (V_i \setminus Y_{j,i})$. Let $\mathcal{Q}_i$ be the common refinement of all of these partitions of V_i (there are at most $2m$ of them, since we may get such a partition for every $X_{i,j}$ and every $Y_{j,i}$). In other words, we let $\mathcal{Q}_i$ be the partition of V_i that cuts up V_i according to the list of, for every j, whether a given vertex of V_i is in $X_{i,j}$ or not. Finally, let $\mathcal{Q}$ be the refinement of $\mathcal{P}$ in which every part V_i of $\mathcal{P}$ is cut up according to $\mathcal{Q}_i$.

We claim that

$$\begin{aligned} q(\mathcal{Q}) &= \sum_{(i,j)\in[m]^2} q(\mathcal{Q}_i, \mathcal{Q}_j) \\ &= \sum_{\substack{(i,j)\in[m]^2 \\ (V_i,V_j)\ \varepsilon\text{-regular}}} q(\mathcal{Q}_i, \mathcal{Q}_j) + \sum_{\substack{(i,j)\in[m]^2 \\ (V_i,V_j)\text{ not }\varepsilon\text{-regular}}} q(\mathcal{Q}_i, \mathcal{Q}_j) \\ &\geq \sum_{\substack{(i,j)\in[m]^2 \\ (V_i,V_j)\ \varepsilon\text{-regular}}} q(V_i, V_j) + \sum_{\substack{(i,j)\in[m]^2 \\ (V_i,V_j)\text{ not }\varepsilon\text{-regular}}} q(\{X_{i,j}, V_i \setminus X_{i,j}\}, \{Y_{i,j}, V_j \setminus Y_{i,j}\}). \end{aligned}$$

Indeed, in the final step, we apply the monotonicity of the mean square density (Lemma 5.2) to both sums. In the first sum, we use the first part of Lemma 5.2, and in the second sum, we use the fact that $\mathcal{Q}_i$ refines the partition $V_i = X_{i,j} \sqcup (V_i \setminus X_{i,j})$ and $\mathcal{Q}_j$ refines the partition $V_j = Y_{i,j} \sqcup (V_j \setminus Y_{i,j})$.

Now, by Lemma 5.3, we know that

$$q(\{X_{i,j}, V_i \setminus X_{i,j}\}, \{Y_{i,j}, V_j \setminus Y_{i,j}\}) \geq q(V_i, V_j) + \varepsilon^4 \frac{|V_i||V_j|}{n^2}.$$

Plugging this in to our earlier computation, we see that

$$\begin{aligned} q(\mathcal{Q}) &\geq \sum_{\substack{(i,j)\in[m]^2 \\ (V_i,V_j)\ \varepsilon\text{-regular}}} q(V_i, V_j) + \sum_{\substack{(i,j)\in[m]^2 \\ (V_i,V_j)\text{ not }\varepsilon\text{-regular}}} \left(q(V_i, V_j) + \varepsilon^4 \frac{|V_i||V_j|}{n^2} \right) \\ &= \sum_{(i,j)\in[m]^2} q(V_i, V_j) + \sum_{\substack{(i,j)\in[m]^2 \\ (V_i,V_j)\text{ not }\varepsilon\text{-regular}}} \frac{\varepsilon^4}{n^2} |V_i||V_j| \\ &= q(\mathcal{P}) + \frac{\varepsilon^4}{n^2} \cdot \sum_{\substack{(i,j)\in[m]^2 \\ (V_i,V_j)\text{ not }\varepsilon\text{-regular}}} |V_i||V_j|. \end{aligned}$$

Finally, by our assumption that $\mathcal{P}$ is not an ε-regular partition, we know that

$$\sum_{\substack{(i,j)\in[m]^2 \\ (V_i,V_j)\text{ not }\varepsilon\text{-regular}}} |V_i||V_j| \geq \varepsilon n^2.$$

Plugging this in, we see $q(\mathcal{Q}) \geq q(\mathcal{P}) + \varepsilon^5$, as claimed. □

With all this prep work, we are finally ready to prove Szemerédi's regularity lemma.

Proof of Szemerédi's regularity lemma. Let $\mathcal{P}_0$ be the "trivial" partition of $V(G)$, namely the one that partitions it into a single part, and let $m_0 = 1$. If $\mathcal{P}_0$ is ε-regular, we are done. If not, then by Lemma 5.4, there is a refinement $\mathcal{P}_1$ of $\mathcal{P}_0$ with at most $m_1 = m_0 2^{2m_0}$ parts and with $q(\mathcal{P}_1) \geq q(\mathcal{P}_0) + \varepsilon^5$. If $\mathcal{P}_1$ is ε-regular, we are done; if not, there is a refinement $\mathcal{P}_2$ of $\mathcal{P}_1$ with $q(\mathcal{P}_2) \geq q(\mathcal{P}_1) + \varepsilon^5$ and at most $m_2 = m_1 2^{2m_1}$ parts.

Continuing in this way, we get a sequence $\mathcal{P}_0, \mathcal{P}_1, \dots$ of partitions of $V(G)$, with $\mathcal{P}_\ell$ having at most $m_\ell = m_{\ell-1} 2^{2m_{\ell-1}}$ parts, and

$$q(\mathcal{P}_\ell) \geq q(\mathcal{P}_{\ell-1}) + \varepsilon^5 \geq q(\mathcal{P}_0) + \ell\varepsilon^5.$$

By (1), we know that $q(\mathcal{P}_0) \geq 0$, and that $q(\mathcal{P}_\ell) \leq 1$. So this process cannot go on forever: after at most $\lceil \varepsilon^{-5} \rceil$ steps, we must have stopped this, meaning that we have found an ε-regular partition of G.

When we stop, the number of parts in the ε-regular partition is bounded by $m_{\lceil \varepsilon^{-5} \rceil}$, where m_ℓ is given by the recurrence $m_\ell = m_{\ell-1} 2^{2m_{\ell-1}}$ above. So by setting $M = m_{\lceil \varepsilon^{-5} \rceil}$, we obtain the claim of Szemerédi's regularity lemma. □

How big is the bound on M we get? Note that for all positive integers x, we have

$$x 2^{2x} \leq 2^{3x} = 8^x.$$

This implies that $m_\ell \leq 8^{m_{\ell-1}}$ for all ℓ, and thus that

$$M \leq \left.8^{8^{\cdot^{\cdot^{\cdot^{8}}}}}\right\} \lceil \varepsilon^{-5} \rceil = \left.2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}\right\} O(\varepsilon^{-5}).$$

This is a truly *enormous* bound. As a consequence, any application of Szemerédi's regularity lemma also has truly terrible bounds involved. For example, in the proof of the triangle removal lemma above, the δ we obtained in terms of ε was defined as

$$\delta = \frac{\varepsilon^6}{10000000 M^3}$$

and thus the bounds we get in the triangle removal lemma is of the form

$$\frac{1}{\delta} \leq \left.2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}\right\} O(\varepsilon^{-15}).$$

Even more amazingly, a result of Gowers shows that such terrible bounds are actually necessary. Namely, Gowers showed that there exists a graph G such that *any* ε-regular partition of $V(G)$ has a number of parts which is of tower-type in a power of ε. In other words, for Szemerédi's regularity lemma, we cannot do any better than the kind of bound we got.

Nonetheless, it turns out that for some applications of the regularity lemma, such as the triangle removal lemma, we *can* do better.

Theorem 5.5 (Fox 2011)**.** *In the triangle removal lemma, we can take*

$$\frac{1}{\delta} \leq \left.2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}\right\} O\left(\log \tfrac{1}{\varepsilon}\right).$$

This is again an *enormous* bound, but it's *less* enormous. Rather than the height of the tower being some power of $1/\varepsilon$, it is merely logarithmic in $1/\varepsilon$.

6 Hypergraph removal and Szemerédi's theorem

We have successfully proved Szemerédi's theorem in the case $k = 3$, i.e. Roth's theorem. What about longer arithmetic progressions?

Szemerédi's original proof involved an extraordinarily complicated inductive argument, plus repeated applications of van der Waerden's theorem. Much later, people realized that roughly the same technique we used to prove Roth's theorem could be used to prove the full Szemerédi's theorem. However, rather than dealing with graphs, we have to deal with *hyper*graphs. If we go back to bare basics, a graph is a collection V of vertices, plus a collection E of edges, which are simply unordered pairs of vertices. Why restrict ourselves to pairs?

Definition 6.1. An *r-uniform hypergraph* (sometimes called an r-graph for short) consists of a finite collection V of *vertices*, as well as a collection E of *r-uniform hyperedges*, which are simply subsets of V of size r.

As with graphs, we say that one r-graph $\mathcal{H}$ is a *subhypergraph* (or simply *subgraph*) of another r-graph $\mathcal{G}$ if we can obtain $\mathcal{H}$ from $\mathcal{G}$ by deleting some vertices and edges. We say that $\mathcal{G}$ is *$\mathcal{H}$-free* if $\mathcal{G}$ does not contain $\mathcal{H}$ as a subgraph (and we also say that $\mathcal{G}$ has no copy of $\mathcal{H}$).

The *complete r-graph* on k vertices, denoted $K_k^{(r)}$, is the r-graph with k vertices whose edge set consists of all subsets of size r.

In many ways, the study of hypergraphs is more or less the same as the study of graphs. Starting in the 1970s and really picking up in the 1980s, various mathematicians started trying to formulate versions of Szemerédi's regularity lemma, the counting lemma, and the removal lemma for hypergraphs. But as it turns out, doing this is *really* hard. Despite people working on this seriously since the early 1980s, the project was not completed until the 2000s.

Theorem 6.2 (Hypergraph removal; Kohayakawa–Nagle–Rödl–Schacht–Skokan 2005; Gowers 2007)**.** *For every $h, r \geq 2$, every h-vertex r-graph $\mathcal{H}$, and every $\varepsilon > 0$, there exists some $\delta > 0$ such that the following holds for all n and any n-vertex r-graph $\mathcal{G}$. If $\mathcal{G}$ has at most $\delta\binom{n}{h}$ copies of $\mathcal{H}$, then $\mathcal{G}$ can be made $\mathcal{H}$-free by deleting at most $\varepsilon\binom{n}{r}$ hyperedges.*

Just as we could use the triangle removal lemma (for ordinary graphs, aka 2-graphs) to prove Roth's theorem, we can use the $K_k^{(k-1)}$ removal lemma for $(k-1)$-graphs to prove Szemerédi's theorem for arithmetic progressions of length k. We'll do the case of $k = 4$ in detail; the general case works in the same way, just with additional annoyances.

Theorem 6.3 (Szemerédi's theorem for $k = 4$)**.** *For any $\varepsilon > 0$, there exists some $N_0 \in \mathbb{N}$ so that for all $N \geq N_0$ and all $S \subseteq [N]$ with $|S| \geq \varepsilon N$, there is a four-term arithmetic progression in S.*

Proof. We will pick N_0 later. For contradiction, suppose that $N \geq N_0$ and $S \subseteq [N]$ is a 4-AP-free set with $|S| \geq \varepsilon N$. We define a 4-graph $\mathcal{G}$ with vertex set $W \sqcup X \sqcup Y \sqcup Z$, where

$W = X = Y = Z = \mathbb{Z}/(6N+1)\mathbb{Z}$. Given $w \in W, x \in X, y \in Y, z \in Z$, we make hyperedges according to the following rules:

$$\begin{aligned} wxy \in E(\mathcal{G}) &\iff 3w + 2x + y \in S \\ wxz \in E(\mathcal{G}) &\iff 2w + x - y \in S \\ wyz \in E(\mathcal{G}) &\iff w - y - 2z \in S \\ xyz \in E(\mathcal{G}) &\iff -x - 2y - 3z \in S, \end{aligned}$$

where we view $S \subseteq [N]$ as a subset of $\mathbb{Z}/(6N+1)\mathbb{Z}$ for this to make sense. Let $n = v(\mathcal{G}) = 4(6N+1)$.

The point of these equations defining the hyperedges of $\mathcal{G}$ is the following: four vertices w, x, y, z will form a copy of $K_4^{(3)}$ if and only if

$$-x - 2y - 3z, w - y - 2z, 2w + x - y, 3w + 2x + y \in S$$

and these four numbers form a (possibly degenerate) four-term arithmetic progression with common difference $d = w + x + y + z$. Setting $d = 0$, we conclude that every element of S gives us exactly $(6N+1)^2$ copies of $K_4^{(3)}$ in $\mathcal{G}$, and these copies are edge-disjoint. So at least $|S|(6N+1)^2 \geq \frac{\varepsilon}{1000}\binom{n}{3}$ hyperedges must be deleted from $\mathcal{G}$ to make it $K_4^{(3)}$-free.

Therefore, by the contrapositive of the $K_4^{(3)}$ removal lemma, we see that $\mathcal{G}$ must have at least $\delta\binom{n}{4}$ copies of $K_4^{(3)}$, for some $\delta > 0$ depending only on ε. If N_0 is large enough, then $\delta\binom{n}{4} > |S|(6N+1)^2$, and thus there is an unaccounted-for copy of $K_4^{(3)}$. This yields a non-trivial four-term arithmetic progression in S. □

Why is it so hard to prove the hypergraph removal lemma? The reason, it turns out, is that it's hard to come up with the "correct" notion of regularity for r-graphs. Basically, one needs a notion of regularity that is *weak* enough so that one can prove a hypergraph regularity lemma, but *strong* enough to imply an associated counting lemma. In the case of graphs, the definition we saw works. But for $r \geq 3$, the natural extension of this is simply too weak to get a counting lemma.

To see this, consider the following example. Fix some parameters $p, q \in (0, 1)$, and let A, B, C, D be vertex sets, each of size N. We build a random graph on $A \cup B \cup C \cup D$ by connecting every pair of vertices in two distinct parts with probability p. Then with high probability, this graph has roughly p^6N^4 copies of K_4, and roughly p^3N^3 triangles between every triple of parts. Now, we build a random 3-graph $\mathcal{G}_1$ as follows: for every triangle x, y, z in G, we make xyz a hyperedge of $\mathcal{G}_1$ with probability q. So with high probability, $\mathcal{G}_1$ has roughly $p^6q^4N^4$ copies of $K_4^{(3)}$ and p^3qN^3 hyperedges between any triple of parts. Moreover, $\mathcal{G}_1$ looks "very random-like", in the sense that the hyperedges of $\mathcal{G}$ are very uniformly distributed; if we take fairly large subsets $X \subseteq A, Y \subseteq B, Z \subseteq C$, then the number of hyperedges between X, Y, Z is very close to $p^3q|X||Y||Z|$. In other words, $\mathcal{G}_1$ is ε-regular for a natural notion of regularity in 3-graphs.

On the other hand, consider the 3-graph $\mathcal{G}_2$ on the same vertex set, where we simply make every triple an edge with probability p^3q. Then again, $\mathcal{G}_2$ looks very regular, and has the same density as $\mathcal{G}_1$. However, the number of $K_4^{(3)}$ in $\mathcal{G}_2$ is roughly $(p^3q)^4N^4 = p^{12}q^4N^4$. Crucially, $p^6q^4 \neq p^{12}q^4$ for $p \in (0,1)$, so this is a different number of copies of $K_4^{(3)}$ than the number in $\mathcal{G}_1$.

So we see that for this natural notion of ε-regularity in 3-graphs, we can't expect a counting lemma for copies of $K_4^{(3)}$: these two 3-graphs are very regular and have the same edge density, but different numbers of copies of $K_4^{(3)}$.

As hinted by this example, the issue is that when dealing with hypergraphs, we can't *only* care about the distribution of hyperedges looking uniform. We also care about edges of lower uniformity: in the case of 3-graphs, we also need to control how *pairs* of vertices lie inside hyperedges. This means that the formulation of a hypergraph regularity lemma in uniformity r is much more complicated: we not only partition the vertices, but also the pairs of vertices, the triples of vertices, ..., up to the $(r-1)$-tuples of vertices.

Once one figures out the appropriate definition of ε-regularity for such a complicated chain of partitions, one has to prove a regularity lemma and a counting lemma. Superficially, the proofs are "the same". For the regularity lemma, one tracks an appropriate notion of mean square density through successive iterations of refining partitions. Similarly, the counting lemma is proved via an induction scheme where one counts the number of embeddings of vertices of $\mathcal{H}$ one at a time. Finally, the removal lemma is proved in exactly the same way: by deleting hyperedges between tuples that are too sparse, or irregular, or contain a part that is too small, one can ensure that we delete few edges and yet all copies of $\mathcal{H}$.

6.1 Quantitative aspects

As we saw, Szemerédi's regularity lemma invokes bounds that are of tower-type, which implies that the proof of Roth's theorem that we got also has tower-type bounds. In other words, the proof we saw implies that if $\varepsilon > 0$ and if

$$N_0 = 2^{2^{\cdot^{\cdot^{2}}}} \Big\} O(\varepsilon^{-5}),$$

then S contains a three-term arithmetic progression for all $N \geq N_0$ and all $S \subseteq [N]$ with $|S| \geq \varepsilon N$. However, this is far from the best bound known: Roth's original Fourier-analytic proof showed that we may instead take $N_0 = 2^{2^{O(1/\varepsilon)}}$, and there were a series of improvements to this over the decades. Very recently, Bloom and Sisask showed that in Roth's theorem, one may take $N_0 = 2^{O(1/\varepsilon^{1-c})}$ for some absolute constant $c > 0$. This was a major breakthrough beyond previous work, and remains the current best known bound.

Szemerédi's original proof of Szemerédi's theorem provided essentially no quantitative bounds at all. In other words, he proved that for every $\varepsilon > 0$ and every $k \in \mathbb{N}$, there is some N_0 so that S contains a k-term arithmetic progression for all $N \geq N_0$ and all $S \subseteq [N]$ with $|S| \geq \varepsilon N$, but his proof said basically nothing about how large N_0 must be.

The proof sketched above, using hypergraph regularity, actually does provide some quantitative bounds. Namely, it turns out that the 3-graph regularity lemma involves bounds of the form

$$\underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}\Big\}\, 2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}\Big\}\cdots\Big\}\, 2^2}_{O(\varepsilon^{-C})}$$

for some absolute $C > 0$. Such a bound is called a *wowzer-type* bound. Just as the tower function is obtained by iterating the exponentiation function, the wowzer function is obtained by iterating the tower function. Similarly, the 4-graph regularity lemma invokes bounds that are of the form "iterate the wowzer function $O(\varepsilon^{-C})$ times". In general, the bounds for r-graph regularity involve r levels of iteration. As such, the bounds we obtain for Szemerédi's theorem are extraordinarily weak.

Nonetheless, much more is known. In the early 200s, Gowers found a new proof of Szemerédi's theorem, which gives a *much* stronger bound.

Theorem 6.4 (Gowers 2001)**.** *For $\varepsilon > 0$ and $k \in \mathbb{N}$, let*

$$N_0 = 2^{2^{1/\varepsilon^{2^{2^{k+9}}}}}.$$

Then for every $N \geq N_0$ and every $S \subseteq [N]$ with $|S| \geq \varepsilon N$, S contains a k-term arithmetic progression.

This remains the best known bound for Szemerédi's theorem, and proving it involved a number of breakthroughs. Note that, although this is a big scary expression, it's *much* smaller than the big scary expressions above. In particular, the dependence on k is "merely" quintuple-exponential, rather than the value of k determining how many times we iterate some huge function.

7 The Green–Tao theorem

To end this class, I want to return briefly to the Green–Tao theorem.

Theorem 7.1 (Green–Tao 2004)**.** *For every $k \in \mathbb{N}$, the set $\mathbb{P}$ of primes contains a k-term arithmetic progression.*

To prove this, Green and Tao actually proved a much stronger and more general theorem, called the *relative Szemerédi theorem.* I can't state it precisely because doing so would involve many more complicated definitions, but here is the gist of it.

"Theorem" 7.2 (Relative Szemerédi theorem)**.** *For every $\varepsilon > 0$ and $k \in \mathbb{N}$, there exists some $N_0 \in \mathbb{N}$ so that the following holds. Suppose $R \subseteq [N]$ is a "very pseudorandom" set, and suppose $S \subseteq R$ satisfies $|S| \geq \varepsilon|R|$. Then S contains a k-term arithmetic progression.*

The reason this is called a "relative" Szemerédi theorem is that the statement is essentially the same as the statement of Szemerédi's theorem, except *relative* to a host set R. As long as R "looks random" in an appropriate sense (which I won't explain), then any subset S

of R with $|S| \geq \varepsilon|R|$ has arbitrarily long arithmetic progressions. In particular, you should believe me that $R = [N]$ is appropriately pseudorandom, so this implies Szemerédi's theorem in the case $R = [N]$.

Note that if $|R| \geq \delta N$ for some fixed $\delta > 0$, then the relative Szemerédi theorem already follows from Szemerédi's theorem. The new ingredient in the relative Szemerédi theorem is that we are allowing $|R|$ to be much smaller than N; for proving the Green–Tao theorem, we will take an R of size roughly $N/\log N$.

Proving the relative Szemerédi theorem is the key part of the proof of the Green–Tao theorem. Once one has it, the Green–Tao is, in some sense, easy[1]. Namely, one defines a set R of *almost primes*, which is simply the set of integers with no prime factors smaller than some fixed integer r. One can then show that this set R is appropriately pseudorandom, and that

$$|\mathbb{P} \cap [N]| \geq \varepsilon|R \cap [N]|$$

for some $\varepsilon > 0$ depending on the choice of r. Once we have this, the Green–Tao theorem follows immediately from the relative Szemerédi theorem.

The proof of the relative Szemerédi theorem is hard, but it boils down to the same kinds of ideas that we've seen in this class: a regularity lemma and a counting lemma. In this context, the regularity lemma is usually called the *dense model theorem*.

"Theorem" 7.3 (Dense model theorem)**.** *Let $R \subseteq [N]$ be appropriately pseudorandom, and let $S \subseteq R$ with $|S| \geq \varepsilon|R|$. There exists a **dense model** $\widehat{S} \subseteq [N]$ so that $|\widehat{S}| \geq \varepsilon N$, and such that the relationship between S and R "looks like" the relationship between $\widehat{S}$ and $[N]$.*

I won't explain what "looks like" means in the statement above, but you should think of it as a version of ε-regularity. In the regularity lemma, we "approximated" a graph G by a "bounded-complexity" object given by the regularity partition. The usefulness of the approximation comes from the notion of ε-regularity, which tells us that this notion of approximation "mean something real". In a similar way, we are approximating $S \subseteq R$ by a new set $\widehat{S} \subseteq [N]$, and the notion of "looks like" is one that says that this is a useful notion of approximation.

What makes it a useful notion? As in the case of graph and hypergraph regularity, the key utility comes from a counting lemma. For a set A, let $t_k(A)$ be the number of k-term arithmetic progressions in A.

"Theorem" 7.4 (Counting lemma)**.** *Let $R \subseteq [N]$ be appropriately pseudorandom, let $S \subseteq R$ be a set with $|S| \geq \varepsilon|R|$, and let $\widehat{S} \subseteq [N]$ be a dense model for S. Then*

$$\frac{t_k(\widehat{S})}{t_k([N])} \approx \frac{t_k(S)}{t_k(R)}.$$

In other words, S and $\widehat{S}$ have roughly the same number of k-term arithmetic progressions, when normalized appropriately, i.e. by dividing out by total possible number of k-term progressions they could have.

[1]By "easy", I mean merely quite hard.

The final step in the proof of the relative Szemerédi theorem is using Szemerédi's theorem as a black box. Since $|\widehat{S}| \geq \varepsilon N$, Szemerédi's theorem implies that $t_k(\widehat{S}) > 0$. But then by the counting lemma, we conclude that $t_k(S) > 0$, which is exactly the statement of the relative Szemerédi theorem.