Graph Theory and Additive Combinatorics

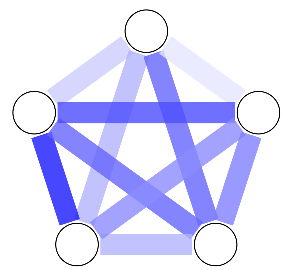
Lecturer: Prof. Yufei Zhao

3
Szemerédi's regularity lemma

3.1 *Statement and proof*

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

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



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

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

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

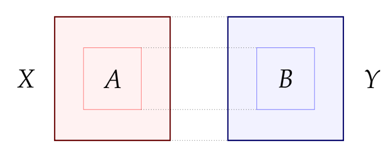
From this, we can define the ***edge density*** between X and Y to be

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

We will drop the subscript G if context is clear.

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

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



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

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

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

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

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

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

We can now state the regularity lemma.

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

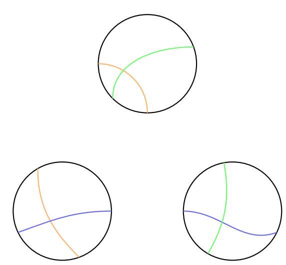
Szemerédi (1978)

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

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

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

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



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The second moment is

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

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

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

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

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

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

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

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

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

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

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

as desired. $\square$

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

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

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

Then

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

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

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

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

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

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

Now we can prove Szemerédi's regularity lemma.

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

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

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

Gowers (1997)

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

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

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

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

3.2 Triangle counting and removal lemmas

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

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

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

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

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

The triangle counting lemma makes this intuition precise.

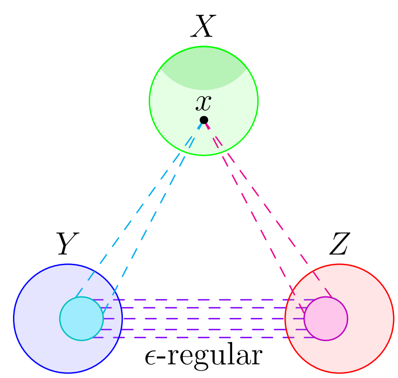
Theorem 3.13 (Triangle counting lemma)**.** *Let G be a graph and X, Y, Z be subsets of the vertices of G such that (X, Y), (Y, Z), (Z, X) are all ϵ-regular pairs some $\epsilon > 0$. Let d_{XY}, d_{XZ}, d_{YZ} denote the edge densities $d(X, Y), d(X, Z), d(Y, Z)$ respectively. If $d_{XY}, d_{XZ}, d_{YZ}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

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

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

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

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



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

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

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

as claimed. ☐

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

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

Ruzsa and Szemerédi (1976)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Indeed, suppose that after following the above procedure and (possibly) removing some edges the resulting graph still has some triangle. Then we can find parts V_i, V_j, V_k (not necessarily distinct) containing each of the vertices of this triangle. Because edges between the pairs described in **(a)** and **(b)** were removed, V_i, V_j, V_k satisfy the hypotheses of the triangle counting lemma. Applying Theorem 3.13 to this triple of subsets implies that the graph still has at least

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

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

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

Then by choosing positive

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

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

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

Fox (2012)

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

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

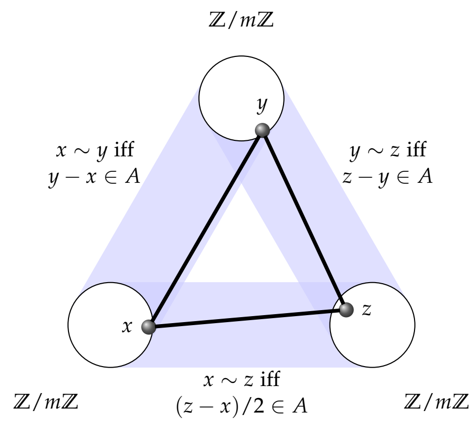
Proof. Let G have m edges. Because each edge lies in one triangle, the number of triangles in G is $m/3$. Since $m < n^2$, this means that G has $o(n^3)$ triangles. By Remark 3.16, we can remove $o(n^2)$ edges to make G triangle-free. However, deleting an edge removes at most one triangle from the graph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

3.3 *Roth's theorem*

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

Roth (1953)

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



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

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

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

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

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

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

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

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

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

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

Sanders (2011)
Bloom (2016)

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

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

3.4 *Constructing sets without 3-term arithmetic progressions*

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

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

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

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

Salem and Spencer (1942)
Behrend (1946)

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

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

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

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

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

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

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

where C is some absolute constant. $\square$

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

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

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

Ajtai and Szemerédi (1975)

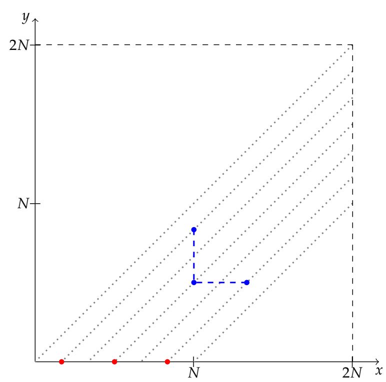
Proof. Consider the sum set $A + A \subseteq [2N]^2$. By the pigeonhole principle, there exists a point $z \in [2N]^2$ such that there are at least $\frac{|A|^2}{(2N)^2}$ pairs of $(a, b) \in A \times A$ satisfying $a + b = z$. Put $A' = A \cap (z - A)$. Then, the size of A' is exactly the number of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

Solymosi (2003)

Now, build a tripartite graph G with parts $X = \{x_1, \ldots, x_N\}$, $Y = \{y_1, \ldots, y_N\}$ and $Z = \{z_1, \ldots, z_{2N}\}$, where each vertex x_i corresponds to a vertical line $\{x = i\} \subseteq \mathbb{Z}^2$, each vertex y_j corresponds to a horizontal line $\{y = j\}$, and each vertex z_k corresponds to a slanted line $\{y = -x + k\}$ with slope -1. Join two distinct vertices of G with an edge if and only if the corresponding lines intersect at a point belonging to A'. Then, each triangle in the graph G corresponds to a set of three lines such that each pair of lines meet at a point of A'. Since A' has no corners with $d \neq 0$, three vertices x_i, y_j, z_k induces a triangle in G if and only if the three corresponding lines pass through the same point of A' and form a trivial corner with $d = 0$. Since there are exactly one vertical line, one horizontal line and one line with slope -1 passing through each point of A', it follows that each edge of G belongs to exactly one triangle. Thus, by Corollary 3.18,

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

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



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

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

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

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

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

Shkredov (2006)

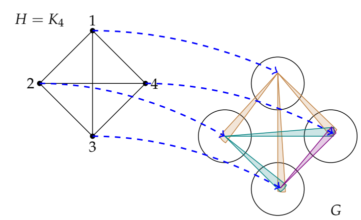
3.5 Graph embedding, counting and removal lemmas

As seen in the proof of the triangle removal lemma Theorem 3.15, one key stepping stone to removal lemmas are counting lemmas. Thus, we would like to generalize the triangle counting lemma to general graphs. To reach our goal, we have two strategies: one is to embed the vertices of a fixed graph one by one in a way that the yet-to-be embedded vertices have lots of choices left, and the other is to analytically remove one edge at a time.

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

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

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



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

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

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

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

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

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

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

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

We will show that

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

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

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

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

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

and

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

so we have

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

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

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

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

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

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

$\square$

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

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

Partition the vertex set using the graph regularity lemma.

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

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

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

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

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

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

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

(a) (V_i, V_j) is not η-regular, or

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

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

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

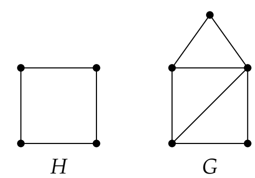
Theorem 3.27 to $\{\tilde{V}_j : j \in [k]\}$, and find that the number of graph homomorphisms from H to G' is at least

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

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

3.6 *Induced graph removal lemma*

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



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

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

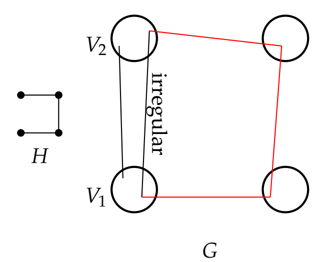
Alon, Fischer, Krivelevich, and Szegedy (2000)

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

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

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

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



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

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

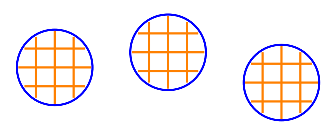
Theorem 3.32 (Colorful graph removal lemma)**.** *For all positive integers k, r, and constant $\epsilon > 0$, there exists a constant $\delta > 0$ so that if $\mathcal{H}$ is a set of r-edge-colorings of K_k, then every r-edge coloring of K_n with less than a δ fraction of its k-vertex subgraphs belonging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

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

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

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

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



The partition $\mathcal{Q}$ in orange refines the partition $\mathcal{P}$ in blue.

Alon, Fischer, Krivelevich, and Szegedy (2000)

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

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

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

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

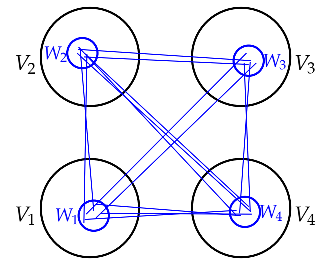
The above version has the same proof as the proof we gave for Theorem 3.5, except instead of starting from the trivial partition, we start from the partition $\mathcal{P}$.

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

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

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

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



A partition with regular subsets.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

of

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

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

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

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

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

Alon and Shapira (2008)

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

3.7 *Property testing*

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

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

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

Alon and Shapira (2008)

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

So far, we have seen that there is a sampling algorithm that tests whether a graph is triangle-free or ϵ-far from triangle-free. Can we find any other properties that are testable? More formally, for which properties $\mathcal{P}$ is there an algorithm such that if we input a graph G that either has property $\mathcal{P}$ or is ϵ-far from having property $\mathcal{P}$, the algorithm determines which of the two cases the graph is in? In particular, for which graphs can this can be done using only an oblivious tester, or in other words by only sampling $k = O(1)$ vertices?

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

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

Alon and Shapira (2008)

3.8 Hypergraph removal lemma

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

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

Rödl et al. (2005)
Gowers (2007)

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

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

Green and Tao (2010)

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

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

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

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

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

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

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

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

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

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

3.9 *Hypergraph regularity*

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

Gowers 2006

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

Definition 3.40 (Naïve definition of 3-graph regularity)**.** Given a 3-graph $G^{(3)}$ and three subsets $V_1, V_2, V_3 \subset V(G^{(3)})$, we say that (V_1, V_2, V_3) is ϵ-regular if, for all $A_i \subset V_i$ such that $|A_i| \geq \epsilon |V_i|$, we have $|d(V_1, V_2, V_3) - d(A_1, A_2, A_3)| \leq \epsilon$. Here, $d(X, Y, Z)$ denotes the fraction of elements of $X \times Y \times Z$ that are in $E(G^{(3)})$.

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

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

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

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

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

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

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

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

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

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

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

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

1. a partition of $E(K_n)$ into graphs such that $G^{(3)}$ sits pseudorandomly on top;

2. a partition of $V(G)$ such that the graphs in the above step are extremely pseudorandom (in a fashion resembling Theorem 3.33).

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

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

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

Gowers (2001)

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

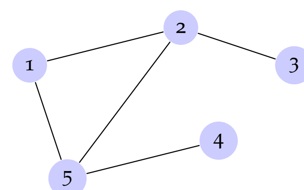
For the multidimensional Szemerédi theorem (Theorem 1.9), the best known bounds generally come from the hypergraph regularity lemma. The first known proof came from ergodic theory, which actually gives no quantitative bounds due to its reliance on compactness arguments. A major motivation for working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

3.10 *Spectral proof of Szemerédi regularity lemma*

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

Tao (2012)

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



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

18.217 Graph Theory and Additive Combinatorics
Fall 2019