> In the beginning God created the heavens and the earth. Now the earth was formless and empty, . . . And God said, "Let there be light," and there was light. . . . and he separated the light from the darkness. God called the light "day," and the darkness he called "night." . . . And God said, "Let there be a vault between the waters to separate water from water." . . . And God said, "Let the water under the sky be gathered to one place, and let dry ground appear." . . . And God said, "Let there be lights in the vault of the sky to separate the day from the night."
>
> Genesis 1:1–14, *New International Version*

1 Numbers

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

Theorem (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*

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

The proof, more or less, is a clever induction argument on both r and k.

van der Waerden's 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. Given a set $S \subseteq \mathbb{N}$, its *(upper) density* is defined as

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

where $[N] = \{1, 2, \ldots, N\}$.

Thus, the set of even numbers has density 1/2, the set of all numbers greater than a million has density 1, and the primes have density 0.

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

$$d(S \cup T) \leq d(S) + d(T)$$

Proof.

$$d(S \cup T) = \limsup_{N \to \infty} \frac{|(S \cup T) \cap [N]|}{N} = \limsup_{N \to \infty} \frac{|S \cap [N]| + |T \cap [N]|}{N} \leq d(S) + d(T)$$

by subadditivity of the lim sup. □

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 (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 (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 (Szemerédi 1969, Szemerédi 1975)**.** *If $S \subseteq \mathbb{N}$ has positive density, then it contains a k-term arithmetic progression for any $k \in \mathbb{N}$. The $k = 4$ case was proven first, with the full case coming six years later.*

Before we discuss proofs, 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 (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 implied from Szemerédi's Theorem, since the primes have density 0, as discussed above. However, both Szemerédi's Theorem and the Green–Tao Theorem are implied by the following conjecture, which remains a major open problem.

Conjecture (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 Graphs

Definition. The Erdős-Rényi random graph $G(n, p)$ is a graph on n vertices, where every pair is connected by an edge randomly with probability p, and all these choices are independent.

Though one might expect a random graph to be, in some sense, maximally unstructured, in other senses random graphs are very structured. For instance, suppose A, B are two disjoint sets of vertices in $G(n, p)$. Then the number of edges we expect to go between them is $p|A||B|$, since there are $|A||B|$ pairs of vertices in $A \times B$, and each one will be an edge with probability p. Moreover, there is a huge range of machinery (generally called *concentration of measure* results) that say that, with extremely high probability, the number of edges we'll see will be very close to its expected value of $p|A||B|$, as long as A, B are reasonably large sets. This is a pretty remarkable fact that is not at all true for graphs in general: the number of edges between sets in a random graph doesn't actually really depend on the sets themselves, only on their sizes. In an arbitrary graph, such results are hopelessly false. An equivalent way of phrasing this result is in terms of the *edge density* $d(A, B)$, defined by

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

where $e(A, B)$ is the number of edges between A and B; thus, the edge density is the fraction of pairs that are edges. In this language, the above just says that in $G(n, p)$, $d(A, B) \approx p$ for all "large" sets A, B.

Similarly, given three disjoint sets A, B, C in $G(n, p)$, the number of triangles with one vertex in each set will be very close to $p^3|A||B||C|$. Again, the expected value is certainly $p^3|A||B||C|$—there are $|A||B||C|$ triples of vertices, and each triple will form a triangle with probability p^3, since we need three independent coin flips to come up heads. And again, the actual number will be very close to its expected value because of concentration of measure results.

As is not surprising, these examples generalize: this very important property of random graphs is that we can count to high accuracy any type of structure contained in it by knowing only global properties (the edge probability p and the sizes of sets we're considering), and not by knowing anything about the actual graph we have.

We can also consider a slight generalization of the Erdős-Rényi model, sometimes called the *stochastic block model.* In this model, we partition our vertices into a collection of blocks, and rather than making all edges appear with the same probability, we can have the probability depend on which blocks the endpoints come from. So for instance, if we have three blocks A, B, C, and have the edge probability be p between A and B, q between B and C, and r between A and C, then the number of triangles with one vertex in each block will be very close to $pqr|A||B||C|$. This is a fairly insignificant generalization (it mostly means there are more parameters), but also in this model we have the crucial property that we can count the instances of substructures by knowing only global parameters.

Of course, if we are simply given a graph, all the above discussion is worthless, since there's no reason to expect it to look anything like a random graph. However, there is a notion due to Szemerédi, called *regularity,* that nevertheless guarantees the same property for a fixed (non-random) graph.

Definition. Given a graph $G = (V, E)$, some $\varepsilon > 0$, and two disjoint subsets of vertices $A, B \subseteq V$, 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.

We will now see that ε-regularity is sufficient to guarantee properties like we had above for random graphs.

Lemma (Counting Lemma)**.** *Let A, B, C be disjoint sets of vertices, and suppose that each of the three pairs $(A, B), (B, C), (A, C)$ is ε-regular. Let*

$$p = d(A, B) \qquad q = d(B, C) \qquad r = d(A, C)$$

Then the number of triangles with one vertex in A, one in B, and one in C is approximately $pqr|A||B||C|$. More formally, if $p, q, r \geq 2\varepsilon$, then the number of such triangles is at least

$$(1 - 2\varepsilon)(p - \varepsilon)(q - \varepsilon)(r - \varepsilon)|A||B||C|$$

and one can also prove a similar upper bound.

Proof. For every $a \in A$, let $d_B(a), d_C(a)$ denote the number of neighbors of a in B, C, respectively. Then we first claim that the number of $a \in A$ such that $d_B(a) \leq (p - \varepsilon)|B|$ is at most $\varepsilon|A|$. For if not, the set of such a would form a subset $A' \subseteq A$ with $|A'| \geq \varepsilon|A|$ and such that $d(A', B) \leq p - \varepsilon$, which contradicts the ε-regularity. Similarly, there are at most $\varepsilon|A|$ vertices $a \in A$ such that $d_C(a) \leq (r - \varepsilon)|C|$.

Now, fix some $a \in A$ with $d_B(a) \geq (p - \varepsilon)|B|, d_C(a) \geq (r - \varepsilon)|C|$. Then let $B' \subseteq B$ be the set of neighbors of a in B, and define $C' \subseteq C$ similarly. Then since we assumed that $p, r \geq 2\varepsilon$, we get that $|B'| \geq \varepsilon|B|, |C'| \geq \varepsilon|C|$, so by regularity of the pair (B, C), we know that

$$|d(B', C') - q| < \varepsilon$$

and thus

$$d(B', C') \geq q - \varepsilon$$

and thus

$$e(B', C') \geq (q - \varepsilon)|B'||C'| \geq (q - \varepsilon)(p - \varepsilon)(r - \varepsilon)|B||C|$$

However, every edge between B' and C' yields a triangle containing a, since a is adjacent to all vertices in B', C'. Now, we sum over the $\geq (1 - 2\varepsilon)|A|$ vertices $a \in A$ that have $d_B(a) \geq (p - \varepsilon)|B|, d_C(a) \geq (r - \varepsilon)|C|$. Doing this, we get that the number of triangles going between A, B, C is at least

$$(1 - 2\varepsilon)(p - \varepsilon)(q - \varepsilon)(r - \varepsilon)|A||B||C|$$

□

Exactly the same idea can be used to derive a counting lemma for structures other than triangles, and the general theme is that ε-regularity guarantees that our graph behaves like a random graph with probabilities given by the edge densities.

The moral of this story is that ε-regularity is a very strong condition to apply to our graphs, since it means that they look like random graphs. ε-regularity allows us to forget any structure our graph actually has, and to remember only a few global parameters, and this suffices to know essentially all there is to know about the graph. Therefore, the following theorem is very surprising:

Theorem (Szemerédi Regularity Lemma). *Let $\varepsilon > 0$. Then there exists some integer M such that we may partition the vertices of G as*

$$V = C_1 \sqcup C_2 \sqcup \cdots \sqcup C_k$$

with $k \leq M$ so that

1. *Each C_i has the same size, plus or minus* 1.
2. *At most $\varepsilon\binom{k}{2}$ of the pairs (C_i, C_j) for $1 \leq i < j \leq k$ are **not** ε-regular.*

In other words, *every* graph can be made to look essentially like a random graph. Moreover, the number of pieces we need depends *only* on the parameter ε, which captures how good we want our approximation to be; in particular, it does *not* depend on the number of vertices of G! Thus, in some sense, all sufficiently large graphs have exactly the same structure, up to some approximation error.

Proof idea. The proof idea is very simple: every time we see any irregularity, we try to improve it, and repeat this iteratively.

We begin with the trivial partition P_0. At each step, we will refine our partition P_i to get a new one P_{i+1}, and will stop when we have the desired properties. To refine the partition, let A, B be two of the blocks in P_i. If the pair (A, B) is ε-regular, we do nothing. If not, there must exist $A' \subseteq A, B' \subseteq B$ with $|A'| \geq \varepsilon|A|, |B'| \geq \varepsilon|B|$, and with $d(A, B)$ significantly different from $d(A', B')$. So we refine the partition by cutting A into $A' \sqcup (A \setminus A')$, and similarly B as $B' \sqcup (B \setminus B')$. By doing this, we've made things a bit more regular.

To prove that this process ever terminates, and that it does so in a bounded number of steps (bounded only in terms of ε), requires introducing a progress measure, called the *mean-squared density.* We won't do the details, but suffice to say that the necessary properties of this progress measure are simple to state and prove (they are all simple consequences of the Cauchy–Schwarz inequality). □

To end this section, we will state and prove one very important consequence of the Szemerédi regularity lemma, which we will later use to prove Roth's theorem:

Lemma (Triangle Removal Lemma, Ruzsa–Szemerédi, 1978). *For every $\varepsilon > 0$, there exists a $\delta > 0$ such that every n and every graph G on n vertices with at most δn^3 triangles, we may remove εn^2 edges from G in order to make it triangle-free.*

Note that G may have up to $\binom{n}{2} \approx n^2/2$ edges, and up to $\binom{n}{3} \approx n^3/6$ 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. There are many ways to think about this result. On the one hand, it says that if a graph has "few" triangles, then they can all be removed by removing "few" edges, which doesn't sound that surprising. On the other hand, the contrapositive of this lemma says that the *only* way to make a graph with few triangles is to start with a triangle-free graph and add a few edges, which perhaps sounds more surprising. Finally, the key fact is that we may remove a *cubic* number of triangles by removing only a *quadratic* number of edges.

Proof. First, using the regularity lemma, we can find some $(\varepsilon/2)$-regular partition of G, namely we can write

$$V(G) = C_1 \sqcup \cdots \sqcup C_k$$

such that all but $\varepsilon\binom{k}{2}/2$ of the pairs (C_i, C_j) are not $(\varepsilon/2)$-regular. Moreover, since refining a partition only makes more pairs more regular, we can assume that $k \geq 2/\varepsilon$. Now, we're going to remove a bunch of edges from G:

1. For every non-$(\varepsilon/2)$-regular pair (C_i, C_j), we will remove all edges between C_i and C_j. Since there are at most $\varepsilon\binom{k}{2}/2$ such pairs, and each pair contains at most $(n/k)^2$ edges (since $|C_i| = n/k \pm 1$), this step removes at most
$$\frac{\varepsilon}{2}\binom{k}{2}\left(\frac{n}{k}\right)^2 \leq \frac{\varepsilon}{2}\frac{k^2}{2}\frac{n^2}{k^2} = \frac{\varepsilon}{4}n^2$$
edges.

2. Within each block C_i, we remove all edges. Since $|C_i| \approx n/k$, this is at most $\binom{n/k}{2}$ edges per cluster, and thus at most
$$k\binom{n/k}{2} \leq k\frac{(n/k)^2}{2} = \frac{n^2}{2k} \leq \frac{\varepsilon}{4}n^2$$
edges.

3. Between every pair of blocks (C_i, C_j) with $d(C_i, C_j) \leq \varepsilon$, we remove all edges. Since there are at most $\binom{k}{2}$ such pairs, and each pair has at most $\varepsilon(n/k)^2$ edges, this step removes at most
$$\binom{k}{2}\frac{\varepsilon n^2}{k^2} \leq \frac{k^2}{2}\frac{\varepsilon n^2}{k^2} = \frac{\varepsilon}{2}n^2$$
edges.

Thus, all in all, we have removed $\leq \varepsilon n^2$ edges to get a subgraph; let's call this subgraph 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.

Since G' has a triangle, and since we deleted all edges internal to every block, such a triangle must go between three blocks C_i, C_j, C_k, and we must have never deleted the edges between any of these pairs. Therefore, since we deleted all edges between irregular pairs, we get that all three of these pairs are $(\varepsilon/2)$-regular. So if we apply the counting lemma we proved earlier to these three $(\varepsilon/2)$-regular pairs, we find that they contain at least

$$(1-\varepsilon)\left(d(C_i, C_j) - \frac{\varepsilon}{2}\right)\left(d(C_j, C_k) - \frac{\varepsilon}{2}\right)\left(d(C_i, C_k) - \frac{\varepsilon}{2}\right)|C_i||C_j||C_k|$$

triangles. Since we removed all edges between pairs that had density $\leq \varepsilon$, we get that all these densities are at least ε. So we have at least

$$(1-\varepsilon)\left(\varepsilon - \frac{\varepsilon}{2}\right)^3\left(\frac{n}{k}\right)^3 = \frac{(1-\varepsilon)\varepsilon^3}{8k^3}n^3 \geq \frac{(1-\varepsilon)\varepsilon^3}{8M^3}n^3$$

triangles in G', and thus at least that many triangles in G. So setting $\delta = (1-\varepsilon)\varepsilon^3/8M^3$ gives us the desired result; note that this value of δ depends only on ε and M, which is itself a function of ε; it does not depend on the graph. □

3 and

Szemerédi used an early version of the regularity lemma to prove Szemerédi's theorem about arithmetic progressions in dense sets. His proof was quite complicated, and there have since been many new techniques and approaches to proving it, each with its own advantages. For our purposes, we will only prove Roth's theorem (the case of 3-term arithmetic progressions), since it's significantly simpler and demonstrates how a result about graphs can be used to conclude a result about numbers. We will prove the following result:

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

Note that this is a slightly different formulation than before, but it certainly suffices; if there were a set $S \subseteq \mathbb{N}$ of positive density but no 3-term arithmetic progression, then taking A to be a sufficiently long initial segment of S would contradict this result (in fact, the two results are equivalent, though the other implication is a bit harder).

Proof (due to Ruzsa–Szemerédi). We will pick N_0 later. From such a set $A \subseteq [N]$, 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 $X \sqcup Y \sqcup Z$. We put no edges inside X or Y or Z. Additionally, we connect $x \in X$ to $y \in Y$ if and only if $y - x \in A$, and we connect $y \in Y$ and $z \in Z$ if and only if $z - y \in A$. Finally, we connect $x \in X$ and $z \in Z$ if and only if $z - x \in 2A$, namely $z - x = 2a$ for some $a \in A$.

Now, for every $x \in [N]$ and $a \in A$, we automatically get a triangle in G, namely the triangle $x \in X, x + a \in Y, x + 2a \in Z$; indeed, by definition, all three of these vertices are pairwise adjacent. Therefore, each $a \in A$ yields at least N triangles in G, so we have at least $N|A| \geq \varepsilon N^2$ 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 εN^2 edges. Since G is a graph on $9N$ vertices, we can apply the contrapositive of the Triangle Removal Lemma (for $\varepsilon/9$) to conclude that there is some $\delta > 0$ such that G has at least δN^3 triangles. Now, we choose N_0 large enough that $\delta N_0^3 > \varepsilon N_0^2$. 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 equal to some a, 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 $a, b, c \in A$. On the other hand, we have that

$$b - a = \frac{z - x}{2} - (y - x) = \frac{z - x - 2y + 2x}{2} = \frac{z + x - 2y}{2}$$

and

$$c - b = (z - y) - \frac{z - x}{2} = \frac{2z - 2y - z + x}{2} = \frac{z + x - 2y}{2}$$

Thus, a, b, c forms a 3-term arithmetic progression contained entirely in A, as desired. □

4 Other stuff

Perhaps the most important result of the Szemerédi regularity lemma is not the lemma itself (though it is used all the time in graph theory today), but rather the rethinking that it brought about. It ushered in an understanding that graphs should be thought of as "very large", which is why ε–δ results have meaning. This analytic perspective has proved very useful, and has really revolutionized the field.

One of the major disadvantages of the Szemerédi regularity lemma is that it provides terrible bounds. Specifically, the number M that it guarantees on the number of parts in a regularity partition is of the form

$$M \leq \underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}_{O(\varepsilon^{-5})}$$

Even worse, Gowers proved that such bounds are unavoidable: there are graphs for which every regularity partition contains a number of parts that is a tower of height $\varepsilon^{-\Omega(1)}$. Therefore, the proof presented above of the triangle removal lemma also gets bounds of the form

$$\frac{1}{\delta} \leq \underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}_{O(\varepsilon^{-5})}$$

and thus one also gets similar bounds for $N_0(\varepsilon)$ in Roth's theorem. For Roth's theorem, significantly better bounds are known, of the form roughly

$$\varepsilon^{-\log(1/\varepsilon)} \lesssim N_0 \lesssim \varepsilon^{-1/\varepsilon}$$

It is a major open problem to improve these bounds (the lower bound is believed to be close to correct). In particular, proving the Erdős conjecture is equivalent to proving (essentially) a sub-exponential upper bound. However, the bounds for triangle removal are much further apart. Running the proof above in reverse, one can use the lower bound for Roth's theorem to get a lower bound for the triangle removal lemma of the same form, namely $1/\delta \gtrsim \varepsilon^{-\log(1/\varepsilon)}$. The upper bound was recently improved by Fox, getting

$$\frac{1}{\delta} \leq \underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}_{O(\log(1/\varepsilon))}$$

In absolute terms, this is a huge improvement, since even reducing the tower height by 1 changes the number enormously. However, it is still a tower-type upper bound, and is very far from the quasipolynomial lower bound.