Szemerédi's Regularity Lemma

One of the most important tools in "dense" combinatorics.

Message: every graph G is the approximate union of constantly many random-like bipartite graph. The number of parts depends only on the error of the approximation constant but **not** the size of G!

For disjoint subsets $X, Y \subseteq V$,

$$d(X, Y) := \frac{|E(X, Y)|}{|X| \cdot |Y|}$$

is the density of the pair (X, Y).

A pair (A, B) of disjoint subsets $A, B \subseteq V$ is called ε-regular pair for some $\varepsilon > 0$ if all $X \subseteq A$, and $Y \subseteq B$ with $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$ satisfy

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

Remark Just like in a random bipartite graph...

Szemerédi's Regularity Lemma

A partition $\{V_0, V_1, \ldots, V_k\}$ of V is called an ε-regular partition if

(i) $|V_0| \leq \varepsilon |V|$

(ii) $|V_1| = \cdots = |V_k|$

(iii) all but at most $\varepsilon\binom{k}{2}$ of the pairs (V_i, V_j), with $1 \leq i < j \leq k^2$, are ε-regular

V_0 is the exceptional set

Regularity Lemma (Szemerédi) $\forall \varepsilon > 0$ and $\forall$ integer $m \geq 1$ $\exists$ integer $M = M(\varepsilon, m)$ such that every graph of order at least m admits an ε-regular partition $\{V_0, V_1, \ldots, V_k\}$ with $m \leq k \leq M$.

Was devised to prove that "dense sets of integers contain an arithmetic progression of arbitrary length".

Proof of the Erdős-Stone Thm

Erdős-Stone Theorem. (Reformulation) For any $\gamma > 0$ and integers $r \geq 2$, $t \geq 1$ there exists an integer $N = N(r, t, \gamma)$, such that any graph G on $n \geq N$ vertices with more than $\left(1 - \frac{1}{r-1} + \gamma\right)\binom{n}{2}$ edges contains $T_{rt,r}$.

Proof strategy:

- Based on an ε-regular partition, build a ”regularity graph“ R of G. (Regularity Lemma)
- Show that R contains a K_r (Turán's Theorem)
- Show that $K_r \subseteq R \;\Rightarrow\; T_{rt,r} \subseteq G$

Regularity graph

Given ε-regular partition $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ of G, $m \leq k \leq M(\varepsilon, m)$,
define the regularity graph $R = R(\mathcal{P}, d)$

$V(R) = \{V_1, \ldots, V_k\}$

$V_i V_j \in E(R)$ if (V_i, V_j) is ε-regular pair with density $d(V_i, V_j) \geq d$

Goal Choose ε, m, d such that "most" edges of G go between the sets V_i and V_j with $V_i V_j \in E(R)$

How many edges are not at the "right place"?

# of edges inside V_i: at most $k\binom{n/k}{2} < \frac{n^2}{k} < \frac{n^2}{m}$

# of edges incident to V_0: at most $\varepsilon n \cdot n = \varepsilon n^2$

# of edges between non-regular pairs:
at most $\varepsilon \binom{k}{2} \left(\frac{n}{k}\right)^2 < \varepsilon n^2$

# of edges between pairs of density $< d$:
at most $\binom{k}{2} d \left(\frac{n}{k}\right)^2 \leq dn^2$

Regularity graph contains an r-clique

Conclusion: If $\varepsilon, m,$ and d is chosen such that

$$d + 2\varepsilon + \frac{1}{m} < \frac{\gamma}{2}$$

then "most" edges of G go between sets V_i and V_j with $V_iV_j \in E(R)$.

"most" means at least $\left(1 - \frac{1}{r-1}\right)\binom{n}{2} + \frac{\gamma}{2}n^2$

On the other hand: # of edges of G going between sets V_i and V_j with $V_iV_j \in E(R)$:

$$\text{at most } |E(R)| \cdot \left(\frac{n}{k}\right)^2$$

Hence

$$\begin{aligned}
\left(1 - \frac{1}{r-1}\right)\binom{n}{2} + \frac{\gamma}{2}n^2 \ &\leq\ |E(R)| \cdot \left(\frac{n}{k}\right)^2 \\
\left(1 - \frac{1}{r-1}\right)\binom{k}{2} + \frac{\gamma}{2}k^2 \ &\leq\ |E(R)|
\end{aligned}$$

Choose $m = m(\gamma)$ such that

$ex(m, K_r) \leq \left(1 - \frac{1}{r-1}\right)\binom{m}{2} + \frac{\gamma}{2}m^2$

Then Turán's Theorem $\Rightarrow$ R contains a K_r

Finding $T_{rt,r}$

There are r classes $V_{i_1}, \ldots, V_{i_r}$ such that (V_{i_j}, V_{i_ℓ}) is an ε-regular pair of density at least d, for every $1 \leq j < \ell \leq r$.
Let $\tilde{n} = |V_{i_j}|$. Then $\frac{n}{k} \geq \tilde{n} \geq \frac{1-\epsilon}{k} n$.

We find a $T_{rt,r}$ in $G[V_{i_1} \cup \cdots \cup V_{i_r}]$.

Lemma
Let (A, B) be an ε-regular pair with $d(A, B) \geq d$
Let $Y \subseteq B$ be a subset with $|Y| \geq \varepsilon |B|$.
Then

$$|\{v \in A : d_Y(v) < (d - \varepsilon)|Y|\}| < \varepsilon |A|.$$

Proof. Otherwise the subsets $Y \subseteq B$ and $\{v \in A : d_Y(v) < (d - \varepsilon)|Y|\} \subseteq A$ would contradict the ε-regularity of (A, B). $\square$

For a set $S \subseteq V$ let $\Gamma(S) = \cap_{v \in S} N(v)$ denote the set of common neighbors of the vertices of S.

Finding $T_{rt,r}$

$$(d-\varepsilon)^{t-1}\tilde{n} \geq \varepsilon\tilde{n}$$
$$(r-1)\varepsilon\tilde{n} + t - 1 < \tilde{n}$$
$$\Downarrow$$
$$\exists S_1 \subseteq V_1,\ \ |S_1| = t$$
$$|\Gamma_{V_i}(S_1)| \geq (d-\varepsilon)^t\tilde{n} \quad \text{for } i = 2, 3, \ldots, r$$

$$(d-\varepsilon)^{2t-1}\tilde{n} \geq \varepsilon\tilde{n}$$
$$(r-2)\varepsilon\tilde{n} + t - 1 < (d-\varepsilon)^t\tilde{n}$$
$$\Downarrow$$
$$\exists S_2 \subseteq V_2,\ \ |S_2| = t$$
$$|\Gamma_{V_i}(S_1 \cup S_2)| \geq (d-\varepsilon)^{2t}\tilde{n} \quad \text{for } i = 3, \ldots, r$$

.

.

.

$$(d-\varepsilon)^{(r-1)t-1}\tilde{n} \geq \varepsilon\tilde{n}$$
$$\varepsilon\tilde{n} + t - 1 < (d-\varepsilon)^{(r-2)t}\tilde{n}$$
$$\Downarrow$$
$$\exists S_{r-1} \subseteq V_{r-1},\ \ |S_{r-1}| = t$$
$$|\Gamma_{V_r}(\cup_{i=1}^{r-1} S_i)| \geq (d-\varepsilon)^{(r-1)t}\tilde{n}$$

Finding $T_{rt,r}$

$\exists S_r \subseteq N_{V_r}(\cup_{i=1}^{r-1} S_i), \quad |S_r| = t$
and thus $G[S_1 \cup \cdots \cup S_r]$ contains a $T_{rt,r}$ provided

$$(d-\varepsilon)^{(r-1)t}\tilde{n} > t-1$$

Strongest of the blue conditions:

$$(d-\varepsilon)^{(r-1)t-1} \geq \varepsilon$$

Let's not forget:

$$d + 2\varepsilon + \frac{1}{m} < \frac{\gamma}{2}$$

Choose, for example: $m = \frac{6}{\gamma}$ *

$$d = \frac{\gamma}{6}$$

$$\varepsilon = \frac{1}{r} \cdot \left(\frac{d}{2}\right)^{t(r-1)}$$

Green conditions are satisfied by choosing a large enough threshold vertex number $N = N(r,t,\gamma)$.

$$r, t, \gamma \rightsquigarrow m, d, \varepsilon \rightsquigarrow M \rightsquigarrow N$$

*This is also a large enough m for using Turán's Theorem earlier.

The Erdős-Turán conjecture

A set S of positive integers is k-AP-free if $\{a, a+d, a+2d, \ldots, a+(k-1)d\} \subseteq S$ implies $d = 0$.

$s_k(n) = \max\{|S| : S \subseteq [n] \text{ is } k\text{-AP-free}\}$

How large is $s_k(n)$? Could it be linear in n?

Erdős-Turán Conjecture (Szemerédi's Theorem)
For every constant k, we have

$$s_k(n) = o(n).$$

Construction (Erdős-Turán, 1936)

$$s_3(n) \geq n^{\frac{\log 2}{\log 3}}.$$

$S = \{s : \text{ there is no 2 in the ternary expansion of } s\}$

S is 3-AP-free. For $n = 3^l$, $|S \cap [n]| = 2^l$

Roth's Theorem (1953) $s_3(n) = o(n)$.

History of Szemerédi's Theorem

Szemerédi's Theorem (1975) For any integer $k \geq 1$ and $\delta > 0$ there is an integer $N = N(k, \delta)$ such that any subset $S \subseteq \{1, \ldots, N\}$ with $|S| \geq \delta N$ contains an arithmetic progression of length k.

Was conjectured by Erdős and Turán (1936).
Featured problem in mathematics, inspired lots of great new ideas and research in various fields;

- Case of $k = 3$: analytic number theory (Roth, 1953; Fields medal)
- First proof for arbitrary k: combinatorial (Szemerédi, 1975)
- Second proof: ergodic theory (Furstenberg, 1977)
- Third proof: analytic number theory (Gowers, 2001; Fields medal)
- Fourth proof: fully combinatorial (Rödl-Schacht, Gowers, 2007)
- Fifth proof: measure theory (Elek-Szegedy, 2007+)

One of the ingredients in the proof of Green and Tao: "primes contain arbitrary long arithmetic progression"

Applications of the Regularity Lemma

Removal Lemma For $\forall\gamma > 0$ $\exists\delta = \delta(\gamma)$ such that the following holds. Let G be an n-vertex graph such that at least $\gamma\binom{n}{2}$ edges has to be deleted from G to make it triangle-free. Then G has at least $\delta\binom{n}{3}$ triangles.

Proof. Apply Regularity Lemma (Homework).

Roth's Theorem For $\forall\epsilon > 0$ $\exists N = N(\epsilon)$ such that for any $n \geq N$ and $S \subseteq [n]$, $|S| \geq \epsilon n$,
there is a three-element arithmetic progression in S.

Proof. Create a tri-partite graph $H = H(S)$ from S.

$$V(H) = \{(i,1) : i \in [n]\} \cup \{(j,2) : j \in [2n]\} \\ \cup \{(k,3) : k \in [3n]\}$$

$(i,1)$ and $(j,2)$ are adjacent if $j - i \in S$
$(j,2)$ and $(k,3)$ are adjacent if $k - j \in S$
$(i,1)$ and $(k,3)$ are adjacent if $k - i \in 2S$

Roth's Theorem — Proof cont'd

$(i,1),(i+x,2),(i+2x,3)$ form a triangle
for every $i \in [n]$, $x \in S$.
These $|S|n$ triangles are pairwise edge-disjoint.

$$\Downarrow$$

At least $\epsilon n^2 \geq \frac{\epsilon}{18}\binom{|V(H)|}{2}$ edges must be removed from H to make it triangle-free.

Let $\delta = \delta\left(\frac{\epsilon}{18}\right)$ provided by the Removal Lemma.
There are at least $\delta\binom{|V(H)|}{3}$ triangles in H.

S has no three term arithmetic progression

$$\Downarrow$$

$\{(i,1),(j,2),(k,3)\}$ is a triangle iff $j-i = k-j \in S$.
Hence the number of triangles in H is equal to
$n|S| \leq n^2 < \delta\binom{6n}{3}$, provided $n > N(\epsilon) := \left\lfloor \frac{1}{\delta} \right\rfloor$. $\square$

Behrend's Construction

Construction (Behrend, 1946)

$$s_3(n) \geq n^{1-O\left(\frac{1}{\sqrt{\log N}}\right)}.$$

Construct set of vectors $\bar{a} = (a_0, a_1, \ldots, a_{l-1})$:

$V_k = \{\bar{a} \in \mathbb{Z}^l \ : \ \|\bar{a}\|^2 = k, \ 0 \leq a_i < \frac{d}{2}$ for all $i < q\}$,

where $\|\bar{a}\| = \sqrt{\sum_{i=0}^{l-1} a_i^2}$.

Interpret a vector $\bar{a} \in \{0, 1, \ldots, d-1\}^l$ as an integer written in d-ary:

$$n_{\bar{a}} = \sum_{i=0}^{l-1} a_i d^i.$$

Let

$$S_k = \{n_{\bar{a}} : \bar{a} \in V_k\}$$

Claim $S_k \subseteq [d^l]$ is 3-AP-free for every k.

Proof. Assume $n_{\bar{a}} + n_{\bar{b}} = 2n_{\bar{c}}$.
Then $a_i + b_i = 2c_i$ for every $i < l$, because $a_i + b_i$ and $2c_i$ are both $< d$ (so there is no carry-over)
So $\bar{a} + \bar{b} = 2\bar{c}$. But

$$\|2\bar{c}\| = 2\|\bar{c}\| = 2\sqrt{k} = \|\bar{a}\| + \|\bar{b}\| \geq \|\bar{a} + \bar{b}\|,$$

and equality happens only if $\bar{a}$ and $\bar{b}$ are parallel. Since they are of the same length, we conclude $\bar{a} = \bar{b}$. □

Take the *largest* S_k. Bound its size by averaging:

$\bar{a} \in \{0, 1, \ldots, d-1\}^l \Rightarrow \|\bar{a}\|^2 < ld^2$,
so there is a k for which

$$|S_k| \geq \frac{|\bigcup_i S_i|}{ld^2} = \frac{(d/2)^l}{ld^2} = \frac{d^{l-2}}{2^l l}$$

For given N, choose $l = \sqrt{\log N}$ and $d = N^{\frac{1}{l}}$.