Motivation: Random Bipartite Graphs

Let $G = (A \cup B, E)$, $|A| = |B| = n$, every possible edge $ab$ is chosen independently with probability $p$, $0 \leq p \leq 1$.

Chernoff Bound. If $Z_1, \ldots, Z_m$ are independent $0/1$-random variables, $\Pr[Z_i = 1] = p$, then

$$\Pr[|\sum_{i} Z_i - pm| > \Delta] < 2e^{-2\Delta^2/m}.$$

Corollary For $X \subseteq A$ and $Y \subseteq B$,

$$\Pr[|e(X, Y) - p|X||Y|| > \varepsilon|X||Y||] < 2e^{-2\varepsilon^2|X||Y||}$$
Random Bipartite Graphs Cont’d

Corollary For $X \subseteq A$ and $Y \subseteq B$,
\[
\Pr[|e(X, Y) - p|X||Y||] > \varepsilon|A||B| < 2e^{-2\varepsilon^2|X||Y|}
\]

Moreover, the number of $X$’s, $Y$’s is $2^n \cdot 2^n = 4^n$.
Thus, with prob. $1 - 4^n \cdot 2e^{-2\varepsilon^2 n^2} \rightarrow 1$ (as $n \rightarrow \infty$),
\[
|e(X, Y) - p|X||Y|| \leq \varepsilon|A||B|
\]
holds for all $X \subseteq A, Y \subseteq B$ with $|X|, |Y| \geq \varepsilon n$.

Remarks. For random graphs, it would actually be enough to require $|X|, |Y| \geq C\sqrt{n}$ for a suitable constant $C$ depending on $\varepsilon$. 

2
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 $\varepsilon$-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 \( \varepsilon \)-regular partition if

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

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

(iii) all but at most \( \varepsilon \left( \frac{k}{2} \right) \) of the pairs \((V_i, V_j)\), with \( 1 \leq i < j \leq k^2 \), are \( \varepsilon \)-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 \( \varepsilon \)-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”.

4
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”
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 $\varepsilon$-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 $\varepsilon$-regular partition $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ of $G$,  
$m \leq k \leq M(\varepsilon, m)$,  
define the regularity graph $R = R(\mathcal{P}, d)$  
$V(R) = \{V_1, \ldots, V_k\}$  
$V_iV_j \in E(R)$ if $(V_i, V_j)$ is $\varepsilon$-regular pair with  
density $d(V_i, V_j) \geq d$

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

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

- # of edges inside $V_i$: at most  
  $k\left(\frac{n}{k}\right) < \frac{n^2}{k} < \frac{n^2}{m}$

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

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

- # of edges between pairs of density $< d$:  
  at most  
  $\left(\frac{k}{2}\right)d\left(\frac{n}{k}\right)^2 \leq dn^2$
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 $(1 - \frac{1}{r-1}) \left( \frac{n}{2} \right) + \frac{\gamma}{2} n^2$

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

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

Hence

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

$$\left( 1 - \frac{1}{r-1} \right) \left( \frac{k}{2} \right) + \frac{\gamma}{2} k^2 \leq |E(R)|$$

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

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

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 $(V_{i_a}, V_{i_b})$ is an $\varepsilon$-regular pair of density at least $d$, for every $1 \leq a < b \leq r$.

W.l.o.g. the classes are $V_1, \ldots, V_r$ (else, relabel).
Set $\ell := |V_1| = \ldots = |V_r|$. Thus, $\frac{1-\varepsilon}{M} n \leq \ell \leq n/m$.

**Goal:** Find a $T_{rt,r}$ in $G[V_1 \cup \cdots \cup V_r]$.

**Lemma**
Let $(A, B)$ be an $\varepsilon$-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 $\varepsilon$-regularity of $(A, B)$. \hfill \Box
Finding $T_{rt,r}$

\[
(d - \varepsilon)^{t-1} \ell \geq \varepsilon \ell \\
(r - 1)\varepsilon \ell \leq \ell - t
\]
\[
\Downarrow
\]
\[
\exists S_1 \subseteq V_1, \ |S_1| = t \\
|N_{V_i}(S_1)| \geq (d - \varepsilon)^t \ell \text{ for } i = 2, 3, \ldots, r
\]

\[
(d - \varepsilon)^{2t-1} \ell \geq \varepsilon \ell \\
(r - 2)\varepsilon \ell \leq (d - \varepsilon)^t \ell - t
\]
\[
\Downarrow
\]
\[
\exists S_2 \subseteq V_2, \ |S_2| = t \\
|N_{V_i}(S_1 \cup S_2)| \geq (d - \varepsilon)^{2t} \ell \text{ for } i = 3, \ldots, r
\]
\[
\ldots
\]

\[
(d - \varepsilon)^{(r-1)t-1} \ell \geq \varepsilon \ell \\
\varepsilon \ell \leq (d - \varepsilon)^{(r-2)t} \ell - t
\]
\[
\Downarrow
\]
\[
\exists S_{r-1} \subseteq V_{r-1}, \ |S_{r-1}| = t \\
|N_{V_r}(\bigcup_{i=1}^{r-1} S_i)| \geq (d - \varepsilon)^{(r-1)t} \ell
\]
Finding $T_{rt,r}$

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

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

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: $m \geq \frac{6}{\gamma}$ *

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

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

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

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

*We also needed large $m$ earlier for using Turán’s Theorem.