Szemerédi’s regularity lemma.

Let $\varepsilon > 0$. Let $A, B \subseteq V(G)$ be disjoint. We say that $(A, B)$ is $\varepsilon$-\textbf{regular} in $G$ if for all $X \subseteq A$ and $Y \subseteq B$ with $|X| \geq \varepsilon |A|$ and $|Y| \geq \varepsilon |B|$ we have

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

where

$$d(C, D) = \frac{|\langle C, D \rangle|}{|C| \cdot |D|}$$

and $\langle C, D \rangle = \{e: e \text{ has one end in } C, \text{ the other in } D\}$.

A partition $(V_0, V_1, ..., V_k)$ of $V(G)$ is $\varepsilon$-\textbf{regular} if

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

(ii) $|V_1| = |V_2| = \cdots = |V_k|$

(iii) all but at most $\varepsilon k^2$ pairs $(V_i, V_j)$ are $\varepsilon$-regular ($1 \leq i, j \leq k$).

\textbf{Theorem} (Szemerédi’s regularity lemma). \forall \varepsilon > 0 \forall \text{ integer } m \exists \text{ integer } M \forall \text{ graph } G \text{ on } n \geq m \text{ vertices } \exists k \text{ with } m \leq k \leq M \text{ and an } \varepsilon\text{-regular partition } (V_0, V_1, V_2, ..., V_k) \text{ of } V(G).
The Erdős-Stone theorem

**Turán’s theorem.** If $G$ has no $K_r$ subgraph, then

$$|E(G)| \leq |E(T_{r-1}(n))|,$$

with equality if and only if $G \cong T_{r-1}(n)$.

Recall that $T_{r-1}(n)$ = complete $(r - 1)$-partite graph on $n$ vertices with color classes as close to each other in size as possible. Let

$$t_{r-1}(n) := |E(T_{r-1}(n))| \approx \frac{r - 2}{r - 1} \cdot \frac{n^2}{2}$$

**Theorem.** (Erdős-Stone) $\forall r, s \forall \varepsilon > 0 \exists n_0 \forall$ graph $G$ on $\geq n_0$ vertices if

$$|E(G)| \geq t_{r-1}(n) + \varepsilon n^2,$$

then $G$ has a $K^r_s := K_{s,s,...,s}$-subgraph.

Equivalently, the hypothesis can be stated as

$$|E(G)| \geq \left(\frac{r - 2}{r - 1} + \varepsilon\right) \frac{n^2}{2}.$$
Definition.

\[ \text{ex}(n, H) := \max\{|E(G)|: |V(G)| = n, G \text{ has no } H \text{ subgraph}\} \]

Example. \( \text{ex}(n, K_r) = t_{r-1}(n) \).

Corollary. \( \lim_{n \to \infty} \frac{\text{ex}(n, H)}{\binom{n}{2}} = \frac{\chi(H)-2}{\chi(H)-1} \) for every \( H \) with \( \geq 1 \) edge.

Proof. Weekly exercise.

---

Definition. The upper density of an (infinite) graph \( G \) is

\[ \lim \sup \left\{ \frac{|E(H)|}{\binom{|V(H)|}{2}} : H \subseteq G, H \text{ finite} \right\} \]

Corollary. The upper density of an infinite graph is

\[ 0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots, \text{ or } 1 \]

Proof. Weekly exercise.
**Proof of Erdős-Stone.** Let $r, s$ and $\gamma > 0$ be given. Want to show \( \exists n_0 \) such that if \( |V(G)| \geq n_0 \), then

\[
|E(G)| \geq \left( \frac{1}{2} \frac{r - 1}{r - 2} + \gamma \right) n^2 \Rightarrow K^r_s \subseteq G
\]

Let \( \varepsilon = \varepsilon(r, s, \gamma) \) and \( m = m(\gamma) \), TBD.

Let \( M \) be as in Szemerédi’s regularity lemma. Let \( G \) have at least \( m \) vertices. By Szemerédi’s regularity lemma \( G \) has an \( \varepsilon \)-regular partition \( (V_0, V_1, \ldots, V_k) \) with \( m \leq k \leq M \).

Note \( (1 - \varepsilon) \frac{n}{k} \leq |V_i| \leq \frac{n}{k} \) for \( i = 1, 2, \ldots, k \).

Let \( R \) be the “regularity graph” defined by \( V(R) := \{1, 2, \ldots, k\} \) and \( i \sim j \) in \( R \) if \( (V_i, V_j) \) is \( \varepsilon \)-regular and the density \( d(V_i, V_j) \) is \( \geq \gamma \).

**Claim 1.** \( \varepsilon, m \) can be chosen so that \( R \) has a \( K_r \) subgraph.

**Pf.** If not, then by Turán’s theorem \( |E(R)| \leq t_{r-1}(k) \leq \frac{1}{2} \frac{r-2}{r-1} k^2 \).

We have that \( |E(G)| \) is at most the sum of:

- # of edges incident with \( V_0 \), which is \( \leq \varepsilon n^2 \)
- # edges in some \( V_i \), which is \( \leq k \frac{1}{2} \left( \frac{n}{k} \right)^2 \leq \frac{1}{2} \frac{n^2}{m} \)
- # of \( V_i-V_j \) edges for \( i, j \in E(R) \), which is \( \leq |E(R)| \left( \frac{n}{k} \right)^2 \leq \frac{1}{2} \frac{r-2}{r-1} n^2 \)
- # of \( V_i-V_j \) edges for \( (V_i, V_j) \) not \( \varepsilon \)-regular, which is \( \leq \varepsilon k^2 \left( \frac{n}{k} \right)^2 \)
- # of \( V_i-V_j \) edges when \( d(V_i, V_j) < \gamma \), which is \( \leq \binom{k}{2} \gamma \left( \frac{n}{k} \right)^2 \)
\[ |E(G)| \leq \varepsilon n^2 + k \frac{1}{2} \left( \frac{n}{k} \right)^2 + |E(R)| \left( \frac{n}{k} \right)^2 + \varepsilon k^2 \left( \frac{n}{k} \right)^2 + \binom{k}{2} \gamma \left( \frac{n}{k} \right)^2 \]

\[
\leq \frac{1}{2} \frac{r - 2}{r - 1} n^2 + \varepsilon n^2 + \frac{1}{2} \frac{n^2}{k} + \varepsilon n^2 + \frac{\gamma}{2} n^2 \leq \\
\leq \frac{1}{2} \frac{r - 2}{r - 1} n^2 + \left( 2\varepsilon + \frac{1}{2m} + \frac{\gamma}{2} \right) n^2
\]

So on choosing \( \varepsilon, m \) such that \( 4\varepsilon + \frac{1}{m} < \gamma \) we get a contradiction.

This proves Claim 1.

Next we show that if \( \varepsilon \) is sufficiently small, then \( K_s^r \subseteq G \). We may assume that \( \{1, 2, \ldots, r\} \) is a clique in \( R \).

For \( 1 \leq i < j \leq r \) we have \( d(V_i, V_j) \geq \gamma \) and so \( d(X, Y) \geq \gamma - \varepsilon \) for every \( X \subseteq V_i \) and \( Y \subseteq V_j \) with \( |X| \geq \varepsilon |V_i| \) and \( |Y| \geq \varepsilon |V_j| \).

We will show:

\( (*) \) Let \( V_i' \subseteq V_i \) satisfy \( |V_i'| \geq \varepsilon |V_i| \) and \( |V_1'| \geq (r - 1) \varepsilon |V_1| + s \).

Then there exist distinct vertices \( v_1, v_2, \ldots, v_s \in V_1' \) and sets \( V_i'' \subseteq V_i' \) such that

- \( \{v_1, \ldots, v_s\} \) is complete to \( V_i'' \) for all \( i = 2, \ldots, r \)
- \( |V_i''| \geq c|V_i'| \), where \( c = c(\gamma, r) \) is independent of \( \varepsilon \)

To prove the theorem assuming \( (*) \), first apply \( (*) \) to \( V_i' = V_i \) to obtain \( v_1, v_2, \ldots, v_s \in V_1' \) and sets \( V_i' \subseteq V_i \). Then apply \( (*) \) to the sets \( V_2', V_3', \ldots, V_r' \) to obtain \( u_1, u_2, \ldots, u_s \in V_2' \) and sets \( V_i'' \subseteq V_i' \). Then apply \( (*) \) to the sets \( V_3'', V_4'', \ldots, V_r'' \) and so on.
Lemma. Let \( Y \subseteq V_2 \) with \(|Y| \geq \varepsilon|V_2|\). Then all but \( \varepsilon|V_1| \) vertices of \( V_1 \) have \( \geq (\gamma - \varepsilon)|Y| \) neighbors in \( Y \).

Proof. Let

\[
X = \{ v \in V_1 : v \text{ is adjacent to } < (\gamma - \varepsilon)|Y| \text{ vertices in } Y \}.
\]

Then \(|\langle X, Y \rangle| < |X| (\gamma - \varepsilon)|Y|\). Thus \( d(X, Y) < \gamma - \varepsilon \), and so \(|X| < \varepsilon|V_1|\), as desired. \( \square \)

By the lemma, all but \((r - 1)\varepsilon|V_1|\) vertices of \( V_1 \) have \( \geq (\gamma - \varepsilon)|Y| \) neighbors in \( Y \) for \( Y = V_2', V_3', ..., V_r' \). Pick \( v_1 \in V_1' \) with this property and let \( W_i \subseteq V_i' \) be such that \(|W_i| \geq (\gamma - \varepsilon)|V_i'|\) and \( v_1 \) is complete to \( W_i \) for all \( i = 2,3,...,r \). Repeat the same argument, but with \( V_2', V_3', ..., V_r' \) replaced by \( W_2, ..., W_r \). We find \( v_2 \in V_1' - \{ v_1 \} \) and \( Z_i \subseteq W_i \) such that \(|Z_i| \geq (\gamma - \varepsilon)|W_i|\) and \( v_2 \) is complete to \( Z_i \) for all \( i = 2,3,...,r \). After \( s \) iterations we will end up with the vertices \( v_1, v_2, ..., v_s \) and sets \( V_2'', V_3'', ..., V_r'' \). \( \square \)
Theorem (Triangle removal lemma) \( \forall \varepsilon > 0 \exists \delta > 0 \exists n_0 \) such that if \( G \) is a graph on \( \geq n_0 \) vertices with at most \( \delta n^3 \) triangles, then \( \exists F \subseteq E(G) \) such that \( |F| \leq \varepsilon n^2 \) and \( G \setminus F \) is triangle-free.

“If a graph has \( o(n^3) \) triangles, then all triangles can be destroyed by removing \( o(n^2) \) edges”

Application to arithmetic progressions

Lemma. \( \forall \varepsilon > 0 \exists n_0 \forall \text{ graph } G \) on \( n \geq n_0 \) vertices such that every edge is in exactly one triangle has \( \leq \varepsilon n^2 \) edges.

Proof. By the Triangle removal lemma \( \exists \delta > 0 \exists n_0 \) such that every graph on \( \geq n_0 \) vertices with \( \leq \delta n^3 \) triangles can be made \( \Delta \)-free by deleting \( \varepsilon n^2 \) edges. Our graph has \( |E(G)| \leq n^2 \) triangles, which is \( \leq \delta n^3 \) for big enough \( n \). So if \( n \) is a big enough, then \( \exists F \subseteq E(G) \) such that \( |F| \leq \varepsilon n^2 \) and \( G \setminus F \) is triangle free. Let \( F' \) consist of the edges of the unique triangle containing \( e \), for all \( e \in F \). Thus \( |F'| \leq 3\varepsilon n^2 \) and \( E(G) \subseteq F' \), as required. \( \square \)

Definition. A corner is a triple of the form

\[ \{(x, y), (x, y + d), (x + d, y)\} \]

for some \( d \), possibly negative.
Corollary. (Ajtai & Szemerédi 1974). Let \( A \subseteq \{1,2,\ldots,N\}^2 \). If \( A \) contains no corner, then \( |A| = o(N^2) \).

Proof. Consider the \( N \times N \) grid.

\( X = \) horizontal lines \( y = i \) for \( i = 1,2,\ldots,N \)
\( Y = \) vertical lines \( x = i \) for \( i = 1,2,\ldots,N \)
\( Z = \) slope \(-1\) lines \( y = -x + i \) for \( i = 1,2,\ldots,2N-1 \)

Then \( |X| = |Y| = N \) and \( |Z| = 2N - 1 \)

Define a tripartite graph \( G \) on \( X \cup Y \cup Z \) by saying that the two lines are adjacent if their intersection belongs to \( A \). Then \( |V(G)| = 4N - 1 \) and \( |E(G)| = 3|A| \). Since \( A \) is corner-free, each edge belongs to a unique triangle. By the previous lemma, \( 3|A| = o(N^2) \), as desired.

\( \square \)

Corollary. (Roth’s theorem) If \( S \subseteq \{1,2,\ldots,N\} \) contains no 3-term arithmetic progression, then \( |S| = o(N) \).

Proof. Let \( A = \{(x,y) : x,y \in \{1,2,\ldots,2N\}, y - x \in S\} \). If \( A \) has a corner, say \( (x,y), (x,y + d), (x + d, y) \in A \), then
\( y - x \in S, y + d - x \in S, y - x - d \in S \), and so \( S \) contains a 3-term arithmetic progression. We may therefore assume that \( A \) has no corner. For every \( \{s,s'\} \subseteq S \) we have \( (s,s + s') \in A \), and hence \( |A| \geq \binom{|S|}{2} \). By the previous corollary \( \binom{|S|}{2} \leq |A| = o(N^2) \), and so \( |S| = o(N) \).
Application to property testing

Definition. A graph $G$ on $n$ vertices is $\varepsilon$-far from triangle-free if for every set $F \subseteq E(G)$ of size at most $\varepsilon n^2$ the graph $G \setminus F$ has a triangle.

Remark. We cannot hope to test whether a graph is triangle-free in constant time, but how about distinguishing triangle-free graphs from those that are $\varepsilon$-far from triangle-free?

Theorem. For every $\varepsilon > 0$ there exists a randomized algorithm which in constant time accepts every triangle-free graph and rejects every graph which is $\varepsilon$-far from triangle-free with probability at least $2/3$.

Proof. Let $\delta$ be as in the Triangle removal lemma. Thus if $G$ is $\varepsilon$-far from triangle-free, then it has more than $\delta n^3$ triangles. Pick $\delta^{-1}$ triples of vertices uniformly independently at random. If none of those triples form a triangle, then accept the graph; otherwise reject. If the graph is a triangle-free, then it will be accepted. If it is $\varepsilon$-far from triangle-free, then the probability of being accepted is at most

$$\left(1 - \frac{\delta n^3}{\binom{n}{3}}\right)^{\delta^{-1}} \leq \frac{1}{3},$$

if $\delta$ is sufficiently small.