

# 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  $\{V_0, V_1, \dots, V_k\}$  of  $V$  is called an  $\varepsilon$ -regular partition if

- (i)  $|V_0| \leq \varepsilon|V|$
- (ii)  $|V_1| = \dots = |V_k|$
- (iii) all but at most  $\varepsilon \binom{k}{2}$  of the pairs  $(V_i, V_j)$ , with  $1 \leq i < j \leq k$ , 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  $\{V_0, V_1, \dots, V_k\}$  with  $m \leq k \leq M$ .

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

## 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  $S \subseteq \{1, \dots, N\}$  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  $\mathcal{P} = \{V_0, V_1, \dots, V_k\}$  of  $G$ ,  
 $m \leq k \leq M(\varepsilon, m)$ ,

define the **regularity graph**  $R = R(\mathcal{P}, d)$

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

$V_i V_j \in E(R)$  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  $V_i V_j \in E(R)$

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  $\binom{k}{2} d \left(\frac{n}{k}\right)^2 \leq d n^2$

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  $V_i V_j \in E(R)$ .

"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  $V_i V_j \in E(R)$ :

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

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  $V_{i_1}, \dots, V_{i_r}$  such that  $(V_{i_j}, V_{i_\ell})$  is an  $\varepsilon$ -regular pair of density at least  $d$ , for every  $1 \leq j < \ell \leq r$

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

### 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  $(A, B)$ .  $\square$

Finding  $T_{rt,r}$

---

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

$$(r - 1)\varepsilon n \leq n - t$$

$\Downarrow$

$$\exists S_1 \subseteq V_1, |S_1| = t$$

$$|N_{V_i}(S_1)| \geq (d - \varepsilon)^t n \quad \text{for } i = 2, 3, \dots, r$$

$$(d - \varepsilon)^{2t-1}n \geq \varepsilon n$$

$$(r - 2)\varepsilon n \leq (d - \varepsilon)^t n - t$$

$\Downarrow$

$$\exists S_2 \subseteq V_2, |S_2| = t$$

$$|N_{V_i}(S_1 \cup S_2)| \geq (d - \varepsilon)^{2t} n \quad \text{for } i = 3, \dots, r$$

$\cdot$   
 $\cdot$   
 $\cdot$

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

$$\varepsilon n \leq (d - \varepsilon)^{(r-2)t} n - t$$

$\Downarrow$

$$\exists S_{r-1} \subseteq V_{r-1}, |S_{r-1}| = t$$

$$|N_{V_r}(\cup_{i=1}^{r-1} S_i)| \geq (d - \varepsilon)^{(r-1)t} n$$

Finding  $T_{rt,r}$

---

$$\exists S_r \subseteq N_{V_r}(\cup_{i=1}^{r-1} S_i), |S_r| = t$$

and thus  $G[S_1 \cup \dots \cup S_r]$  contains a  $T_{rt,r}$  provided

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

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.

## Applications of the Regularity Lemma\_\_\_\_\_

**Counting 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.



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 Counting Lemma.

There are at least  $\delta \binom{|V(H)|}{3}$  triangles in  $H$ .

$S$  has no three term arithmetic progression



$\{(i, 1), (j, 2), (k, 3)\}$  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  $n > N(\epsilon) := \left\lfloor \frac{1}{\delta} \right\rfloor$ . □

## Applications — Property testing\_\_\_\_\_

Input is extremely large: even a linear time algorithm could take too long.

Have: **random access** to every entry of the input.  
One read from the input is called a **query**.

Want: **some** information in constant time.

An input, given as function  $f : \mathcal{D} \rightarrow F$ , is  **$\varepsilon$ -close to satisfying property  $P$**

if there exists a function  $f' : \mathcal{D} \rightarrow F$  that

- satisfies  $P$  and
- differs from  $f$  at no more than  $\varepsilon|\mathcal{D}|$  places.

An input that is not  $\varepsilon$ -close to satisfying  $P$  is called  **$\varepsilon$ -far from satisfying  $P$** .

Let  $P$  be a property and  $n$  be the input size.

An  **$\varepsilon$ -test for  $P$  with  $q = q(\varepsilon, n)$  queries** is a randomized algorithm that reads the input up to  $q$  places and with probability at least  $\frac{2}{3}$  distinguishes between the case that the input satisfies  $P$  and the case that the input is  $\varepsilon$ -far from satisfying  $P$ .

## Testing linearity

---

Property  $P$  is called **testable** if it is  $\varepsilon$ -testable with  $q = q(\varepsilon)$  queries for any  $\varepsilon > 0$ .

Let  $\mathbb{F}$  be an arbitrary finite field.

**Proposition** Linearity of a function is testable.

It is possible to test with  $q = O(\varepsilon^{-1})$  queries whether a function  $\mathbb{F} \rightarrow \mathbb{F}$  is linear.

**Proposition** It is possible to test with  $q = O(k + \varepsilon^{-1})$  queries whether a function  $\mathbb{F} \rightarrow \mathbb{F}$  is a polynomial of degree at most  $k$ .

*Homework.*

## Triangle-freeness is testable\_\_\_\_\_

Let  $\varepsilon$  be given.

Let  $\delta = \delta(\varepsilon)$  be from the Counting Lemma.

Let  $G = (V, E)$  be an input graph.

Choose uniformly and independently  $\ell = \left\lceil \frac{2}{\delta} \right\rceil$  vertex triplets  $T_1, \dots, T_\ell \subseteq V$ ,  $|T_i| = 3$ .

If a triangle is found **reject**  $G$ , otherwise **accept** it.

What is the probability that we reject a triangle-free input  $G$ ? **0**

What is the probability that we accept an input  $G$  which is  $\varepsilon$ -far from being triangle-free? **at most  $\frac{1}{3}$**

By Counting Lemma

$$\Pr[T_i \text{ forms a triangle in } G] \geq \delta$$

$$\begin{aligned} \Pr[\text{none of } T_1, \dots, T_\ell \text{ is a triangle}] &\leq (1 - \delta)^\ell \leq e^{-\delta\ell} \\ &\leq e^{-2} < \frac{1}{3} \end{aligned}$$