Lecture 11. Applications of regularity lemma.

Szemerédi regularity lemma.

For every $\varepsilon > 0$ and every integer $m \geq 1$. there exists an integer $M = M(\varepsilon, m)$ s.t. any graph of size ≥ 1 has an ε-regular partition $V = V_0 \cup V_1 \cup \cdots \cup V_k$ with $m \leq k \leq M$. such that

① $|V_0| \leq \varepsilon |V|$

② $|V_1| = |V_2| = \cdots = |V_k|$

③ at most εk^2 pairs (V_i, V_j) are not ε-regular. where (X, Y) are ε-regular pair if $\forall A \subseteq X$. $B \subseteq Y$ with $|A| \geq \varepsilon |X|$. $|B| \geq \varepsilon |Y|$.

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

Application: triangle removal lemma.

$\forall \varepsilon > 0$. $\exists \delta = \delta(\varepsilon) > 0$. s.t. if a graph G contains at most δn^3 triangles. then all triangles can be destroyed by removing εn^2 edges.

Proof. We claim that, $\forall X. Y. Z \subseteq V$ (not necessarily disjoint). if $d(X, Y)$. $d(Y. Z)$. $d(Z, X) \geq 2\varepsilon$. then # of triangles in $X \times Y \times Z$ is $\geq (1-2\varepsilon)(d(X,Y)-\varepsilon)(d(Y,Z)-\varepsilon)(d(Z,X)-\varepsilon)|X||Y||Z|$.

Note that if G is a random graph. $\mathbb{E}[\#\text{ of }\Delta] = p^3|X||Y||Z|$

Now applying regularity lemma. $\exists\ \varepsilon/4$-regular partition $V_1 \cup \cdots \cup V_k$.

Then remove "bad edges": 1. $e \in V_i \times V_j$. (V_i, V_j) not $\varepsilon/4$-regular.

2. $e \in V_i \times V_j$. $d(V_i, V_j) \leq \varepsilon/2$. 3. $|V_i|$ or $|V_j| \leq \varepsilon n/4k$.

$\#$ of removed edges $\leq \frac{\varepsilon}{4} \cdot k^2 \cdot \frac{n^2}{k^2} + \binom{k}{2} \frac{\varepsilon}{2} \frac{n^2}{k^2} + k^2 \cdot \frac{\varepsilon n}{4k} \cdot \frac{n}{k} \leq \varepsilon n^2$.

If $\exists$ remaining triangle $\in V_i \times V_j \times V_k$ then $\#$ of triangles $\geq$

$(1-\frac{\varepsilon}{2})(\frac{\varepsilon}{4})^3(\frac{\varepsilon}{4} \cdot \frac{n}{k})^3$. Choose $\delta < (1-\frac{\varepsilon}{2})(\frac{\varepsilon}{4})^3(\frac{\varepsilon}{4k})^3$. $\square$

Triangle counting lemma.

Let $G=(V,E)$ be a graph and $X. Y. Z \subseteq V$. Suppose $(X.Y)$. $(Y.Z)$ $(Z.X)$ are ε-regular pair with density $\geq 2\varepsilon$. Then the number of triangles in $X \times Y \times Z$ is at least

$$(1-2\varepsilon)(d(X,Y)-\varepsilon)(d(Y,Z)-\varepsilon)(d(Z,X)-\varepsilon)|X||Y||Z|.$$

Remark. $X. Y. Z$ are not necessarily disjoint.

Proof. Let $\alpha = d(X,Y)$. $\beta = d(Y,Z)$. $\gamma = d(Z,X)$. $d_S(v) = |N(v) \cap S|$

Then. the number of $x \in X$ such that $d_Y(x) \leq (\alpha-\varepsilon)|Y|$ is at most $\varepsilon|X|$. Similarly $|\{x \in X : d_Z(x) \leq (\gamma-\varepsilon)|Z|\}| \leq \varepsilon|X|$.

If $x \in X$ has $d_Y(x) \geq (\alpha - \varepsilon)|Y|$ and $d_Z(x) \geq (\gamma - \varepsilon)|Z|$.

applying ε-regularity to $N(x) \cap Y$ and $N(x) \cap Z$, it gives that

the number between $N(x) \cap Y$ and $N(x) \cap Z$ is at least

$$(\beta - \varepsilon)\, d_Y(x)\, d_Z(x) \geq (\alpha - \varepsilon)(\beta - \varepsilon)(\gamma - \varepsilon)|Y||Z|. \qquad \square$$

A motivation of the triangle removal lemma is the study of Roth theorem.

Corollary. Suppose G is a graph on n vertices such that every edge lies in a unique triangle. Then G has $o(n^2)$ edges.

Proof: Let m be the number of edge. So # of triangles $= m/3 = o(n^3)$.

So we could remove $o(n^2)$ edges to obtain a triangle-free graph.

However, deleting an edge removes at most one triangle. Thus only $o(n^2)$ triangles are removed, which implies $m = o(n^2)$. $\square$

Theorem (Roth, 1953).

Every subset of integers with positive upper density contains a 3-AP.

(or equivalently, for all $\delta > 0$, $\exists n_0$ s.t. for $n \geq n_0$ any subset of $[n]$ of size $\geq \delta n$ contains an arithmetic progression of length 3).

Proof. $\forall S$ that does not contain AP of length 3, we show $|S| = o(n)$.

To simplify the proof, let $M = 2n+1$ and $S \subseteq \mathbb{Z}/M\mathbb{Z}$. ($M = 2n+1$ to avoid wraparound). So S does not contain 3-term AP mod M.

Now we construct a tripartite graph G whose vertices X, Y and Z are all copies of $\mathbb{Z}/M\mathbb{Z}$. $\forall x \in X$, $y \in Y$, $z \in Z$, connect (x, y) if $y - x \in S$, connect (y, z) if $z - y \in S$, and connect (z, x) if $(z-x)/2 \in S$. Note that M is odd so $(z-x)/2$ always exists.

If (x, y, z) forms a triangle, then $y-x$, $\frac{z-x}{2}$, $z-y \in S$, which forms an AP of length 3. By assumption, $y - x = \frac{z-x}{2} = z - y$. So x, y, z forms an AP of length 3. Conversely, if $x \sim y$ for some $x \in X$, $y \in Y$, clearly $y \sim 2y - x \in Z$, and $x \sim 2y - x \in Z$. Similarly for edges between Y, Z or Z, X. Thus every edge in G belongs to exactly 1 triangle. By corollary, there are $o(n^2)$ edges. But we know the number of edges is $3M|S|$. Hence $|S| = o(n)$ $\square$.

Now we are going to prove the regularity lemma. The idea is to refine the partition until the partition is ε-regular. To show that the refinement can terminate, we apply a technique called energy increment argument.

Definition (Mean square density). Let $G=(V,E)$ and $n=|V|$.

Suppose $X, Y \subseteq V$. Define $q(X,Y) = \frac{|X||Y|}{n^2} d(X,Y)^2$.

Given two partitions $\mathcal{P}_X : X = X_1 \cup \cdots \cup X_k$ and $\mathcal{P}_Y : Y = Y_1 \cup \cdots \cup Y_\ell$.

define $q(\mathcal{P}_X, \mathcal{P}_Y) = \sum_{i=1}^{k} \sum_{j=1}^{\ell} q(X_i, Y_j)$. Given $\mathcal{P} : V = V_1 \cup \cdots \cup V_k$.

define $q(\mathcal{P}) = \sum_{i=1}^{k} \sum_{j=1}^{k} q(V_i, V_j) = \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2$

Remark. $q(\mathcal{P}) \leq \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{|V_i||V_j|}{n^2} = 1$.

The following lemma shows that q is not decreasing under refinement.

Lemma. Given $X, Y \subseteq V$ and any partitions $\mathcal{P}_X, \mathcal{P}_Y$. it holds that

$$q(\mathcal{P}_X, \mathcal{P}_Y) \geq q(X, Y).$$

Proof. Suppose $\mathcal{P}_X : X = X_1 \cup \cdots \cup X_k$ and $\mathcal{P}_Y : Y = Y_1 \cup \cdots \cup Y_\ell$.

Choose $x \in X$ and $y \in Y$ uniformly at random. Assume $x \in X_i$, $y \in Y_j$.

Let $Z = d(X_i, Y_j)$ be a random variable. Then the expectation is

$$\mathbb{E}[Z] = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{|X_i||Y_j|}{|X||Y|} d(X_i, Y_j) = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{e(X_i, Y_j)}{|X||Y|} = \frac{e(X,Y)}{|X||Y|} = d(X,Y).$$

The second moment is $\mathbb{E}[Z^2] = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{|X_i||Y_j|}{|X||Y|} d(X_i, Y_j)^2 = \frac{q(\mathcal{P}_X, \mathcal{P}_Y)}{|X||Y|} \cdot n^2$

The result follows from $\mathbb{E}[Z^2] \geq \mathbb{E}[Z]^2$. $\square$

Corollary. Given $\mathcal{P} : V = V_1 \cup \cdots \cup V_k$ and any refinement $\mathcal{P}'$. $q(\mathcal{P}') \geq q(\mathcal{P})$.

The following lemma shows that if (X, Y) is not ε-regular, refinement can increase the mean square density.

Lemma. If (X, Y) is not ε-regular, then there exist partitions $X = X_1 \cup X_2$, $Y = Y_1 \cup Y_2$, s.t. $q(\mathcal{P}_X, \mathcal{P}_Y) \geq q(X, Y) + \varepsilon^4 \frac{|X||Y|}{n^2}$

Proof. Since (X, Y) is not ε-regular, there exist $X_1 \subseteq X$, $Y_1 \subseteq Y$ s.t. $|X_1| \geq \varepsilon |X|$, $|Y_1| \geq \varepsilon |Y|$ and $|d(X_1, Y_1) - d(X, Y)| > \varepsilon$.

Let $X_2 = X \setminus X_1$, $Y_2 = Y \setminus Y_1$, $x \in X$, $y \in Y$ and $Z = d(X_i, Y_i)$.

Then $\mathbb{E}[Z] = d(X, Y)$, $Var[Z] = \frac{n^2}{|X||Y|}(q(\mathcal{P}_X, \mathcal{P}_Y) - q(X, Y))$

It implies that $Var[Z] = \mathbb{E}[(Z - \mathbb{E}[Z])^2]$

$$\geq (d(X_1, Y_1) - d(X, Y))^2 \Pr[Z - \mathbb{E}[Z] = d(X_1, Y_1) - d(X, Y)]$$

$$\geq \varepsilon^2 \cdot \frac{|X_1||Y_1|}{|X||Y|} \geq \varepsilon^4.$$

□

Finally, if a partition is not ε-regular, refinement can increase q sufficiently.

Lemma. If the partition $\mathcal{P}: V = V_1 \cup \cdots \cup V_k$ is not ε-regular, then there exists a refinement $\mathcal{P}'$ of $\mathcal{P}$, where each V_i is partitioned into at most 2^k parts, such that $q(\mathcal{P}') \geq q(\mathcal{P}) + \varepsilon^5$.

Remark. Here "not ε-regular" means $\sum_{(V_i, V_j) \text{ not } \varepsilon\text{-regular}} |V_i||V_j| \leq \varepsilon n^2$.

Proof. Let $I = \{(i,j) : (V_i, V_j) \text{ is not } \varepsilon\text{-regular}\}$ and $q = q(\mathcal{P})$.

For each $(i,j) \in I$, $\exists$ partitions $V_i = V_i^{j_1} \cup V_i^{j_2}$ and $V_j = V_j^{i_1} \cup V_j^{i_2}$

s.t. $q(\mathcal{P}_{V_i}^{j}, \mathcal{P}_{V_j}^{i}) \geq q(V_i, V_j) + \varepsilon^4 \frac{|V_i||V_j|}{n^2}$.

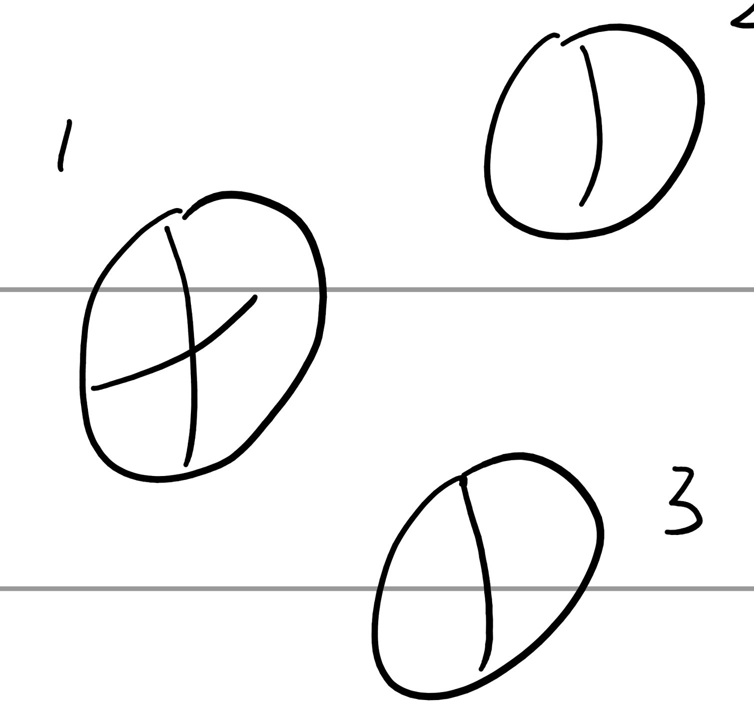
Now consider their common refinement, V_i are divided into at most

2^{k-1} parts. Let $\mathcal{P}'$ be the common refinement. We have that

$$q(\mathcal{P}) = \sum_{S,T} \sum_{i=1}^{k} \sum_{j=1}^{k} q(\mathcal{P}_{V_i}^{S}, \mathcal{P}_{V_j}^{T})$$

$$\geq \sum_{i=1}^{k} \sum_{j=1}^{k} q(V_i, V_j) + \varepsilon^4 \sum_{(V_i, V_j) \text{ not regular}} \frac{|V_i||V_j|}{n^2}$$

$$\geq q + \varepsilon^5.$$



$\square$

Combining all of them together we can prove the regularity lemma.

At each step, k parts $\to k \cdot 2^k \leq 2^{2^k}$ parts, at most $1/\varepsilon^5$ steps.

Introduction to the container method.

$G(n,p)$ has triangles w.h.p. Question: # of edges in Δ-free subgraph = ?

Lower bound: $G(n,p) \cap K_{n/2, n/2} \Rightarrow \mathbb{E}[\# \text{ of edges}] = \frac{n^2}{4} p$.

Upper bound: compute X_m = # of m-edge Δ-free subgraphs.

We hope $\mathbb{E}[X_m] = o(1)$. Unfortunately, $\mathbb{E}[X_m] \geq \binom{n^2/4}{m} p^m \approx \left(\frac{epn^2}{4m}\right)^m >> 1$.

Container theorem for Δ-free graphs.

There exists a collection $\mathcal{G}$ of graphs on $[n]$ such that.

1. (container) for all Δ-free H, $\exists G \in \mathcal{G}$, $H \subseteq G$.

2. (exponentially fewer) $|\mathcal{G}| \le n^{O(n^{3/2})}$ (compared to $2^{e(K_{n/2,n/2})} = 2^{O(n^2)}$).

3. (almost Δ-free). $\forall G \in \mathcal{G}$, G contains $o(n^3)$ triangles.

(with supersaturation) $e(G) \le n^2/4 + o(n^2)$.

Supersaturation : triangle removal lemma.

$\forall \varepsilon > 0$. $\exists \delta > 0$. if $e(G) \ge (\frac{1}{4} + \varepsilon) n^2$. G contains $\ge \delta n^3$ triangles.

Counting # of Δ-free subgraphs :

$\le \sum_{G \in \mathcal{G}} 2^{e(G)} \le |\mathcal{G}| \cdot 2^{n^2/4 + o(n^2)} = 2^{O(n^{3/2} \log n)} \cdot 2^{n^2/4 + o(n^2)} = 2^{n^2/4 + o(n^2)}$.

Asymptotically optimal since $K_{n/2, n/2}$ has $2^{n^2/4}$ subgraphs.

Bounding probability of size-m Δ-free subgraphs in $G(n,p)$:

Let H be a Δ-free subgraph with $m \ge (\frac{1}{4} + \varepsilon) p n^2$ edges.

$\forall G \subseteq \mathcal{G}$. $\Pr[H \subseteq G \cap G(n,p)] \le \Pr[\mathrm{Bin}(e(G), p) \ge m] \le e^{-\Omega(pn^2)}$

by Chernoff bound. Thus. $\Pr[H \subseteq G(n,p)] \le |\mathcal{G}| \cdot e^{-\Omega(pn^2)} = o(1)$.