Lecture 6 (summary)

In this lecture, we will prove Szemerédi Regularity Lemma. We first recall the statement.

Lemma (Szemerédi Regularity Lemma, 1978)**.** *For every $\varepsilon > 0$ and $k_0 \in \mathbb{N}$, there exists $K_0 \in \mathbb{N}$ such that every graph G with at least k_0 vertices has a vertex partition $V_0 \dot\cup V_1 \dot\cup \cdots \dot\cup V_k$, $k_0 \le k \le K_0$, such that*

- *$|V_0| \le \varepsilon|V(G)|$ and $|V_1| = \cdots = |V_k|$, and*
- *all pairs of parts V_i and V_j, $1 \le i < j \le k$, except for at most εk^2 pairs satisfy that*

$$\left| \frac{e(A,B)}{|A|\,|B|} - \frac{e(V_i,V_j)}{|V_i|\,|V_J|} \right| \le \varepsilon$$

holds for all subsets $A \subseteq V_i$ and $B \subseteq V_j$ with $|A| \ge \varepsilon|V_i|$ and $|B| \ge \varepsilon|V_j|$,

where $e(X,Y)$ denotes the number of edges between sets X and Y.

Recall that the pairs V_i and V_j that satisfy the property given in the second bullet point of the lemma are referred to as *ε-regular.*

Fix $\varepsilon \in (0, 1/2)$ and $k_0 \in \mathbb{N}$; the value of K_0 will follow from the proof. The construction of the desired partition of the vertex set of a given graph G proceeds in rounds and the partition of the vertex set of G is refined in each round. Throughout the proof, the number of vertices of G will be denoted by n.

The partition after m rounds will be denoted by $V_0^m \dot\cup V_1^m \dot\cup \cdots \dot\cup V_{k_m}^m$ and it will satisfy that $|V_0^m| \le \frac{(m+1)\varepsilon^6 n}{32}$, $|V_1^m| = \cdots = |V_{k_m}^m|$, and, if $m > 0$, $k_m \le K_m$ for a value of K_m to be determined later. We define the following quantity for a pair of disjoint subsets A and B of $V(G)$ (the intuition behind the quantity that it is a "weighted" square of the density of the pair A and B):

$$q(A,B) = \frac{|A|\,|B|}{n^2} \left(\frac{e(A,B)}{|A|\,|B|} \right)^2 = \frac{e(A,B)^2}{|A|\,|B|\,n^2}.$$

For a partition $\mathcal{P}$ of the vertex set of G, we define $q(\mathcal{P})$ to be the sum of all $q(A,B)$ over all pairs of distinct parts A and B of $\mathcal{P}$. Observe that

$$q(\mathcal{P}) = \sum_{A \ne B \in \mathcal{P}} q(A,B) = \sum_{A \ne B \in \mathcal{P}} \frac{e(A,B)^2}{|A|\,|B|\,n^2} \le \sum_{A \ne B \in \mathcal{P}} \frac{|A|^2 |B|^2}{|A|\,|B|\,n^2} = \sum_{A \ne B \in \mathcal{P}} \frac{|A|\,|B|}{n^2} < 1.$$

Throughout the proof, we will understand the partition $V_0^m \dot\cup V_1^m \dot\cup \cdots \dot\cup V_{k_m}^m$ to be the partition into $|V_0^m| + k_m$ sets that are all one-element subsets of V_0^m and k_m sets $V_1^m, \ldots, V_{k_m}^m$. In each round, the quantity $q(\cdot)$ will increase by at least $\varepsilon^5/16$ and so the number of rounds will be at most $M = \lfloor 16\varepsilon^{-5} \rfloor$.

Before proceeding further, we analyze the difference $q(A_1,B)+q(A_2,B)-q(A,B)$ where $A=A_1\dot\cup A_2$:

$$\begin{aligned} q(A_1,B)+q(A_2,B)-q(A,B) &= \frac{e(A_1,B)^2}{|A_1|\,|B|\,n^2}+\frac{e(A_2,B)^2}{|A_2|\,|B|\,n^2}-\frac{e(A,B)^2}{|A|\,|B|\,n^2}\\ &= \frac{|A_1|\,|A_2|\,|B|}{|A|n^2}\left(\frac{|A|\,e(A_1,B)^2}{|A_1|^2|A_2|\,|B|^2}+\frac{|A|\,e(A_2,B)^2}{|A_2|^2|A_1|\,|B|^2}-\frac{e(A,B)^2}{|A_1|\,|A_2|\,|B|^2}\right)\\ &= \frac{|A_1|\,|A_2|\,|B|}{|A|n^2}\left(\frac{e(A_1,B)^2}{|A_1|^2|B|^2}+\frac{e(A_2,B)^2}{|A_2|^2|B|^2}-\frac{2e(A_1,B)e(A_2,B)}{|A_1|\,|A_2|\,|B|^2}\right)\\ &= \frac{|A_1|\,|A_2|\,|B|}{|A|n^2}\left(\frac{e(A_1,B)}{|A_1|\,|B|^2}-\frac{e(A_2,B)}{|A_2||B|}\right)^2\ge 0. \end{aligned}$$

In particular, it follows that refining a partition $\mathcal{P}$ never decreases $q(\mathcal{P})$.

We now resume the proof of the Szemerédi Regularity Lemma. The initial partition is obtained by splitting the vertices of G into k_0 sets $V_1^0,\dots,V_{k_0}^0$ each with $\lfloor n/k_0\rfloor$ vertices and placing the remaining k_0 vertices to the set V_0^0. If $k_0\le\varepsilon^5 n/32$, then the initial partition has the properties to start the iterative proof; otherwise, $n\le 32k_0\varepsilon^{-5}$ and by setting $K_0\ge 32k_0\varepsilon^{-5}$, the partition sought in the lemma can be the partition of $V(G)$ into n one-element sets.

We now describe the iterative step, which starts with a partition $V_0^m\dot\cup V_1^m\dot\cup\cdots\dot\cup V_{k_m}^m$ described as above. If all but εk_m^2 pairs $1\le i<j\le k_m$ satisfy that the pair V_i^m and V_j^m is ε-regular, then we have the desired partition and stop. Observe that if $X\subseteq V_i$ and $Y\subseteq V_j$ witness that the pair V_i and V_j is not ε-regular, then

$$\left|\frac{e(X,V_j)}{|X|\,|V_j|}-\frac{e(V_i\setminus X,V_j)}{|V_i\setminus X|\,|V_j|}\right|\ge\frac{\varepsilon}{2}\text{ or }\left|\frac{e(X,Y)}{|X|\,|Y|}-\frac{e(X,V_j\setminus Y)}{|X|\,|V_j\setminus Y|}\right|\ge\frac{\varepsilon}{2}.$$

It follows that the difference $q(X,Y)+q(V_i\setminus X,Y)+q(X,V_j\setminus Y)+q(V_i\setminus X,V_j\setminus Y)-q(V_i,V_j)$ is at least

$$\frac{\varepsilon^2}{2}\cdot\frac{|V_i|\,|V_j|}{n^2}\left(\frac{\varepsilon}{2}\right)^2=\frac{\varepsilon^4}{8}\cdot\frac{|V_i|\,|V_j|}{n^2}\ge\frac{\varepsilon^4}{32}\frac{1}{k_m^2}.$$

For every pair V_i and V_j that is not ε-regular fix a partition as defined above and for every ε-regular pair, fix any partition of each set into two disjoint sets. Consider the refinement of the partition $V_1^m\dot\cup\cdots\dot\cup V_{k_m}^m$ to a partition with $2^{k_m-1}k_m$ parts such that two vertices are in the same part iff they belong to the same V_i and they are in the same part for any partition involving V_i. Note that the value $q(\cdot)$ increased by at least $2\varepsilon k_m^2\frac{\varepsilon^4}{32}\frac{1}{k_m^2}=\frac{\varepsilon^5}{16}$. Split each of the obtained parts into parts of size exactly $\left\lfloor\frac{\varepsilon^6}{32}\cdot\frac{n}{2^{k_m-1}k_m}\right\rfloor$ and move the vertices that do not fit to the garbage part, which now becomes the set V_0^{m+1}. Note that the size of V_0^{m+1} is at most $|V_0^m|+2^{k_m-1}k_m\cdot\frac{\varepsilon^6}{32}\cdot\frac{n}{2^{k_m-1}k_m}\le\frac{(m+2)\varepsilon^6 n}{32}$. We set $K_{m+1}=2^{K_m+4}K_m\varepsilon^{-6}$ and note that the number k_{m+1} of the obtained parts is at most K_{m+1}. Since the value of $q(\cdot)$ increases by at least $\frac{\varepsilon^5}{16}$ in each iteration, the whole process ends after at most M iterations, at which points we have at most $K_0:=\max\{K_M,32k_0\varepsilon^{-5}\}$ parts in addition to the garbage part, which has size at most $\frac{(M+1)\varepsilon^6 n}{32}\le\varepsilon n^2$.