Extremal and Probabilistic Graph Theory

Instructor: Jie Ma, Scribed by Tianchi Yang

March 4th 2020, Wednesday

1 Lecture 4. Szemerédi's regularity Lemma

In order to state the regularity lemma precisely, we need some definitions. Let G be a graph. For disjoint sets $X, Y \subseteq V(G)$, the *edge-density* between X and Y is $d(X,Y) = \frac{e(X,Y)}{|X||Y|}$. A bipartite graph (A, B) is called *ε-regular*, if for any $A' \subseteq A$ and $B' \subseteq B$ with $|A'| \geq \varepsilon|A|$, $|B'| \geq \varepsilon|B|$, we have $|d(A', B') - d(A, B)| < \varepsilon$. A partition of $V(G)$ into sets $V_1, V_2, \cdots, V_k$ is an *equipartition* if the sizes of all parts only differ by at most 1. The *order* of the partition is the number of sets in it. An equipartition $V(G) = V_1 \cup V_2 \cup \cdots \cup V_k$ is called *ε-regular* if all but εk^2 of the pairs (V_i, V_j) are ε-regular.

Theorem 1.1 (Szemerédi's Regularity Lemma, 1975)**.** *For any $\varepsilon > 0$, there exists $T(\varepsilon)$ such that any graph G has an ε-regular equipartition of order k, where $\frac{1}{\varepsilon} \leq k \leq T(\varepsilon)$.*

Before proving it, let us introduce some other notation. Assume $P = \{V_1, V_2, \cdots, V_k\}$ is a partition of G with sizes $a_1 n, a_2 n, ..., a_k n$, where $0 \leq a_i \leq 1$, and $\sum_i a_i = 1$. Let $q(V_i, V_j) = a_i a_j d^2(V_i, V_j)$. Denote the *potential function* by

$$q(P) = \sum_{i,j} q(V_i, V_j).$$

Note that $0 \leq q(P) \leq 1$. We say a partition $P' = \{V_1', V_2', \cdots, V_\ell'\}$ *refines* the partition $P = \{V_1, V_2, \cdots, V_k\}$ if each V_i' is contained in some V_j.

Define a function $\text{Tow}(\cdot)$ by $\text{Tow}(k) = 2^{\text{Tow}(k-1)}$ and $\text{Tow}(0) = 1$. As we will see later on, the function $T(\varepsilon)$ from the statement of the regularity lemma is given by $\text{Tow}(\frac{4}{\varepsilon^5})$. The key in proving the regularity lemma is the following lemma.

Lemma 1.2 (Key Lemma)**.** *If $P = \{V_1, \cdots, V_k\}$ is an equipartition of an n-vertex graph which is not ε-regular and $k \geq \frac{1}{\varepsilon^6}$, then there exists a refinement of P of order k^*, say P^*, such that $q(P^*) \geq q(P) + \frac{\varepsilon^5}{2}$ and $k^* \leq k^2 4^k \leq 2^{2^k}$.*

The proof of the regularity lemma now follows easily by repeated applications of the key lemma.

Proof of regularity lemma. (Assuming the key lemma) We start with an arbitrary equipartition of $V(G)$ of order $\frac{1}{\varepsilon^6}$. Then repeatedly apply the key lemma until we get an ε-regular partition. As $0 \leq q(P) \leq 1$ and each application of the key lemma increases $q(P)$ by at least $\frac{\varepsilon^5}{2}$, this process must stop after at most $\frac{2}{\varepsilon^5}$ iterations. This will give an ε-regular equipartition of order $k \leq T(\varepsilon) \triangleq \text{Tow}(\frac{4}{\varepsilon^5})$. ∎

All we need now is to prove the key lemma.

Proposition 1.3. *Let A, B be disjoint subsets of sizes an, bn. Let $A = A_1 \cup ... \cup A_t$ and $B = B_1 \cup ... \cup B_s$, where $|A_i| = x_i an$, $|B_j| = y_j bn$, $\sum_i x_i = \sum_j y_j = 1$ and $0 \le x_i, y_j \le 1$. Suppose that $d(A_i, B_j) = d(A, B) + \varepsilon_{ij}$, then we have*

$$\sum_{i,j} q(A_i, B_j) = q(A, B) + ab \sum_{i,j} x_i y_j \varepsilon_{ij}^2.$$

Proof. Let $d = d(A, B)$. Note that

$$d = \frac{e(A,B)}{|A||B|} = \frac{\sum_{i,j} e(A_i, B_j)}{abn^2} = \sum_{i,j} x_i y_j d(A_i, B_j) = \sum_{i,j} x_i y_j (d + \varepsilon_{ij}) = d + \sum_{i,j} x_i y_j \varepsilon_{ij},$$

implying that $\sum_{i,j} x_i y_j \varepsilon_{ij} = 0$. Thus we have

$$\begin{aligned}\sum_{i,j} q(A_i, B_j) &= \sum_{i,j} x_i a y_j b \cdot d^2(A_i, B_j) = ab \sum_{i,j} x_i y_j (d + \varepsilon_{ij})^2 \\ &= abd^2 + 2abd \sum_{i,j} x_i y_j \varepsilon_{ij} + ab \sum_{i,j} x_i y_j \varepsilon_{ij}^2 = q(A, B) + ab \sum_{i,j} x_i y_j \varepsilon_{ij}^2.\end{aligned}$$

∎

This proposition gives the following facts.

Fact 1.4. *For $A = A_1 \cup ... \cup A_t$ and $B = B_1 \cup ... \cup B_s$, we have*

$$\sum_{i,j} q(A_i, B_j) \ge q(A, B).$$

Fact 1.5. *Assume (A, B) is not ε-regular, where $|A| = |B| = \frac{n}{k}$. Then there exist $A_1 \subseteq A, B_1 \subseteq B$ with $|A_1| \ge \varepsilon|A|, |B_1| \ge \varepsilon|B|$ such that $|d(A_1, B_1) - d(A, B)| > \varepsilon$. Set $A_2 = A \backslash A_1$ and $B_2 = B \backslash B_1$. Then we have*

$$\sum_{1 \le i,j \le 2} q(A_i, B_j) \ge q(A, B) + \frac{\varepsilon^4}{k^2}.$$

Now we are ready to prove the key lemma.

Proof of the key lemma. Assume $P = \{V_1, \cdots, V_k\}$ is not ε-regular. Let I be the set of pairs (i, j) such that (V_i, V_j) is not ε-regular. So $|I| \ge \varepsilon k^2$.

First we show there exists a partition (not equipartition) of order k_1, say P_1, which refines P with $q(P_1) \ge q(P) + \varepsilon^5$ and $k_1 \le k 2^{k-1}$. Let $(i, j) \in I$. As (V_i, V_j) is not ε-regular, there exist $V_1^{i,j} \subseteq V_i$, $V_1^{j,i} \subseteq V_j$ witnessing Fact 1.5. Set $V_2^{i,j} = V_i \backslash V_1^{i,j}$ and $V_2^{j,i} = V_j \backslash V_1^{j,i}$. Define $\mathcal{A}^{i,j} = \{V_1^{i,j}, V_2^{i,j}\}$ and $\mathcal{A}^{j,i} = \{V_1^{j,i}, V_2^{j,i}\}$. We see $\mathcal{A}^{i,j}$ is a partition of V_i for $(i, j) \in I$. Let $\mathcal{A}^i$ be the unique minimal partition of V_i, which refines each $\mathcal{A}^{i,j}$ for $(i, j) \in I$. Thus $\mathcal{A}^i$ has at most 2^{k-1} parts. Let $P_1 = \cup_{i=1}^k \mathcal{A}^i$ be the partition. For $(i, j) \notin I$, by Fact 1.4, we know

$$\sum_{A \in \mathcal{A}^i, B \in A^j} q(A, B) \ge q(A, B).$$

For $(i,j) \in I$, as $\mathcal{A}^i$ refines $\mathcal{A}^{i,j}$ and $\mathcal{A}^j$ refines $\mathcal{A}^{j,i}$, we see

$$\sum_{A \in \mathcal{A}^i, B \in A^j} q(A,B) \geq \sum_{A \in \mathcal{A}^{i,j}, B \in A^{j,i}} q(A,B) \geq q(V_i, V_j) + \frac{\varepsilon^4}{k^2}.$$

Putting the above together, we obtain

$$q(P_1) = \sum_{A,B \in P_1} q(A,B) \geq \sum_{i,j} q(V_i, V_j) + \frac{\varepsilon^4}{k^2}|I| \geq q(P) + \varepsilon^5.$$

In the remaining step, we will turn P' in to an equipartition P^* with $q(P^*) \geq q(P_1) - 10/k_1$, where k_1 is the order of P_1 and $1/\varepsilon^6 \leq k \leq k_1 \leq k2^k$. We start by partitioning each $A \in P_1$ into subsets of size exactly $\frac{n}{k_1^2}$ (with at most one exceptional subset of size at most $\frac{n}{k_1^2}$). Let P_2 be the resulting refinement of P_1, say $P_2 = \{V_1, V_2, ..., V_{k_2}, U_1, U_2, ..., U_{k_1}\}$, where $|V_j| = \frac{n}{k_1^2}$ and $|U_i| \leq \frac{n}{k_1^2}$. Note that $\sum_i |U_i| \leq \frac{n}{k_1}$ and $k_1^2 \geq k_2 \geq k_1^2 - k_1$. Check that, one can turn P_2 into an equipartition P^* of order k_2 by simply "distribute" the vertices of $U_1 \cup ... \cup U_{k_1}$ equally among the sets $V_1, ..., V_{k_2}$ such that $q(P^*) \geq q(P_2) - 10/k_1 \geq q(P_1) - 10/k_1$.

Finally, we conclude that P^* is the desired equipartition in the key lemma, since

$$q(P^*) \geq q(P_1) - \frac{10}{k_1} \geq q(P_1) - 10\varepsilon^6 \geq q(P) + \varepsilon^5 - 10\varepsilon^6 \geq q(P) + \frac{\varepsilon^5}{2},$$

and the order k_2 satisfies $k_2 \leq k_1^2 \leq (k2^k)^2 = k^2 4^k$. ∎