Lecture 1 - The regularity lemma

In this lecture, we will prove the famous regularity lemma of Szemerédi. Roughly speaking, Szemerédi's regularity lemma says that any graph may be partitioned into a finite number of sets such that most of the bipartite graphs between different sets are random-like. To be absolutely precise, we will need some notation and some definitions.

Let G be a graph and let A and B be subsets of the vertex set. If we let $E(A, B)$ be the set of edges between A and B, the density of edges between A and B is given by

$$d(A, B) = \frac{|E(A, B)|}{|A||B|}.$$

Definition 1 *Let G be a graph and let A and B be two subsets of the vertex set. The pair (A, B) is said to be ϵ-regular if, for every $A' \subset A$ and $B' \subset B$ with $|A'| \geq \epsilon|A|$ and $|B'| \geq \epsilon|B|$,*

$$|d(A', B') - d(A, B)| \leq \epsilon.$$

We say that a partition $V(G) = X_1 \cup X_2 \cup \cdots \cup X_k$ is ϵ-regular if

$$\sum \frac{|X_i||X_j|}{n^2} \leq \epsilon,$$

where the sum is taken over all pairs (X_i, X_j) which are not ϵ-regular.

That is, a bipartite graph is ϵ-regular if all small induced subgraphs have approximately the same density as the full graph and a partition of the vertex set of a graph G into smaller sets is ϵ-regular if almost every pair forms a bipartite graph which is ϵ-regular. The regularity lemma is now as follows.

Theorem 1 (Szemerédi's regularity lemma) *For every $\epsilon > 0$, there exists an M such that, for every graph G, there is an ϵ-regular partition of the vertex set of G with at most M pieces.*

In order to prove the regularity lemma, we will associate a function, known as the mean square density, with each partition of $V(G)$. We will prove that if a particular partition is not ϵ-regular it may be further partitioned in such a way that the mean square density increases. But, as we shall see, the mean square density is bounded above by 1, so we eventually reach a contradiction.

Definition 2 *Let G be a graph. Given a partition $V(G) = X_1 \cup X_2 \cup \cdots \cup X_k$ of the vertex set of G, the mean square density of this partition is given by*

$$\sum_{1 \leq i,j \leq k} \frac{|X_i||X_j|}{n^2} d(X_i, X_j)^2.$$

We now observe that since $\sum_{1 \leq i,j \leq k} \frac{|X_i||X_j|}{n^2} = 1$ and $0 \leq d(X_i, X_j) \leq 1$, the mean square density also lies between 0 and 1.

Lemma 1 *For every partition of the vertex set of a graph G, the mean square density lies between 0 and 1.*

Another important property of mean square density is that it cannot increase under refinement of a partition. That is, we have the following.

Lemma 2 *Let G be a graph with vertex set $V(G)$. If $X_1, X_2, \ldots, X_k$ is a partition of $V(G)$ and $Y_1, Y_2, \ldots, Y_\ell$ is a refinement of $X_1, X_2, \ldots, X_k$, then the mean square density of $Y_1, Y_2, \ldots, Y_\ell$ is at least the mean square density of $X_1, X_2, \ldots, X_k$.*

Proof Because the Y_i partition is a refinement of the X_i partition, every X_i may be rewritten as a disjoint union $X_{i1} \cup \cdots \cup X_{ia_i}$, where each $X_{ia_i} = Y_j$, for some j. Now, by the Cauchy-Schwarz inequality,

$$\begin{aligned} d(X_i, X_j)^2 &= \left(\sum_{s,t} \frac{|X_{is}||X_{jt}|}{|X_i||X_j|} d(X_{is}, X_{yt}) \right)^2 \\ &\leq \left(\sum_{s,t} \frac{|X_{is}||X_{jt}|}{|X_i||X_j|} \right) \left(\sum_{s,t} \frac{|X_{is}||X_{jt}|}{|X_i||X_j|} d(X_{is}, X_{yt})^2 \right) \\ &= \sum_{s,t} \frac{|X_{is}||X_{jt}|}{|X_i||X_j|} d(X_{is}, X_{yt})^2. \end{aligned}$$

Therefore,

$$\frac{|X_i||X_j|}{n^2} d(X_i, X_j)^2 \leq \sum_{s,t} \frac{|X_{is}||X_{jt}|}{n^2} d(X_{is}, X_{yt})^2.$$

Adding over all values of i and j implies the lemma. □

An analogous result also holds for bipartite graphs G. That is, if G is a bipartite graph between two sets X and Y, $\cup_i X_i$ and $\cup_i Y_i$ are partitions of X and Y and $\cup_i Z_i$ and $\cup_i W_i$ refine these partitions, then

$$\sum_{i,j} \frac{|X_i||Y_j|}{n^2} d(X_i, Y_j)^2 \leq \sum_{i,j} \frac{|Z_i||W_j|}{n^2} d(Z_i, W_j)^2.$$

We will now show that if X and Y are two sets of vertices and the graph between them is not-regular then there is a partition of each of X and Y for which the mean square density increases.

Lemma 3 *Let G be a graph and suppose X and Y are subsets of the vertex set $V(G)$. If $d(X, Y) = \alpha$ and the graph between X and Y is not ϵ-regular then there are partitions $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2$ such that*

$$\sum_{1 \leq i,j \leq 2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i, Y_j)^2 \geq \alpha^2 + \epsilon^4.$$

Proof Since the graph between X and Y is not ϵ-regular, there must be two subsets X_1 and Y_1 of X and Y, respectively, with $|X_1| \geq \epsilon|X|$, $|Y_1| \geq \epsilon|Y|$ and $|d(X_1, Y_1) - \alpha| > \epsilon$. Let $X_2 = X \backslash X_1$, $Y_2 = Y \backslash Y_1$

and $u(X_i, Y_j) = d(X_i, Y_j) - \alpha$. Then

$$\begin{aligned}
\epsilon^4 &\leq \sum_{1 \leq i,j \leq 2} \frac{|X_i||Y_j|}{|X||Y|} u(X_i, Y_j)^2 \\
&= \sum_{1 \leq i,j \leq 2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i, Y_j)^2 - 2\alpha \sum_{1 \leq i,j \leq 2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i, Y_j) + \alpha^2 \sum_{1 \leq i,j \leq 2} \frac{|X_i||Y_j|}{|X||Y|} \\
&= \sum_{1 \leq i,j \leq 2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i, Y_j)^2 - \alpha^2.
\end{aligned}$$

Note that the second line holds since

$$\sum_{1 \leq i,j \leq 2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i, Y_j) = d(X, Y) = \alpha.$$

The result therefore follows. □

To complete the proof of the regularity lemma, we need to prove that if a partition is not ϵ-regular there is a refinement of this partition which has a higher mean square density. This is taken care of in the following lemma.

Lemma 4 *Let G be a graph and let $X_1 \cup X_2 \cup \cdots \cup X_k$ be a partition of the vertices of G which is not ϵ-regular. Then there is a refinement $X_{11} \cup \cdots \cup X_{1a_1} \cup \cdots \cup X_{k1} \cup \cdots \cup X_{ka_k}$ such that every a_i is at most 2^{2k} and the mean square density is at least ϵ^5 larger.*

Proof Let $I = \{(i,j) : (X_i, X_j)$ is not ϵ-regular$\}$. Let α^2 be the mean square density of G with respect to $X_1 \cup \cdots \cup X_k$.
For each $(i,j) \in I$, the previous lemma gives us partitions $X_i = A_1^{ij} \cup A_2^{ij}$ and $X_j = B_1^{ij} \cup B_2^{ij}$ for which

$$\sum_{1 \leq p,q \leq 2} \frac{|A_p^{ij}||B_q^{ij}|}{|X_i||X_j|} d(A_p^{ij}, B_q^{ij})^2 \geq d(X_i, X_j)^2 + \epsilon^4.$$

For each i, let $X_{i1} \cup \cdots \cup X_{ia_i}$ be the partition of X_i which refines all partitions which arise from partitioning X_i or X_j into A_is or B_is. Note that this partition has at most 2^{2k} pieces, that is, $a_i \leq 2^{2k}$. Moreover, since refining bipartite partitions does not decrease the mean square density, we have

$$\sum_{p=1}^{a_i} \sum_{q=1}^{a_j} \frac{|X_{ip}||X_{jq}|}{|X_i||X_j|} d(X_{ip}, X_{jq})^2 \geq d(X_i, X_j)^2 + \epsilon^4,$$

for all $(i,j) \in I$. Multiplying both sides of the equation by $\frac{|X_i||X_j|}{n^2}$ and summing over all (i,j), we have

$$\begin{aligned}
\sum_{1 \leq i,j \leq k} \sum_{p=1}^{a_i} \sum_{q=1}^{a_j} \frac{|X_{ip}||X_{jq}|}{n^2} d(X_{ip}, X_{jq})^2 &\geq \sum_{1 \leq i,j \leq k} \frac{|X_i||X_j|}{n^2} d(X_i, X_j)^2 + \epsilon^4 \sum_{(i,j) \in I} \frac{|X_i||X_j|}{n^2} \\
&\geq \alpha^2 + \epsilon^5.
\end{aligned}$$

The result follows. □

We now have all the ingredients necessary to finish the proof.

Proof of Szemerédi's regularity lemma Start with a trivial partition into one set. If it is ϵ-regular, we are done. Otherwise, there is a partition into at most 4 sets where the mean square density increases by ϵ^5.

If, at stage i, we have a partition into k pieces and this partition is not ϵ-regular, there is a partition into at most $k2^{2k} \leq 2^{2^k}$ pieces whose mean square density is at least ϵ^5 greater. Because the mean square density is bounded above by 1, this process must end after at most ϵ^{-5} steps. The number of pieces in the final partition is at most a tower of 2s of height $2\epsilon^{-5}$. □

The tower function $t_i(x)$ is defined by $t_0(x) = x$ and, for $i \geq 0$, $t_{i+1}(x) = 2^{t_i(x)}$. The bound given in the proof above is $t_{2\epsilon^{-5}}(2)$, which is clearly enormous. Surprisingly, as was shown by Gowers, there are graphs where, to get an ϵ-regular partition, one needs roughly that many pieces in the partition.

We note that sometimes the regularity lemma is stated with the additional condition that the partition is equitable. This means that each piece in the partition is of the same order or, more accurately, that $||X_i| - |X_j|| \leq 1$ for all i and j. When applying the lemma, it is occasionally useful to assume this extra condition.