Lecture 4

The main aim of the next two lectures will be to prove the famous regularity lemma of Szemerédi. This was developed by Szemerédi in his work on what is now known as Szemerédi's theorem. This astonishing theorem says that for any $\delta > 0$ and $k \geq 3$ there exists an n_0 such that, for $n \geq n_0$, any subset of $\{1, 2, \ldots, n\}$ with at least δn elements must contain an arithmetic progression of length k. The particular case when $k = 3$ had been proven earlier by Roth and is accordingly known as Roth's theorem.

Our initial purpose in proving the regularity lemma will be to give another proof of the Erdős-Stone-Simonovits theorem. However, its use is widespread throughout extremal graph theory and we will see a number of other applications in the course.

Roughly speaking, Szemerédi's regularity lemma says that no graph is entirely random because every graph is at least somewhat random. More precisely, the 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 $\sum_{1\le i,j\le k}\frac{|X_i||X_j|}{n^2}=1$ and $0\le d(X_i,X_j)\le 1$, 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,

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

Therefore,

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

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

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

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*

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

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 $X_2 = X \setminus X_1$, $Y_2 = Y \setminus Y_1$ 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. □