Lecture 5

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 1 *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 $I = \{(i,j) : (X_i, X_j) \text{ is not } \epsilon\text{-regular}\}$. 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

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

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, $a_i \le 2^{2k}$. Moreover, since refining bipartite partitions does not decrease the mean square density, we have

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

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

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

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 $k2^{2k} \le 2^{2^k}$ 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 $i \ge 0$, $t_{i+1}(x) = 2^{t_i(x)}$. 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.

In this and the next lecture, we will prove a beautiful consequence of the regularity lemma, the triangle removal lemma, and show how one may deduce Roth's theorem from it. The triangle removal lemma says that if a graph contains very few triangles, then one may remove all such triangles by removing very few edges. Though this lemma sounds simple, its proof is surprisingly subtle. We begin with what is known as a counting lemma.

Lemma 2 *Let G be a graph and let X, Y, Z be subsets of the vertex set $V(G)$. Suppose that (X, Y), (Y, Z) and (Z, X) are ϵ-regular and that $d(X, Y) = \alpha$, $d(Y, Z) = \beta$ and $d(Z, X) = \gamma$. Then, provided $\alpha, \beta, \gamma \geq 2\epsilon$, the number of triangles xyz with $x \in X$, $y \in Y$ and $z \in Z$ is at least*

$$(1 - 2\epsilon)(\alpha - \epsilon)(\beta - \epsilon)(\gamma - \epsilon)|X||Y||Z|.$$

Proof For every x, let $d_Y(x)$ and $d_Z(x)$ be the number of neighbours of x in Y and Z, respectively. Then the number of $x \in X$ with $d_Y(x) < (\alpha - \epsilon)|Y|$ is at most $\epsilon|X|$. Suppose otherwise. Then there will be a subset X' of X of size at least $\epsilon|X|$ such that the density of edges between X' and Y is less than $\alpha - \epsilon$. But this would contradict regularity. We may similarly show that there are at most $\epsilon|X|$ values of x for which $d_Z(x) < (\gamma - \epsilon)|Z|$. If $d_Y(x) > (\alpha - \epsilon)|Y|$ and $d_Z(x) > (\gamma - \epsilon)|Z|$, the number of edges between $N(x) \cap Y$ and $N(x) \cap Z$, and consequently the number of triangles containing x, is at least

$$(\alpha - \epsilon)(\beta - \epsilon)(\gamma - \epsilon)|Y||Z|.$$

Summing over all $x \in X$ gives the result. $\square$

We will complete the proof of the removal lemma in the next lecture.