Lecture 6

We are now ready to prove the triangle removal lemma.

Theorem 1 (Triangle removal lemma) *For every $\epsilon > 0$ there exists $\delta > 0$ such that, for any graph G on n vertices with at most δn^3 triangles, it may be made triangle-free by removing at most ϵn^2 edges.*

Proof Let $X_1 \cup \cdots \cup X_M$ be an $\frac{\epsilon}{4}$-regular partition of the vertices of G. We remove an edge xy from G if

1. $(x, y) \in X_i \times X_j$, where (X_i, X_j) is not an $\frac{\epsilon}{4}$-regular pair;
2. $(x, y) \in X_i \times X_j$, where $d(X_i, X_j) < \frac{\epsilon}{2}$;
3. $x \in X_i$, where $|X_i| \leq \frac{\epsilon}{4M} n$.

The number of edges removed by condition 1 is at most $\sum_{(i,j) \in I} |X_i||X_j| \leq \frac{\epsilon}{4} n^2$. The number removed by condition 2 is clearly at most $\frac{\epsilon}{2} n^2$. Finally, the number removed by condition 3 is at most $Mn\frac{\epsilon}{4M} n = \frac{\epsilon}{4} n^2$. Overall, we have removed at most ϵn^2 edges.

Now, suppose that some triangle remains in the graph, say xyz, where $x \in X_i$, $y \in X_j$ and $z \in X_k$. Then the pairs (X_i, X_j), (X_j, X_k) and (X_k, X_i) are all $\frac{\epsilon}{4}$-regular with density at least $\frac{\epsilon}{2}$. Therefore, since $|X_i|, |X_j|, |X_k| \geq \frac{\epsilon}{4M} n$, we have, by the counting lemma that the number of triangles is at least

$$\left(1 - \frac{\epsilon}{2}\right) \left(\frac{\epsilon}{4}\right)^3 \left(\frac{\epsilon}{4M}\right)^3 n^3.$$

Taking $\delta = \frac{\epsilon^6}{2^{20} M^3}$ yields a contradiction. □

We now use this removal lemma to prove Roth's theorem. We will actually prove the following stronger theorem.

Theorem 2 *Let $\delta > 0$. Then there exists n_0 such that, for $n \geq n_0$, any subset A of $[n]^2$ with at least δn^2 elements must contain a triple of the form $(x, y), (x + d, y), (x, y + d)$ with $d > 0$.*

Proof The set $A + A = \{x + y : x, y \in A\}$ is contained in $[2n]^2$. There must, therefore, be some z which is represented as $x + y$ in at least

$$\frac{(\delta n^2)^2}{(2n)^2} = \frac{\delta^2 n^2}{4}$$

different ways. Pick such a z and let $A' = A \cap (z - A)$ and $\delta' = \frac{\delta^2}{4}$. Then $|A'| \geq \delta' n^2$ and if A' contains a triple of the form $(x, y), (x + d, y), (x, y + d)$ for $d < 0$, then so does $z - A$. Therefore, A will contain such a triple with $d > 0$. We may therefore forget about the constraint that $d > 0$ and simply try to find some non-trivial triple with $d \neq 0$.

Consider the tripartite graph on vertex sets X, Y and Z, where $X = Y = [n]$ and $Z = [2n]$. X will correspond to vertical lines through A, Y to horizontal lines and Z to diagonal lines with constant values of $x + y$. We form a graph G by joining $x \in X$ to $y \in Y$ if and only if $(x, y) \in A$. We also join x and z if $(x, z - x) \in A$ and y and z if $(z - y, y) \in A$.

If there is a triangle xyz in G, then $(x, y), (x, y+(z-x-y)), (x+(z-x-y), y)$ will all be in A and thus we will have the required triple unless $z = x + y$. This means that there are at most $n^2 = \frac{1}{64n}(4n)^3$ triangles in G. By the triangle removal lemma, for n sufficiently large, one may remove $\frac{\delta}{2}n^2$ edges and make the graph triangle-free. But every point in A determines a degenerate triangle. Hence, there are at least δn^2 degenerate triangles, all of which are edge disjoint. We cannot, therefore, remove them all by removing $\frac{\delta}{2}n^2$ edges. This contradiction implies the required result. $\square$

This implies Roth's theorem as follows.

Theorem 3 (Roth) *For all $\delta > 0$ there exists n_0 such that, for $n \geq n_0$, any subset A of $[n]$ with at least δn elements contains an arithmetic progression of length 3.*

Proof Let $B \subset [2n]^2$ be $\{(x, y) : x - y \in A\}$. Then $|B| \geq \delta n^2 = \frac{\delta}{4}(2n)^2$ so we have $(x, y), (x + d, y)$ and $(x, y + d)$ in B. This translates back to tell us that $x - y - d$, $x - y$ and $x - y + d$ are in A, as required. $\square$

To prove Szemerédi's theorem by the same method, one must first generalise the regularity lemma to hypergraphs. This was done by Gowers and, independently, by Nagle, Rödl, Schacht and Skokan. This method also allows you to prove the following more general theorem.

Theorem 4 (Multidimensional Szemerédi) *For any natural number d, any $\delta > 0$ and any subset P of $\mathbb{Z}^d$, there exists an n_0 such that, for any $n \geq n_0$, every subset of $[n]^d$ of density at least δ contains a homothetic copy of P, that is, a set of the form $k.P + \ell$, where $k \in \mathbb{Z}$ and $\ell \in \mathbb{Z}^d$.*

The theorem proved above corresponds to the case where $d = 2$ and $P = \{(0, 0), (1, 0), (0, 1)\}$. Szemerédi's theorem for length k progressions is the case where $d = 1$ and $P = \{0, 1, 2, \ldots, k - 1\}$.