Triangle deletion

Ernie Croot

February 3, 2010

1 Introduction

The purpose of this note is to give an intuitive outline of the "triangle deletion theorem" of Ruzsa and Szemerédi, which says that if $G = (V, E)$ has $n = |V|$ vertices, $e = |E|$ edges, and $o(n^3)$ triangles, one can destroy all these triangles through removing only $o(n^2)$ edges. A triangle here is just a triple of edges of the form $\{x, y\}, \{y, z\}, \{x, z\}$.

Another way of stating this theorem, which is better as far as our proof is concerned, is as follows: If G has at most $\varepsilon_0 n^3$ triangles, then one can destroy them all through removing at most δn^2 edges from the graph, where $\delta = \delta(\varepsilon_0) \to 0$ as $\varepsilon_0 \to 0$.

As we saw last semester, this result is what allowed Solymosi to prove his theorem on 2D corners in an $n \times n$ grid: For every $0 < \gamma \leq 1$, and all n sufficiently large, we have that any subset S of the $n \times n$ grid of integer lattices points $\{1, 2, ..., n\}^2$ satisfying $|S| \geq \gamma n^2$, necessarily contains a "corner". Recall that a corner is a triple of points $(x, y), (x, y + z), (x + z, y + z)$ (they form the vertices of a 45-45-90 right triangle).

2 Szemerédi Regularity Lemma

As we discussed, the proof of triangle deletion makes heavy use of the Szemerédi Regularity Lemma, so we state it here for future reference. First, we need to define the notions of "edge density" and "regular".

Definition. Suppose that a graph $G = (V, E)$ is an undirected graph, and suppose that $X, Y \subseteq V$. Then, define the *edge density function*

$$d(X,Y) \ := \ \frac{e(X,Y)}{|X| \cdot |Y|},$$

where $e(X,Y)$ denotes the number of edges of the form $\{x,y\}$, where $x \in X$ and $y \in Y$. Note that if there are no edges between X and Y, then $d(X,Y) = 0$; and, if the graph has all possible edges between X and Y, then $d(X,Y) = 1$.

Definition. Suppose that a graph $G = (V,E)$ is an undirected graph, and that $X, Y \subseteq V$ are two sets of vertices of G. We say that $\{X,Y\}$ is an *ε-regular pair* if for every

$$X' \subseteq X,\ Y' \subseteq Y, \text{ satisfying } |X'| \geq \varepsilon|X|,\ |Y'| \geq \varepsilon|Y|,$$

we have that

$$|d(X',Y') - d(X,Y)| \ \leq \ \varepsilon.$$

Remark. Note that, intuitively, "regularity" is some sort of "continuity property" for graphs. It is basically saying that the edge density can't vary too much if you pass to large subsets of the vertex sets X and Y.

Theorem 1 (Szemerédi's Regularity "Lemma") *For all real numbers $\varepsilon > 0$ and integers $m \geq 1$, there exist integers $N, M \geq 1$, such that the following holds: If $G = (V,E)$ is a graph having $n = |V| \geq N$ vertices, then V can be partitioned into $m' + 1$ vertex sets*

$$V_0,\ V_1,\ V_2,\ ...,\ V_{m'},$$

where

- $m \leq m' \leq M$
- $|V_0| \leq \varepsilon n$.
- $|V_1| = |V_2| = \cdots = |V_{m'}|$.
- *All but at most $\varepsilon(m')^2$ pairs $\{V_i, V_j\}$ of vertex sets are ε-regular.*

Remark. We think of the set V_0 as the "error vertices". The remaining vertex sets $V_1, ..., V_{m'}$ will constitute the bulk of the vertices of G – they will contain $(1-\varepsilon)n$ of the n vertices of G.

3 Proof of triangle deletion

3.1 Invoking regularity

We take our graph G, and apply the Regularity Lemma with m and ε chosen so that $(\varepsilon/M)^3/4$ is just larger than ε_0. Such a choice is always possible since we know that M depends upon m and ε. What we will ultimately show is that we only need to remove cn^2 edges where c depends on ε and M, where c shrinks to 0 as $\varepsilon_0 \to 0$. [1]

So, we have a partition of G into the vertex sets

$$V_0,\ V_1,\ ...,\ V_{m'},$$

and "lots" of regular pairs (at least if $\varepsilon > 0$ is small, which it will be).

3.2 Exceptional triangles

Triangles in the graph either connect only vertices within a single vertex set (in other words, $x, y, z \in V_i$ for some i); or connect two vertex sets (e.g. $x, y \in V_i$ while $z \in V_j$); or connect three vertex sets. It turns out that we can easily destroy triangles with two of x, y, or z belonging to the same vertex set, by removing very few edges. Basically, we just remove all edges connecting a vertex set V_i to itself, and there are at most

$$\sum_{i=0}^{m'} |V_i|^2 \ < \ (\varepsilon^2 + 1/m')n^2$$

such edges. So far, so good, because this is of the form $c'n^2$, where c' depends on ε and m', and shrinks to 0 with ε_0.

3.3 Removing edges between irregular, and low-density, pairs

We also remove from the graph all edges connecting a vertex of V_i to a vertex of V_j if $\{V_i, V_j\}$ fails to be regular, or if i or j is 0. Since there are at most

[1] Note that it is well known (by a theorem of Gowers), that M has tower-type dependence on ε and m. So, the number of edges we will need to delete, to remove all triangles, will be quite large; though, it still will be a vanishingly small fraction of n^2 provided ε_0 is small enough.

$\varepsilon(m')^2$ potential irregular pairs like this, we have that this accounts for only (after a little work) at most $2\varepsilon n^2$ edges. So, removing them all, we still have not removed many edges from G (as noted in the previous subsection).

Lastly, we remove all edges between a pair V_i and V_j if the edge density is smaller than 2ε (i.e. $d(V_i, V_j) < 2\varepsilon$). If we do this, then we remove at most

$$\sum_{1\leq i,j\leq m'} 2\varepsilon|V_i|\cdot|V_j| \ \leq \ 2\varepsilon n^2,$$

edges.

3.4 Surprisingly, we have actually already deleted all the triangles!

It turns out that the above edge deletions have completely removed all the triangles for our graph. To see this, note that if there is even a single remaining triangle in the graph, then it must connect three vertex sets V_i, V_j, V_k, such that all pairs have "high" edge density, and are regular. What we will now show is that if this is the case, then in fact our graph has more than $\varepsilon_0 n^3$ triangles, which is a contradiction.

So suppose we indeed have such a triple V_i, V_j, V_k. Then, we first claim that at most $|V_i|/3$ vertices of V_i connect to fewer than $\varepsilon|V_j|$ vertices of V_j; and similarly, at most $|V_i|/3$ vertices connect to at most as many in V_k. To see this we make use of regularity: If there were *more* than $|V_i|/3$ vertices connecting to fewer than $\varepsilon|V_j|$, then letting $A \subseteq V_i$ denote these vertices and just letting $B := V_j$, we would get that

$$d(A,B) \ = \ \frac{e(A,B)}{|A|\cdot|B|} \ \leq \ \varepsilon.$$

But this, along with the fact $d(V_i, V_j) \geq 2\varepsilon$, contradicts the regularity inequality

$$|d(V_i, V_j) - d(A,B)| \ < \ \varepsilon.$$

(Note: We also need the condition that $|V_i|/3 = |A| > \varepsilon|V_i|$ – that is, $\varepsilon < 1/3$ – in order for regularity to apply; but that is guaranteed by having $\varepsilon_0 > 0$ small enough, which makes $\varepsilon > 0$ small enough.)

We conclude that there are at least $|V_i|/3$ vertices of V_i connected to at least $\varepsilon|V_j|$ vertices of V_j, and to at least $\varepsilon|V_k|$ vertices of V_k.

Now we make use of regularity one last time: For each "good vertex" $v \in V_i$ (connected to many vertices of V_i and to many of V_j), let $X_v \subseteq V_j$ denote the vertices in V_j it is connected to; similarly, let $Y_v \subseteq V_k$. Now, since $\{V_j, V_k\}$ is regular, we have that

$$|d(V_i, V_j) - d(X_v, Y_v)| \ < \ \varepsilon,$$

This then implies that

$$d(X_v, Y_v) \ > \ \varepsilon,$$

and what this means is that V_i, V_j, V_k contains at least

$$\varepsilon|X_v| \cdot |Y_v| \ \geq \ \varepsilon^3|V_j| \cdot |V_k|$$

triangles passing through v. Summing over all the $\geq |V_i|/3$ "good vertices" of V_i, this gives at least

$$(\varepsilon^3/3)|V_i| \cdot |V_j| \cdot |V_k| \ \gtrsim \ (\varepsilon^3/3)(n/m')^3$$

triangles in G in total. But this will exceed $\varepsilon_0 n^3$ for ε large enough (though still as close to 0 as needed, provided ε_0 is small enough) and m' small enough, so we have reached a contradiction. Our theorem is now proved.