CS 590 Sublinear Algorithms February 4th, 2021

Lecture 6

Lecturer: Elena Grigorescu *Scribe: Tamalika Mukherjee*

The goal for this lecture is to prove the Triangle-removal lemma, i.e., if graph G is ϵ-far from triangle-freeness it has at least $\delta(\epsilon)\binom{n}{3}$ triangles.

1 Recap

Last time we saw the proof of Szemeredi's regularity lemma. We first recall the main statement and a few important definitions below.

Definition 1 (density) *The density of edges between two sets of vertices X, Y is defined as*

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

Definition 2 (ϵ-regular) *Two sets of vertices X, Y are ϵ-regular if*

$$|d(X,Y) - d(V_X, V_Y)| < \epsilon \ \forall V_X \subseteq X \ : |V_X| > \epsilon \cdot |X|, \ \forall V_Y \subseteq Y \ : |V_Y| > \epsilon \cdot |Y|$$

Lemma 3 (Szemeredi's Regularity Lemma) *For every $\epsilon > 0$, there exists $M = M(\epsilon)$ such that $\forall$ G on n vertices, $n > M$, G has an equipartition $\{V_1, \ldots, V_k\}$ where $5/\epsilon < k < M(\epsilon)$ such that $(1-\epsilon)\binom{k}{2}$ pairs $G[V_i, V_j]$ are ϵ-regular.*

2 Komlos-Simonovits Lemma

Before stating and proving the Komlos-Simonovits Lemma, let us first look at a simpler claim.

Claim 4 *The number of triangles in a random 3-partite graph (A, B, C) is $\Omega(n^3)$.*

Proof We assume that each edge is sampled with probability η. Define the indicator variable I_{uvw} for every vertex triplet $u \in A, v \in B, w \in C$ as follows:

$$I_{uvw} = \begin{cases} 1 & \text{if } u, v, w \text{ form a triangle.} \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned}
&\text{Exp}[\text{number of triangles in } (A,B,C)] \\
&= \text{Exp}[\sum_{(u,v,w)\in(A,B,C)} I_{uvw}] \\
&= \sum_{(u,v,w)} \text{Exp}[I_{uvw}] \\
&= \sum_{(u,v,w)} \Pr[u, v, w \text{ forms a triangle}] \\
&= \eta^3 \binom{n}{3} = \Theta(n^3)
\end{aligned}$$

■

Komlos-Simonovits essentially states that if we relax the requirement of the edges being sampled uniformly at random to the pairs of partite sets being regular and dense, then we will get the same number of triangles.

Lemma 5 (Komlos-Simonovits Lemma) *Given $G = A \cup B \cup C$ such that $(A, B), (A, C)$ and (B, C) are $\eta/2$-regular and η-dense then the number of triangles in G is at least $\delta(n)|A||B||C|$*

Proof Given a 3-partite graph $G = A \cup B \cup C$ we will look at vertices $a \in A$ that have large neighborhoods in B and C denoted by B_a and C_a and show that there are many edges between B_a and C_a. Next, we will show that there exist many such a's in A. It is clear that if we combine the two claims we will have shown that there are many triangles in G.

More formally, fix a vertex $a \in A$, and let B_a be the neighborhood of a in B and let C_a be the neighborhood of a in C. In the sequel, we will consider vertices $a \in A$ such that $|B_a| \geq \frac{\eta}{2}|B|$, and $|C_a| \geq \frac{\eta}{2}|C|$. The claim below shows that there are "many edges" between the neighborhoods of a in B and C.

Claim 6 $e(B_a, C_a) > \left(\frac{\eta}{2}\right)^3 \cdot |B| \cdot |C|$.

Proof Since $|B_a| \geq \frac{\eta}{2}|B|$, and $|C_a| \geq \frac{\eta}{2}|C|$, and (B, C) is $\eta/2$-regular, by definition of regularity, we have that

$$|d(B, C) - d(B_a, C_a)| \leq \frac{\eta}{2} \tag{1}$$

Combining Relation 1 with the fact that (B, C) is η-dense, i.e., $d(B, C) \geq \eta$, we get

$$d(B_a, C_a) > \frac{\eta}{2}$$

Now recall that $e(B_a, C_a) = d(B_a, C_a) \cdot |B_a| \cdot |C_a|$, thus

$$e(B_a, C_a) \geq \frac{\eta}{2} \cdot \frac{\eta}{2}|B| \cdot \frac{\eta}{2}|C| = \left(\frac{\eta}{2}\right)^3 \cdot |B| \cdot |C|$$

■ The following claim shows that the number of a's in A that have large neighborhoods in B and C is large.

Claim 7 *Let $a \in A$ such that $|B_a| \geq \frac{\eta}{2}|B|$ and $|C_a| \geq \frac{\eta}{2}|C|$. Then the number of such a's is at least $(1 - \eta)|A|$.*

Proof We define $A_B := \{a \; : \; |B_a| \leq (\eta/2) \cdot |B|\}$ and $A_C := \{a \; : \; |C_a| \leq (\eta/2) \cdot |C|\}$.

Let $A' := A - (A_B \cup A_C)$, note that by definition A' is exactly the set of vertices $a \in A$ that satisfy the assumption of our claim, thus we want to show that $|A'| \geq (1 - \eta)|A|$.

By definition of density between (A_B, B) and the definition of A_B itself, we have

$$d(A_B, B) = \frac{e(A_B, B)}{|A_B||B|} \leq \frac{(\eta/2)|A_B||B|}{|A_B||B|} \leq \frac{\eta}{2}$$

Note that

$$|d(A,B) - d(A_B,B)| \geq d(A,B) - d(A_B,B) \geq \eta - \eta/2 = \eta/2$$

Because (A,B) is η-regular and $|B|$ is large, this implies that $|A_B| \leq \frac{\eta}{2} \cdot |A|$, since if $|A_B|$ was large, then $|d(A,B) - d(A_B,B)| < \eta/2$.

Using a similar argument, we can deduce that $|A_C| \leq \frac{\eta}{2} \cdot |A|$. From these two relations, it is easy to see that,

$$|A_B \cup A_C| \leq \eta \cdot |A|$$

Thus, $|A'| = |A - (A_B \cup A_C)| \geq (1-\eta)|A|$. ■

■

3 Triangle-Removal Lemma

Lemma 8 (Triangle Removal Lemma) *If graph G is ϵ-far from triangle-freeness it has at least $\delta(\epsilon)\binom{n}{3}$ triangles.*

The main idea behind the proof is to first apply Szemeredi Regularity Lemma to the graph G and obtain an equipartition of the vertices where most of the edges are accounted for between these partitions. Then we will remove a certain number of edges so that we can apply Komlos Simonovits Lemma to obtain a large number of triangles.

Proof Let G be ϵ-far from triangle-freeness. We will apply Szemeredi's Regularity Lemma to G where $\epsilon_{sz} = \epsilon/8$ and obtain an equipartition $V_1, \ldots, V_k$ and $40/\epsilon \leq k \leq M(\epsilon)$. Assume that $k|n$. From the lemma statement, know that $(1-\epsilon/8)\binom{k}{2}$ many $G[V_i, V_j]$ are $\epsilon/8$-regular.

Now we need to remove edges in a way that sets us up to apply Komlos Simonovits Lemma. Thus we do the following operations

1. Remove all edges within each V_i.
2. Remove all edges between V_i, V_j that are not $\epsilon/8$-regular.
3. Remove all edges between V_i, V_j that are not $\epsilon/4$-dense.

Claim 9 *The total number of edges removed in Item 1 is less than $(\epsilon/8)\binom{n}{2}$.*

Proof Since $|V_i| = n/k$ for $1 \leq i \leq k$, there can be at most $(n/k) - 1$ many edges between a vertex $v \in V_i$ to all other vertices in V_i. And since there n vertices in total, the total number of edges removed within each V_i is at most $\left(\frac{n}{k} - 1\right) n \leq \frac{\epsilon}{8} \cdot \frac{n(n-1)}{2}$.

■

Claim 10 *The total number of edges removed in Item 2 is less than $(\epsilon/8)\binom{n}{2}$.*

Proof By Szemeredi Regularity Lemma, the number of non-regular pairs is at most $\frac{\epsilon}{8}\binom{k}{2}$. And since each partition contains n/k many vertices, the number of edges between any V_i, V_j is at most $\left(\frac{n}{k}\right)^2$. Therefore, the total number of edges removed between V_i, V_j that are not $\epsilon/8$-regular is $\frac{\epsilon}{8} \cdot \frac{k(k-1)}{2} \cdot \frac{n^2}{k^2} \leq \frac{\epsilon}{8} \cdot \binom{n}{2}$. ■

Claim 11 *The total number of edges removed in Item 3 is less than* $(\epsilon/4)\binom{n}{2}$.

Proof Since V_i, V_j are not $\epsilon/4$-dense, we know that $d(V_i, V_j) = \frac{e(V_i,V_j)}{\|V_i\|\|V_j\|} < \frac{\epsilon}{4}$. Therefore $e(V_i, V_j) < \frac{\epsilon}{4}\left(\frac{n}{k}\right)^2$ for any pair V_i, V_j.

The total number of edges removed is thus at most $\binom{k}{2}\frac{\epsilon}{4}\left(\frac{n}{k}\right)^2 < \frac{\epsilon}{4} \cdot \frac{n(n-1)}{2}$. ■

Combining Claims 1,2,3 we have removed a total of *at most* $\left(\frac{\epsilon}{8} + \frac{\epsilon}{8} + \frac{\epsilon}{4}\right)\binom{n}{2} = \frac{\epsilon}{2}\binom{n}{2}$ many edges. Note that by the definition of triangle-freeness, we need to remove *at least* $\epsilon\binom{n}{2}$ edges to remove *all* triangles. Now we need to count the number of triangles remaining after all these edges have been removed.

Since we have graph G with k parts and each $G[V_i, V_j]$ is $\epsilon/8$-regular and $\epsilon/4$-dense, we can apply Komlos Simonovits Lemma to conclude that there are at least $\delta(\epsilon)\left(\frac{n}{k}\right)^3$ many triangles for every triplet set of vertices. Now $k < M(\epsilon)$ by assumption, and the total number of triangles in G can be given by $\delta(\epsilon)\left(\frac{1}{M(\epsilon)}\right)^3\binom{n}{3}$. ■

4 Concluding Remarks

In this lecture, we saw how a powerful tool like Szemeredi Regularity Lemma can be used to analyse the tester for a graph property like triangle-freeness. In fact, almost every graph property tester has to use some variant of partitioning the graph into partite sets and using Szemeredi Regularity Lemma in its analysis. It seems like an inherent tool for any graph tester that uses a constant number of queries.

Montone graph properties are closed under the removal of edges and vertices, e.g., bipartiteness, triangle-freeness, 3-colorability. Alon and Shapira showed that all monotone graph properties are testable with 1-sided error and constant number of queries.

Hereditary graph properties are closed under removal of vertices, e.g., induced H-freeness. We know that 2-sided testing in dense graphs inherently relies on Szemeredi Regularity Lemma.