Lecture 4,5: Triangle Freeness and Szemeredi's Regularity Lemma

Lecturer: Elena Grigorescu *Scribe: Rohan Garg*

In this text, we will discuss how to test for triangle freeness in a graph and Szemeredi's Regularity Lemma. We will also give the proof for Szemeredi's Regularity Lemma (SRL).

1 Testing Triangle Freeness

We will first discuss how to test for triangle freeness in dense graphs. As the name suggests, a graph G is triangle free if it contains no triangles, i.e. a clique of three nodes. We will consider dense graphs with $\Theta(n^2)$ edges. In the dense graph model, we would expect to see many triangles. We will show that triangle freeness is testable with a constant number of queries.

Theorem 1 *Triangle freeness is testable with a constant number of queries.*

We will design a test on ϵ, G on n vertices and $\Theta(n^2)$ edges. We want the test to,

1. accept, if G is triangle free
2. reject with probability $\frac{2}{3}$ if G is $\epsilon-$far from being triangle free.

Definition 2 *A graph G is $\epsilon-$far from triangle free if we need to remove $\Theta(\epsilon n^2)$ edges in order to get a triangle free graph.*

In layman's terms, we need to remove an ϵ fraction of edges to have a triangle free graph. So, lets think of a basic test.

Test 1: We will sample three random vertices v_1, v_2, v_3. If, these three vertices form a triangle, then we know the graph is not triangle free and we can reject. Else, we accept.

Let's analyze this basic test.

1. **Completeness:** If G is triangle free, then $Pr[accept] = 1$.
2. **Soundness:** If G is $\epsilon-$far from being triangle free, we want that $Pr[reject] = \delta(\epsilon, n)$.

Notice that if we run the basic test $\frac{2}{\delta}$ times, then $Pr[rejects] > \frac{2}{3}$. We can set p in the expression $(1-(1-p)^{\frac{2}{p}})$ to get 2/3.

$$Pr[reject] = \frac{\#\ triangles\ in\ G}{total\ \#\ of\ possible\ triangles} = \frac{\#\ triangles\ in\ G}{\binom{n}{3}}$$

We want $\delta(\epsilon, n)$ to be $o(n)$. But we can get it to be a function of ϵ which is independent of n and thus it is a constant! We will show:

$$\frac{1}{\delta(\epsilon, n)} = 2^{2^{2^{\cdots}}} \quad (poly(\frac{1}{\epsilon})\ times)$$

So, we get $\Theta(1/\epsilon^2)$ from Szemeredi's Regularity Lemma.

2 Szemeredi's Regularity Lemma

Now we will discuss Szemeredi's Regularity Lemma. Szemeredi's Regularity Lemma argues that no matter the size of a graph G, we can partition G into a bounded number of parts such that the inter-part edges behave randomly. Before we state the Lemma, let us first give some preliminaries and definitions.

Definition 3 *The density of edges between two sets of vertices V_1 and V_2 is defined as*

$$d(V_1, V_2) = \frac{e(V_1, V_2)}{|V_1| \cdot |V_2|}$$

Definition 4 *Two sets of vertices V_1 and V_2 are ϵ-regular if*

$$|d(V_1, V_2) - d(x, y)| < \epsilon \quad \forall x \subseteq V_1 : |x| > \epsilon \cdot |V_1|, \forall y \subseteq V_2 : |y| > \epsilon \cdot |V_2|$$

Definition 5 *A partition P of V into $\{V_1, V_2, \dots V_n\}$ of a graph G=(V,E) is an Equipartition if*

$$||V_i| - |V_j|| \leq 1 \quad \forall\ i \neq j$$

Definition 6 *We use the following notation for the subgraph induced by two sets S_1 and S_2 on a graph G.*

$$G[S_1, S_2]$$

Now we can state Szemeredi's Regularity Lemma and give its proof.

Lemma 7 ***Szemeredi's Regularity Lemma***

$\forall \epsilon, m > 1, \exists K > 0$ such that $\forall G$ on $n > k$ vertices, $\exists m'$ for $m < m' < M$ such that $\exists$ partition of G into $V_0, V_1, \dots V_{m'}$ such that:

1. *$|V_0| < \epsilon |V|$ (exceptional set)*
2. *$|V_1| = |V_2| = \cdots = |V_{m'}|$*
3. *All but $\leq \epsilon (m')^2$ of pairs (V_i, V_j) for $i \neq j \neq 0)$ are ϵ-regular*

Now we can give a sketch of the proof for Szemeredi's Regularity Lemma.

Given a partition $P : V_0 \cup V_1 \cup \cdots \cup V_{m_0}$ with $|V_i| = |V_j|$ for $i \neq j \neq 0$, define the potential of P as:

$$f(P) = f(V_0, V_1, \dots, V_{m_0}) = \frac{1}{m_0^2} \cdot \sum_{1<i,j<m_0} d^2(V_i, V_j) \quad (l_2^2 \text{ norm of density vector})$$

Claim 8 $\forall P$

$$f(P) < \frac{1}{2}$$

Proof

$$f(P) = \frac{1}{m_0^2} \cdot \binom{m_0}{2} = \frac{m_0(m_0-1)}{2m_0^2} < \frac{1}{2}$$

■

Lemma 9 *(Main Lemma:)* $\exists P_2$ *s.t.*

$$f(P_2) > f(P_1) + \epsilon^5$$

Main Lemma $\Longrightarrow$ Main Theorem as follows. We'll start with some arbitrary partition P_1. If it is not regular, we will refine it to get P_2. So,

$$f(P_2) > f(P_1) + \epsilon^5$$

Recall: $f(P) < \frac{1}{2}$. We can refine P_1 to P_2 to P_3 and so on till we get a partition P_k such that $f(P_k) = \frac{1}{2}$. We can refine $O(\frac{1}{\epsilon^5})$ many times to get to the partition P_k.

Idea: Start with some arbitrary partition $V_0 \cup V_1 \cup \cdots \cup V_{m_0}$ where $|V_0| \leq \frac{\epsilon n}{2}$ and $|V_i| = |Vj|$ for $i \neq j \neq 0$. If this partition is not regular, then there exists a non regular pair of vertex-sets V_i and V_j such that for a subset $X_i(j) \subset V_i$ and subset $X_j(i) \subset V_j$ such that:

$$|d(V_i, V_j) - d(X_i(j), X_j(i)| > \epsilon$$

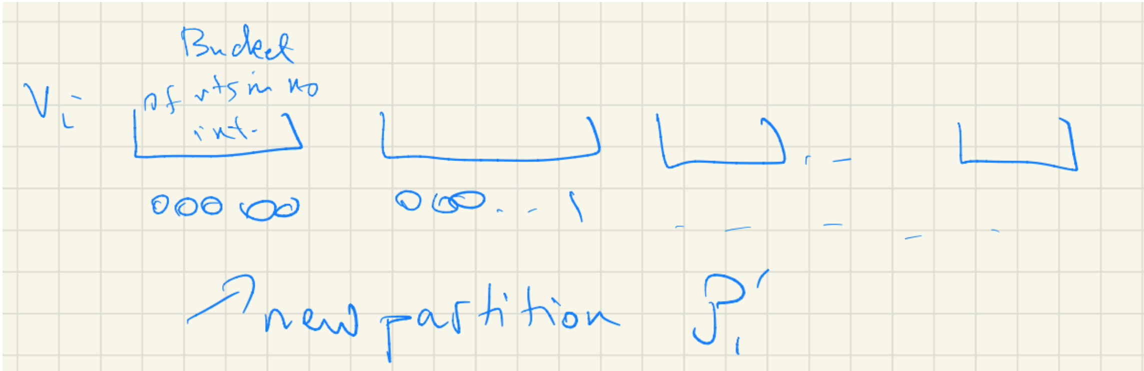
We will call the subsets $X_i(j) \subset V_i$ and subset $X_j(i) \subset V_j$ *witness sets* for irregularity. From these witness sets of V_i and V_j, we will build an intermediate partition P_1'. Let's focus on V_i and analyze its interaction with the remaining sets V_j for $i \neq j$ and more specifically the witness sets for irregularity between V_i and any other set V_j.

Let P_i'

On V_i, consider sets indexed by

$$X_i(1), X_i(2) \ldots X_i(i-1), X_i(i+1) \ldots X_i(m_0)$$

We will create a 0-1 vector for each possible set placement for each vertex. So, for vertex x, we associate x to each of these X_i's as follows. Following the indexing order given above, we have a 1 in the vector if $x \in X_i(j)$ and 0 if $x \notin X_i(j)$. This $m_0 - 1$ length vector represents one set. i.e for each $v \in V_i$ place v in the set indexed by one of the 2^{m_0-1} characteristic vectors. Each new part is denoted by $V_{i,r}$ where $r = 1 \ldots 2^{m_0-1}$. You can think of this as us having 2^{m_0-1} many buckets, and we place vertices in their corresponding bucket. The drawing below shows how the vectors correspond to buckets for the new partition P'.



P_2 : We will break each $V_{i,r}$ into sets of size s for:

$$S = \lceil \frac{n}{m_0^2 \cdot 2^{m_0}} \rceil$$

From this *breaking* process, we get:

$$P_2 = \cup_{i,r} triangles\ in\ each\ V_{i,r}$$

Notice that each $V_{i,r}$ will contribute some *scrap* vertices.

$$\#scrap\ vertices\ in\ V_i = 2^{m_0} \cdot \frac{n}{m_0 \cdot 2^{m_0}} = \frac{n}{m_0^2}$$

To find the total number of scrap vertices in V, we have that:

$$\#scrap\ vertices\ in\ V < m_0 \cdot \frac{n}{m_0^2} = \frac{n}{m_0}$$

Recall: $|V_i|m_0 + |V_0| = n$ and $|V_0| \leq \frac{\epsilon n}{2}$. So,

$$\forall i \geq 1, \quad |V_i| = \frac{n}{m_0} \cdot (1 - \frac{\epsilon}{2})$$

So the number of scrap vertices in total is about the size of one V_i. In the new partition P_2, we will set $V_0' = V_0 \cup scrap\ vertices$. So,

$$|V_0'| = \frac{\epsilon n}{2} + \frac{n}{m+0} = n \cdot (\frac{\epsilon}{2} + \frac{1}{m_0}) < n\epsilon$$

Lets analyze the number of tiny cells.

$$\#tiny\ cells\ \approx \frac{n}{s} = \frac{n \cdot m_0^2 \cdot 2^{m_0}}{n} = m_0^2 \cdot 2^{m_0} \approx 2^{m_0 \log(m_0)}$$

This expression on the right hand side is the number of parts in P_2. In each step, we are doing $\frac{1}{\epsilon^5}$ many steps. This is how we get the following tower:

$$2^{2^{2\cdots}} \quad \frac{1}{\epsilon^5}\ many\ times$$

Now, we can move on to proving the Main Lemma: We want to compare:

$$f(P_1) \ vs. \ f(P_2))$$

$$f(P_1) = \sum_{1 \leq i,j \leq m_0} d^2(V_i, V_j)$$

$$f(P_2) = \sum_{i,j} \sum_{1 \leq r_1, r_2 \leq 2^{m_0-1}, t_1, t_2} d^2(V_{i,r_1,t_1}, V_{j,r_2,t_2})$$

Claim 10 *If (V_i, V_j) is a regular pair, then*

$$\sum_{r_1,r_2,t_1,t_2} d^2(V_{i,r_1,t_1}, V_{j,r_2,t_2}) \geq \frac{|V_i|^2}{S^2}(d^2(V_i, V_j) - O(\frac{1}{m_0}))$$

Claim 11 *If (V_i, V_j) is a non-regular pair, then*

$$\sum_{r_1,r_2,t_1,t_2} d^2(V_{i,r_1,t_1}, V_{j,r_2,t_2}) \geq \frac{|V_i|^2}{S^2}(d^2(V_i, V_j) + \epsilon^4 - O(\frac{1}{m_0}))$$

What we are saying here, is that, on average, the density between these tiny non-regular pairs, the density contributes this ϵ^4 factor. We will now use Claim 10 and Claim 11 to show the Main lemma.

$$f(P_2) = \frac{1}{T^2} \cdot (\sum_{regular \ i,j} d^2(V_{i,r_1,t_1}, V_{j,r_2,t_2}) + \sum_{non\text{-}regular \ i,j} d^2(V_{i,r_1,t_1}, V_{j,r_2,t_2})$$

By the above two claims,

$$= \frac{1}{T^2} \cdot (\sum_{regular \ i,j} d^2(V_i, V_j) + \sum_{non\text{-}regular \ i,j} d^2(V_i, V_j) + \epsilon^4 - O(\frac{1}{m_0}$$

$$= \frac{1}{T^2} \cdot \frac{|V_i|^2}{S^2} \cdot m_0^2 \cdot (f(P_1) + \epsilon^4 - O(\frac{1}{m_0}))$$

If we want to compute this equality, we use:

$$T = \frac{n}{S}, \quad |V_i| = \frac{n}{m_0} \cdot (1 - \frac{\epsilon}{2})$$

This gives us:

$$\frac{S^2}{n^2} \cdot \frac{1}{S^2} \cdot \frac{n^2}{m_0^2} \cdot m_0^2 = 1$$

$$\implies \textit{Main Lemma}$$

If we can prove the claims above, we will have $f(P_2) > f(P_1) + \epsilon^5$.
To see how the ϵ appears in Claim 11. To prove Claims 10 and 11, see Claim 12.

Claim 12

$$\sum_{non\text{-}regular\ V_i,V_j} d(V_{i,r_1,t_1}, V_{j,r_2,t_2}) \geq \frac{1}{S^2} \cdot |X_i(j)| \cdot |X_j(i)| \cdot (d(V_i, V_j) + \epsilon - O(\frac{1}{m_0}))$$

Recall that:

$$|X_i(j)| \geq \epsilon |V_i| \ and \ |X_j(i)| \geq \epsilon |V_j|$$

.

Let us look at the portion of $X_i(j)$ and $X_j(i)$ that is not *scrap*. We will let $X = X_i(j) - scrap$ and $Y = X_j(i) - scrap$. We can get to Claim 11 from Claim 12 by using Cauchy-Schwartz Inequality. Recall that Cauchy-Schwartz states:

$$(\sum_{i=1}^{n} a_i)^2 \leq n(\sum_{i=1}^{n} a_i^2)$$

For us, these a_i's will be the densities $d(V_{i,r_1,t_1}, V_{j,r_2,t_2})$.

For claim 11, $\sum d(V_{i,r_1,t_1}, V_{j,r_2,t_2})$ for $V_{i,r_1,t_1} = X$ (without scrap) and $V_{j,r_2,t_2} = Y$ (without scrap), we have:

$$\sum d(V_{i,r_1,t_1}, V_{j,r_2,t_2}) = \frac{1}{S^2} \sum e(V_{i,r_1,t_1}, V_{j,r_2,t_2})$$

where $e(V_1, V_2)$ is the number of edges between the two sets V_1 and V_2. So, we have:

$$= \frac{1}{S^2} e(X, Y) = \frac{1}{S^2} \cdot e(X_i(j), X_j(i))(1 - \frac{1}{m_0})$$

$$= \frac{1}{S^2} d(X_i(j), X_j(i)) \cdot |X_i(j)| \cdot |X_j(i)| \cdot (1 - \frac{O(1)}{m_0})$$

$$> \frac{1}{S^2} (d(V_i, V_j) + \epsilon) \cdot |X_i(j)| \cdot |X_j(i)| \cdot (1 - \frac{O(1)}{m_0})$$

The above holds because V_i and V_j are a non-regular pair. Lastly, recall that:

$$|X| = |X_i(j)| \cdot (1 - O(\frac{1}{m_0}))$$

and

$$|Y| = |X_j(i)| \cdot (1 - O(\frac{1}{m_0}))$$

.

Many of the finer details such as verifying with Cauchy-Schwartz have been left out of these notes. We urge the reader to take the time to verify it carefully. In summary, we have presented Szemeredi's Regularity Lemma, its proof, and its application towards testing triangle freeness.

For another set of notes on Szemeredi's Regularity lemma, see these notes by E. Croot.