Lectures on Szemerédi's Regularity Lemma

Lecturer: Ronitt Rubinfield *Scribe: Yuchong Pan*

1 Randomness and Regularity

Random graphs have many nice properties, and various questions can be asked in a random graph. For instance, the following question asks for the expected number of triangles in a random tripartite graph:

Problem 1. *How many triangles are in a random tripartite graph with partition classes A, B, C and density η?*

For all $u \in A, v \in B, w \in C$, let

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

Then for all $u \in A, v \in B, w \in C$,

$$\mathbb{E}\left[\sigma_{u,v,w}\right] = \mathbb{P}\left[\sigma_{u,v,w} = 1\right] = \eta^3.$$

Hence,

$$\mathbb{E}[\# \text{ triangles}] = \mathbb{E}\left[\sum_{\substack{u \in A \\ v \in B \\ w \in C}} \sigma_{u,v,w}\right] = \eta^3 |A||B||C|.$$

Can we make a weaker assumption and still obtain reasonable bounds? In other words, what if the edges are not completely independent? We introduce the notions of *desntity* and *regularity* of set pairs to describe behaviors like those of a random graph.

Definition 2 (density and regularity of set pairs)**.** Let $G = (V, E)$ be a graph. Let $A, B \subset V$ be such that $A \cap B = \emptyset$ and $|A| > 1, |B| > 1$. Let $e(A, B)$ be the number of edges between A and B. Let the *density* of (A, B) be defined to be $d(A, B) = e(A, B)/(|A||B|)$. We say that (A, B) is γ*-regular* if for all $A' \subset A$ and $B' \subset B$ such that $|A'| \geq \gamma|A|$ and $|B'| \geq \gamma|B|$,

$$\left|d\left(A', B'\right) - d(A, B)\right| < \gamma.$$

In other words, the fraction of edges between A' and B' is roughly the same as the fraction of edges between A and B.

2 Triangle Counting Lemma

In this example, we demonstrate the power of the notion of regularity of set pairs by proving the following lemma. Informally, we show that three disjoint subsets A, B, C of vertices, each pair of which is γ-regular, contain many triangles. A random graph would have $\eta^3|A||B||C|$ triangles, and γ-regular set pairs give $\eta^3/16|A||B||C|$ triangles, which is only slightly worse within a constant factor than a random graph.

Lemma 3 (triangle counting lemma)**.** *Let $G = (V, E)$ be a graph. For all η (called the* density*), there exists $\gamma > 0$ (called the* regularity parameter*) and $\delta > 0$ (the fraction of triangles) such that if A, B, C are disjoint subsets of V such that each pair is γ-regular with densities greater than η, then G contains at least $\delta|A||B||C|$ distinct triangles with vertices in each of A, B, C.*

We remark that $\gamma = \gamma^{\triangle}(\eta) := \eta/2$ and $\delta = \delta^{\triangle}(\eta) := (1-\eta)\eta^3/8$ are functions of η independent of the number of vertices or edges.

Proof. Let A^* be the vertices in A with at least $|\eta - \gamma||B|$ neighbors in B and at least $|\eta - \gamma||C|$ neighbors in C. We claim that $|A^*| \geq (1 - 2\gamma)|A|$. To see this, let A' be the set of "bad" vertices with respect to B, i.e., with fewer than $|\eta - \gamma||B|$ neighbors in B, and let A'' be the set of "bad" vertices in C, i.e., with fewer than $|\eta - \gamma||C|$ neighbors in C. Then

$$d\left(A', B\right) < \frac{|A'| \cdot |\eta - \gamma| \cdot |B|}{|A'| \cdot |B|} = |\eta - \gamma| = \eta - \gamma, \qquad d(A, B) = \eta.$$

It follows that the difference betweeen $d(A', B)$ and $d(A, B)$ is greater than γ. Since $B \geq \gamma|B|$ and since (A, B) is γ-regular, then $|A'| < \gamma|A|$. Similarly, $|A''| < \gamma|A|$. Since $A^* = A \setminus (A' \cup A'')$, then

$$|A^*| \geq |A| - \left|A'\right| - \left|A''\right| \geq |A| - 2\gamma|A| = (1 - 2\gamma)|A|.$$

For each $u \in A^*$, let B_u be the set of neighbors of u in B, and let C_u be the set of neighbors of u in C. Since $\gamma = \eta/2$, then $\eta - \gamma \geq \gamma$. Hence,

$$|B_u| \geq |\eta - \gamma||B| \geq \gamma|B|,$$
$$|C_u| \geq |\eta - \gamma||C| \geq \gamma|C|.$$

Note that the number of edges between B_u and C_u equals the number of triangles involving u. See Figure 1 for an illustration.

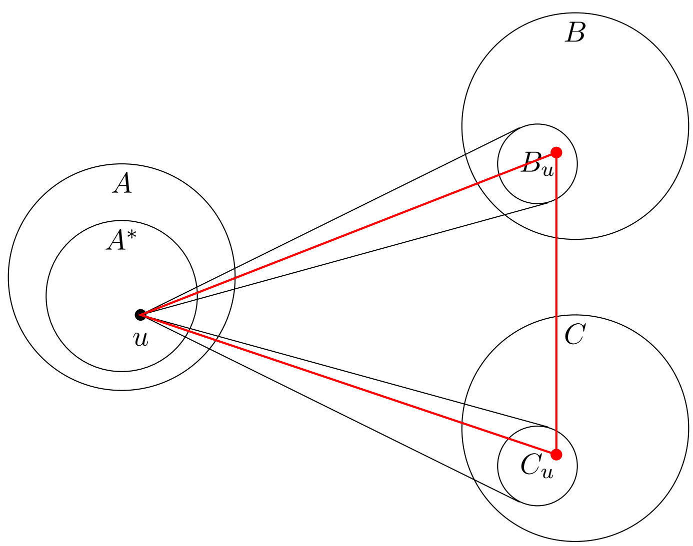

Figure 1: The number of edges between B_u and C_u equals the number of triangles involving u.

Since (B, C) is γ-regular, then $d(B_u, C_u) \geq \eta - \gamma$. Hence,

$$e\left(B_u, C_u\right) \geq (\eta - \gamma)\left|B_u\right|\left|C_u\right| \geq (\eta - \gamma)(\eta - \gamma)|B|(\eta - \gamma)|C| = (\eta - \gamma)^3|B||C|.$$

It follows that the total number of triangles is at least

$$|A^*|\,(\eta-\gamma)^3|B||C| \geq (1-2\gamma)(\eta-\gamma)^3|A||B||C| \geq (1-\eta)\left(\frac{\eta}{2}\right)^3|A||B||C| = \delta|A||B||C|.$$

This completes the proof. □

3 Szemerédi's Regularity Lemma

Informally, Szemerédi's regularity lemma says that *every* graph can be partitioned into a *constant* number of γ-regular pairs. We state Szemerédi's regularity lemma without giving a proof due to time constraints.

Theorem 4 (Szemerédi's regularity lemma)**.** *For all $m > 0$ and $\varepsilon > 0$, there exists $T = T(m, \varepsilon)$ such that for any graph $G = (V, E)$ with $|V| > T$ and any equi-partition $\mathcal{A}$ of V, there exists an equipartition $\mathcal{B}$ into k sets which refines $\mathcal{A}$ such that $m \leq k \leq T$ and at most $\varepsilon\binom{k}{2}$ set pairs are not ε-regular.*

Historically, Szemerédi's regularity lemma was first studied to prove a conjecture of Erdős and Turán that sequences of integers have long artihmetic progressions.

A very rough idea for the proof of Szemerédi's regularity lemma is as follows: We introduce the notion of the *variance* of a partition of the vertices in a graph. Starting with an initial partition, whenever a partition violates regularity, we refine it such that the variance grows significantly, i.e., by approximately ε^c for some constant c. Therefore, in fewer than $1/\varepsilon^c$ refinements, we have a good partition.

How big is T? The above construction shows that T is in the order of

$$\left.2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}\right\}\text{height } 1/\varepsilon^c.$$

It is amazing that this is a constant independent of the number of vertices, although this is very large and not practical algorthmically.

4 Triangle-Freeness Testing

An application of Szemerédi's regularity lemma is triangle-freeness testing in a graph:

Problem 5 (triangle-freeness testing)**.** *Let G be a graph (not necessarily tripartite) and $\varepsilon > 0$. If G is triangle-free, then accept. If one needs to delete at least εn^2 edges to make G triangle-free (i.e., G is ε-far from being triangle-free), then reject.*

This model is interesting only in dense graph. We give an algorithm for triangle-free testing in Algorithm 1.

Theorem 6. *For all $\varepsilon > 0$, there exists $\delta > 0$ such that any graph $G = (V, E)$ ε-far from being triangle-free contains at least $\delta\binom{|V|}{3}$ distinct triangles.*

Proof sketch. Apply Szemerédi's regularity lemma to G, obtaining an equi-partition of V into k sets, where $5/\varepsilon \leq k \leq T$ (i.e., $\varepsilon n/5 \geq n/k \geq n/T$). Let $\varepsilon' = \min(\varepsilon/5, \delta^{\triangle}(\varepsilon/5))$. Then at most $\varepsilon'\binom{k}{2}$ pairs of sets in the equi-partition are not ε'-regular. Delete edges that are

(i) internal to the sets in the equi-partition;

```
repeat O(1) times
    pick v_1, v_2, v_3 ∈ V
    if v_1, v_2, v_3 form a triangle then
        reject
accept
```

Algorithm 1: An algorithm for triangle-free testing in a graph $G = (V, E)$.

(ii) between non-regular pairs of sets in the equi-partition;

(iii) between low-density (i.e., less than $\varepsilon/5$) pairs of sets in the equi-partition.

We can show that we have deleted fewer than εn^2 edges, so the resulting graph G' contains at least one triangle. Moreover, any triangle in G' satisfies the following:

(i) the three vertices forming the triangle are in three distinct sets in the equi-partition;

(ii) each pair of these three sets are regular;

(iii) the density between each pair of these three sets is not low.

Finally, applying the triangle counting lemma (i.e., Lemma 3) gives a lower bound on the number of distinct triangles in G' and hence G, showing that indeed many triangles remain. $\square$