6.842 Randomness and Computation	March 10, 2014
Lecture 10	
Lecturer: Ronitt Rubinfeld	*Scribe: Tana Wattanawaroon*

1 Topics

- Szemerédi's Regularity Lemma
- Testing the property of triangle-freeness on dense graphs.

2 Triangle Counting in a Random Tripartite Graph

Consider a random tripartite graph with "density" η. More precisely, let $G = (V, E)$ be a graph with vertex partitions A, B and C. Between each pair of vertices from different partitions, there is an edge between the vertices with probability η (independently). We shall count the number of triangles in this random tripartite graph.

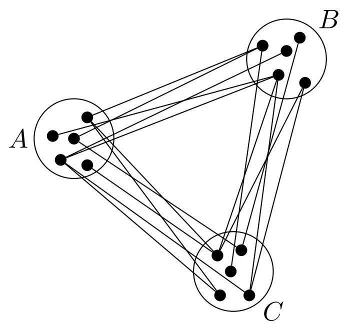

Figure 1: a tripartite graph with vertex partitions A, B and C

For $u \in A$, $v \in B$, and $w \in C$, define the indicator variable

$$\sigma_{u,v,w} = \begin{cases} 1 & \text{if } (u,v),(u,w),(v,w) \in E \\ 0 & \text{otherwise} \end{cases}$$

It follows that $\mathrm{E}[\sigma_{u,v,w}] = \Pr[u,v,w \text{ forms a triangle}] = \eta^3$. Therefore,

$$\mathrm{E}[\text{number of triangles}] = \mathrm{E}[\sum_{\substack{u\in A\\ v\in B\\ w\in C}} \sigma_{u,v,w}] = \sum_{\substack{u\in A\\ v\in B\\ w\in C}} \mathrm{E}[\sigma_{u,v,w}] = \eta^3 \cdot |A||B||C|$$

3 Triangle Counting in a Regular Dense Graph

Here, we achieve a similar bound to above without requiring the graph to be random. The graph has a more relaxed assumption based on *density* and *regularity*.

Definition 1 (Density and Regularity) *For $A, B \in V$ such that $A \cap B = \emptyset$ and $|A|, |B| > 1$, let $e(A,B)$ denote the number of edges between A and B, and let $d(A,B) = \frac{e(A,B)}{|A||B|}$ be the* density.

(A,B) is γ-regular *if it has the following property: for all $A' \subseteq A$ and $B' \subseteq B$, if $|A'| \geq \gamma|A|$ and $|B'| \geq \gamma|B|$, then $|d(A',B') - d(A,B)| < \gamma$.*

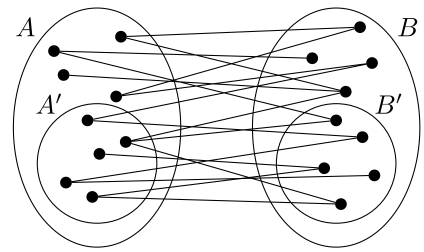

Figure 2: Density $d(A, B)$ and $d(A', B')$ should not differ by much.

Less formally, for large-enough subsets $A' \subseteq A$ and $B' \subseteq B$, the density between A' and B' should be close (within γ) to the density between A and B.

Lemma 2 (Komlós-Simonovits [2]) *For all density $\eta > 0$, there exists a regularity parameter γ and number of triangles δ such that if A, B, C are disjoint subsets of V, each pair δ-regular with density greater than η, then G has at least $\delta \cdot |A||B||C|$ distinct triangles with vertices from each of A, B and C.*

Both γ and δ are parameters of η only. For triangle counting in particular, we can choose parameters $\gamma = \gamma^{\triangle}(\eta) = \frac{\eta}{2}$ and $\delta = \delta^{\triangle}(\eta) = (1-\eta) - \frac{\eta^3}{8}$. Note that if $\eta < \frac{1}{2}$, then $\delta \geq \frac{\eta^3}{16}$. Therefore, for $\eta < \frac{1}{2}$ the bound is within a factor of 16 of the random graph.

Proof (Alon, Fischer, Krivelevich, Szegedy [1]) Let A^* be a set of vertices in A with a lot of neighbors in B and C. More precisely, each vertex in A^* has at least $(\eta-\gamma)|B|$ neighbors in B and at least $(\eta-\gamma)|C|$ neighbors in C.

Claim 3 $|A^*| \geq (1-2\gamma)|A|$

Proof of Claim Let A' be the "bad" nodes of A with respect to B, i.e. they have fewer than $(\eta-\gamma)|B|$ neighbors in B. Likewise, Let A'' be the "bad" nodes of A with respect to C, i.e. they have fewer than $(\eta-\gamma)|C|$ neighbors in C.

By definition, $A^* = A \setminus (A' \cup A'')$. We would like to show that A' and A'' cannot be too big. That is, we would like $|A'| \leq \gamma|A|$ and $|A''| \leq \gamma|A|$, which would imply that $|A^*| \geq |A| - 2\gamma|A| = (1-2\gamma)|A|$.

To show that $|A'| \leq \gamma|A|$, we assume to the contrary that $|A'| > \gamma|A|$. Consider (A', B). Because of γ-regularity of (A, B), $d(A', B) \geq \eta - \gamma$. However, because each vertex in A' has fewer than $(\eta-\gamma)|B|$ neighbors, $d(A', B) < |A'| \cdot \frac{(\eta-\gamma)|B|}{|A| \cdot |B|} \leq \eta - \gamma$, a contradiction. The same proof holds for A''. ■

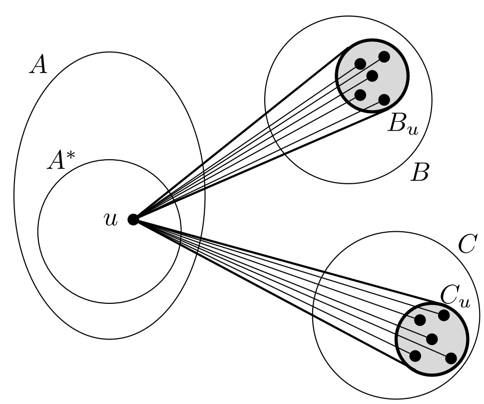

Figure 3: B_u and C_u

For $u \in A^*$, define B_u to be neighbors on u in B, and define C_u to be neighbors on u in C. Note that $\sum_u$(number of edges between B_u and C_u) gives a lower bound on the number of distinct triangles. Also, $|B_u| \geq (\eta - \gamma)|B|$ and $|C_u| \geq (\eta - \gamma)|C|$ by the definition of A^*.

Since γ is chosen as $\frac{\eta}{2}$, $\eta - \gamma = \gamma$. Therefore, $|B_u| \geq \gamma|B|$ and $|C_u| \geq \gamma|C|$. Because (B, C) is γ-regular with density at least η,

$$\begin{aligned} d(B, C) &\geq \eta \\ d(B_u, C_u) &\geq \eta - \gamma \\ e(B_u, C_u) &\geq (\eta - \gamma) \cdot |B_u||C_u| \\ e(B_u, C_u) &\geq (\eta - \gamma)^3 \cdot |B||C| \end{aligned}$$

Thus, $(\eta - \gamma)^3 \cdot |B||C|$ is the lower bound on the number of triangles with $u \in A^*$. Therefore, the total number of triangles in the graph can be lower-bounded by $|A^*| \cdot (\eta - \gamma)^3 \cdot |B||C| = (1 - \eta)(\eta - \gamma)^3 \cdot |A||B||C| = (1 - \eta)\frac{\eta^3}{8} \cdot |A||B||C|$. ■

4 Szemerédi's Regularity Lemma

This lemma was first developed to prove properties of integer sets without arithmetic progressions [3]. The idea of the lemma is that every graph "large enough" can be "approximated" by a constant number of sets of random graphs.

Consider graph $G = (V, E)$ with $|V| = n$, where V is partitioned into k sets of almost equal size (differing by at most one). The edges internal to a partition is not important. Looking at edges across partitions, each pair of partitions is somewhat similar to a random bipartite graph. Partitioning is trivial for $k = 1$, where all edges become internal edges, and for $k = n$, where each vertex has its own partition.

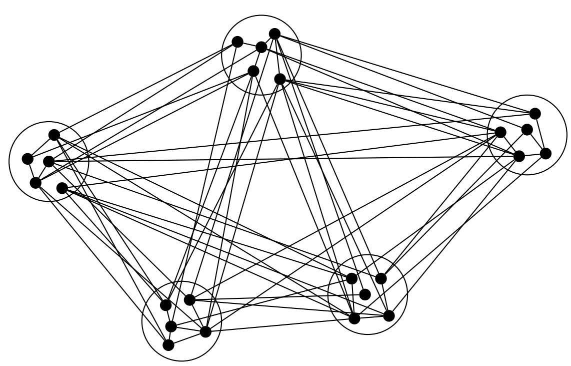

Figure 4: a graph divided into five partitions of equal size

Lemma 4 (Szemerédi's Regularity Lemma [4]) *For all m and $\epsilon > 0$, there exists $T = T(m, \epsilon)$ such that given $G = (V, E)$ where $|V| > T$ and an equipartition $\mathcal{A}$ of V into m sets, there exists an equipartition $\mathcal{B}$ into k sets which refines $\mathcal{A}$ such that $m \leq k \leq T$ and at most $\epsilon\binom{k}{2}$ set pairs are not ϵ-regular.*

$T(m, \epsilon)$ is actually quite big.

$$T(m, \epsilon) \approx 2^{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}$$

where there are $\frac{1}{\epsilon^c}$ levels of exponents.

Proof Idea The following is a very rough idea of the actual proof. Let

$$\text{ind}(V_1, \ldots, V_k) = \frac{1}{k^2} \sum_{i=1}^{k} \sum_{j=i+1}^{k} d^2(V_i, V_j) \leq \frac{1}{2}$$

If a partition violates the property, we can refine into a new parition $V_1', \ldots, V_{k'}'$ such that $\text{ind}(V_1', \ldots, V_{k'}')$ grows significantly, by approximately ϵ^c. We achieve a good partition after $\frac{1}{\epsilon^c}$ refinements. ■

5 Testing Triangle-Freeness of a Dense Graph

This is an application of Szemerédi's Regularity Lemma.

Given graph G in the adjacency matrix format, we would like a one-sided-error randomized algorithm that determines if G is triangle-free. In particular, if G is triangle-free, it should always output PASS. If G is *ϵ-far* from being triangle-free, i.e. at least ϵn^2 edges must be removed from G for it to become triangle-free, it should output FAIL with probability at least $\frac{2}{3}$.

This can be achieved in $O(n^3)$ running time using naive matrix multiplication, or $O(n^\omega)$ with $\omega < 3$ using smarter matrix multiplication. However, this can actually be achieved in $O(1)$, or more accurately

$$O(2^{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}) \text{ (with } \tfrac{1}{\epsilon^c} \text{ levels of exponents)}$$

The following simple algorithm actually gives the desired bound.

```
for O(δ^{-1}) times
    do pick v1, v2, v3
       if v1, v2, v3 forms a triangle
          then output FAIL and halt
output PASS
```

Note that the algorithm always output PASS if the graph is triangle-free. However, it is not obvious that being ϵ-far from triangle-free implies that there are many triangles, enough for the algorithm to find at least one. The following theorem shows that this is actually the case.

Theorem 5 *For all ϵ, there exists δ such that if G is a graph with $|V| = n$ and G is ϵ-far from triangle-free, then G has at least $\delta\binom{n}{3}$ distinct triangles.*

The theorem implies that

$$\Pr[\text{the algorithm fails to find a triangle}] \leq (1-\delta)^{c/\delta} \leq e^{-c}$$

which is less then $\frac{1}{3}$ for $c > \ln 3$.

Proof Let $\mathcal{A}$ be any equipartition of V into $\frac{5}{\epsilon}$.

We use the Szemerédi's Regularity Lemma with $\epsilon' = \min\left\{\frac{\epsilon}{5}, \gamma^{\triangle}\left(\frac{\epsilon}{5}\right)\right\}$ to get a refinement such that

$$\frac{5}{\epsilon} \leq k \leq T\left(\frac{5}{\epsilon}, \epsilon'\right)$$

That is, we use $m = \frac{5}{\epsilon}$. Equivalently,

$$\frac{\epsilon n}{5} \geq \frac{n}{k} \geq \frac{n}{T\left(\frac{5}{\epsilon}, \epsilon'\right)}$$

In addition, the refined partitioning has at most $\epsilon'\binom{k}{2}$ set pairs not ϵ'-regular.

For simplicity, assume that $\frac{n}{k}$, the number of vertices per partition, is an integer. We define G' to be a cleaned-up version of G by doing the following to G:

- Delete edges internal to any V_i. There are n vertices, each with at most $\frac{n}{k}$ neighbors in the same partition. Therefore, the number of edges deleted is at most

$$n \cdot \frac{n}{k} \leq \frac{\epsilon}{5} \cdot n^2$$

- Delete edges between non-regular pairs. There are at most $\epsilon' \binom{k}{2}$ pairs not ϵ'-regular, each with at most $\left(\frac{n}{k}\right)^2$ edges. Therefore, the number of edges deleted is at most

$$\epsilon' \binom{k}{2} \left(\frac{n}{k}\right)^2 \leq \epsilon' \cdot \frac{k^2}{2} \cdot \frac{n^2}{k^2} \leq \frac{\epsilon}{10} \cdot n^2$$

- Delete edges between low-density pairs, where density is less than $\frac{\epsilon}{5}$. First, note that

$$\sum_{\substack{\text{low density} \\ \text{pair}}} \left(\frac{n}{k}\right)^2 \leq \binom{n}{2}$$

  Therefore, the number of edges deleted is at most

$$\sum_{\substack{\text{low density} \\ \text{pair}}} \frac{\epsilon}{5} \left(\frac{n}{k}\right)^2 \leq \frac{\epsilon}{5} \binom{n}{2} \leq \frac{\epsilon}{10} \cdot n^2$$

If the partition sizes were not exactly equal, the number of vertices would be more safely bounded by $\frac{n}{k} + 1$. Nevertheless, the total number of edges deleted is less then ϵn^2. Because we assumed that G is ϵ-far from triangle-free, G' still contains a traingle. In fact, G' has a triangle between V_i, V_j and V_k, for distinct i, j and k, where each pair is ϵ'-regular with density at least $\frac{\epsilon}{5}$.

The idea here is that the existence of one triangle in G' implies the existence of many more triangles because of density. From above, there exists distinct i, j and k such that $x \in V_i$, $y \in V_j$ and $z \in V_k$ where V_i, V_j and V_k all form pairs of density $\eta \geq \frac{\epsilon}{5}$ and γ'-regular where $\gamma' \geq \gamma^{\triangle}\left(\frac{\epsilon}{5}\right) \geq \frac{\eta}{2} \geq \frac{\epsilon}{10}$.

By the triangle counting lemma, there are at least

$$\delta^{\triangle}\left(\frac{\epsilon}{5}\right) \cdot |V_i||V_j||V_k| \geq \frac{\delta^{\triangle}\left(\frac{\epsilon}{5}\right) n^3}{\left(T\left(\frac{5}{\epsilon}, \epsilon'\right)\right)^3} > \delta' \binom{n}{3}$$

triangles in G', and thus in G, for $\delta' = \frac{6\delta^{\triangle}\left(\frac{\epsilon}{5}\right)}{\left(T\left(\frac{5}{\epsilon}, \epsilon'\right)\right)^3}$. ■

6 Other Applications

The technique explained here can be used to test not only for triangles, but also for other constant-sized subgraphs. In addition, almost as-is, this can be used to test properties such as first-order graph properties.

References

[1] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[2] J. Komls and M. Simonovits. Szemerdi's regularity lemma and its applications in graph theory, 1996.

[3] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975. Collection of articles in memory of Juriĭ Vladimirovič Linnik.

[4] E. Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.