0368.416701 Sublinear Time Algorithms December 14, 2015

Lecture 9

Lecturer: Ronitt Rubinfeld Scribe: Elior Rahmani, Lotan Faygler, Alon Brutzkus, Nadav Bar

Lecture overview

1. Testing triangle-freeness in dense graphs
2. Szemerédi Regularity Lemma

1 Testing triangle-freeness of dense graphs

Recall that property testing algorithms in dense graphs use an adjacency matrix representation. While we focus here on triangles, everything that follows can be generalized to other constant size structures as well.

Definition 1 *Given a graph $G = (V, E)$, the adjacency matrix of G is a symmetric matrix A s.t. for every $u \in A, v \in A$:*

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

Definition 2 *$G = (V, E)$ is $\triangle$-free (triangle-free) if $\nexists x, y, z \in V$ s.t. $A(x, y) = A(y, z) = A(z, x) = 1$.*

Claim 3 *If there is a property testing algorithm for $\triangle$-freeness then there is an algorithm that works as follows: pick x, y, z at random from V; fail if $A(x, y) = A(y, z) = A(x, z) = 1$ and otherwise continue (HW3).*

The question remains - how many triplets of vertices do we need to sample? As we later show, only a (large) constant number of queries is needed.

1.1 Random tripartite graphs

Definition 4 *A graph $G = (V, E)$ is tripartite if its vertices can be divided into three disjoint non-empty sets of nodes $A, B, C \subseteq V$ s.t. $V = A \cup B \cup C$ and the nodes of each edge $e \in E$ are either in A and B, or A and C, or B and C. More formally, $(u, v) \in E \Longrightarrow (u \in A, v \in B) \vee (u \in A, v \in C) \vee (u \in B, v \in C)$.*

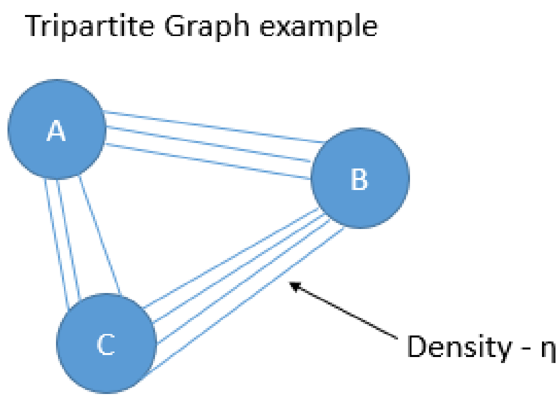
In a random tripartite graph $G = (V, E)$, assuming that the edges are i.i.d., the probability of some edge to be in the graph is determined by a density parameter η such that:

$$\forall u \in A, \forall v \in B : Pr[(u, v) \in E] = \eta$$
$$\forall u \in A, \forall v \in C : Pr[(u, v) \in E] = \eta$$
$$\forall u \in B, \forall v \in C : Pr[(u, v) \in E] = \eta$$

Put differently,

$$E\left[\#A, B\text{-edges}\right] = \eta \cdot |A| \cdot |B|$$
$$E\left[\#A, C\text{-edges}\right] = \eta \cdot |A| \cdot |C|$$
$$E\left[\#B, C\text{-edges}\right] = \eta \cdot |B| \cdot |C|$$

Figure 1: A random tripartite graph with density parameter η.



These settings are demonstrated in Figure 1.

We are interested in the number of triangles in such random graphs. Denote σ_{uvw} to be an indicator variable as follows:

$$\sigma_{uvw} = \begin{cases} 1 & \text{if uvw is } \triangle \\ 0 & otherwise \end{cases}$$

then we get

$$\forall u \in A, v \in B, w \in C : Pr[uvw \text{ is } \triangle] = \eta^3 = E[\sigma_{uvw}]$$

and therefore

$$E[\#\triangle\text{'s}] = E\left[\sum_{u\in A, v\in B, w\in C} \sigma_{uvw}\right] = |A| \cdot |B| \cdot |C| \cdot \eta^3$$

The last equality holds since the edges are assumed to be i.i.d.

1.2 Density & Regularity

Let $A, B \subseteq V$ s.t.:

1. $A \cap B = \emptyset$
2. $|A|, |B| > 1$

we denote $e(A, B)$ to be the number of edges between A and B.

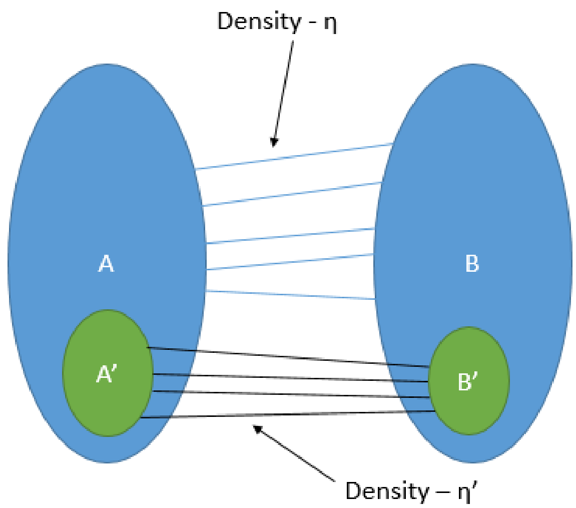
Definition 5 *The density of A, B is $d(A, B) = \frac{e(A,B)}{|A| \cdot |B|}$.*

Note that $|A| \cdot |B|$ is the number of "possible slots" for edges between A and B.

Definition 6 *We say that (A, B) is γ-regular if $\forall A' \subseteq A, B' \subseteq B$ s.t. $|A'| \geq \gamma \cdot |A|, |B'| \geq \gamma \cdot |B|$ we get:* $|\underbrace{d(A', B')}_{\eta'} - \underbrace{d(A, B)}_{\eta}| < \gamma$

From Definition 6 it follows that if (A,B) is γ-regular then (A,B) "behaves" like a random graph (see Figure 2). Note that we could parameterize the difference between the densities and the ratio between the sets and subsets differently. However, for simplicity, we use a single parameter γ.

Figure 2: (A,B) is γ-regular in a graph with density η.



1.3 The Komlós-Simonovits lemma [K-S]

Lemma 7 (K-S) $\forall\eta > 0$:

1. $\exists\gamma \equiv \gamma^{\triangle}(\eta) = \frac{\eta}{2}$ *(regularity parameter)*
2. $\exists\delta \equiv \delta^{\triangle}(\eta) = (1-\eta)\cdot\frac{\eta^3}{8} \underbrace{\geq}_{for\ \eta<\frac{1}{2}} \frac{\eta^3}{16}$

s.t. if A,B,C are disjoint subsets of V and each pair is γ-regular with density $> \eta$, then G contains $\geq \delta\cdot|A|\cdot|B|\cdot|C|$ distinct $\triangle$'s.

Proof Let $A^* \subseteq A$ be the nodes in A with $\geq (\eta-\gamma)\cdot|B|$ neighbors in B and $\geq (\eta-\gamma)\cdot|C|$ neighbors in C (see Figure 3).

Claim 8 $|A^*| \geq (1-2\gamma)\cdot|A|$

Proof of claim: Let A' be the set of nodes in A that have $< (\eta-\gamma)|B|$ neighbors in B, and let A'' be the set of nodes in A that have $< (\eta-\gamma)|C|$ neighbors in C. We get that $|A'| \leq \gamma|A|$, $|A''| \leq \gamma|A|$. In order to see that:

$$d(A',B) < \frac{(\eta-\gamma)\,|B||A'|}{|A'||B|} = \eta-\gamma$$

but, $d\,(A,B) > \eta$, thus we get

$$|d(A',B) - d\,(A,B)| > \eta - (\eta-\gamma) = \gamma$$

which contradicts the γ-regularity.

The proof for A'' is similar. Define $A^* \equiv A \backslash (A' \cup A'')$, we get

$$|A^*| \geq |A| - 2\gamma|A| = (1-2\gamma)\,|A|$$

as claimed. $\square$

For each $u \in A^*$, let B_u, C_u be the set of u's neighbors in B and C, respectively (see Figure 4). We get that:

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

We choose γ to be $\frac{\eta}{2}$, therefore

$$|B_u| \geq \gamma|B|$$
$$|C_u| \geq \gamma|C|$$

The number of edges between B_u and C_u is a lower bound for the number of distinct $\triangle$'s with u. Finally, since $d\,(B, C) \geq \eta$, and since $|B_u| \geq \gamma|B|$ and $|C_u| \geq \gamma|C|$, and B, C are $\gamma-$regular, we get that

$$\begin{aligned}
& d\,(B_u, C_u) \geq \eta - \gamma \\
\Longrightarrow \quad & e\,(B_u, C_u) \geq (\eta - \gamma)\,|B_u||C_u| \geq (\eta-\gamma)^3\,|B||C| \\
\Longrightarrow \quad & \#\triangle\text{'s} \geq (1-2\gamma)\,|A|\,(\eta-\gamma)^3\,|B||C| \\
& = (1-\eta)\cdot\frac{\eta^3}{8}|A||B||C|
\end{aligned}$$

$\blacksquare$

Figure 3: A^* and its neighbors in B and C.

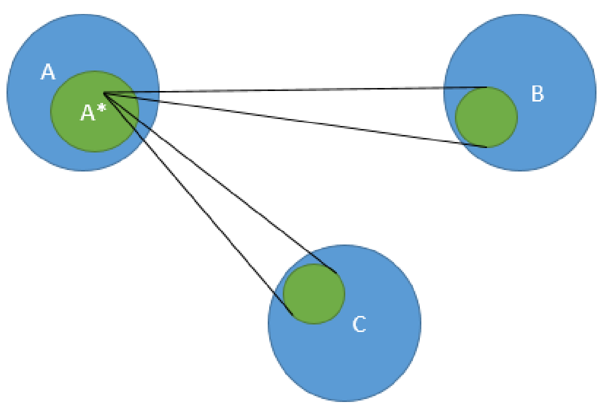
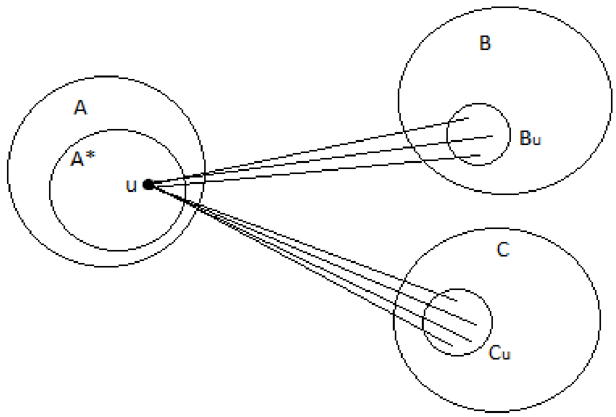


Figure 4: Each edge between B_u and C_u represents a distinct triangle.

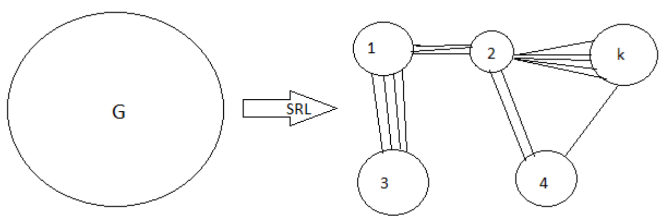


2 Szemerédi Regularity Lemma (SRL)

2.1 Intiution

Given a regularization parameter, we can perform a partition of each graph with a large enough number of nodes, such that each one of the partition's pairs of sets will be almost regular (only ϵ fraction will not be regular; demonstrated in Figure 5).

Figure 5: Given a graph G, SRL refines its partition into k sets s.t. every pair acts as a bipartite graph with a different density.



2.2 The Lemma

Lemma 9 (SRL) *$\forall m, \epsilon > 0 : \exists T = T(m, \epsilon)$ s.t. given $G = (V, E)$ with $|V| > T$ and A, an equipartition of V into m sets, there is some equipartition B of V into k sets which refines A s.t $m \leq k \leq T$ and at most $\epsilon\binom{k}{2}$ set pairs are not ϵ-regular.*

Notes:

- Using the regularity lemma, we can partition any large enough graph into a constant number of parts (depends in ϵ). Each graph "behaves" like a random bipartite graph.
- SRL was studied to prove a conjecture by Erdős and Turán: sequence of integers must always contain long arithmetic progressions.

2.3 Application of SRL to Triangle-Freeness

We will now show how SRL can be utilized to derive an efficient tester for the triangle-freeness property.

Given a graph in adjacency matrix representation, we require that a tester for the triangle-freeness property satisfies the following:

- If G is $\triangle$-free, output PASS with probability 1.
- If G is ϵ-far from $\triangle$-free (i.e., we need to delete at least ϵn^2 edges to make it $\triangle$-free), output FAIL with probability $\geq \frac{3}{4}$.

Algorithm 1: Triangle-Freeness Property Tester

```
Do O(1/δ) times:
    Randomly choose v1, v2, v3 ∈ V.
    If these vertices form a triangle, output FAIL and halt.
Output PASS
```

We will show that Algorithm 1 has the desired behavior. The algorithm runs in constant time. Note, however, that the constant may be huge in terms of ϵ, as we shall see.

The crux of our argument will be that, following SRL, a graph which is ϵ-far from $\triangle$-free contains three disjoint sets of vertices, such that each pair of these sets is dense and "behaves" like a random bipartite graph. Hence, the graph has a sufficient amount of triangles by the K-S triangle-counting lemma.

Theorem 10 *$\forall \epsilon : \exists \delta$ s.t. any graph $G = (V, E)$ with $|V| = n$ that is ϵ-far from $\triangle$-free, has at least $\delta \binom{n}{3}$ triangles.*

Corollary 11 *Algorithm 1 is a tester for the triangle-freeness property.*

Proof of corollary: ssume that the algorithm runs for $\frac{c}{\delta}$ iterations. If G is $\triangle$-free then clearly the algorithm outputs PASS with probability 1. If G is ϵ-far from $\triangle$-free, then by theorem 10 it contains at least $\delta \binom{n}{3}$ triangles. Therefore, the probability that we won't sample a triangle in all of the $\frac{c}{\delta}$ iterations is at most $(1-\delta)^{\frac{c}{\delta}} \leq e^{-c} \leq \frac{1}{4}$ for a sufficiently large c, and the conclusion follows. $\square$

Proof of Theorem

Let A be an arbitrary equipartition of $|V|$ into $\frac{5}{\epsilon}$ sets. Let $\epsilon' = \frac{\epsilon}{10}$, then by SRL there is a an equipartition $B = \{V_1, ..., V_k\}$ that refines A such that $\frac{\epsilon}{5} \leq k \leq T(\frac{\epsilon}{5}, \epsilon')$, and at most $\epsilon' \binom{k}{2}$ pairs of sets in B are ϵ'-regular. We assume, without loss of generality, that all sets in B are of equal size. Hence, each set is of size $\frac{n}{k}$ for which $\frac{n}{T(\frac{\epsilon}{5}, \epsilon')} \leq \frac{n}{k} \leq \frac{\epsilon n}{5}$.

We will now construct a new graph G' which is a subgraph of G, and we will prove that the claim holds for G' and thus holds for G as well. We construct G' by taking G and deleting the following edges:

1. Edges which have both ends in some set V_i. The number of edges to delete is at most

$$\sum_{i=1}^{k} \sum_{v \in V_i} |V_i| = \sum_{i=1}^{k} \sum_{v \in V_i} \frac{n}{k} \leq \frac{\epsilon n^2}{5}$$

2. Edges that connect two partition sets that are not ϵ'-regular. The number of edges to delete is at most

$$\epsilon' \binom{k}{2} \left(\frac{n}{k}\right)^2 \leq \frac{\epsilon}{5} \cdot \frac{k^2}{2} \cdot \frac{n^2}{k^2} = \frac{\epsilon n^2}{10}$$

3. Edges between partition sets with low density (at most $\frac{\epsilon}{5}$). The number of edges to delete is at most

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

   The first inequality holds since

$$\sum_{\text{low density pairs}} \left(\frac{n}{k}\right)^2 \leq \binom{k}{2}\left(\frac{n}{k}\right)^2 = \frac{k-1}{k}\cdot\frac{n^2}{2} \leq \frac{n-1}{n}\cdot\frac{n^2}{2} = \binom{n}{2}$$

The total number of edges deleted from G is at most ϵn^2. Since G is ϵ-far from $\triangle$-free, G' must contain a triangle. Let a, b, c be the vertices of the triangle. Due to the aforementioned construction, there exist three distinct partition sets $\{V_i, V_j, V_k\}$ such that $a \in V_i$, $b \in V_j$, $c \in V_k$ and each pair has density $\eta > \frac{\epsilon}{5}$ and is $\gamma^{\triangle}(\frac{\epsilon}{5})$-regular (note that $\gamma^{\triangle}(\frac{\epsilon}{5}) = \frac{\epsilon}{10} = \epsilon'$).

By the K-S triangle-counting lemma, we conclude that the number of triangles in G' is at least

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

where $\delta = \frac{6\delta^{\triangle}(\frac{\epsilon}{5})}{T(\frac{5}{\epsilon},\epsilon')^3}$ (note that $\delta^{\triangle}(\frac{\epsilon}{5}) = (1-\frac{\epsilon}{5})\frac{(\frac{\epsilon}{5})^3}{8} \geq \frac{1}{2}\frac{\epsilon^3}{1000} = \frac{\epsilon^3}{2000}$). ■

Remark Although we only focused on the traingle-freeness property, these results can be extended to derive efficient testers for any H-free property, where H is a constant sized graph.