6.889 Sublinear Time Algorithms March 4, 2019

Lecture 8

Lecturer: Ronitt Rubinfeld *Scribe: William Loucks*

1 Testing "Triangle Freeness" for Dense Graphs

Definition 1 *Triangle Freeness.*

Graph G is triangle free, or Δ-free, if there does not exist an x, y, x such that $A(x, y) = A(y, z) = A(x, z) = 1$.

Claim: If there exists a property testing algorithm for Δ-freeness, then there exists an algorithm that works as follows:

1. Pick random x, y, z
2. Test if $A(x, y) = A(y, z) = A(x, z) = 1$

However, we need to show how many times we must query the above instructions.

2 Detour

Let's first determine how many triangles are in a random tripartite graph and then illustrate tools to assess triangle freeness.

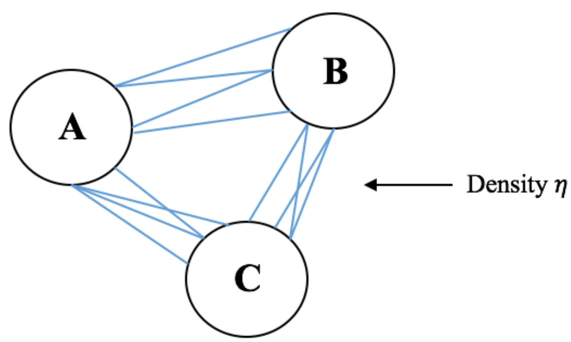

Figure 1: Random tripartite graph with density η

Assume that the density of edges between all subgraphs, or sets, above is η and $\Delta_{a,b,c}$ is an indicator variable such that:

$$\Delta_{a,b,c} = \begin{cases} 1 & \text{if there exists a triangle connecting } a, b, c \\ 0 & \text{otherwise} \end{cases}$$

Now, $\forall\ a \in A, b \in B, c \in C$, the probability that there exists a triangle connecting some a, b, c and the expected value of the indicator are the following:

$$Pr[\Delta_{a,b,c}] = \eta^3$$
$$E[\Delta_{a,b,c}] = \eta^3$$

Further, the expected number of triangles connecting the three subgraphs above is computed as:

$$E[\#\Delta s] = |A||B||C| \cdot \eta^3$$

Now, let's define the density and regularity of set pairs.

Definition 2 *Regular Pairs. (i.e. γ-regular)*

Let $A, B \subseteq V$ such that $A \cap B = \emptyset$, $|A| > 1$, and $|B| > 1$. Let $e(A, B) =$ the number of edges between A and B, with density defined as:

$$d(A, B) = \tfrac{e(A,B)}{|A||B|}$$

We say that A, B are γ-regular if $\forall\ A' \subseteq A$ and $\forall\ B' \subseteq B$ where

$$|A'| \geq \gamma \cdot |A| \text{ and } |B'| \geq \gamma \cdot |B|,$$

the difference in densities between the pairs is:

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

Thus, the graphs A' and B' must first be large enough to behave like random graphs, and then the densities between the pairs must be less than γ. Note, the γ values above – indicating the size of the subsets and the difference in density – do not have to be the same. Here, we simply use the same variable to reduce the number of parameters.

Lemma 3 *Triangle Counting Lemma (Komlós and Simonovits). $\forall \eta > 0$, there exists $\gamma = \gamma^{\Delta}(\eta) = \frac{1}{2} \cdot \eta$ and $\delta = \delta^{\Delta}(\eta) = (1 - \eta) \cdot \frac{\eta^3}{8} \geq \frac{\eta^3}{16}$ (if $\eta < \frac{1}{2}$), such that if A, B, and C are disjoint subsets of V, and each pair is γ-regular with density $> \eta$, then G contains $\geq \delta \cdot |A||B||C|$ triangles with a node in each of A, B, and C.*

Proof We aim to prove the Triangle Counting Lemma. Note, such a lemma exists for all sizes of subgraphs. Let $A^* =$ the nodes in A with $\geq (\eta - \gamma)|B|$ neighbors in B and $\geq (\eta - \gamma)|C|$ neighbors in C.

In order to proceed, consider the following claim:

Claim 4 $|A^*| \geq (1 - 2\gamma)|A|$

> **Proof** To prove the above claim, we know that if A' is the number of *bad* nodes of A with respect to B and A'' is the number of *bad* nodes of A with respect to C – in other words, there are $< (\eta - \gamma)|B|$ neighbors in B and $< (\eta - \gamma)|C|$ neighbors in C, respectively – then $|A'| \leq \gamma \cdot |A|$ and $|A''| \leq \gamma \cdot |A|$.
>
> For contradiction, assume this is not true, i.e. $|A'| > \gamma \cdot |A|$. Then
>
> $$d(A', B) = \tfrac{|A'| \cdot (\eta - \gamma) \cdot |B|}{|A'||B|} = (\eta - \gamma)$$

However, we know that $d(A, B) > \eta$ (by definition in the lemma), causing

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

which contradicts the assumed γ-regularity. Note, B is large enough to behave as a random graph, by definition, and A' is at least A by the assumption, leading A' to be large enough to also behave as a random graph. One can make a similar argument for A''.

Observe that $A^* = A \setminus (A' \cup A'')$, since A^* does not contain *bad* nodes. So

$$\begin{aligned} A^* &\geq |A| - |A'| - |A''| \\ &\geq |A| - 2\gamma \cdot |A|, \text{ since we showed that } |A'| \leq \gamma \cdot |A| \text{ and } |A''| \leq \gamma \cdot |A| \\ &\geq (1 - 2\gamma)|A| \end{aligned}$$

■

To complete the proof of the Triangle Counting Lemma, for each $u \in A^*$, define B_u to be the neighbors of u in B and C_u to be the neighbors of u in C. Thus, if $\gamma \leq \frac{\eta}{2}$:

$$\begin{aligned} |B_u| \geq (\eta - \gamma) \cdot |B| \geq \gamma \cdot |B| \\ |C_u| \geq (\eta - \gamma) \cdot |C| \geq \gamma \cdot |C| \end{aligned}$$

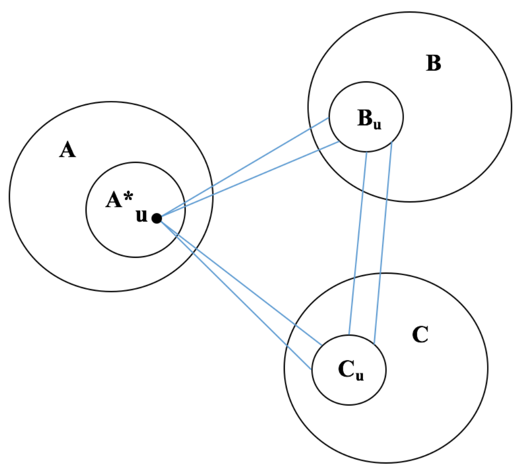

Figure 2: Tripartite graph with $u \in A^*$, where B_u and C_u are neighbors of u.

As a result, $|B_u|$ and $|C_u|$ are large enough. Further, note that we assume $d(B, C) = \eta$ in the lemma. Thus,

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

This gives a lower bound on the number of triangles that contain u as an endpoint. The total number of triangles with a node in each of A, B, and C is then as follows:

$$\begin{aligned}\text{total \# of triangles} &\geq \sum_{u \in A^*} (\eta - \gamma)^3 \cdot |B||C| \\ &\geq (1-2\gamma)|A| \cdot (\eta-\gamma)^3 \cdot |B||C| \\ &\geq (1-2\gamma) \cdot (\eta-\gamma)^3 \cdot |A||B||C|, \text{ and since we choose } \gamma \leq \frac{\eta}{2}, \\ &\geq (1-\eta) \cdot \frac{\eta^3}{8} \cdot |A||B||C|\end{aligned}$$

■

3 Szemerédi's Regularity Lemma (SRL)

We would like to equipartition the nodes in a graph into sets $V_1, ..., V_k$ such that all (or most) pairs (V_i, V_j) are ϵ-regular.

Lemma 5 *$\forall m$ and $\epsilon > 0$, there exists $T(m, \epsilon)$ such that given $G = (V, E)$ with $|V| > T$ where A is an equipartition of V into ($m << T$) sets, then there exists an equipartition B of V into k sets which refine A such that $m \leq k \leq T$ and $\leq \binom{k}{2}$ set pairs are not ϵ-regular.*

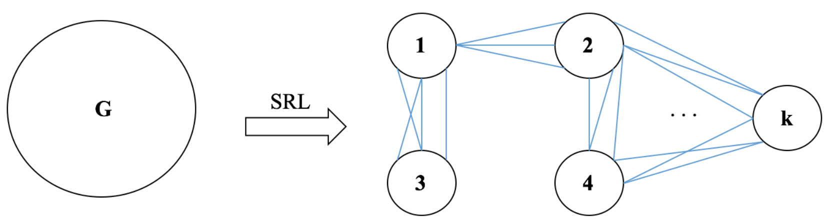

Figure 3: Apply SRL to refine G into a constant number of partitions such that the pairs behave like a random bipartite graph and are mostly regular.

In other words, given an arbitrary starting point, we can refine A so that the graph is ϵ-regular and all subgraphs have roughly the same number of nodes. Further, we can partition the graph into a constant number of partitions such that each pair of sets behaves like a random bipartite graph.

3.1 Property Testing

Property testing is an application of the SRL. Given a graph in adjacency matrix form, we would like to construct an algorithm which outputs PASS if the graph is triangle free and FAIL with probability $\geq \frac{3}{4}$ if the graph is ϵ-far from triangle free. Note, if the graph is ϵ-far from triangle free, one must add $\epsilon \cdot n^2$ edges to transform the graph to be triangle free. A possible algorithm is the following:

Algorithm 1: Triangle Freeness

Input : Graph G in adjacency matrix form

for $O(\delta^{-1})$ *iterations* **do**

    pick V_1, V_2, V_3

    if Δ, halt and output FAIL

Return PASS

To assess the behavior of the above algorithm, consider the theorem:

Theorem 6 *If G is ϵ-far from Δ-free, then G has $\geq \delta \cdot \binom{n}{3}$ distinct Δs.*

As a result, $O(\frac{1}{\delta})$ loops of the algorithm finds a possible triangle with high probability.

Corollary 7 *The algorithm accepts with probability 1 if the graph is triangle free. If the graph is ϵ-far from triangle free, meaning there are more than $\delta \cdot \binom{n}{3}$ triangles,*

$$\begin{aligned} Pr[\textit{do not find a tringle in } \frac{c}{\delta} \textit{ loops}] &\leq (1-\delta)^{c/\delta} \\ &\leq e^{-c} \\ &< \frac{1}{4} \textit{ for big enough } c \end{aligned}$$

Proof Given the corollary, we need to prove Theorem 6. With this, we can construct the algorithm to test if the graph is ϵ-far from triangle free with failure probability less than $\frac{1}{4}$. First, we use the SRL to obtain $\{V_1, ..., V_k\}$ such that $\frac{5}{\epsilon} \leq k \leq T(\frac{5}{\epsilon}, \epsilon')$ for $\epsilon' = min\{\frac{\epsilon}{5}, \gamma^{\Delta}(\frac{\epsilon}{5})\}$ such that less than $\epsilon' \cdot \binom{k}{2}$ pairs are not ϵ'-regular. The aforementioned is equivalent to $\frac{\epsilon \cdot n}{5} \geq \frac{n}{k} \geq \frac{n}{T(\frac{5}{\epsilon}, \epsilon')}$, representing the number of nodes per partition.

To clean up G, we assume that $\frac{n}{k}$ (the number of nodes per partition) is an integer. G' is the result after performing the following:

1. Delete edges internal to any V_i. This amounts to $\leq \frac{n}{k} \cdot n \leq \frac{\epsilon \cdot n^2}{5}$ deleted edges. Note, we multiplied by n to sum over all of the nodes.
2. Delete edges between non-regular pairs. This amounts to $\leq \epsilon' \cdot \binom{k}{2} \cdot (\frac{n}{k})^2 \leq \frac{\epsilon}{5} \cdot \frac{k^2}{2} \cdot \frac{n^2}{k^2} \leq \frac{\epsilon \cdot n^2}{10}$.
3. Remove low density ($< \frac{\epsilon}{5}$) pairs. This amounts to $\leq \sum_{\text{low density pairs}} \frac{\epsilon}{5} (\frac{n}{k})^2 \leq \frac{\epsilon}{5} \binom{n}{2} \leq \frac{\epsilon \cdot n^2}{10}$

Therefore, the total number of deleted edges is $\frac{\epsilon \cdot n^2}{5} + \frac{\epsilon \cdot n^2}{10} + \frac{\epsilon \cdot n^2}{10} < \epsilon \cdot n^2$. Thus, since G was ϵ-far from triangle free, G' must still have a triangle. By the way we constructed G', we know the remaining triangles between some V_i, V_j, V_k contain:

1. Distinct endpoints, since we removed all edges within the partitions
2. Regular pairs, since we removed all non-regular pairs
3. Dense pairs, since we removed all low density pairs

In the end, nodes in each one of V_i, V_j, V_k comprise distinct triangles which have $\geq \frac{\epsilon}{5}$ density and $\delta^{\Delta}(\frac{\epsilon}{5})$-regular pairs.

To determine the number of triangles in G', we invoke the Triangle Counting Lemma.

$$
\begin{aligned}
&\geq \delta^{\Delta}(\frac{\epsilon}{5})|V_i||V_j||V_k| \ \Delta s \text{ remain in } G' \\
&\geq \frac{(\frac{\epsilon}{5})^3}{16}|V_i||V_j||V_k| \\
&\geq \frac{(\frac{\epsilon}{5})^3}{16} \cdot \frac{n^3}{T(\frac{5}{\epsilon}, \epsilon')^3} \text{ since } k \leq T(\frac{5}{\epsilon}, \epsilon') \\
&\geq \delta \cdot \binom{n}{3}
\end{aligned}
$$

■

Now that we have proven Theorem 6, we can use the previously mentioned algorithm for triangle freeness, which fails with probability less than $\frac{1}{4}$ after $O(\frac{1}{\delta})$ iterations when the graph is ϵ-far from Δ-free.