0368.416701 Sublinear Time Algorithms December 22, 2014

Lecture 6

Lecturer: Ronitt Rubinfeld Scribe: Noam Touitou, Yaniv Sabo and Aviel Atias

1 Today

Testing triangle freeness in dense graphs.

2 Some definitions

Definition 1 *G is $\triangle$-free if $\nexists x, y, z$ such that $A(x,y) = A(x,z) = A(y,z) = 1$ where A is the adjacency matrix of G.*

Claim 2 *(left for homework) If there is a property testing algorithm for $\triangle$-freeness then there is an algorithm that works as follows:*

- *pick random x, y, z*
- *test if $A(x,y) = A(y,z) = A(x,z) = 1$*

The claim states that using more samples, one can turn a non-adaptive algorithm into adaptive.

Definition 3 *In a random graph, for each edge we flip a coin in order to determine if it exists in the graph. We denote the probability of the coin to say "yes" by η and call this value the "graph density"*

Definition 4 $\triangle_{uvw} = \begin{cases} 1 & \text{if } A(u,v) = A(v,w) = A(u,w) = 1 \\ 0 & \text{otherwise} \end{cases}$

3 Number of triangles in a dense graph

Detour 5 *How many $\triangle$'s in a random tripartite graph?*

$$\forall u \in A, v \in B, w \in C : \Pr[\triangle_{uvw} = 1] = \eta^3$$

$$\mathrm{E}[\triangle_{uvw}] = \eta^3$$

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

Definition 6 *For $A, B \subseteq V$ such that (1) $A \cap B = \emptyset$ (2) $|A|, |B| > 1$ let $e(A, B) =$ number of edges between A, B.*

$$\textit{density: } d(A, B) = \frac{e(A, B)}{|A||B|}$$

Say (A, B) is γ-regular if:

$$\forall A' \subseteq A, B' \subseteq B \text{ such that } |A'| \geq \gamma|A| \text{ and } |B'| \geq \gamma|B|$$

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

Lemma 7 *(Triangle Counting) Komlos Simonovitz*
$\forall \eta > 0,$

$$\exists \gamma, \delta \text{ such that if } A, B, C \text{ disjoint subsets of } V,$$
$$\text{and each pair is } \gamma\text{-regular with respect to density } \eta$$
$$\gamma(\text{depends on } \eta) = 1/2\eta = \gamma^{\triangle}(\eta)$$
$$\delta(\text{depends on } \eta) = (1-\eta) \cdot \frac{\eta^3}{8} \geq \frac{\eta^3}{16} = \delta^{\triangle}(\eta)$$
$$(\text{the last inequailty holds whenever } \eta < 1/2)$$
$$\text{then } G \text{ contains } \geq \delta \cdot |A||B||C| \text{ distinct } \triangle\text{'s}$$
$$\text{with nodes from each } A, B, C.$$

Proof (simplification of [Alon Fischer Krivelevich Szegedy])
$A^* \leftarrow$ nodes in A with $\geq (\eta - \gamma)|B|$ neighbors from B and with $\geq (\eta - \gamma)|C|$ neighbors from C.

Claim 8 $|A^*| \geq (1 - 2\gamma)|A|$
Proof
$A' \leftarrow$ nodes in A that have $< (\eta - \gamma)|B|$ nodes in B
$A'' \leftarrow$ nodes in A that have $< (\eta - \gamma)|C|$ nodes in C
$|A'| \leq \gamma|A|$, $|A''| \leq \gamma|A|$
why? if not, assume $|A'| > \gamma|A|$. Consider pair (A', B). $|A'| \geq \gamma|A|$, and since $\gamma \leq 1$ then $|B| \geq \gamma|B|$. So:

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

since γ-regularity, $|d(A', B) - d(A, B)| < \gamma$, but $d(A, B) > \eta$ so $|d(A', B) - d(A, B)| > \eta - (\eta - \gamma) = \gamma$ which contradicts γ-regularity.
The proof for A' is similar.

So:

$$A^* \equiv A \setminus (A' \cup A'')$$
$$|A^*| \geq |A| - 2\gamma|A|$$
$$= (1 - 2\gamma)|A|$$

■

For each $u \in A^*$, define:
B_u = neighbors of u in B
C_u = neighbors of u in C
Then:

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

If we make assumption on γ choice ($\gamma < \frac{\eta}{2}$), we have $\eta - \gamma \geq \gamma$ so:

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

Number of edges between B_u and C_u $\Rightarrow$ lower bound on number of distinct $\triangle$'s with u as a vertex.

$$d(B, C) \geq \eta$$
$$\Rightarrow d(B_u, C_u) \geq \eta - \gamma \text{ (Since } |B_u|, |C_u| \text{ big enough, and } B, C \text{ are } \gamma\text{-regular.)}$$
$$\Rightarrow e(B_u, C_u) \geq (\eta - \gamma)|B_u||C_u|$$
$$\geq (\eta - \gamma)^3|B||C|$$
$$\Rightarrow \text{total number of } \triangle\text{'s } \geq (1 - 2\gamma) \cdot |A| \cdot (\eta - \gamma)^3 \cdot |B||C|$$
$$= (1 - \eta) \cdot \frac{\eta^3}{8} \cdot |A||B||C| \text{ (choosing } \gamma = \eta/2)$$

■

4 Szemerédi Regularity Lemma (SRL)

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

4.1 Notes

- Using the regularity lemma, we can partition any graph into a "constant" number of parts, i.e it only depends on ϵ. Each pair 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.

4.2 An application of the SRL

Given a graph G in adjacency matrix format we would like an algorithm which has this behavior:

- If G is $\triangle$-free output pass.
- If G is ϵ-far from $\triangle$-free (i.e, we need to delete at least ϵn^2 edges to make it $\triangle$-free) then output fail with probability 3/4.

Definition 10 *Our algorithm*
Do $O(1/\delta)$ times: pick random v_1,v_2,v_3 in V. If it is a triangle reject and halt. If no such triangle was found, accept.

Theorem 11 *$\forall \epsilon > 0$, $\exists \delta$ s.t $\forall G$ with $|V| = n$ and G is ϵ-far from $\triangle$-free then G has at least $\delta\binom{n}{3}$ distinct triangles.*

Corollary 12 *Algorithm has desired behavior*

Proof If G is triangle free, the algorithm accepts with probability 1. If G is ϵ-far, then there are at least $\delta\binom{n}{3}$ triangles and so the probability we won't sample a triangle in c/δ loops is at most $(1-\delta)^{c/\delta} \le e^{-c} < 1/4$ for a large enough c. ∎

Proof (of theorem)
Use regularity to get an equipartition $\{V_1, V_2, ..., V_k\}$ s.t $\frac{5}{\epsilon} \le k \le T(\frac{5}{\epsilon}, \epsilon')$ (use A, an arbitrary equipartition into $5/\epsilon$ sets).
The number of nodes in each part is n/k and so $\frac{n}{T(\frac{5}{\epsilon},\epsilon')} \le \frac{n}{k} \le \frac{\epsilon n}{5}$
We will choose $\epsilon' = min\left[\frac{\epsilon}{5}, \gamma^{\triangle}(\frac{\epsilon}{5})\right]$ s.t at most $\epsilon'\binom{k}{2}$ set pairs are not ϵ'-regular.
We need the number of parts to be large enough s.t the number of edges inside each part isn't too big.
Assume n/k is an integer. We'll define a new graph G' as follows:
Take G and:

1. Delete edges of G internal to any element V_i of the partition. The number of edges we have deleted is

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

2. Delete edges between ϵ'-nonregular pairs(note that $\epsilon' \le \frac{\epsilon}{5}$). The number of edges deleted is
$$\le \epsilon' \binom{k}{2} (\frac{n}{k})^2 \le \frac{\epsilon}{5} \cdot \frac{k^2}{2} \cdot \frac{n^2}{k^2} = \frac{\epsilon n^2}{10}$$

3. Delete edges between low density pairs (pairs of density $\le \frac{\epsilon}{5}$). The number of edges deleted is
$$\le \sum_{\text{low density pairs}} \frac{\epsilon}{5} \cdot (\frac{n}{k})^2 \le \frac{\epsilon}{5} \cdot \binom{n}{2} \approx \frac{\epsilon n^2}{10}$$
[Note that $\sum_{\text{low density pairs}} (\frac{n}{k})^2 \le \binom{k}{2} (\frac{n}{k})^2 = \frac{k(k-1)}{2} (\frac{n}{k})^2 = \frac{k-1}{k} \cdot \frac{n^2}{2} \le \frac{n-1}{n} \cdot \frac{n^2}{2} = \binom{n}{2}$]

The total number of edges deleted from G is $< \epsilon n^2$. G was ϵ-far from triangle-free, and thus G' still has a triangle.

Let a, b, c be the nodes of the triangle. Due to the aforementioned edge removal, $\exists i, j, k$ that are distinct s.t. $a \in V_i, b \in V_j, c \in V_k$ and each pair from $\{V_i, V_j, V_k\}$ is both a high density pair(i.e., has density $\ge \frac{\epsilon}{5}$) and $\gamma^{\triangle}(\frac{\epsilon}{5})$-regular.

Due to the triangle-counting lemma, we have that the number of triangles in G' is
$$\ge \delta^{\triangle}(\frac{\epsilon}{5}) |V_i||V_j||V_k| \ge \delta^{\triangle}(\frac{\epsilon}{5}) \cdot \frac{n^3}{T(\frac{5}{\epsilon}, \epsilon')} \ge \delta' \cdot \binom{n}{3}$$
for $\delta' = 6\delta^{\triangle}(\frac{\epsilon}{5})(T(\frac{5}{\epsilon}, \epsilon'))^{-3}$ [Notice that $\delta^{\triangle}(\frac{\epsilon}{5}) = (1 - \frac{\epsilon}{5})\frac{(\frac{\epsilon}{5})^3}{8} \ge \frac{1}{2} \cdot \frac{\epsilon^3}{1000} = \frac{\epsilon^3}{2000}$].

■