6.896 Sublinear Time Algorithms March 22, 2007

Lecture 14

Lecturer: Ronitt Rubinfeld *Scribe: Huy Ngoc Nguyen*

1 First order properties testing:

First order graph property is a property that can be expressed in the following form:

$$\exists u_1, u_2, ...u_k \forall v_1, v_2, ...v_l \qquad \underbrace{R(u_1, ..., u_k, v_1, ..., v_l)}_{\text{may include } \wedge, \vee, \neg \text{ or other logical operations)}}$$

In this lecture, we will show a special case of the following theorem.

Theorem 1 *: Alon, Fischer, Krivilevich and Szegedy*
All first order properties of $\exists$, $\forall$ type are testable in adjacent matrix model with queries number independent of n.

In fact, recent results have given characterizations of the graph properties testable in constant time in the adjacency matrix representation, though we will not cover those proofs in lecture. For example, such properties include all monotone graph properties.

1.1 Δ-free testing

In this lecture, we will consider a case of first order property testing,

$$\Delta - free : \forall u_1, u_2, u_3 R(u_1, u_2, u_3) = \neg(u_1\ u_2 \wedge u_2\ u_3 \wedge u_1\ u_3)$$

Definition 1 *G is "ε-far" from Δ-free if we must add/delete at least εn^2 edges to make it Δ-free.*

We will present a 1-sided Δ-free tester which has the following behavior:

1. If G is Δ-free, it PASSES
2. If G is ε-far from Δ-free, it FAILS with constant probability.

1.2 Regularity

Definition 2 *For any pair of disjoint subsets A, B in V s.t $|A|, |B| > 1$. Let $e(A, B) \leftarrow$ number of edges between A and B,* **density** $d(A, B) = \frac{e(A,B)}{|A||B|}$.

Definition 3 *A, B is* **"γ"-regular** *if $\forall A' \subset A, B' \subset B$ s.t*

1. $|A'| \geq \gamma |A|$
2. $|B'| \geq \gamma |B|$
3. $|d(A', B') - d(A, B)| < \gamma$

Lemma 2 *Szemeredi's Regularity Lemma - will be proved in the next lecture*
$\forall m$ and $\varepsilon > 0$, $\exists$ $\mathcal{T} (= \mathcal{T}(m, \varepsilon))$ s.t if $G = (V, E)$ with $|V| > \mathcal{T}$ and $\mathcal{A}$ is equipatition of V into m sets then there exists a equipartition $\mathcal{B}$, a refinement of $\mathcal{A}$ into $\mathcal{K}$ sets ($m \leq \mathcal{K} < \mathcal{T}$) s.t at most $\varepsilon \binom{\mathcal{K}}{2}$ set pairs that not ε-regular.

Note that this is only a theoretical result as the value of $\mathcal{T}$ is normally very large (tower of 2's of length polynomial to $\frac{1}{\varepsilon}$).

1.3 Number of triangles in a tripartite graph

Consider a random tripartite graph G (That is, a graph has a equipartition into 3 sets V_1, V_2 and V_3 s.t there are no internal edges in each subgraph V_i and the probability for each pair of nodes (u,v) from 2 different subgraphs to be an edge is η) . For any (u,v,w) s.t $u \in V_1, v \in V_2, w \in V_3$ if we let

$$\sigma_{u,v,w} = \begin{cases} 1 \text{ if u,v,w is triangle} \\ 0 \text{ ow} \end{cases}$$

$\Rightarrow E\left[\sigma_{u,v,w}\right] = Pr\left[\sigma_{u,v,w} = 1\right] \geq \eta^3$
$\Rightarrow E\left[\Sigma\sigma_{u,v,w}\right] = \eta^3(\frac{n}{3})^3$
Therefore, we can expect the number of triangles in a random tripartite graph is $(\eta\frac{n}{3})^3$. In case of regular graphs, the following lemma shows that there are also as many $(\Omega(\eta n^3))$ triangles.

Lemma 3 $\forall\eta > 0, \exists\gamma = \gamma^{\Delta}(\eta) < \frac{1}{2}\eta$ & $\delta \equiv \delta^{\Delta}(\eta) \geq \frac{\eta^3}{16}$ *st if A, B, C are disjoint subset of V s.t each pair $\gamma - regular$ with density at least η then G contains at least $\delta|A||B||C|$ triangles.*

Proof $A^* \leftarrow$ nodes in A with

1. at least $(\eta-\gamma)|B|$ neighbors in B and
2. at least $(\eta-\gamma)|C|$ neighbors in C.

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

Proof
$|A'| \leftarrow$ nodes with less than $(\eta-\gamma)|B|$ nbrs in B, consider the pair (A', B):

$$\begin{aligned} d(A',B) &= \frac{e(A',B)}{|A||B|} \\ &\leq \frac{(\eta-\gamma)|B||A'|}{|A'||B|} \\ &= \eta-\gamma \\ &\leq d(A,B)-\gamma \end{aligned}$$

Therefore, by definition of $\gamma - regular$,

$$|A'| < \gamma|A|$$

Similarly, $A'' \leftarrow$ nodes in A width less than $(\eta-\gamma)|C|$ nbrs in C, we can show that $|A''| < \gamma|A|$.
But $A^* = A \setminus (A' \cup A'') \Rightarrow A^* \geq |A| - 2\gamma|A| = (1-2\gamma)|A|$ ∎

For each $u \in A^*$, let $B_u \leftarrow$ nbrs of u in B and $C_u \leftarrow$ nbrs of u in C.
From the definition of A^*, $|B_u| \geq (\eta-\gamma)|B|$ and $|C_u| \geq (\eta-\gamma)|C|$.
Since each edge (y,z) between B_u, C_u gives a distinct triangle (u,y,z)

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

$$\begin{aligned} \text{Total number of triangles is} &\geq (1-2\gamma)(\eta-\gamma)^3|A||B||C| \\ &\geq (1-\eta)\frac{\eta^3}{8}|A||B||C| \end{aligned}$$

∎

1.4 Δ-free tester

Theorem 5 *$\forall \varepsilon$, $\exists \delta$ $(= \delta(\varepsilon))$ s.t $\forall$ G s.t $|V| = n$ and G is ε-far from Δ-free in adjacency matrix model, then G has at least $\delta \binom{n}{3}$ distinct triangles.*

Corollary 6 *1 sided tester for Δ-free*

Do $O(\delta^{-1}(\varepsilon))$ times

1. *Pick v_1, v_2, v_3.*
2. *If (v_1, v_2, v_3) is a triangle, FAIL.*

Proof Sketch proof for the theorem

Start with an arbitrary $\frac{5}{3}$-equipartition, use regularity lemma to get equipartition $\{V_1, V_2, ..., V_{\mathcal{K}}\}$ s.t

$$\tfrac{5}{3} \leq \mathcal{K} \leq \mathcal{T}$$

use $\varepsilon' = min(\frac{\varepsilon}{5},\, \delta^{\Delta}(\frac{\varepsilon}{5}))$
Let $G' \leftarrow G$

1. Delete all edges intenal to partitions. There are at most $n\frac{\varepsilon n}{5} \leq \frac{\varepsilon n^2}{5}$ such edges .
2. Delete all non-regular pair edges. There are at most $\varepsilon' \binom{\mathcal{K}}{2} \left(\frac{n}{\mathcal{K}}\right)^2 \leq \frac{\varepsilon n^2}{10}$ such edges.
3. Delete edges between low density pair. There are at most $\frac{1}{5}\varepsilon \binom{n}{2} \leq \frac{1}{10}\varepsilon n^2$

The total number of removed edges is at most $\frac{2\varepsilon n^2}{5}$, at least $\Omega(\varepsilon n^2)$ violating edges left in G'. There must be at least a triangle left in G' whose edges are between pairwise-regular high densities sets. Therefore, there must be much more other triangles remaining (and will be dectected by the tester). ■