6.896 Sublinear Time Algorithms April 12, 2007

Lecture 18

Lecturer: Ronitt Rubinfeld *Scribe: Ivaylo Riskov*

1 Recap

Theorem 1 *1-sided error testing for triangle-free property requires* $\Omega((\frac{c}{\epsilon})^{c \log \frac{c}{\epsilon}})$.

Theorem 2 *Let P be any graph property in the adjacency matrix model. Let T be a tester for P with query complexity $q(n, \epsilon)$. Then there is a tester T' for P, such that T' has form:*

1. *Selects random subset of $2q(n, \epsilon)$ nodes.*
2. *Queries all pairs.*
3. *Outputs decision.*

If T has 1-sided error, so does T'.

Theorem 3 $\forall m \exists X \subset M = \{1, \ldots, m\}$ *of size* $|X| \geq \frac{m}{e^{10\sqrt{\log m}}}$ *with no nontrivial solutions to* $x_1 + x_2 = x_3$.

2 Construction of a graph that is hard to test

In this section we present two attempts to construct a graph that is far from triangle-free, but is also hard to test.

2.1 First attempt

Consider the following sets of vertices $V_1 = \{1 \ldots m\}, V_2 = \{1 \ldots 2m\}, V_3 = \{1 \ldots 3m\}$. The edges are defined as follows:

- For every $j \in V_1$ put an edge $(j, j + x)$ from V_1 to V_2, $\forall x \in X$. i.e. this is the set of edges $\{(j, j + x_1), (j, j + x_2), \ldots\}$.
- For every $l \in V_2$ put an edge $(l, l + x)$ from V_1 to V_3, $\forall x \in X$.
- For every $k \in V_2$ put an edge $(k, k + 2x)$ from V_2 to V_3, $\forall x \in X$.

The number of nodes is $6m$ and the number of vertices is $\Theta(m.|X|) = \Theta\left(\frac{m^2}{e^{10\sqrt{\log m}}}\right)$.
By construction there are triangles of the form:

$$j, j + x, j + 2x, j \in V_1, j + x \in V_2, j + 2x \in V_3, \forall j \in \{1, \ldots, m\} x \in X$$

The number of these intended triangles is $\Theta\left(\frac{m}{e^{10\sqrt{\log m}}}\right)$.

Claim 4 *There are no other triangles*

Consider a triangle with edges $(j, j + x_1), (j + x_1, j + x_1 + x_2), (j, j + 2x_3)$. Then $x_1 + x_2 = 2x_3$ and from the lemma this triangle is intended.■

Claim 5 *All triangles are edge disjoint*

Consider the following 2 triangles $j, j+x_1, j+x_1+y$; and $k, k+x_2, k+x_2, k+x_2+y$ that share the common edge with endpoints $j+x_1 = k+x_2$ and $j+x_1+y = k+x_2+y$. Since these triangles are intended $x_1 = y$ and $x_2 = y$ and so $x_1 = x_2$ and $j = k$, i.e. this is the same triangle.$\blacksquare$

When all triangles are edge disjoint, we have to remove one edge from each triangle to get to triangle-free. In other words, we have lots of triangles, but we are also far from triangle-free. Recall that the algorithm is a 1-sided tester and the only way it could fail is to find a triangle so we want to show that it is hard to find such triangle. However, this version of the argument does not work because of the $e^{10\sqrt{\log m}}$ factor in the denominator. We need to bound the number of edges by a constant.

2.2 Second(Final) attempt

Definition 6 (s-blowup) *An s-blowup of graph G is the graph $G^{(s)}$ for which a vertex in G goes to an independent set of size s in the new graph. Edges in the old graph go to a complete bipartite subgraph.*

For $G^{(s)}$ we have $|V^{(s)}| = |V|\dot{s} = ms$. The number of edges in the s-blowup is $\approx ms^2|X|$. The number of triangles is $\approx ms^3|X|$, because for each starting node $[1 \dots s]$, there are s ways to pick an outgoing edge to another independent set vertex and from that node there are again s ways to go to the third vertex.

Finally, the number of edge disjoint triangles is at least $ms^2|X|$. This is in the optional homework and is easy to prove.

Given ϵ, pick m to be the largest integer satisfying $\epsilon \leq \frac{1}{3^8 e^{10\sqrt{\log m}}}$. This value of m satisfies $m \leq \left(\frac{c}{\epsilon}\right)^{c \log \frac{c}{\epsilon}}$.

Also pick $s = \frac{n}{6m}$ so that we get n nodes in the blowup graph. The number of edges is ϵn^2 and the number of triangles is at most $\left(\frac{\epsilon}{c}\right)^{c \log \frac{c}{\epsilon}}$.

Finally, if we take sample of size q, the expected number of triangles in the sample is:

$$E[\#triangles] \leq \binom{q}{3} \left(\frac{\epsilon}{c}\right)^{c \log \frac{c}{\epsilon}} << 1$$

By Markov's inequality this is also the probability that we see a triangle. This probability is very low unless $q > \left(\frac{c'}{\epsilon}\right)^{c' \log \frac{c'}{\epsilon}}$, for some constant c'. In order to output pass we need this many queries, otherwise we reject a graph that is far from triangle-free with high probability.

3 Graph Isomorphism

We now turn to the problem of testing graph isomorphism, but first we need to define the problem and a distance metric.

Definition 7 *Graphs G and H are isomorphic if there exists a permutation $\delta : V_G \to V_H$, such that it preserves the edge relationship, i.e. $(x, y) \in E_G \Leftrightarrow (\delta(x), \delta(y)) \in E_H$.*

Definition 8 *For permutation $\delta : V_G \to V_H$, the distance between G and H* under δ, *d_δ is the symmetric distance between the adjacent matrices of $\delta(G)$ and H divided by $\binom{n}{2}$.*

Definition 9

$$d(G, H) = \min_{\delta} d_\delta(G, H)$$

The following variations of the problem exist:

1. One of the graph is known.

2. Both graphs are known.

3. Both graphs are unknown.

We will assume that G is known and H is unknown.

Graph isomorphism is not a Hereditary property, so it is not testable in constant time. The following table summarizes the best known results:

	H known	H unknown
1-sided error	$\tilde{\Theta}(n)$	$\tilde{\Theta}(n^{3/2})$
2-sided error	$\tilde{\Theta}(\sqrt{n})$	Not really known. At least $\Omega(n)$, probably $\tilde{\Theta}(n^{5/4})$

We give a proof sketch of the lower bound for a 2-sided tester, when H is known.

Sketch of Proof We show graphs for which $\Omega(\sqrt{n})$ queries are required . Next time we will sketch the proof that if we allow $\tilde{\Theta}(\sqrt{n})$ queries in general, it is possible to test graph isomorphism when one of the graphs is known.

Let G is a random graph with n nodes and each edge is inserted with probability $1/2$ i.i.d. Let H be either G or a graph close to $G^{(W)}$, i.e. a graph is close if its nodes are renamed according to a permutation π.

Definition 10 **($G^{(W)}$)** *For G and $w \subseteq V_G$, such that $|W| = \frac{|V_G|}{2}$*

- $V(G^{(w)}) \leftarrow W \cup \{w' | w \in W\}$
- $E(G^{(w)}) \leftarrow \{(u,v),(u,v'),(u',v),(u',v') | (u,v) \in E(G),(u,v) \in W\}$

In other words we consider the restriction of G to W and then just put another copy of it.

Lemma 11 *$\forall W \subseteq V_G$ such that $|W| = \frac{|V_G|}{2}$ with probability at least $1 - o(1)$ G and $G^{(w)}$ are $1/8$-far. i.e. they differ by at least $1/8\binom{n}{2}$ edges under symmetric distance.*

We are not going to prove this lemma.

We can show that if no u, u' pair is seen, the distribution on query answers is identical in the $G, H = \pi_R(G)$ case and the $G, H = \pi_R(G^{(W)})$ case. (π_R is a random permutation)

Finally, for every set of $o(\sqrt{n})$ queries it is unlikely that a u, u' pair is queried. ■