6.896 Sublinear Time Algorithms	April 19, 2007

Lecture 20

Lecturer: Ronitt Rubinfeld *Scribe: Erdong Chen*

1 Graph Isomorphism $\cong$

We talked about a lower bound last time. Recall from last lecture (all are based on adjacent matrix model):

Thm	H known	H unknown
1-sided error	$\tilde{\Theta}(n)$	$\tilde{\Theta}(n^{\frac{3}{2}})$
2-sided error	$\tilde{\Theta}(\sqrt{n})$	$\Omega(n), \tilde{O}(n^{\frac{5}{4}})$

Difference between the case that H is known and H is unknown: when H is known, query complexity to H is zero. We will discuss two upper bounds in today's class.

1.1 1-Sided Tester

In this subsection, we will discuss the upper bound of 1-sided testers when H is known.

Lemma 1 *When H is known, there is a 1-sided tester with $\tilde{\Theta}(n)$ queries.*

Proof

Algorithm:

1 $Q \leftarrow$ pick each edge slot with probablity $\frac{\ln n}{\varepsilon n}$ i.i.d. (no queries yet!)

2 if $|Q| \le \frac{10 n \ln n}{\varepsilon}$, ACCEPT, else ask queries in Q of G.

3 if there is a isomorphism from G to H consistent with Q, ACCEPT

runtime: $O(\frac{n! n \ln n}{\varepsilon})$ (because $n!$ different permutations exist)
query complexity:$O(\frac{n \ln n}{\varepsilon})$
Behavior:

- If $G \cong H$, algorithm always accepts.
- IF G is ε-far from isomorphic to H

$$\forall \text{ permuation } \pi, d(\pi(G), H) = \frac{\text{symmetric distance of adjacent matrix of } \pi(G) and H}{n^2} \ge \varepsilon$$

Thus, Pr[accept in step 2] is $o(1)$ (smaller than any constant).

By Chernoff bounds, we will have:

$$E[|Q|] = \frac{n^2 \ln n}{\varepsilon n} = \frac{n \ln n}{\varepsilon}$$

If reach Step 3:

$\forall$ permuation $\pi : V_G \to V_H$
$\exists E_\pi s.t. |E_\pi| \ge \varepsilon n^2 \& \forall (u,v) \in E_\pi, (u,v) \in E_{\pi(G)} \triangle E_H$,

where $\triangle$ is the operator of symmetric difference, and (u,v) is denoted as a witness against π.

$$\forall (u,v) \in E_\pi, \Pr[(u,v) not \in Q] = 1 - \frac{\ln n}{\varepsilon n}$$

$$\Pr[\forall (u,v) \in E_\pi, (u,v) not \in Q] = (1 - \frac{\ln n}{\varepsilon n})^{\varepsilon n^2}$$

By union bound,

$$\Pr[\exists \pi \text{ s.t. } \forall (u,v) \in E_\pi, (u,v) not \in Q] \leq n!(1 - \frac{\ln n}{\varepsilon n})^{\varepsilon n^2}$$

$$\leq n!(e^{-\frac{\ln n}{\varepsilon n}})^{\varepsilon n^2} \leq n!\frac{1}{n^n} = o(1)$$

Thus, Pr[accept] is smaller than any constant.

By the analysis above, the algorithm with $\tilde{\Theta}(n)$ queries is correct. ■

1.2 2-Sided Tester

In this subsection, we will discuss the upper bound on 2-sided testers when H is known. We will discuss this based on a special case, and please refer the paper for full proof.

Lemma 2 *When H is known, a 2-sided tester needs only $\tilde{O}(\sqrt{n})$ queries.*

Very special case: Assume a graph H is s.t. $\forall u, v, d_{u,v} \equiv \frac{1}{n}|\{N(u) \vartriangle N(v)\}| > \epsilon/4$, where $\vartriangle$ is the operator of symmetric difference, $N(v)$ is the set of all neighbors of v .

Definition 3 *For any $C \subsetneqq V$, define C-labeling of any $v \in V$: set of neighbors $N(v) \cap C$ is represented by a string of length $|C|$, while the string is denoted by $label_c(v)$.*

Note: There are at most $2^{|C|}$ possible labels. Note that an n-node graph can have at most n distinct labels.

In our special case, since we have $\forall u, v, d_{u,v} > \frac{\varepsilon}{4}$, if we pick $c = \log n^2$ random vertices (call it a "core set"), then

$\forall u, v, \Pr[\text{label}_c(u) = \text{label}_c(v)] \leq (1 - \frac{\varepsilon}{4})^{\log n^2} = n^{-\frac{\varepsilon}{4}\log n}$

and thus,$\Pr[\forall u, v, \text{label}_c(u) = \text{label}_c(v)] \leq (1 - \frac{\varepsilon}{4})^{\log n^2} = n^2 n^{-\frac{\varepsilon}{4}\log n} = o(1)$

Idea of Algorithm:

1(a) pick $C(= \log n^2)$ nodes in unknown graph

1(b) query all edges in C's subgraph

2 Pick $n^{\frac{1}{8}}$ (much less for the special case) nodes in G, and find all the C-labels.

3 Pair up $n^{\frac{1}{8}}$ nodes, and then query edges of each pair in G.

4 For each C', and for each mapping $\pi_{C,C'}$

let $\tilde{\pi} \leftarrow$ permutation from G to H consistent with $\pi_{C,C'}$. (We will prove that only one such mapping exists later)

$\forall$ pairs (u,v) in Step 3, if $\exists$ edge between exactly one of $(u,v) \in \left\{\pi(\tilde{u}), \pi(\tilde{v})\right\}$, reject $\tilde{\pi}$ and $\pi_{C,C'}$. Such pair (u,v) is called a witness against $\tilde{\pi}$.

Claim: In our special case, $\tilde{\pi}$ is unique if it exists at all(maybe no way to map from C to C')

- If $\text{label}_C(X) = \text{label}_{\pi_{C,C'}}(X)$, then $\tilde{\pi}$ maps x to y.
- If no such y, then $G \not\cong H$.
- if there is more than one such y, then $G \not\cong H$ or we were unlucky with choice of labels.

Behavior:
if $G \cong H$, $\tilde{\pi}$ exists that no witness is found
if G is ε-far from H,
$\forall \tilde{\pi}$ in Step 4, $\Pr[\tilde{\pi}$ passes Step 4$] \leq (1-\varepsilon)^{n^{\frac{1}{8}}}$ (each edge passes with probability $1-\varepsilon$)

$$\leq e^{-\varepsilon n^{\frac{1}{8}}} = o(\alpha^{-\log n^3})$$

Thus, $\Pr[$any $\tilde{\pi}$ passes$] \leq O(n^{\log n^2} \alpha^{-\log n^3}) = O(\alpha^{\log n^2} \alpha^{-\log n^3}) \leq \frac{1}{4}$.