6.896 Sublinear Time Algorithms March 20, 2007

Lecture 13

Lecturer: Ronitt Rubinfeld *Scribe: Huy Ngoc Nguyen*

1 Bipartiteness testing on dense graph

Definition 1 *G is "$\epsilon - far$" from bipartiteness if must remove at least ϵn^2 edges to make it bipartite.*

Definition 2 *(equivalent to definition 1) $\forall$ partitions (V_1, V_2) of V there are at least ϵn^2 edges (w, v) s.t either $u, v \in V_1$ or $u, v \in V_2$ (in this case (u, v) is a* **violating pair***).*

Algorithm (1): :
TestBipartite(A,G) *(A - adjacency matrix for G)*

1 choose sample S uniformly s.t $|S| = \Theta(\frac{1}{\epsilon^2}\log\frac{1}{\epsilon})$
2 query $A(u, v)$ $\forall u, v \in S$.
3 FAIL if subgraph not bipartite else PASS.

Note: total number of queries : $O(\frac{1}{\varepsilon^4}\log^2\frac{1}{\varepsilon})$

Theorem 1 *The algorithm is a property tester for bipartiteness. More precisely,*

(1) if G is bipartite, TestBipartite PASSES.

(2) if G is $\varepsilon - far$ from bipartiteness, Pr [TestBipartite FAILS] $\geq \frac{3}{4}$.

Proof Consider the following algorithm,
Algorithm (2):

1 Choose U uniformly, s.t $|U| = O\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)$
2 Choose a bunch of pairs $(u_i, v_i) \to W$ $(|W| = m = O\left(\frac{1}{\varepsilon^2}\log\frac{1}{\varepsilon}\right))$
3 Query (u, v) $\forall u, v \in U$
4 Query $(u_i, v_i) \in W$
5 Query all edges (w, u_i), (w, v_i) for $w \in U$ and $(u_i, v_i) \in W$.
6 FAIL if no bipartite partition possible.

Since $|U|$ and $|W|$ is $O\left(\frac{1}{\varepsilon^2}\log\frac{1}{\varepsilon}\right)$, we can consider U and W in this algorithm as subsets of S in algorithm (1). Therefore, to prove algorithm (1), it is sufficient to prove that the algorithm (2) is a property tester for bipartiteness.

Firstly, consider a special case when all nodes in V connected to U, U is "bad" if there is no partition of U into (U_1, U_2) that induces a consistent partition of (V_1, V_2) in V. It will be shown that the algorithm will output FAIL with high probability when U is "'bad".

Consider a bipartition (U_1, U_2) of U, a bipartition of V can be induced from (U_1, U_2) by the following algorithm
Bipartition Algorithm:

1 Put U_1 into V_1
2 Put U_2 into V_2
3 $\forall v \in V \setminus (U_1 \cup U_2)$:

- if v is adjacent to a node in U_1 place v in V_2
- if v is adjacent to a node in U_2 place v in V_1
- if placed in both, choose arbitrarily.

Note: Every node get placed in time $O(|U|)$. Therefore, given any edge (u, v), we can detect if it violates (U_1, U_2) (i.e showing that (U_1, U_2) not leading to bipartition) in $O(|U|)$ time.

Definition 3 *W is "not compatible" with (U_1, U_2) if cant partition nodes in W (touched by edges in W) into $\tilde{W} = (\tilde{W_1}, \tilde{W_2})$ s.t $((\tilde{W_1} \cup U_1), (\tilde{W_2} \cup U_2))$ is a bipartition*

$\Rightarrow$ Algorithm (1), (2) reject if there is no compatible W for any bipartition (U_1, U_2) of U.

Definition 4 .

(1) A node w is "witness" against (U_1, U_2) if $\exists u_1 \in U_1, u_2 \in U_2$ s.t $(w, u1)$ and $(w, u_2) \in E$.

(2) An edge (w_1, w_2) is "witness" against (U_1, U_2) if $\exists u_1, u_2$ s.t :

- *$u_1, u_2 \in U_1$ or $u_1, u_2 \in U_2$.*
- *$(w, u_1), (w, u_2) \in E$.*

Therefore, $\forall$ bipartition (U_1, U_2), if there are at least εn^2 violating edges, each gives witness against (U_1, U_2):

$$Pr\left[\text{ catch a pair in } W \text{ witness against } (U_1, U_2)\right] \geq \varepsilon.$$

Thus,

$$\begin{aligned} Pr\left[(U_1, U_2) \text{ survives } W\right] &\leq (1-\varepsilon)^{|W|} \\ &\leq \frac{1}{8} 2^{-|W|} \end{aligned}$$

Since $|W| = \Theta(\frac{|U|}{\varepsilon})$

$$\begin{aligned} Pr\left[\text{ any partition of } U \text{ survives } W\right] &\leq \frac{1}{8} 2^{-|U|} 2^{|U|} \\ &\leq \frac{1}{8} \end{aligned}$$

Although the argument above based on the assumption that U is connected to all nodes in V, it is still valid if U is adjacent to "enough" $(\Omega(\varepsilon n^2))$ violating edges . In the remaining of this lecture, we will argue that with high probability when $|U| = \Omega(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon})$, U is adjacent to at least $\frac{\varepsilon n^2}{2}$ violating edges.

Definition 5 *$v \in V$ is a "high degree" node if $deg(v) \geq \frac{\varepsilon}{4} n$ else v is "low degree".*

Lemma 2 *By choosing $|U| = \frac{4}{\varepsilon} \log \frac{32}{\varepsilon}$,*

$$Pr_U\left[\text{ at most } \tfrac{\varepsilon}{4} n \text{ "high degree" in } V \text{ dont have enough neighbors in } U\right] \geq \tfrac{7}{8}$$

Proof For each "high degree" v node in V,

$$\begin{aligned} Pr\left[\text{v has no nbr in U}\right] &\leq (1 - \frac{\varepsilon n}{4})^{|U|} \\ &\approx \frac{1}{e}^{\log \frac{32}{\varepsilon}} \\ &= \frac{\varepsilon}{32} \end{aligned}$$

Thus,

$$E\left[\text{number of "high degree" nodes in V that not adjacent to U }\right] \le \frac{\varepsilon n}{32}$$

By Markov inequality,

$$Pr\left[\text{number of "high degree" nodes in V that not adjacent to U} \ge \frac{\varepsilon n}{4}\right] \le \frac{1}{8}$$

■

How many edges adjacent either to low degree node or high degree node not adjacent to U?

(1) adjacent to low degree node $\le n\frac{\varepsilon}{4}n$

(2) adjacent to high degree non-U-neighbor $\le \frac{\varepsilon n}{4}n$

$\Rightarrow$ total $\le \frac{\varepsilon}{2}n^2$ edges!
$\Rightarrow$ at least $\frac{\varepsilon}{2}n^2$ violating edges adjacent to U. ■