6.896 Sublinear Time Algorithms	September 28, 2004
	Lecture 5
Lecturer: Ronitt Rubinfeld	*Scribe: Ben Wagner*

1 From last lecture

In the last lecture we began to analyze a sublinear-time property tester for bipartiteness. We will refer to several things explained in the last lecture, summarized here.

We defined ϵ-far from bipartite by two equivalent definitions. G is ϵ-far from bipartite if we need to remove $\geq \epsilon n^2$ edges to make G bipartite. Equivalently, for all partitions (V_1, V_2) of V, $|\{(u,v) : u, v \in V_1 \text{ or } u, v \in V_2\}| \geq \epsilon n^2$.

We saw an algorithm for Test-Bipartite that queries all edges in a small sample of random nodes, and checks if the sample is bipartite. The algorithm passes if the sample is bipartite, and fails otherwise.

It is clear that if the graph is bipartite, this algorithm will output Pass. We want to show that if the graph is ϵ-far from bipartite, the algorithm outputs Fail with probability at least 2/3.

2 Overview

We ended the last lecture with a summary of the proof idea. Here it is again in more detail.
Proof Idea Our proof idea consists of four parts.

1. We view the sample as two parts, U and S.
$$|U| = m_1 = \Theta\left(\frac{\log \frac{1}{\epsilon}}{\epsilon}\right)$$
$$|S| = m_2 = \Theta\left(\frac{\log \frac{1}{\epsilon}}{\epsilon^2}\right)$$
(Note that $m_1 < m$.)

2. For all partitions of U, consider the partition induced on the rest of the graph.

3. If the graph is ϵ-far from bipartite, the many violating edges will cause S to contain a violating edge.

4. The violating edge will cause the algorithm to fail.

■

Why is it that we look only at partitions of U? This makes the combinatorics easier—if we consider partitions of U, we get 2^{m_1} partitions, while there are 2^n partitions of G. The intuition is that we can rule out many potential partitions of G by looking at just a few nodes.

3 Influential vertices

Definition 1 (Influential) *We say a vertex v is influential if the degree of v in G is $\geq \frac{\epsilon}{4}n$.*

(Notice that there are $\leq n$ noninfluentials, each with degree $\leq \frac{\epsilon}{4}n$, so the noninfluentials can contribute $\leq \frac{\epsilon}{4}n^2$ violating edges to *any* partition. Thus there are $\geq \frac{3\epsilon}{4}n^2$ violating edges from influential nodes.)

Definition 2 (Covers) *$\forall v \in V, U \subseteq V$, U covers v if v has at least one neighbor in U.*

Lemma 1 *With probability $\geq \frac{5}{6}$ (over choice of U), there are $\leq \frac{\epsilon}{4}n$ influential vertices that are not covered by U.*

Proof

$$\forall v \text{ influential}, \Pr[v \text{ not covered by } U] \leq \left(1 - \frac{\epsilon}{4}\right)^{m_1} = \left(1 - \frac{\epsilon}{4}\right)^{\frac{c \log 1/\epsilon}{\epsilon}} < \frac{\epsilon}{24}$$

by picking c large enough. (Note that the $\frac{\epsilon}{4}$ comes from $\frac{\epsilon}{4}n \cdot \frac{1}{n}$.

We define an indicator random variable σ_v as follows:

$$\sigma_v = \begin{cases} 1 & \text{if } v \text{ not covered by } U \\ 0 & \text{if } v \text{ is covered by } U \end{cases}$$

Thus, the number of not-covered influentials is

$$\sum_{v \text{ influential}} \sigma_v$$

and the expectation of the number of not-covered influentials is

$$\mathrm{E}\left[\sum_{v \text{ influential}} \sigma_v\right] = \sum_{v \text{ influential}} \mathrm{E}[\sigma_v] < \frac{\epsilon}{24}n$$

(since $\Pr[\sigma_v = 1] = \mathrm{E}[\sigma_v]$).

Therefore, by applying Markov's inequality, we find that

$$\Pr\left[> \frac{\epsilon}{4}n \text{ not-covered influentials}\right] = \Pr\left[\text{number of not-covered influentials } \geq 6 \cdot \text{expectation}\right] \leq \frac{1}{6}$$

■

4 Graph partitions

Let C be the set of vertices covered by U, and define R as

$$R = V \setminus (C \cup U)$$

We define I as the set of influentials. Note the following:

- There are no edges between U and R.
- I could have vertices from any of U, C, and R.
- C does not contain any vertices without edges to U.
- R does not contain many influentials, which leads to our definition of a good sample.

Definition 3 (Good) *We say U is good if there are $\leq \frac{\epsilon}{4}n$ influential vertices that are not covered by U.*

We will assume for the following that U is good.

A partition (U_1, U_2) of U induces a partition (C_1, C_2) of C:

$$C_2 = \{v \in V : v \text{ adjacent to any node in } U_1\}$$

$$C_1 = C \setminus C_2$$

Thus, there are no edges between U_1 and C_1. There could be edges between C_1 and C_2, or edges between U_2 and C_2. Now we can explicitly define what we mean by a violating edge.

Definition 4 (Violating edge) *$\forall$ partitions (U_1, U_2), this partition induces a partition (C_1, C_2) of C and a partition (R_1, R_2) of R. An edge is violating if it connects $u, v \in U_1 \cup C_1 \cup R_1$ or connects $u, v \in U_2 \cup C_2 \cup R_2$.*

We now count the number of violating edges incident to R without regard to the particular partition induced in R. We know that there are $\leq \frac{\epsilon}{4}n$ influential vertices in R, all with degree $\leq n$. In addition, there are $\leq n$ non-influential vertices in R, all with degree $\leq \frac{\epsilon}{4}n$. Thus the total number of edges incident to R is

$$\begin{array}{rl} \text{influentials} & \frac{\epsilon}{4}n \cdot n \\ + \text{ non-influentials} & n \cdot \frac{\epsilon}{4}n \\ \hline & \frac{\epsilon}{2}n^2 \end{array}$$

with respect to *any* partition (R_1, R_2). The number of violating edges incident to U is $\leq m_1 n = \frac{c \log \frac{1}{\epsilon}}{\epsilon} n = o(n^2)$. Thus the number of violating edges incident to $U \cup R$ is $\leq \frac{2\epsilon}{3}n^2$. This implies that the number of violating edges in C_1 or C_2 is $\geq \frac{\epsilon}{3}n^2$, since there are at least ϵn^2 violating edges in G.

5 Violating edges in sample

Now we look at the other part of our sample, S, and show that S contains a violating edge, and that edge is in C.

Lemma 2 *With probability $\geq 1 - \frac{2^{-|U|}}{6}$ over choice of S, for all partitions (S_1, S_2) of S, there is an edge between vertices in $U \cup S$ that violates the partition $(U_1 \cup S_1, U_2 \cup S_2)$.*

We will complete the proof in the next lecture.