Sublinear Algorithms	October 6, 2015

Lecture 11

Lecturer: Sofya Raskhodnikova *Scribe(s): Tao-yang Fu*

1 Testing Bipartiteness (continued)

1.1 GGR Algorithm

Algorithm 1: GGR Algorithm.

input : ϵ, access to adjacency matrix of a graph $G = (V, E)$, $n = |V|$
output: accept or reject

1 Pick uniformly and independently at random a sample set S of size $\theta(\frac{1}{\epsilon^2} \times log\frac{1}{\epsilon})$.
2 Query subgraph G induced by S.
3 If G' is bipartite, accept; otherwise, reject.

1.2 Analysis

Break S into two sets:

1. Learning set L of size $\theta(\frac{1}{\epsilon} \times log\frac{1}{\epsilon})$.
2. Testing set T of size $\theta(\frac{1}{\epsilon^2} \times log\frac{1}{\epsilon})$.

Definition 1. *A node v in V is covered by L if v has a neighbor in L.*

Let C be the set of nodes covered by L. To separate L to L_1 and L_2, and C_1, C_2 are nodes adjacent by nodes of L_1, L_2 respectively. $C = C_1 \bigcup C_2$. Figure. 1 shows the relation between L and C.

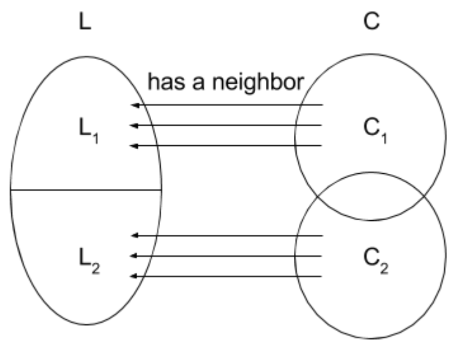

Figure 1: Learning set L and the set of nodes covered by L

The idea is that to try to catch a violating edge (w.r.t. every partition of L) with both endpoints in C.

Definition 2. *A node v is influential if its degree $deg(v) \geq \frac{\epsilon n}{8}$*

Claim 3. *$Pr[> \frac{\epsilon n}{8}$ influential nodes are not covered by $L] \leq \frac{1}{6}$*

Table 1: Upperbounds or Lowerbounds of sets of nodes

Violating edges incident to	# of nodes	degree	# of violating edges
influential nodes in R	$\leq \frac{\epsilon n}{8}$	$\leq n$	$\leq \frac{\epsilon n^2}{8}$
non-influential nodes in R	$\leq n$	$\leq \frac{\epsilon n}{8}$	$\leq \frac{\epsilon n^2}{8}$
nodes in L	$\theta(\frac{1}{\epsilon} \times log\frac{1}{\epsilon})$	$\leq n$	$O(n)$ $\leq \frac{\epsilon n^2}{8}$
nodes in C			$\geq \frac{\epsilon n^2}{2} - \frac{3\epsilon n^2}{8}$ $\geq \frac{\epsilon n^2}{8}$

Claim 4. *Let BAD_1 be the event that $> \frac{\epsilon n}{8}$ influential nodes are not covered by L. If BAD_1 does not happen then every partition of L induces $\geq \frac{\epsilon n^2}{8}$ violating edges with both end-points in C*

Proof. By observation, w.r.t. every partition, there are at least $\geq \frac{\epsilon n^2}{8}$ violating edges. □

The upperbounds or lowerbounds of each set of nodes are shown in Table. 1

Fix a partition of L. If defines C_1 and C_2. View samples from T as pairs $(v_1,\, v_2)$, $(v_3,\, v_4)$, , $(v_{\frac{|T|}{2}-1},$ $V_{\frac{|T|}{2}})$.

Assuming BAD_1 does not happen:

$$
\begin{aligned}
Pr[\text{ no pairs } (v_i,\, v_{i+1}) \text{ are violating edges with both endpoints in } C] &\leq (1 - \tfrac{\epsilon}{8})^{\frac{|T|}{2}} \\
&\leq e^{-\frac{\epsilon}{8} \times \frac{|T|}{2}} \\
&\leq e^{-\frac{\epsilon}{16} \times \frac{|T|}{|L|} \times |L|} \\
&\leq \frac{2^{-|L|}}{6}
\end{aligned}
$$

By a union bound over all partition of L, since there are $2^{|L|}$ partitions,

$$Pr[BAD_2] \leq 2^{|L|} \times \frac{2^{-|L|}}{6},$$

where BAD_2 = there exists a partition of L with no violating edge sampled.

$$\leq \frac{1}{6}$$

G is ϵ-far,

$$
\begin{aligned}
Pr[G \text{ is accepted }] &\leq Pr[BAD_1] + Pr[BAD_2|\overline{BAD_1}] \times Pr[\overline{BAD_1}] \\
&\leq \tfrac{1}{6} \times \tfrac{1}{6}
\end{aligned}
$$