6.842 Randomness and Computation	February 26, 2014

Lecture 7

Lecturer: Ronitt Rubinfeld — *Scribe: Payut Pantawongdecha*

Topics

- More random walks
- Cover time for graphs
- UST-Conn $\in$ RL

1 Review

In the previous lecture, we talked about theorems related to Markov chains and cover time for graphs. Here is a review about topics from the last class.

Given initial distribution $\Pi^0 = (\Pi_1^0, \Pi_2^0, \ldots, \Pi_n^0)$, where $\Pi_i^0 = \Pr[\text{start at node } i]$, we have t-step distribution being just a matrix multiplication $\Pi^t = \Pi^0 P^t$.

Definition 1 *An* **ergordic** *Markov chain is a Markov chain with transition matrix P such that $\exists t_0 \forall t > t_0, \forall x, y, P^t(x, y) > 0$.*

Theorem 1 *If a Markov chain M is ergodic, then M has a unique stationary distribution.*

Theorem 2 *The stationary distribution of any doubly stochastic and ergodic Markov chain is uniform.*

Definition 2 *The* **hitting time** *h_{ij} of a graph G is the expected number of steps for a random walk on G that starts at node i and reaches node j.*

Definition 3 *The* **return time** *h_{ii} of a graph G is the expected number of steps for a random walk on G that starts at node i and returns to node i.*

Theorem 3 *If a random walk is ergodic, then $h_{ii} = \frac{1}{\Pi_i}$, where Π is the stationary distribution.*

Definition 4 *The* **cover time** *$C_u(G)$ of a graph G at node u is the expected number of steps of a random walk that starts at u and hits every node in G. $C(G)$ is the maximum cover time, i.e., $C(G) = \max_{u \in V} C_u(G)$.*

Definition 5 *The* **commute time** *c_{ij} is the expected number of steps of a random walk that starts at i, hits j and comes back to i.*

Fact 1 $c_{ij} = h_{ij} + h_{ji}$.

Theorem 4 *Let G be a connected graph with m edges and n nodes, then $C(G)$ is $O(mn)$.*

Lemma 1 For any edge $(u, v) \in E$, $c_{uv} \leq O(m)$.

We have already shown that the above lemma implies Theorem 4. In this lecture, we will prove the lemma.

2 Bound for cover time

We will start by proving the following lemma.

Lemma 1 For any edge $(u, v) \in E$, $c_{uv} \leq O(m)$.

Proof. The key idea of this proof is that instead of looking at how we traverse the nodes, we look at how we traverse the edges. If we traverse edge (u, v) twice, then we must have traveled from u to v and back to u, hence a commute. Therefore, if we are able to show that the expected number of steps between visits to (u, v) is $O(m)$, then c_{uv} must also be $O(m)$.

First, create an undirected graph G' from G by adding $d(u)$ self-loops to each node u. Each node in G' stays in place with probability $\frac{1}{2}$ and walk according to G with probability $\frac{1}{2}$. Next, we claim that $C(G') = 2C(G)$, because we can transform any path in G' to a path in G by removing self-loops and the expected number of self-loops is exactly half of the length of the path.

Next, we construct a directed graph G'' that represents walks on directed edges of G'. The set of vertices V'' is the set of edges E' in both directions in G' except self-loops. For example, edge $(1, 2)$ in E' becomes vertices $(1, 2)$ and $(2, 1)$ in G''. The set of edges E'' is the set of paths of length two in G'.

Here is the example of G, G' and G''. Suppose P is the transition matrix of G and

$$P = \begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 0 & 1 \\ 2 & 1 & 0 \end{array}$$

Then we have

$$P' = \begin{array}{c|cc} & 1 & 2 \\ \hline 1 & \frac{1}{2} & \frac{1}{2} \\ 2 & \frac{1}{2} & \frac{1}{2} \end{array}$$

and

$$P'' = \begin{array}{c|cccc} & (1,1) & (1,2) & (2,1) & (2,2) \\ \hline (1,1) & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ (1,2) & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ (2,1) & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ (2,2) & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{array}$$

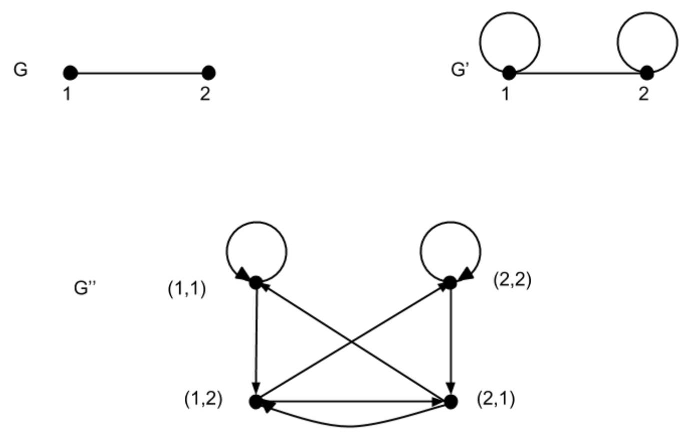

Figure 1: Illustration of how to construct G' and G'' from a graph G.

Now we will show that G'' must be doubly stochastic. Let P' be the transition matrix for G'. Consider the transition matrix Q of G'', we have

$$Q_{(u,v)(z,w)} = \begin{cases} P'_{vw} & \text{if } z = v \text{ and } (u,v),(v,w) \in E' \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $\forall (v,w) \in E'$, the column sum of Q is

$$\sum_{\substack{(u,v) \text{ s.t.} \\ (u,v)(v,w) \in E'}} Q_{(u,v)(v,w)} = \sum_{(u,v)\in E'} \frac{1}{d(v)} = 1.$$

We can conclude that G'' is doubly stochastic. Also, note that G'' is ergodic, and we will outline its proof here: G'' is ergodic if and only if G'' is irreducible and aperiodic from a theorem last time. G'' is irreducible because G' is irreducible. G'' is aperiodic because of self-loops in G'. These two properties of G'' imply that the stationary distribution $\Pi(G'')$ is uniform from theorem 2.

Since there are at most $4m$ nodes in G'', we can conclude that

$$\Pi_{u \in G''} G'' \geq \frac{1}{4m}.$$

Therefore, for any node (u,v) in G'', we have $h_{(u,v)(u,v)} \leq 4m$ from theorem 3, which concludes the proof. □

3 The Undirected Connectivity Problem (UST-Conn)

We can use the properties of random walk that we just showed to check whether an undirected graph is connected. In this section, we will consider the uniredted connectivity problem.

Definition 6 *The* **undirected connectivity problem (UST-Conn)** *is to decide whether nodes s, t of an undirected graph G are in the same component.*

Definition 7 *The complexity class* **RL** *is the class of problems solvable in logarithmic-space with probabilistic Turing machines with one-sided error.*

Theorem 5 UST-Conn $\in$ RL.

Proof. We consider the following algorithm:

1. Start at node s.
2. Take a random walk for cn^3 steps (c will be specified later).
3. If ever see t output "yes". Otherwise, output "no".

Next, we will analyze the correctness of this algorithm.

If s, t are not connected, the algorithm will never always output "no".
If s, t are connected, let G_s be connected component of s in G, from theorem 4 we have

$$h_{st} \leq C(G_s) \leq c'nm.$$

Pick $c = 4c'$, then Pr[output "no"] = Pr[start at s, walk more that $4h_{st}$ steps and don't reach t] $\leq \frac{1}{4}$ by Markov's inequality. Therefore, we have a one-sided error randomized algorithm.

Since we only need to keep track of the number of nodes we have visited so far, we only use logarithmic space. Therefore, UST-Conn is in RL. □

So far we have shown that UST-Conn is in L and that $RL \subseteq L^{3/2}$, the class of problems solvable in $(\log n)^{3/2}$ space. However, it is still open whether $RL = L$.