CS 830: Randomness and Complexity 2/22/05

Lecture 11: Random Walks

Instructor: Jin-Yi Cai Scribe: Aparna Das, Giordano Fusco

1 Overview

Today's lecture is on the new topic of Random Walks. We define random walks and commutes and prove some interesting properties about random commutes. We end by showing how random walks can be used to decide ST-connectivity on undirected connected graphs in $O(\log n)$ space and linear time.

2 Introduction to Random Walk

We will consider random walks on undirected graphs. It is easy to extend this treatment to directed graphs. The graph may also have self loops and/or multiple edges.

A random walk starts at some vertex v_0, which may be fixed or selected randomly. Once a vertex v_i is defined, the vertex $v_{i+1} = u$ is chosen with probability $1/\deg(v_i)$ among the nighbours u of v_i.

Random walks are related to Markov chains. A Markov chain is defined over a set of states. For each state there is a certain probability to move to a neighbor state. A random walk can be thought as a Markov chain in which the probabilities are uniformly distributed on the neighbors of the state.

A fundamental concept about random walks is the *random commute*.

Definition 2.1. *A* commute *from i to j is a sequence of edges starting at i and returning to i for the first time after having visited j. A* random commute *is a random walk that forms a commute.*

Note that this definition allows multiple visits to i before arriving at j, and then multiple (but at least one) visits to j before returning back to i. Once we get back to i the commute is finished.

In the graph shown in figure 1, if for example $i = 2$ and $j = 3$, then $2, 3, 4, 5, 4, 3, 4, 5, 4, 3, 2$ is a random commute and $2, 3, 2, 3$ is not a commute.

We are now going to define a random variable that will be useful throughout the lecture.

Definition 2.2. *In an undirected graph $G = (V, E)$, for every $i, j, u, v \in V$, with $(u, v) \in E$ and $u \neq v$, let c_{ijuv} be a random variable that counts the number of times the edge (u, v) is traversed in the direction from u to v in a random commute from i to j.*

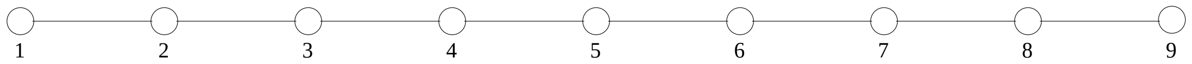

Figure 1: Example of a graph.

To became more familiar with this concept, consider the following examples. In the graph in figure 1, let $i = 2$, $j = 3$, $u = 2$, $v = 3$. Then $c_{2323} = 1$ because in a commute from vertex 2 to vertex 3 we are only allowed to traverse the edge $(2, 3)$ once.

As a second example we compute c_{2343}. The random walk will start at vertex 2. Any initial moves on the left (for example $2, 1, 2, 1, 2$), does not affect the value of c_{2343}. The first interesting move is the one made form 2 to 3. Note that we may make many moves before we cross the edge from 2 to 3, but with probability one we will make this move at some point if the random walk is continued long enough. After 3 is reached there are two possibilities each with probability 1/2 of being selected: we can either go back to 2 or we can go to 4. If we go back to 2, the commute is finished and $c_{2343} = 0$. In the second case, we would go to vertex 4. Any additional move to the right from 4 will not affect the final count of c_{2343} (the argument is the same as above). If we wait long enough, the random walk will eventually go back to 3 again. Then there are two choices: go to 2 or go to 4. If the walk goes to 2, then $c_{2343} = 1$ and the commute is finished. Otherwise, if we go to 4 we are back to the start of the analysis of the second case. In general every time we are at 3, with probability 1/2, we move to 2, and we complete the commute. With probability 1/2 we go to 4 and increase the value of c_{2343} next time on the way back to 3.

Let the random variable X be the number of tries needed to get back to 2 from 3. Then X is an exponentially distributed random variable with p, the probability of success, equal to $\frac{1}{2}$. From the discussion above it is clear that c_{2343} is the number of times the random walk fails to arrive back to 2 from 3, before it finally succeeds. In other words:

$$c_{2343} = X - 1$$

So, $E[c_{2343}] = E[X] - 1$. Since $E[X] = 1/p = 2$, then $E[c_{2343}] = 1$.

Let $\theta_{ijuv} = E[c_{ijuv}]$. In the next section we show that for any fixed connected undirected graph and for any two vertices i, j, θ_{ijuv} is independent of u, v.

3 θ_{ijuv} is independent of u, v

We will show the independence of θ_{ijuv} from u, v by giving the proofs of two intermediate lemmas. The first lemma says that θ_{ijuv} is the same for all neighbors of u, the second says that we can reverse the uv to vu, i.e. $\theta_{ijuv} = \theta_{ijvu}$. We conclude the section, by applying the lemmas repeatedly to get the result.

Lemma 3.1. *$\theta_{ijuv} = \theta_{ijuv'}$ where v' is any neighbor of u.*

Proof. The following definition extends a random commute into an infinite process.

Definition 3.2. *In an undirected graph $G = (V, E)$, $\forall i, j, u, v \in V$, with $u \neq v$ and $(u, v) \in E$, define the random variables Y_0, Y_1, ... in the following way*

$$Y_k = \begin{cases} v_k & \text{if at step } k \text{ the random commute from } i \text{ to } j \text{ has not yet finished} \\ i & \text{otherwise} \end{cases}$$

A common trick to compute the expectation of a complicated random event is to decompose it into simpler events and define an indicator variable for each of the simpler events. Then using the linearity of the expectation, the expectation of each of the simpler events are summed together to get the expectation of the original complicated event.

So define the following indicator variables

$$X_k = \begin{cases} 1 & \text{if the } k\text{-th step is from } u \text{ to } v \\ 0 & \text{otherwise} \end{cases}$$

We have that $c_{ijvu} = \sum_{k=0}^{\infty} X_k$, and now using the linearity of the expectation

$$\begin{aligned} \theta_{ijuv} &= E[c_{ijvu}] = \sum_{k=0}^{\infty} E[X_k] = \sum_{k=0}^{\infty} \Pr[Y_k = u \wedge Y_{k+1} = v] = \\ &= \sum_{k=0}^{\infty} \Pr[Y_k = u] \cdot \Pr[Y_{k+1} = v \mid Y_k = u] \end{aligned}$$

Since $\Pr[Y_{k+1} = v \mid Y_k = u] = \frac{1}{\deg(u)}$, it is the same for all neighbors, of u. So we can replace v with any other neighbor, v', of u, thus $\theta_{ijuv} = \theta_{ijvu}$. □

LEMMA 3.3.

$$\theta_{ijuv} = \theta_{ijvu}$$

Proof. Consider the commute shown in figure 2. The commute starts at vertex i. We can think of the commute in the four distinct parts. In the first part of the commute we may visit vertex i several times before we get to the last i. We never visit j in the first part of the commute because otherwise the commute could be over. The second segment of the commute starts after we have seen the last i and before we have seen the first j. Thus we don't encounter i or j in the second segment. The third segment of the commute starts after we have seen the first j and goes until we encounter the last j. We don't visit i in the third segment as this would end the commute. The

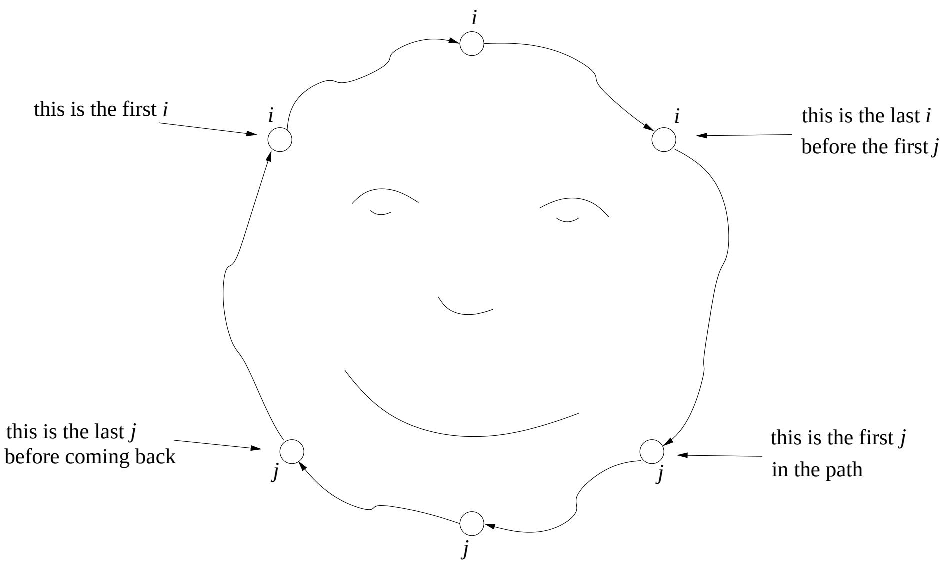

Figure 2: A generic commute.

forth segment of the commute occurs after we have visited the last j and before we get back to i. Vertex i and j don't appear in the forth segment.

To formalize these idea we define a commute in the following way:

DEFINITION 3.4. *A commute is:*

$$C = (i, w_1, \ldots, w_a, i, x_1, \ldots, x_b, j, y_1, \ldots, y_c, j, z_1, \ldots, z_d, i)$$

where $\forall k$, $w_k \neq j$, $x_k \neq i, j$, $y_k \neq i$ *and* $z_k \neq i, j$.

It is possible to visit i *only once before arriving to the first* j *and to visit* j *only once before returning back to* i.

Now we define the reverse commute C^R as follows:

DEFINITION 3.5. *A reverse commute is:*

$$C = (i, w_a, \ldots, w_i, i, x_b, \ldots, x_i, j, y_c, \ldots, y_1, j, z_d, \ldots, z_i, i)$$

where $\forall k$, $w_k \neq j$, $x_k \neq i, j$, $y_k \neq i$ *and* $z_k \neq i, j$.

Note that $((C^R)^R) = C$.

Intuitively, C^R starts at the last i in C and traverses all edges in C backwards, i.e. for every edge $(u, v) \in C$, there is an edge $(v, u) \in C^R$. Thus C and C^R visit the same vertices. Also each vertex is visited the same number of times in C and C^R. Since the probability of a commute is the product of the degrees of the vertices on the commute, $\Pr[C^R] = \Pr[C]$. This implies the following:

$$\begin{aligned}
\theta_{ijuv} &= \sum_{C \text{ is an } i,j \text{ commute}} \Pr[C] \cdot c_{ijuv} \,|_C \\
&= \sum_{C^R \text{ is an } i,j \text{ commute}} \Pr[C^R] \cdot c_{ijuv} \,|_{C^R} \\
&= \theta_{ijvu}
\end{aligned}$$

□

For example applying lemma 3.1 and 3.3 to the graph in figure 3 we get

$$\begin{aligned}
\theta_{ijab} &= \theta_{ijba} && \text{(lemma 2)} \\
&= \theta_{ijbc} && \text{(lemma 1)} \\
&= \theta_{ijcb} && \text{(lemma 2)} \\
&= \theta_{ijcd} && \text{(lemma 1)} \\
&\;\vdots
\end{aligned}$$

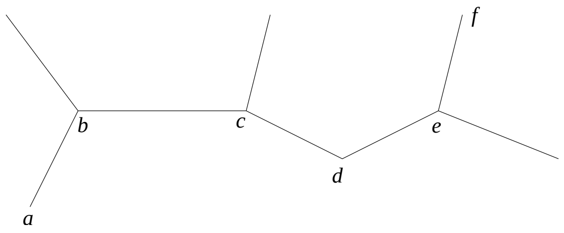

Figure 3: Another example graph.

Since we can apply the above technique to any graph we have shown that for a connected, undirected graph and for any two vertices i and j, θ_{ijuv} is independent of u and v

This implies that for any graph $\theta_{ijuv} = \theta_{ijij}$. Also by definition of a commute $c_{ijij} \leq 1$. Thus $\theta_{ijij} \leq 1$.

4 ST-Connectivity using Random Walks

In this section we will show how random walks give us a probabilistic method to decide ST-Connectivity on undirected graphs using $O(\log n)$ space and polynomial time.

Assume that $G = (V, E)$ is a connected, undirected graph, where $n = |V|$ and $e = |E|$. We will show that the expected number of steps for a random walk to visit every vertex in G is at most $4ne$.

Consider any spanning tree of G. Fix an in-order traversal of the spanning tree and denote it as I. Let N be the number of steps taken in a random walk which starts from s and visits every vertex in G. Our goal is to show that $E[N] \leq 4ne$. Let N' be the number of steps in a random walk which starts from s and visits every vertex in the sequence of the order given by the in-order traversal I. It is easy to see that $N \leq N'$.

We can define N' in terms of adjacent vertices in I. Let T_{ij} be the number of steps for a random walk to get from vertex i to vertex j, where i and j are adjacent in the I. We have that

$$N' = \sum_{i,j \text{ s.t. } (i,j) \in I} T_{ij}$$

T_{ij} can be bounded by the number of steps in a random commute from i to j. Since the random commute might traverse the edges of I in either directions we have that

$$T_{ij} \leq \sum_{(u,v) \in V \times V,\ (u,v) \in E} c_{ijuv}$$

So the expectation of T_{ij} is:

$$\begin{aligned}
E[T_{ij}] &\leq \textstyle\sum_{(u,v) \in V \times V,\ (u,v) \in E} \theta_{ijuv} \\
&= \textstyle\sum_{(u,v) \in V \times V,\ (u,v) \in E} \theta_{ijij} && (\theta_{ijuv} \text{ is independent of} u, v) \\
&\leq \textstyle\sum_{(u,v) \in V \times V,\ (u,v) \in E} 1 && (\theta_{ijij} \leq 1 \text{ by definition of commute so } \theta_{ijij} \leq 1) \\
&= 2e && (\text{since each edge is counted twice in the sum})
\end{aligned}$$

Returning to the calculation of $E[N]$, we have that :

$$\begin{aligned}
E[N] &\leq E[N'] \\
&= E \textstyle\sum_{i,j \text{ s.t. } (i,j) \in I} T_{ij} \\
&= \textstyle\sum_{i,j \text{ s.t. } (i,j) \in I} E[T_{ij}] \\
&\leq \textstyle\sum_{i,j \text{ s.t. } (i,j) \in I} 2e \\
&= 2(n-1) \cdot 2e && (\text{each of the } n-1 \text{ edges appear twice in } I) \\
&\leq 2e
\end{aligned}$$

Using Markov's inequality we have that:

$$\Pr[N \geq 8ne] < \frac{1}{2}$$

Thus Random walks can be used to solve ST-Connectivity in probabilistic log-space and linear time.