6.895 Randomness and Computation April 2, 2008

Lecture 15

Lecturer: Ronitt Rubinfeld *Scribe: Jeremy Fineman*

1 Random Walks

1.1 Markov Chains

Let Ω be a set of states (for the purposes of this class, Ω is always finite, so we can think of it as nodes in a graph). A ***Markov chain*** is a sequence of random variables $X_0, X_1, \ldots, X_t \in \Omega$ that obey the "Markovian property", that is,

$$\Pr[X_{t+1} = y | X_0 = x_0, X_1 = x_1, \ldots, X_t = x_t] = \Pr[X_{t+1} = y | X_t = x_t].$$

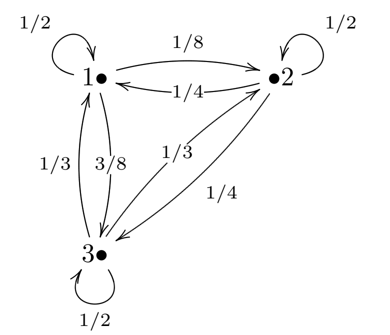
One can think of X_i's as states visited in consecutive steps. The Markovian property essentially says that the transitions between states are historyless—the probability of transitioning to the next state depends only on the current state, not on any of the other previous states.

Without loss of generality, we also assume that transitions are independent of time. More formally, there exists some $P(x, y)$ such that

$$P(x, y) = \Pr[X_{t+1} = y | X_t = x] ,$$

for all t. This assumption is without loss of generality because we can simply create a new set of states $\Omega \times [t]$, having a different set of states associated with each timestep.

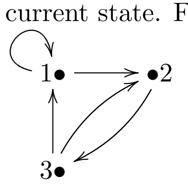
The transition probabilities $P(x, y)$ can be represented either by a graph with probabilities on edges, or by a ***transition matrix*** P. For example, both of the following represent the same transitions.



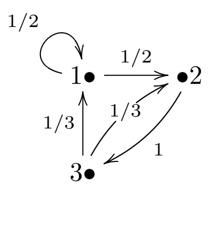
P		1	2	3
	1	1/2	1/8	3/8
	2	1/4	1/2	1/4
	3	1/3	1/3	1/3

1.2 Random walk on a graph

A random walk on a graph $G = (V, E)$ is a special case of a Markov chain. Here, we pick the next state uniformly from among the neighbors of the current state. For example, if we have the following graph



then the transition probabilities are given by



P		1	2	3
	1	1/2	1/2	0
	2	0	0	1
	3	1/2	1/2	0

For a random walk on a graph, $P(i,j)$ is easy to compute. Let $\deg_{\text{out}}(i)$ denote the number of outedges from a node i. Then we have

$$P(i,j) = \begin{cases} \frac{1}{\deg_{\text{out}}(i)} & \text{if } (i,j) \in E \\ 0 & \text{otherwise .} \end{cases}$$

We note that $\forall i, \sum_j P(i,j) = 1$, which is good because rows of P specify a probability of transition.

1.3 The t-step distribution

We call the initial probability distribution over states the ***initial distribution***, denoted by Π^0. The ***t-step distribution*** is the distribution after taking t steps from the starting distribution, given by $\Pi^t = \Pi^0 P^t$, where P^t means the transition matrix P raised to the tth power. To see that this formulation is correct, we show that $P^t(x,y)$ is the probability of getting from x to y in t steps. This fact is easily exhibited by considering a t-step path from x to y as first taking a single to step to some vertex z, and then taking $t-1$ steps to y. Thus, we have

$$P^t(x,y) = \begin{cases} P(x,y) & \text{if } t = 1 \\ \sum_z P(x,z)P^{t-1}(z,y) & \text{for } t > 1 \text{ .} \end{cases}$$

In the case for $t > 1$, it is clear that P^t is just the matrix product of P and P^{t-1}.

1.4 Nice properties for Markov chains

Let's define some properties for finite Markov chains. Aside from the "stochastic" property, there exist Markov chains without these properties. However, possessing some of these qualities allows us to say more about a random walk.

- ***stochastic*** (always true): rows in the transition matrix sum to 1.
- ***doubly stochastic***: rows and columns sum to 1 in the transition matrix. An example of a doubly stochastic graph is one where the $\deg_{\text{in}}(i) = \deg_{\text{out}}(i) = d$, for all nodes $i \in V$. For undirected graphs, a d-regular graph is doubly stochastic.
- ***aperiodic***: $\forall x \in \Omega . \gcd\{t : P^t(x,x) > 0\} = 1$, i.e., the graph is not k-partite for any k. Usually we'll make a graph aperiodic by adding self loops to every node.
- ***irreducible*** (roughly means "strongly connected"): $\forall x, y . \exists t = t(x,y)$ such that $P^t(x,y) > 0$. In other words, for any pair of states, there is some positive probability of transitioning from the first to the second in some number of steps.
- ***ergodic***: $\exists t_0$ such that $\forall t > t_0 . \forall x, y . P^t(x,y) > 0$. This property is strictly stronger than irreducibility.

Ergodicity may seem like a strong property, and it may also seem difficult to prove. The following theorem states that ergodicity is equivalent to irreducibility and aperiodicity.

Theorem 1 *A finite Markov chain is ergodic if and only if it is aperiodic and irreducible.*

1.5 Stationary distributions

A ***stationary distribution*** is one such that $\forall y \in \Omega$, we have

$$\Pi(y) = \sum_x \Pi(x) P(x, y) \ ,$$

or equivalently $\Pi P = \Pi$.

An important class of Markov chains is one in which a stationary distribution Π exists and is unique. It turns out ergodicity is sufficient to guarantee a stationary distribution, as stated by the following theorem.

Theorem 2 *Every ergodic Markov chain has a stationary distribution that is unique.*

Note that if a graph is bipartite, you may never arrive at a stationary distribution simply due to oscillations between two sets. If a graph is unconnected, there may be many stationary distributions.

For an *undirected* graph that is connected and not bipartite, the stationary distribution is given by

$$\Pi(x) = \frac{\deg(x)}{2\,|E|} \ , \tag{1}$$

where $\deg(x)$ is the degree of vertex x.

1.6 Cover time

First, we define the ***hitting time*** of i to j, denoted by h_{ij}, to be the expected time to reach state j when starting from state i. For the special case of the hitting time of a state to itself, we have $h_{ii} = \frac{1}{\Pi(i)}$.

We now define the ***cover time*** of a graph (we focus on undirected graphs) to be

$$\begin{aligned} \mathcal{C}_u(G) &= E[\# \text{ steps to reach all nodes in } G \text{ on walk that starts at } u] \text{ , and} \\ \mathcal{C}(G) &= \max_u \mathcal{C}_u(G) \ . \end{aligned}$$

Let's consider some examples of cover times for simple graphs.

- $\mathcal{C}(K_n^*) = \Theta(n \log n)$, where K_n^* is the complete graph on n nodes that included self loops. The bound follows from the coupon collector.
- $\mathcal{C}(L_n) = \Theta(n^2)$, where L_n is the line graph on n nodes.
- $\mathcal{C}(n\text{-node lollipop}) = \Theta(n^3)$, where an n-node lollipop is a $L_{n/2}$ with a $K_{n/2}^*$ at one of the ends. For intuition, the worst thing to do is start in the clique. Look at how many times you must hit the start of the line before getting all the way to the end. Roughly speaking, it's $\Theta(n^2)$ times, and you only escape the clique with probability $1/n$.

It turns out that the $\Theta(n^3)$ bound is the worst possible for cover time. We will prove something stronger in a moment.

First, from here on, we assume, without loss of generality, that G is aperiodic. This assumption *is* without loss of generality because for any walk in the loopy graph that covers the graph and follows a self loop, we can remove the self loops from the walk only getting even shorter walks.

Before we can get to the main theorem, we need a definition and a lemma. We define the ***commute time*** from i to j, denoted by C_{ij}, to be the expected number of steps for a random walk starting at i to hit j and then return to i. Thus, we have $C_{ij} = h_{ij} + h_{ji}$ by linearity of expectation.

Lemma 3 *For all $(u, v) \in E$, we have $C_{uv} \leq 2m$.*

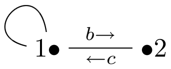
Proof The key idea is to consider a walk of the form $v \to u \rightsquigarrow v \to u$. We will show that

$$E[\text{time between 2 visits to the directed edge } (v,u)] \le 2m \ .$$

Note that this bounds C_{uv}. If we are at u (and we can assume that we just came from v), then after we visit $v \to u$ again, we have commuted from u to v, and to u again.

Given $G = (V, E)$, we construct a $G' = (V', E')$ representing walks on *edges* of G. In particular, the set V' is the set of *directed* edges in G, that is, for every undirected edge between x and y in E, we have two edges (x, y) and (y, x) in V'. The set of edges is $E' = \{((u, v), (v, w)) | (u, v), (v, w) \in V'\} \subseteq V'^2$.

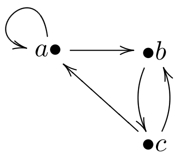
For example, consider the following graph G, transition matrix, and example of a walk.



		1	2
A	1	1/2	1/2
	2	1	0

$1 \to 1 \to 2 \to 1 \to 1$

We would transform the graph into G', with the transition matrix and walk shown below.



		a	b	c
A	a	1/2	1/2	0
	b	0	0	1
	c	1/2	1/2	0

$a \to b \to c \to a$

G' is called the "line graph" of the graph G. In G' our goal is now to figure out what the hitting time of $h_{(u,v)(u,v)}$ is.

Note that G' is doubly stochastic because $P'_{(u,v)(v,w)} = P_{vw} = \frac{1}{\deg(v)}$ if and only if $(u, v), (v, w) \in E$ (once you get to node v, it doesn't matter how you got there), and for all $(v, w) \in E$, we have $\sum_{u:((u,v),(v,w))\in E'} P'_{(u,v)(v,w)} = \sum_{u:(u,v)\in E} \frac{1}{\deg(v)} = 1$.

We apply the fact that G' is doubly stochastic implies Π' is uniform to get

$$\Pi'_{(v,u)} = \frac{1}{|V'|} = \frac{1}{2m}$$

which implies that

$$h'_{(v,u)(v,u)} = \frac{1}{\Pi'_{(v,u)}} = 2m \ .$$

Therefore, the expected time between two visits of an edge in the same direction is at most $2m$. ∎

Now we prove the main theorem.

Theorem 4 *For any graph $G = (V, E)$, we have $\mathcal{C}(G) = O(mn) < O(n^3)$.*

Proof Pick any start vertex v_0, and construct any spanning tree of G rooted at v_0. Note that the number of edges in T is exactly $n - 1$.

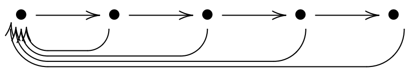
Let $v_0, v_1, v_2, \ldots, v_{2n-2}$ be a depth-first traversal of the spanning tree T. Notice that $v_{2n-2} = v_0$, and each edge of T appears exactly twice, once in each direction.

We conclude that

$$\begin{aligned}
\mathcal{C}(G) &\le \sum_{j=0}^{2n-3} h_{v_j v_{j+1}} \\
&= \sum_{(u,v)\in T} C_{uv} \\
&\le \sum_{(u,v)\in T} 2m && \text{from Lemma 3} \\
&= O(nm) \ .
\end{aligned}$$

$\blacksquare$

We conclude by observing that this theorem does not hold for directed graphs. In particular, consider the graph



Here, the cover time $\mathcal{C}(G) = \Theta(2^n)$ which can be exhibited by starting at the leftmost node.