0368.4163 Randomness and Computation April 22, 2009

Lecture 6

Lecturer: Ronitt Rubinfeld *Scribe: Ido Trivizki and Mille Gandelsman*

1 Today's Lecture

Markov Chains:

- The stationary distribution.
- Special Case of random walks on graphs.
- Cover Times.
- An application to complexity: $STConn \in RL$.

2 Definition of Markov Chains

Definition 1 *Let Ω be the ground set of states. A **Markov chain** (in the following, referred to as MC) is a sequence of random variables $X_0, X_1, \ldots, X_t \in \Omega$ that fulfill the 'Markovian Property':*

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

Ω can be infinite in the general case, but for the purposes of this class is always finite, so we can think of it as the set of nodes in a graph (usually denoted by V).

The definition mathematically describes a situation where given the present state, future states are not reliant upon past states.

Without loss of generality we can assume that the transitions are independent in time. This is convenient because it allows discussing the chain in the terms of the joint probability function $P(x, y)$ defined for every $x, y \in \Omega$ as:

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

The set of transition probabilities $P(x, y)$ can be described as a directed graph, where the edges are labeled by the probabilities of going from one state to the other, or by a ***transition matrix***:

Definition 2 *The **transition matrix** P of a set of transition probabilities $P(x, y)$ is a matrix P where P(x,y) denotes the probability that a system in state x transitions to state y.*

3 t-step Distribution

We call the initial probability distribution over all states ***initial distribution***, and denote it by Π^0.

Definition 3 *The **t-step distribution**, denoted by Π^t, is the distribution obtained after taking t steps from the initial distribution.*

Using matricidal notation: $\Pi^t = \Pi^0 P^t$, where P^t is the transition matrix raised to the t-th power. It is easy to show that $P^t(x, y)$ is exactly the probability to start at x and end at y after taking t steps. Indeed, consider a t-step path from x to y by first taking a single step to some vertex z, and then taking $t - 1$ steps to y. Thus we have:

$$P^t(x, y) = \begin{cases} P(x, y) & \text{if } t = 1 \\ \sum_{z \in \Omega} P(x, z) P^{t-1}(z, y) & \text{if } t > 1 \end{cases}$$

4 The Stationary Distribution

Definition 4 *The* ***stationary distribution*** *is a distribution* Π *such that for every* $y \in \Omega$*:*

$$\Pi(y) = \sum_{x \in \Omega} \Pi(x) P(x, y)$$

Note that the last is naturally not true for a general distribution.
Comments:

1. Not all graphs have a stationary distribution.
2. Some graphs have more than one stationary distribution.
3. Different graphs have different stationary distributions.

5 Which Graphs Have Unique Stationary Distributions?

Definition 5 *A Markov chain is said to be* ***ergodic*** *if* $\exists t_0$ *such that* $\forall t > t_0$ *and* $\forall x, y \in \Omega$*:* $P^t(x, y) > 0$.

Let us mention without a proof the following theorem, which states a couple of conditions which are sufficient and essential for a Markov chain to be ergodic:

Theorem 6 *A finite Markov chain is* ***ergodic*** *if and only if it is* ***irreducible*** *and* ***aperiodic****.*

Recall that:

Definition 7 *A Markov chain is said to be* ***irreducible*** *if* $\forall x, y \in \Omega$*:* $\exists t = t(x, y)$ *such that* $P^t(x, y) > 0$.

Definition 8 *A Markov chain is said* ***aperiodic*** *if* $\forall x \in \Omega$*:* $\gcd\{t | P^t(x, x) > 0\} = 1$.

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

Theorem 9 *Every ergodic Markov chain has a unique stationary distribution.*

6 Random Walk on a Graph

Definition 10 *A* ***random walk*** *on a graph* $G = (V, E)$ *is a sequence of nodes* $s_0, s_1, ...$ *where* s_0 *is the* ***start node*** *and at each step* i, s_{i+1}*is picked by choosing a transition uniformly at random from* $N(s_i)$*, the set of edges outgoing form* s_i*.*

Note that we pick the next state uniformly at random from the edges outgoing from the current state and not from the neighbors of the current state (it is not the same, as one can see for example in the case of double edges or self loops).

For a random walk on a graph, $P(i, j)$ is easy to compute. Let$E(i, j)$ denote the number of edges outgoing from i to j and let $\deg_{\text{out}}(i)$ denote the total number of outgoing edges from a node i. Then:

$$P(i, j) = \frac{|E(i, j)|}{\deg_{\text{out}}(i)}$$

Note that indeed $\forall i$: $\sum_j P(i, j) = 1$, or in other words P is stochastic, which is required because the rows of P specify a probability of transition.

Reminder:

- **Definition 11** *A matrix is said to be* ***stochastic*** *if the entries in each row sum to one.*
- **Definition 12** *A matrix is said to be* ***doubly stochastic*** *if both the entries in each row sum to one and the entries in each column sum to one.*

Definition 13 *A Graph is said to be* ***doubly stochastic*** *if it's transition matrix is doubly stochastic.*

An example of a doubly stochastic directed graph is one where $\deg_{\text{in}}(i) = \deg_{\text{out}}(i) = d$, where d is the same for all nodes $i \in V$. For undirected graphs, a d-regular graph is doubly stochastic.

Note, by the way, that the stationary distribution for the random walk on an undirected graph looks like $(\frac{\deg(x_1)}{2m}, \frac{\deg(x_2)}{2m}, ...)$, where m is the total number of edges.

Theorem 14 *A stationary distribution of a Markov chain with a doubly stochastic transition matrix is the uniform distribution.*

7 Cover Times

For the general case, we define the ***hitting time of i to j*** as follow:

Definition 15 *The* ***hitting time of i to j****, denoted by h_{ij} is the expected time to reach the state j when starting from the state i.*

Theorem 16 *The expected time for a random walk on an ergodic Markov chain starting at state i to return to i, denoted by h_{ii}, is* $\frac{1}{\Pi(i)}$

Definition 17 *Denote:*

$$\mathcal{C}_u(G) = \mathbb{E}[\textit{number of steps to visit all the nodes in } G \textit{ when start at } u]$$

And based on that, we'll define the ***cover time of a graph*** *to be:*

$$\mathcal{C}(G) = max_u \mathcal{C}_u(G)$$

7.1 Examples

1. K_n^* - the complete graph on n nodes and self loops at each node.

   Fact 18 (Cover time of K_n^*)
   $$\mathcal{C}(K_n^*) = \Theta(n \cdot \ln n)$$

   The calculation is the same as n *coupons collection*: Let us define a new markov chain as follow:

   - Let state j be the state at which you've already collected j coupons.
   - Let I_j be the expected number of coupons you should collect in order to get from state j to state $j+1$.
   - $I_j = 1/\mathbb{P}[\text{the next coupon you'll take will be new when you have } j \text{ coupons}] = \frac{n}{n-j}$
   - $\mathcal{C}(K_n^*) = \sum_{j=0}^{n-1} I_j = n \cdot \sum_{j=1}^{n} \frac{1}{j}$
   - We known that $\sum_{i=1}^{n} \frac{1}{i} = \Theta(\ln n)$
   - Therefore $\mathcal{C}(K_n^*) = \Theta(n \cdot \ln n)$

2. L_n - *line n-graph*

Fact 19 (Cover time of L_n)

$$\mathcal{C}(L_n) = \Theta(n^2)$$

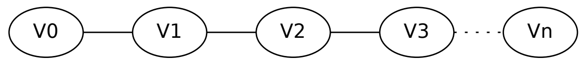

Figure 1: A line n-graph

3. n *nodes lollipop-graph* - a $L_{\frac{n}{2}}$ graph connected at one end to a $K^*_{\frac{n}{2}}$ graph.

Fact 20 (Cover time of n *nodes lollipop-graph*)

$$\mathcal{C}(n \text{ nodes lollipop-graph}) = \Theta(n^3)$$

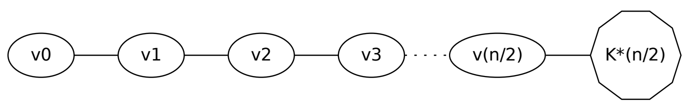

Figure 2: A lollipop graph

7.2 An upper bound on the cover time of an undirected graph

Theorem 21 *Let $G(V, E)$ be an undirected, connected graph with no multiedges, with m undirected edges, and n nodes. Then:*

$$\mathcal{C}(G) \leq 2(m+n)(n-1) \leq 4m(n-1) = O(n^3)$$

Notice that:

$$2(m+n)(n-1) = O(n^3)$$

follows from the fact that $m = O(n^2)$.

Observation 22 *If G is an undirected graph, and δG is G with a self-loop on every node, the cover time of δG is at least the cover time of G, since walks on δG are the same as walks on G, except they might "spend" some steps on the self loops.*

By using the observation, we can prove the theorem by assuming G has a self-loop on every node, and showing a bound on its cover time that is $2m(n-1)$. Then, notice that for the general case, we could look at δG, which has at most $m+n$ edges, and the upper bound on its cover time, which is $2(m+n)(n-1)$ is an upper bound on the cover time of G.

Proof of Theorem 21

Definition 23

$$C_{i,j} = \mathbb{E}[\textit{number of steps for a random-walk starting at } i \textit{ to hit } j \textit{ and return to } i]$$

By linearity of expectations we can see that:

$$C_{i,j} = h_{ij} + h_{ji}$$

Lemma 24

$$\forall (u,v) \in E : C_{u,v} \le 2m$$

Proof (of lemma) A traversal on the graph in which (u,v) is traversed twice: $(u,v),(v,w)\dots(u,v)$ is a commute from u to v and back to u. Therefore, an upper bound on the expected length of such a traversal will also be an upper bound on $C_{u,v}$.
We now want to show that

$$\mathbb{E}[\text{time between two visits to } (u,v)] \le 2m$$

Denote P as the transition matrix for a random walk over G. Observe that $P_{u,v} = \frac{1}{d(u)}$.
Let's define a new directed graph $G'(V',E')$:

- V' will have two nodes for each undirected edge $(u,v) \in E$: a node (u,v), and a node (v,u).
- E' will have edges to connect each pair of adjacent egdes in G. i.e. $(a,b),(c,d) \in E' \iff b = c$'.
- Q is a transition matrix for G' defined as follows:

$$Q_{(u,x),(v,w)} = \begin{cases} P_{v,w} = \frac{1}{d(v)}, & \text{if } (u,x),(v,w) \in E' \\ 0, & \text{otherwise} \end{cases}$$

Let's take a look at an example of building G' from G and Q from P.
Let G be the graph:

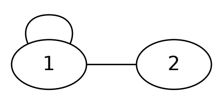

Figure 3: Example undirected graph G

The random walk transition matrix for G, P will be:

	1	2
1	$\frac{1}{2}$	$\frac{1}{2}$
2	1	0

Q will therefore be:

	(1,1)	(1,2)	(2,1)
(1,1)	$\frac{1}{2}$	$\frac{1}{2}$	0
(1,2)	0	0	1
(2,1)	$\frac{1}{2}$	$\frac{1}{2}$	0

And G':

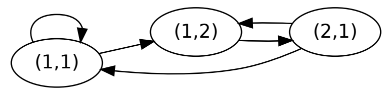

Figure 4: G' as built according to G

It's easy to see that G' defined that way is a Markov Chain. We can also show that Q is *doubly stochastic*:

- rows sum to 1: Fix $(u,v) \in E$

$$\sum_{a \in V, b \in \Gamma(a)} Q_{(u,v),(a,b)} = \sum_{b \in \Gamma(v)} Q_{(u,v),(v,b)} = \sum_{b \in \Gamma(v)} P_{v,b} = \sum_{b \in \Gamma(v)} \frac{1}{d(v)} = d(v) \cdot \frac{1}{d(v)} = 1$$

- columns sum to 1: Fix $(u,v) \in E$

$$\sum_{b \in V, a \in \Gamma(b)} Q_{(a,b),(u,v)} = \sum_{a \in \Gamma(u)} Q_{(a,u),(u,v)} = \sum_{a \in \Gamma(u)} P_{u,v} = \sum_{a \in \Gamma(u)} \frac{1}{d(u)} = d(u) \cdot \frac{1}{d(u)} = 1$$

Since by assumption, G has a self loop on every node, G' also has a self loop on every node (by construction), and therefore, since it's irreducible and aperiodic, it's ergodic.
Therefore, according to Theorem 16 in G':

$$h_{(a,b),(a,b)} = 2m$$

Which means that the expected time between two visits of (u,v) is $2m$, as we wanted to show.
■**(of lemma)**

In order to continue with the proof of the theorem, let T be some spanning tree of G, with root $v_0 \in V$. T has $n-1$ edges. A *depth first search* traversal on T would give us $v_0, v_1, \ldots, v_{2n-2}$, s.t $v_{2n-2} = v_0$, and the traversal traverses every edge of T exactly once in each direction. In addition, $\forall v \in V : v \in \{v_0, v_1, \ldots, v_{2n-2}\}$. Therefore, if we look at a random walk that visits $\{v_0, v_1, \ldots, v_{2n-2}\}$ (in that order), its expected length is an upper bound on $\mathcal{C}_{v_0}(G)$. So:

$$\mathcal{C}_{v_0}(G) \leq \sum_{j=0}^{2n-3} h_{v_j, v_j+1} = \sum_{(u,v) \in T} C_{u,v}$$

But according to Lemma 24,

$$C_{u,v} \leq 2m$$

And therefore, as T has $n-1$ edges:

$$\mathcal{C}_{v_0}(G) \leq 2m(n-1)$$

That's true for every choice of v_0, and thus:

$$\mathcal{C}(G) \leq 2m(n-1)$$

■

- Notice that the theorem doesn't work for graphs that aren't connected, as those can't be covered.
- Also note that we demanded G to be undirected. For directed graphs the bound isn't true, and moreover, there are examples of directed graphs with exponential cover time.
  For example, see the graph in Figure 5.

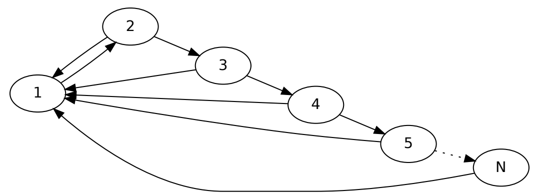

Figure 5: A directed graph with exponential cover time

However, for *d-regular* directed graphs, the theorem does hold.

8 An application to complexity: $STConn \in RL$

$STConn$ is the problem of determining, given a graph G and two vertexes s and t, whether s and t are connected. The problem can be solved in linear time. The question is, however, whether we could solve it in logarithmic space.
$USTConn$ is $STConn$ for an undirected graph G.

Definition 25 ***RL*** *is the class of problems that are solvable in logarithmic space, with random bits.*

Given an algorithm $\mathcal{A}(G, s, t)$, we'll say that it solves $USTConn$ in RL if:

1. $\mathbb{P}[\mathcal{A}(G, s, t) = \text{yes}] \geq \frac{3}{4}$ if s and t are connected in G.
2. $\mathbb{P}[\mathcal{A}(G, s, t) = \text{no}] = 1$ otherwise.

Theorem 26

$$USTConn \in RL$$

Proof We'll use the following algorithm:

1. Given $G(V, E)$, $s, t \in V$, take a random walk, starting at s for $4 \cdot B$ steps, with $B = 4m(n-1)$.
2. If the random walk visited t, output "yes", otherwise, output "no".

The algorithm only uses space to save the current state of the walk, and to count the steps, and therefore it uses logarithmic space.

If s and t are not connected, the random walk will never visit t and "no" will be returned. Therefore: $\mathbb{P}[\mathcal{A}(G, s, t) = \text{no}] = 1$
.

If s and t are connected, "no" is returned if t was not visited, therefore:

$$\mathbb{P}[\mathcal{A}(G, s, t) = \text{no}] \leq \mathbb{P}[\text{ we didn't cover the graph in 4B } steps]$$

But according to Theorem 21: $4B \geq 4 \cdot \mathbb{E}$[number of steps to cover G]. And therefore, using Markov's inequality:

$$\mathbb{P}[\mathcal{A}(G,s,t) = \text{no}] \leq \frac{1}{4}$$

Therefore: $\mathbb{P}[\mathcal{A}(G,s,t) = \text{yes}] \geq \frac{3}{4}$

$\blacksquare$ A very recent important result has shown that this algorithm can be de-randomized."