6.842 Randomness and Computation March 4, 2022

Lecture 10

Lecturer: Ronitt Rubinfeld *Scribe: Angelos Assos*

1 Introduction

In this lecture we do an introduction in Markov Chains. In the first section, we define what is a Markov chain, its properties and Random Walks on Graphs. We continue by studying what happens to the probability distribution of the Markov Chain after t steps and see what properties need to hold in order to eventually achieve unique stationary distributions. After we develop our basic machinery we introduce the hitting times and cover times and use the machinery we developed to provide a bound for cover times in undirected graphs.

2 Markov Chains

We start setting up the stage with the definition of Markov Chains.
A **Markov chain** is a process which moves among the elements of a set Ω (Ω is a set of possibly infinite states $x_1, x_2, ...$) in the following manner: when at $X_t \in \Omega$, the next position X_{t+1} is chosen according to a fixed probability distribution $P(X_t, \cdot)$ depending only on X_t. What we said in the last sentence is essentially the **Markovian Property**, which we can write down formally:

Definition 1 *Markovian Property:*

$$\Pr[X_t = y | X_1, X_2, ..., X_{t-1}] = \Pr[X_t = y | X_{t-1}]$$

The above property is essential and is basically the building block upon which we will construct all our theorems and proofs later on. From now on we continue with the following simplified notation:

Definition 2 $P(x, y) := \Pr[X_{t+1} = y | X_t = x]$

Note that in order to define a Markov Chain we just need the states, and how we move between those states, i.e. the transition matrix. Below we provide such an example:

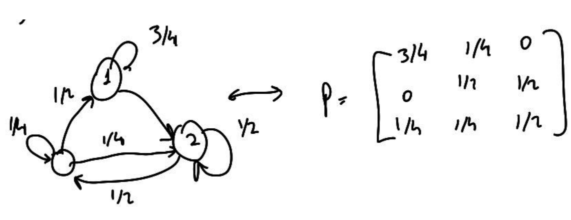

Figure 1:

Next we examine the case where our Markov Chain is a directed Graph.

Definition 3 *A Random Walk on a graph $G = (V, E)$ is a sequence $s_1, s_2, ..., s_t$ for which $s_{i+1} \in N(s_i)$, where $N(v)$ is the neighborhood of v. Note that the transition matrix of a graph is going to have:*

$$P(x, y) = \frac{1}{d_x}$$

where d_x is the degree of vertex x.
The transition matrix of any Markov Chain is stochastic, i.e. the entries in every row add up to 1. A transition matrix is **doubly stochastic** if the rows and the columns add to 1 (for example the matrix in figure 1 is doubly stochastic).

We are generally interested in the question of: given that we start on a vertex v, where will we end up being after a number of steps in our Markov Chain? What is the probability that if a long time passes we will end up in another vertex u? What if we do not start at v, but we start with a probability distribution across all vertices? We will try to answer these questions in the next section.

2.1 Distribution after t steps

Suppose we start with a distribution $\pi^{(0)} = (\pi_1^{(0)}, \pi_2^{(0)}, ..., \pi_n^{(0)})$, i.e. we start at state i with probability $\pi_i^{(0)}$. We want to know our probability distribution after t steps, i.e. $\pi^{(t)}$. Let us focus for $t = 1$ for now. We want to know what $\pi_i^{(1)}$ is. If we start from j, the probability of ending up at i is going to be $\pi_j^{(0)} P(i, j)$. Therefore we have:

$$\pi_i^{(1)} = \sum_j \pi_j^{(0)} P(j, i)$$

Indeed one can see that we can get the following formula:

$$\pi^{(1)} = \pi^{(0)} P$$

and inductively, we see that the distribution at time t can be computed from the distribution at time t-1 via matrix multiplication::

$$\pi^{(t)} = \pi^{(0)} P^t$$

Now that we determined our state after t steps, we are interested in knowing whether or not the distribution will converge after some time and under which conditions.

2.2 Stationary distributions

Definition 4 *A stationary distribution π^* has the following property:*

$$\forall x, \pi^*(x) = \sum_y \pi^*(y) P(y, x)$$

or even better:

$$\pi^* = \pi^* P$$

In order for a Markov Chain to have a unique and stationary distribution we require 2 properties: **aperiodicity** and **irreducibility** which we define below. To get a between grasp on those two properties, we can imagine instead the following two examples: On the left we have a graph of two vertices that is a periodic Markov Chain. That is not a stationary distribution because at any given time we are going to be either at the white or the black node. For the right example the triangle graph will never reach the other two vertices and therefore again we can have various different stationary distributions depending on our initial distribution.
Here we give the formal definitions for aperiodicity and irreducibility.

Definition 5 *A Markov Chain is aperiodic if:*

$$\forall x, gcd\{t : P^t(x, x) > 0\} = 1$$

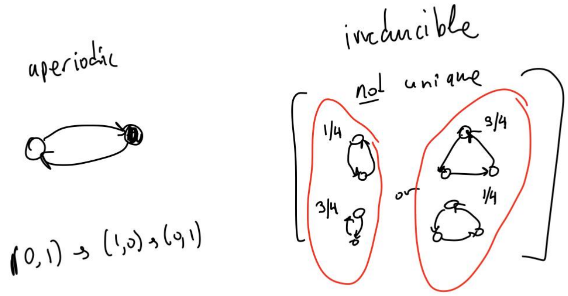

Figure 2:

Definition 6 *A Markov Chain is irreducible if:*

$$\forall x, y, \exists t = t(x, y) : P^{t'}(x, y) > 0, \forall t' > t$$

We can capture the above two properties into one bigger one, ergodicity:

Definition 7 *A Markov Chain is ergodic if:*

$$\exists t^* : \forall t > t^*, \Pr(x, y) > 0, \forall x, y$$

Theorem 8 *If a Markov Chain is ergodic then it is going to have a unique stationary distribution no matter what the initial distribution was.*

For undirected graphs, the stationary distribution is going to be:

$$\pi^* = (\frac{d(v_1)}{2|E|}, \frac{d(v_2)}{2|E|}, ..., \frac{d(v_n)}{2|E|})$$

For d-regular graphs, i.e. graphs where the degrees are all equal to d, we have that:

$$\pi^* = (\frac{1}{d}, \frac{1}{d}, ..., \frac{1}{d})$$

3 Cover and hitting times

Definition 9 *We define the Hitting time of x to y as:*

$$h_{x,y} := E[\# \textit{ steps to go from } x \textit{ to } y]$$

Note that for $x = y$ we have the expected number of steps we are going to take to return to x. We call that **recurrence times**. We are going to prove the following theorem:

Theorem 10

$$h_{x,x} = \frac{1}{\pi^*(x)}$$

Proof (Sketch) Consider all the states that we visit along the way. The fraction of positions that are going to be x, is $\pi^*(x)$. Therefore we expect the distance between them to be

$$h_{x,x} = \frac{1}{\pi^*(x)}$$

■

Definition 11 *We define the cover time of a vertex v of a graph G as follows:*

$$C_v(G) = E[\#\ \textit{steps to visit all vertices when we start at } v]$$

We define the cover time of a graph G as:

$$C(G) = \max_{v \in G} C_v(G)$$

We end the lecture by proving the following upper bound for the cover times of G.

Theorem 12 *For a connected undirected graph G the cover time of the Graph can be bounded as follows:*

$$C(G) \leq O(nm)$$

More specifically we have:

$$C(G) \leq m(n-1)$$

Below we provide some examples for the complete graph, the line graph and the lollipop graph:

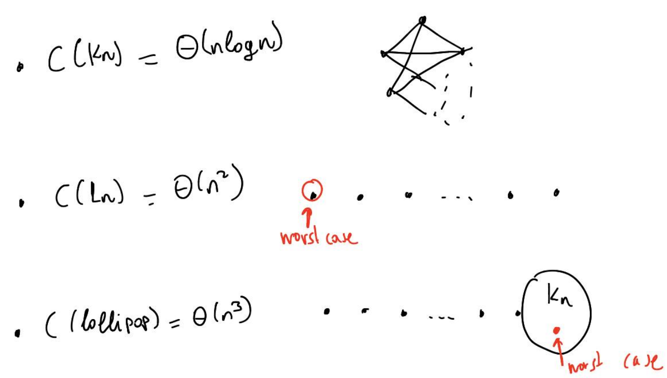

Figure 3: Complete, line and lollipop graph

Note that the theorem we are going to prove givs very weak bounds for several cases. One of them is the K_n graph: the actual cover time for the complete graph is $O(n \log n)$ however, the theorem is giving us a bound of $O(n^2 \cdot n) = O(n^3)$.
We continue with the proof of the theorem:
Proof We begin by creating a new graph G' by multiplying the transition matrix of G with $\frac{1}{2}$ and adding a self loop from all the vertices with probability $\frac{1}{2}$. We begin by proving the following claim:

Claim 13

$$C(G) = \frac{C(G')}{2}$$

Proof Suppose we have a path in G'. In order to convert it back to a path in G,we have to remove all self loops of G' that we intentionally added. However, each loop happens with probability $\frac{1}{2}$ therefore we expect to erase half of the path and therefore the result follows.
■ We introduce one last definition:

Definition 14 *We define commute time from x to y the expected number of steps it takes to go from x to y and back to x*

$$C_{x,y} = E[\# \textit{ of steps to go from } x \textit{ to } y \textit{ and back to } x] = h_{x,y} + h_{y,x}$$

where the last equality follows from the linearity of expectation

We prove the following lemma which is crucial to the proof of the theorem:

Lemma 15

$$C_{x,y} \leq O(m)$$

Proof Here, it is sufficient for us to bound the expected time for one way of doing a commute; there might be other ways of doing the commute as well that are cheaper which can only make the bound we have better, so it does not bother us. Write down again all the vertices we visit, $s_1, s_2, \ldots$. We are looking for consecutive occurrences of x, y. The probability of having an x is going to be exactly $\pi^*(x)$ (Theorem 10). The probability of having y right after x is going to be exactly $\frac{1}{\deg(x)}$ since after visiting x we have $deg(x)$ many options, only one of which is y. Therefore:

$$\Pr[(s_i, s_{i+1} = (x, y)] = \Pr[s_i = x] \cdot \Pr[s_{i+1} = y | s_i = x] = \pi^*(x) \frac{1}{deg(x)} = \frac{deg(x)}{2m} \cdot \frac{1}{deg(x)} = \frac{1}{2m}$$

Now, since each edge $x - y$ appears with probability $x - y$, the average distance between 2 $x - y$ edges is going to be

$$\frac{1}{\frac{1}{2m}} = 2m$$

from theorem 10, and hence we got $C_{x,y} \leq O(m)$. ■

Now we have everything we need to finish the proof.
Consider a spanning tree T of G', and a traversal of the tree T:

$$v_0, v_1, v_2, \ldots, v_{2n-2}$$

We have:

$$C(G) = \frac{C(G')}{2} \leq \frac{1}{2} = \sum_{j=0}^{2n-2} h_{v_j, v_{j+1}}$$

We have that

$$\sum_{j=0}^{2n-2} h_{v_j, v_{j+1}} = \frac{1}{2} \sum_{(u,v) \in T} C_{(u,v)}$$

since $C_{(u,v)} = h_{u,v} + h_{v,u}$ by definition and both the terms appear in the spanninng tree

$$\frac{1}{2} \sum_{(u,v) \in T} C_{(u,v)} = \frac{1}{2} 2m \cdot (n-1) = m(n-1) \leq O(nm)$$

as wanted. ■

Note: Note that following the traversal shown in the theorem is only one way of covering the graph and it is sufficient for proving cover times, but for some graph (e.g. K_n, as discussed above), this may give a very weak bound.