CS 388R: Randomized Algorithms, Fall 2021 October 28th, 2021

Lecture 18: Markov Chains

Prof. Eric Price *Scribe: Georgios Smyrnis and Michael Jaber*

NOTE: THESE NOTES HAVE NOT BEEN EDITED OR CHECKED FOR CORRECTNESS

1 Overview

In this lecture, we covered some basic definitions and properties of Markov chains. We also examined a lemma corresponding to the expected commute time from one state of the Markov chain to another, and provided some first steps for its proof.

2 Basic Definitions

In this lecture, we defined a Markov chain as a *discrete* and *memoryless* stochastic process. These definitions correspond to the following:

- **Discrete:** A Markov chain consists of a random walk on a discrete set of n states.
- **Memoryless:** This means that the next state of a Markov chain only depends on the previous state. In other words:

$$\mathbb{P}[X_{t+1} = j \mid X_0 = i_0, X_1 = i_1, \ldots, X_t = i_t] = \mathbb{P}[X_{t+1} = j \mid X_t = i_t] \tag{1}$$

Definition 1 (Transition Matrix)**.** *We define the transition matrix of a Markov chain as the matrix* $P = p_{ij}$*, where:*

$$p_{ij} = \mathbb{P}[X_{t+1} = j \mid X_t = i] \tag{2}$$

Using the transition matrix, we can write the following relationship regarding the distribution across states[1] $q^{(t)} = [\mathbb{P}[X_t = 1], \ldots, \mathbb{P}[X_t = n]]$:

$$q^{(t+1)} = q^{(t)} P \tag{3}$$

Definition 2 (Stationary Distribution)**.** *We call a distribution* π *over the states of a Markov chain stationary if it satisfies:*

$$\pi P = \pi \tag{4}$$

Equivalently[2]*, it is an eigenvector of* P^T*, corresponding to eigenvalue 1.*

[1]Scribe note: Note that this is defined as a row vector.

[2]Scribe note: Note that this implies that a Markov chain always has some stationary distribution. Indeed, the matrix P has rows summing to 1, and so $P\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the all-ones vector. This means that P has an eigenvalue equal to 1, and so does P^T, with π being the corresponding eigenvector.

(a) This Markov chain is periodic: if we start from a single state with certainty, we will only ever reach that state again after 3 steps.

(b) This Markov chain is reducible: there is no path from states 2 or 3 to state 1.

Figure 1: Examples of state sequences that are non-ergodic.

Definition 3 (Ergodic Markov Chain)**.** *We call a Markov chain ergodic if it satisfies the following two properties:*

- *It has a* unique *stationary distribution* π, *with* $\pi_i > 0$, $\forall i \in [n]$.
- *For any initial state distribution* $q^{(0)} = q$, $q^{(t)} = qP^t$ *converges to* π. *In other words:*

$$\forall q, \quad \lim_{t \to \infty} (qP^t) = \pi \tag{5}$$

Theorem 4 (Fundamental Theorem of Markov Chains)**.** *A Markov chain is ergodic if the following hold:*

1. *It is* finite*:* $n \in \mathbb{N}$.
2. *It is* irreducible*: For any pair of states* i *and* j*, there exists a path from* i *to* j *with non-zero probability.*
3. *It is* aperiodic*: The greatest common denominator of all paths from a node* i *back to itself is equal to* 1.

In Figure 1 we can see some examples of Markov chains which do not satisfy these properties, and are thus non-ergodic. We also note here that, for irreducible Markov chains, the aperiodic condition needs to only be checked with respect to a single state i (for any other state, we can move to i and back, since the chain is irreducible).

Now, we define as $N(i,t)$ the number of times that i is reached in the first t steps. The following holds:

Lemma 5. *For an ergodic Markov chain, we have:*

$$\lim_{t \to \infty} \frac{N(i,t)}{t} = \pi_i \tag{6}$$

In other words, the number of times we expect to reach state i *in our random walk should converge to the probability of that state in the stationary distribution (since the distribution converges to* π *for all states.*

Definition 6 (Hitting Time)**.** *For any two states* u *and* v*, we define as* h_{uv} *the expected number of steps we will take to reach* v*, if we start from* u*. In the case of* $u = v$*, we define this time to be the expected time to reach state* u *again (so* $h_{uu} > 0$*).*

3 Walks on Undirected Graphs

From here on, we shall consider our Markov chains to model walks on undirected graphs. More specifically, given a graph $G = (V, E)$, the transition matrix P of our markov chains has the following form:

$$P_{uv} = \begin{cases} \frac{1}{\deg(u)}, & (u, v) \in E \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

for all vertices $u, v \in V$. In this case, the following statements are equivalent:

- A Markov chain defined using the graph G is ergodic.
- The graph G is connected and contains an odd cycle.
- The graph G is connected and not bipartite.

The stationary distribution of the above Markov chain is $\pi_v = \frac{\deg(v)}{2m}$. Indeed:

- $\sum_{v \in V} \pi_v = \frac{1}{2m} \sum_{v \in V} \deg(v) = 1$.
- $(\pi P)_v = \sum_{u:(u,v) \in E} \pi_u P_{uv} = \sum_{u:(u,v) \in E} \frac{\deg(u)}{2m} \frac{1}{\deg(u)} = \frac{\deg(v)}{2m} = \pi_v$

Definition 7 (Commute Time)**.** *We define the commute time c_{uv} between two vertices as the expected time required to move from u to v and back to u. In other words:*

$$c_{uv} = h_{uv} + h_{vu} \tag{8}$$

In the following, we shall examine the hitting and commute times for a few specific graphs:

- **Clique of n vertices:** In this example, all n vertices are connected to each other. For any two vertices u and v, we will either move from one to the other immediately with probability $\frac{1}{n-1}$, or take one essentially "useless" step. So:

  $$h_{uv} = \frac{1}{n-1} + \frac{n-2}{n-1}(h_{uv} + 1) \Rightarrow h_{uv} = n - 1 \tag{9}$$

  This is an expected result - since we have so many options to move to, it will take a lot of time to finally reach our destination.

- **Connected path of n vertices:** In this case, we can consider a random walk Z, with $Z_i = 1$ with probability $1/2$ and $Z_i = -1$ with probability $1/2$ as well. In this case, we have that $\text{Var}\left[\sum_{i=1}^{t} Z_i\right] = t$. This means that we need $\Theta(n^2)$ steps to reach a standard deviation of $\Theta(n)$ (so that we have a likely chance to reach the one side of the path from the other). This means that, if u and v are the endpoints of the path, we have $h_{uv} = \Theta(n^2)$.

- **"Lollipop" graph:** This graph consists of a clique of $\frac{n}{2}$ vertices, one of which is also connected to a path of $\frac{n}{2}$ vertices. Note here that this case is not a simple combination of the above two: in fact, the commute time from a vertex u in the clique and the other endpoint of the path v is increased by the fact that it is likely that we make circles within the clique, never actually progressing on the path. In the following, we shall show that in this case, $c_{uv} = \Theta(n^3)$.

We shall use the following lemma to analyze the final graph presented:

Lemma 8. *Let us consider a random walk on a graph $G = (V, E)$. For any two vertices u and v, the following holds:*

$$c_{uv} = 2mR_{uv} \tag{10}$$

where R_{uv} is the effective resistance between the two vertices.

Using the above lemma, we can derive the commute time for the lollipop graph. We have the following (assuming that w is the vertex where the clique and the path are connected):

- The effective resistance of the clique is $R_{uw} \leq 1$ (since all vertices are connected, and connecting more paths in the clique only lowers the resistance).
- The effective resistance of the path is $R_{wv} = \Theta(n)$ (equivalent to resistors being added in series).
- The total effective resistance is $R_{uv} = \Theta(n)$ (since the two graphs are connected in series).

Now, since the entire graph has $m = \Theta(n^2)$ (due to the clique), using the lemma we have:

$$c_{uv} = 2mR_{uv} = \Theta(n^3) \tag{11}$$

Note that, as stated above, this means that the commute time is much higher than when the two graphs are examined independently.

4 Start of Proof for Lemma 8

We shall now start the proof of Lemma 8, which shall be concluded in the next lecture. We shall make use of the physical analogy of a graph to a resistor network, with each edge representing a resistor between two nodes, and each node having an assigned voltage.

Let us define the following current, going through the graph:

$$i = \begin{bmatrix} \deg(v_1) \\ \vdots \\ \deg(u) \\ \vdots \\ \deg(v_n) \end{bmatrix} - \begin{bmatrix} 0 \\ \vdots \\ 2m \\ \vdots \\ 0 \end{bmatrix} \tag{12}$$

The physical equivalent of this is introducing current $\deg(v_i)$ into all nodes, and retrieving it all from node u. We shall show that, if x_v is the voltage at node v, then $x_v - x_u = h_{vu}$.

As a reminder, the graph has a Laplacian matrix $L = D - A$, and the transition probabilities satisfy $P = D^{-1}A$ (we can move to any adjacent node, with probability inversely proportional to the degree of the node we are currently in). Combining the above, we get that $L = D - A = D(I - P)$. Thus, we have:

$$i^{(u)} = Lx = D(I - P)x \tag{13}$$

For now, let us alter our definitions of hitting times temporarily, so that we have $h_{uu} = 0$. Doing so, we have:

$$\begin{aligned} h_{vu} &= \sum_{w \in N(v)} \frac{1}{\deg(v)}(1 + h_{wu}) \\ &= 1 + \sum_{w \in N(v)} \frac{1}{\deg(v)} h_{wu} \Leftrightarrow \\ \deg(v) &= \sum_{w \in N(v)} (h_{vu} - h_{wu}) \end{aligned} \tag{14}$$

Now, notice that if we set $x_v = h_{vu}$, then on the edge (u, v) we have current equal to $x_u - x_v$. This means that:

$$\sum_{w \in N(v)} (x_v - x_w) = \deg(v), \ \forall v \neq u \tag{15}$$

which means that the current input into the nodes $v \neq u$ is precisely that defined by $i^{(u)}$. Moreover, given that the total current must be 0, the current into u is $\deg(u) - 2m$, which is precisely the one defined by $i^{(u)}$. This means that choosing $x_v = h_{vu}$ is consistent with our choice for $i^{(u)}$, and since $h_{uu} = 0$ (as mentioned above, temporarily) we have $x_v - x_u = h_{vu}$.

In the above, we showed that $x = L^+ i^{(u)}$ satisfies $x_v = h_{vu}$. In the following lecture, we shall complete this definition by taking advantage of the effective resistance $R_{uv} = (e_u - e_v)^T L^+ (e_u - e_v)$ of the network.