CS 388R: Randomized Algorithms, Fall 2019 November 19th, 2019

Lecture 24: Markov Chains

Prof. Eric Price *Scribe: Supawit Chockchowwat, Ruizhe Zhang*

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

1 Definition and Properties

Markov chain is a discrete, memoryless stochastic process. It has finite number of states $x \in [n]$ and transition probability associated on each pair of states p_{ij} as followed,

$$p_{ij} = \mathbb{P}[x_{t+1} = j | x_t = i] = \mathbb{P}[x_{t+1} = j | x_t = i, x_{t-1}, \ldots, x_0] \tag{1}$$

where the last equality describes its Markovian property. We often writes the transition probabilities together in a form of transition matrix $P \in [0,1]^{n \times n}$ such that $P_{ij} = p_{ij}$ and $\sum_{j=1}^{n} P_{ij} = 1$ for all i. Let $q^{(t)} \in \mathbb{R}^n$ be a distribution over n states at time t. Markov chain evolves this distribution by

$$q^{(t+1)} = q^{(t)} P = q^{(0)} P^{t+1} \tag{2}$$

Stationary distribution $\pi \in \mathbb{R}^n$ is a stationary distribution if $\pi = \pi P$, i.e. the distribution does not change as time moves forward. It corresponds to an eigenvector of P having eigenvalue 1.

1.1 Fundamental Theorem of Markov Chain

Definition 1. *A Markov chain is ergodic if its stationary distribution π is unique and*

$$\lim_{t \to \infty} q^{(0)} P^t = \pi$$

for all possible distribution $q^{(0)}$.

Intuitively, ergodicity guarantees that every state distribution will eventually converge to a unique distribution asymptotically.

Theorem 2. *If a Markov chain satisfies the following conditions, then it is ergodic.*

1. *Finite: the number of states $n < \infty$.*
2. *Irreducible: $\forall i, j \ \exists$ path $i \rightsquigarrow j$.*
3. *Aperiodic: the $\gcd$ of distances of all loops l_i (path from i to i) is 1 for all $i \in [n]$.*

These guarantees that every other eigenvalue has magnitude less than 1 with only one eigenvalue 1 associated with the stationary distribution. When applying $P^(t)$, the direction to this eigenvector dominates the evolution, and so every initial state converges the unique π.

Hitting time $h_{ij} = \mathbb{E}[\text{number of steps from } i \text{ to reach } j]$. Note that $h_{ii} = \frac{1}{\pi_i}$. Closely related, define $N(i,t)$ as the number of times visiting state i in t steps. Then we have $\lim_{i\to\infty} \frac{N(i,t)}{t} = \pi_i$.

Commute time $C_{uv} = h_{uv} + h_{vu}$, the number of steps for $u \rightsquigarrow v \rightsquigarrow u$.

Cover time $C_u(G) = \mathbb{E}[\text{number of steps to visit all vertices}]$.

2 Random Walks on Undirected Graph

Let $G = (V, E)$ be an undirected graph with $|V| = n$ and $|E| = m$. We can consider it as a Markov chain with n states and define the transition probabilities on the edges as follows:

$$P_{uv} = \begin{cases} \frac{1}{d(u)} & \text{if } (u,v) \in E, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

If the graph is non-bipartite, G is an ergodic Markov chain. Since it is an undirected graph, any edge contains a loop of length 2, so if there is a cycle of odd length (non-bipartite), the gcd of loops is 1.

We also know that $\pi_u = \frac{d(u)}{2m}$ and hitting time $h_{uu} = \frac{2m}{d(u)}$, for any $u, v \in V$, where $d(u)$ is the degree of u. Some interesting examples are

- **Clique**: $h_{uv} = \Theta(n), c_{uv} = \Theta(n)$ follows the recurrence of identical Bernoulli trials. $C_u(G) = \Theta(n \log n)$ follows from the coupon collector problem.
- **Line**: $h_{uv} = \Theta(n^2), c_{uv} = \Theta(n^2), C_u(G) = \Theta(n^2)$ from random walks on a discrete line.
- **Lollipop**: a graph with a clique containing $\frac{n}{2}$ vertices, and a line extending out with length $\frac{n}{2}$. If u is in the clique and v is the vertex at the far end of the line (from the clique), $h_{uv} = \Theta(n^3), h_{vu} = \Theta(n^2), C_{uv} = \Theta(n^3)$, proved by the following lemma.

Lemma 3. *For any $u, v \in V$, the commute time C_{uv} is*

$$C_{uv} = 2m \cdot R_{eff}(u, v),$$

where $R_{eff}(u,v)$ is the effective resistance between vertices u and v

Let's first recall some spectral graph notations from last lecture. Let $U \in \mathbb{R}^{m\times n}$ be a collection of m edges such that each k-th row represents a directed edge (u, v) with $U_{ku} = -1$ and $U_{kv} = 1$ and zero everywhere else. Let D be the diagonal matrix filled with degree of vertices $D_{uu} = d(u)$, A be the adjacency matrix, and $L = D - A$ be the Laplacian of the graph. We can show that $L = D - A = U^\top U$. The effective resistance is defined on every pair $(u, v) \in V^2$ to be $R_{eff}(u, v) = (e_v - e_u)^\top L^\dagger (e_v - e_u)$ where $e_u \in \mathbb{R}^n$ is a standard basis vector with 1 entry on u-th component and 0 everywhere else, and $L^\dagger$ is the pseudo-inverse of L.

To prove the lemma, we need the following claim:

Claim 4. *For $u \in V$, let $\mathbf{i}_u$ be a vector in $\mathbb{R}^n$ such that*

$$\mathbf{i}_u := \begin{bmatrix} d(1) \\ \vdots \\ d(u) \\ \vdots \\ d(n) \end{bmatrix} - \begin{bmatrix} 0 \\ \vdots \\ 2m \\ \vdots \\ 0 \end{bmatrix} = \sum_{v \in V} d(v)(e_v - e_u). \tag{4}$$

Let $x = L^\dagger \mathbf{i}_u$. Then, we have

$$x_v - x_u = h_{vu}.$$

Intuitively, given a vertex u, suppose we inject electrical current through all other vertices $v \neq u$ with $d(v)$ amps and through vertex u with $d(u) - 2m$ (since $d(u) < 2m$, we technically draw current out of u). Let $\mathbf{i}_u$ be the vector of such currents, Consider the voltage $X \in \mathbb{R}^n$. Ohm's law gives $X = L^\dagger \mathbf{i}_u$ assuming $X_u = 0$. The claim states that we can evaluate the hitting times by Ohm's law.

Proof. Define $h_{vv} = 0$, for any $v \neq u$, and let $N(v)$ be the set of neighbors of v. Then, we have

$$\begin{aligned} h_{vu} &= \sum_{w \in N(v)} \frac{1}{d(v)}(1 + h_{wu}) \\ &= 1 + \sum_{w \in N(v)} \frac{1}{d(v)} h_{wu} \end{aligned}$$

Multiplying both sides with $d(v)$,

$$\begin{aligned} d(v)h_{vu} - \sum_{w \in N(v)} h_{wu} &= d(v) \\ \sum_{w \in N(v)} (h_{vu} - h_{wu}) &= d(v) \end{aligned} \tag{5}$$

The vector of hitting time ending at u, $h_{*u} \in \mathbb{R}^n$, must satisfy the equation (5), written concisely as $Lh_{*u} = \mathbf{i}_u$. Hence, the solution by taking pseudo-inverse is $h_{*u} = L^\dagger \mathbf{i}_u = X$; in other words, $h_{vu} = X_v$ by applying currents according to $\mathbf{i}_u$. □

Now, we are ready to prove the lemma:

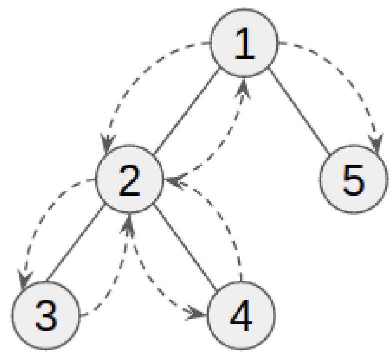

Figure 1: DFS traversal of the spanning tree.

Proof of Lemma 3. For any pair $u, v \in V$:

$$\begin{aligned}
C_{uv} &= h_{uv} + h_{vu} \\
&= (L^\dagger i_v)_u + (L^\dagger i_u)_v \\
&= (e_v - e_u)^\top (L^\dagger i_v + L^\dagger i_u) && ((L^\dagger i_u)_u = X_u = 0) \\
&= (e_v - e_u)^\top L^\dagger (i_u - i_v) \\
&= (e_v - e_u)^\top L^\dagger \left(\sum_{w \in V} d(w)(e_v - e_u) \right) && \text{(from equation (4))} \\
&= 2m \cdot (e_v - e_u)^\top L^\dagger (e_v - e_u) \\
&= 2m \cdot R_{eff}(u, v)
\end{aligned}$$

□

Cover time To upper bound the cover time $C_u(G)$, we can consider a spanning tree of G, say T_G. Then, we can start from u and traverse the whole tree by DFS. Note that the DFS path goes through each edge twice (See Figure 1). Therefore, the expected total time for visiting all vertices is

$$\begin{aligned}
C_u(G) &= \sum_{(v,v') \in T_G} h_{vv'} + h_{v'v} \\
&= \sum_{(v,v') \in T_G} c_{vv'} \\
&= \sum_{(v,v') \in T_G} 2m \cdot R_{eff}(v, v') \\
&\leq \sum_{(v,v') \in T_G} 2m && (R_{eff}(v, v') \leq 1 \text{ for any edge } (v, v') \in E.) \\
&= 2m \cdot (n - 1) && (T_G \text{ has } n - 1 \text{ edges.})
\end{aligned}$$