6.842 Randomness and Computation	March 7, 2022

Lectures on Random Walks

Lecturer: Ronitt Rubinfield — *Scribe: Yuchong Pan*

1 Markov Chains and Random Walks

Definition 1. Let Ω be a set of states (throughout this note, Ω is finite). A sequence of random walks $X_0, X_1, \ldots \in \Omega$ is a *Markov chain* if it satisfies the *Markovian property*, i.e., for each $t \in \mathbb{N}$ and for all $x_1, \ldots, x_t, y \in \Omega$,

$$\mathbb{P}\left[X_{t+1} \mid X_1 = x_1, \ldots, X_t = x_t\right] = \mathbb{P}\left[X_{t+1} = y \mid X_t = x_t\right].$$

WLOG, we assume that transitions are independent of time. For $x, y \in \Omega$, let

$$P(x, y) = \mathbb{P}\left[X_{t+1} = y \mid X_t = x\right].$$

Interpreted as a matrix, P is called the *transition matrix* of the Markov chain. We can also interpret the transition matrix P as a weighted directed graph with vertex set Ω such that the weight on $(i, j) \in \Omega^2$ equals $P(i, j)$.

A random walk on a directed graph is a special case of Markov chains.

Definition 2. A *random walk* on a directed graph $G = (V, E)$ is a sequence $S_1, S_2, \ldots \in V$ such that S_{t+1} is picked uniformly in $N^+(S_t)$, i.e., the transition matrix P is defined so that for $x, y \in V$,

$$P(x, y) = \begin{cases} \frac{1}{d^+(x)}, & \text{if } (x, y) \in E, \\ 0, & \text{otherwise.} \end{cases}$$

Definition 3. Let P be an $n \times n$ matrix. Then P is said to be *stochastic* if for all $i \in [n]$,

$$\sum_{j=1}^{n} P(i, j) = 1.$$

Moreover, P is said to be *doubly stochastic* if P is stochastic and for all $j \in [n]$,

$$\sum_{i=1}^{n} P(i, j) = 1.$$

For each $t \in \mathbb{N}$, let $P_t(x, y)$ ve the transition probability from x to y for t steps. Then for all $x, y \in \Omega$ and $t \in \mathbb{N}$,

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

Interpreted as matrix multiplication, for each $t \in \mathbb{N}$ with $t > 1$,

$$P^t = P \cdot P^{t-1}.$$

Let $\pi^{(0)} = (\pi_1^{(0)}, \ldots, \pi_n^{(0)})$ be the initial distribution, where $\pi_i^{(0)}$ is the probability of starting at vertex i for each $i \in [n]$.[1] Let $\pi^{(t)}$ be the distribution after t steps for each $t \in \mathbb{N}$. For each $t \in \mathbb{N}$,

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

[1] WLOG, we assume $\Omega = [n]$.

Definition 4. A distribution π^* is called a *stationary distribution* of a Markov chain with state set Ω and transition matrix P if for all $x \in \Omega$,

$$\pi^*(x) = \sum_{y \in \Omega} \pi^*(y) P(y, x).$$

Definition 5. A Markov chain with state set Ω and transition matrix P is said to be *irreducible* if for all $x, y \in \Omega$, there exists $t \in \mathbb{N}$ such that $P^t(x, y) > 0$.

Definition 6. A Markov chain with state set Ω and transition matrix P is said to be *aperiodic* if for all $x \in \Omega$,

$$\gcd \left\{ t \in \mathbb{N} : p^t(x, x) > 0] \right\} = 1.$$

Definition 7. A Markov chain with state set Ω and transition matrix P is said to be *ergodic* if there exists $t^* \in \mathbb{N}$ such that for all $t \in \mathbb{N}$ with $t > t^*$ and for all $x, y \in \Omega$, we have $P^t(x, y) > 0$.

Theorem 8. *Every ergodic Markov chain has a unique stationary distribution.*

In the special case of a random walk on an undirected graph $G = (V, E)$, the stationary distribution $\pi^* = (\pi_1^*, \ldots, \pi_n^*)$ is given by $\pi_i^* = d(i)/(2|E|)$ for all $i \in [n]$. Therefore, for a random walk on a d-regular graph or on a directed graph with each in-degree and each out-degree equal to d, the stationary distribution is uniform; this is not true in general directed graphs.

2 Hitting Time, Cover Time and Commute Time

Definition 9. Consider a random walk on a graph $G = (V, E)$. For $x, y \in V$, the *hitting time* $H_{x,y}$ is defined to be the expected number of steps to go from x to y. For each $x \in V$, we call $H_{x,x}$ the *recurrence time* for x.

Theorem 10. *Consider a random walk on a graph $G = (V, E)$ with stationary distribution π^*. For each $x \in V$,*

$$h_{x,x} = \frac{1}{\pi_*(x)}.$$

Proof sketch. Consider a very long walk. Then a $\pi^*(x)$ fraction of the positions are x. Then the average gap between the occurrences of x is $h_{x,x} = \pi^*(x)^{-1}$. $\square$

Definition 11. Consider a random walk on a graph $G = (V, E)$. For $u \in V$, the *cover time* $C_u(G)$ is defined to be the expected steps from u to visit all states in Ω. Define $C(G) = \max_{u \in V} C_u(G)$.

Following are several examples of the cover time:

- $C(K_n) = \Theta(n \log n)$, where K_n is the complete graph on n vertices with a self-loop at each vertex. This can be proved by a coupon collector argument.
- $C(L_n) = \Theta(n^2)$, where L_n is the n-vertex line graph with a self-loop at each vertex.
- $C(\text{lollipop}_n) = \Theta(n^3)$, where lollipop_n is an n-vertex lollipop vertex formed by $L_{n/2}$ and $K_{n/2}$ joined at a vertex. This is illustrated in Figure 1.

Theorem 12. *Let G be an undirected graph. Then*[2]

$$C(G) \leq O(mn).$$

[2] When the context is clear, we denote $m = |E|$ in a graph $G = (V, E)$.

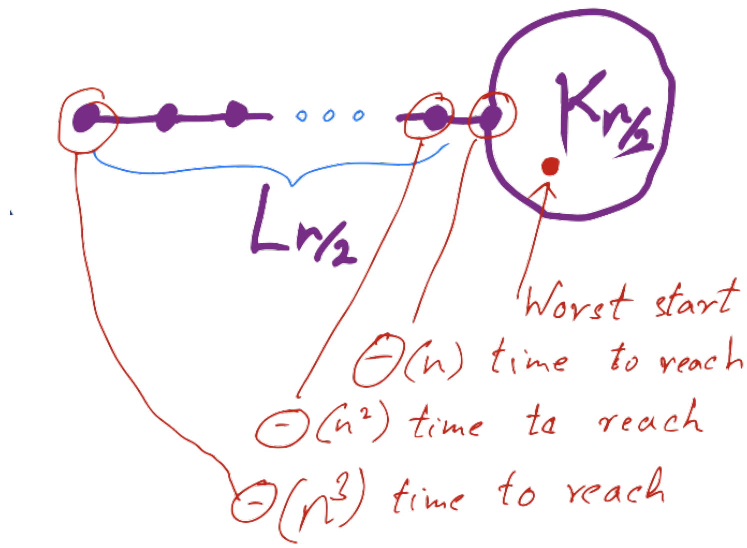

Figure 1: A lollipop graph lollipop$_n$ and its cover time.

Definition 13. Consider a random walk on a graph $G = (V, E)$. For $x, y \in V$, the *commute time* $C_{x,y} = C_{x,y}(G)$ is defined to be the expected number of steps for the random walk to start at x, hit y and return to x.

Proposition 14. *For $x, y \in V$,*

$$C_{x,y} = h_{x,y} + h_{y,x}.$$

Proof. This is due to linearity of expectation. □

Lemma 15. *Consider a random walk on a connected undirected graph $G = (V, E)$. For each $(x, y) \in E$,*

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

Proof. Construct a graph G' by adding a self-loop at each vertex with probability $1/2$. Let $x, y \in V$. We claim that $C_{x,y}(G') = 2C_{x,y}(G)$. To see this, for each path from x to y in G', removing the self-loops in the path gives a path in G, and the expected fraction of self-loops in the path is $1/2$. Then G' is ergodic. This implies that there exists a unique stationary distribution π^*.

Consider a walk $u_1, u_2, \ldots$, where $u_i \in V$ and $(u_i, u_{i+1}) \in E$ for each $i \in \mathbb{N}$. We look for commutes of the form

$$x \to y \to \ldots \to x \to y.$$

For each $i \in \mathbb{N}$,

$$\mathbb{P}[u_i = x, u_{i+1} = y] = \mathbb{P}[u_i = x] \cdot \mathbb{P}[u_{i+1} = y \mid u_i = x] = \frac{d(x)}{2m} \cdot \frac{1}{d(x)} = \frac{1}{2m}.$$

Therefore, the expected fraction of $x \to y$ equals $1/(2m)$. This implies that the expected gap between the $(x \to y)$'s equals $2m$. This proves that $C_{x,y}(G) = O(m)$. □

Proof of Theorem 12. Let T be a spanning tree of G. Let $(v_0, v_1, \ldots, v_{2n-2})$ be a DFS traversal of T. For instance, $(1,2,3,2,4,2,1,5,1)$ is a DFS traversal of the tree given in Figure 2. Then

$$C(G) \le \sum_{i=0}^{2n-3} h_{v_i,v_{i+1}} = \sum_{(x,y)\in E(T)} C_{x,y} \le (n-1)\cdot O(m) = O(mn).$$

This completes the proof.

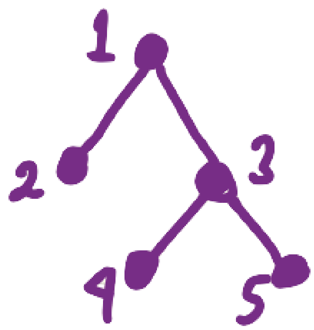

Figure 2: $(1,2,3,2,4,2,1,5,1)$ is a DFS traversal of the tree in the figure.

□

3 Unconnected s-t Connectivity

Problem 16 (undirected s-t connectivity, UST-Conn)**.** *Given an undirected graph $G = (V, E)$ and $s, t \in V$, output "yes" if s and t are in the same connected component of G, and "no" otherwise.*

Definition 17. Let RL be the class of problems solvable by randomized log-space computations.

The computational model that we consider consists of a read-only input tape of n bits and a read-write tape of $O(\log n)$ bits.

Theorem 18. UST-Conn $\in$ RL.

Proof. Let $G = (V, E)$ be an undirected graph. By Theorem 12, $C(G) = O(nm) = O(n^3) = cn^3$ for some constant c. We give Algorithm 1 for some parameter k.

```
starting at s, take a random walk for k · cn^3 steps
if ever see t then
    return "yes"
else
    return "no"
```

Algorithm 1: A randomized algorithm for UST-Conn on an undirected graph $G = (V, E)$ and vertices $s, t \in V$.

The running time of Algorithm 1 is $O(n^3)$ times the time to pick a random neighbor (which depends on the specific data structure used). For the space of Algorithm 1, we need to keep track of the step counter, which uses $O(\log n)$ space, and need to pick a random neighbor, which uses $O(\log n)$. Therefore, Algorithm 1 uses $O(\log n)$ space in total.

Now we analyze the behavior of Algorithm 1. If s and t are not connected, then the algorithm never outputs "yes." If s and t are connected, then $h_{s,t} \leq C(G_S) \leq n^3$, where G_S is the connected component of G that contains s and t. Therefore,

$$\begin{aligned}\mathbb{P}[\text{output "no"}] &\leq \mathbb{P}[\text{start at } s\text{, walk at least } k \cdot h_{s,t} \text{ steps and still don't see } t] \\ &= \mathbb{P}[\text{start at } s\text{, walk at least } k \cdot \mathbb{E}[\# \text{ steps to start at } s \text{ and see } t] \text{ steps, not see } t] \\ &\leq \frac{1}{k}.\end{aligned}$$

Note that the last inequality follows from Markov's inequality. This completes the proof. □

4 Mixing Time

We first review some definitions and results from linear algebra.

Definition 19. A vector v is said to be an *eigenvector* of a matrix A with corresponding *eigenvalue* λ if $vA = \lambda v$.

Definition 20. The L_2*-norm* of a vector $v = (v_1, \ldots, v_n)$ is defined to be $\sqrt{\sum_{i=1}^n v_i^2}$.

Definition 21. A set of vectors $v^{(1)}, \ldots, v^{(m)}$ is said to be *orthonormal* if for all $i, j \in [m]$,

$$v^{(i)} \cdot v^{(j)} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases}$$

where $v^{(i)} \cdot v^{(j)}$ is the *inner product* of $v^{(i)}$ and $v^{(j)}$, defined to be $\sum_{\ell=1}^n v_\ell^{(i)} v_\ell^{(j)}$.

Let P be the transition matrix of the random walk on a d-regular undirected graph. Then P is doubly stochastic. Therefore,

$$\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \cdot P = 1 \cdot \left(\frac{1}{n}, \ldots, \frac{1}{n}\right),$$
$$\left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right) \cdot P = 1 \cdot \left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right).$$

This shows that $(1/n, \ldots, 1/n)$ and $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$ are eigenvectors of P with eigenvalue 1. Note that the L_2-norm of $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$ equals 1.

Theorem 22. *Let P be an $n \times n$ transition matrix that is real and symmetric. Then there exist eigenvectors $v^{(1)}, \ldots, v^{(n)}$ that form an orthonormal basis with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ such that*

$$1 = \lambda_1 \geq |\lambda_2| \geq \ldots |\lambda_n|,$$

and that

$$v^{(1)} = \frac{1}{\sqrt{n}}(1, \ldots, 1).$$

Proposition 23. *Let P be a matrix with all positive entries, vectors $v^{(1)}, \ldots, v^{(n)}$, and corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$.*

(i) For all $\alpha \in \mathbb{R}$, αP has eigenvectors $v^{(1)}, \ldots, v^{(n)}$ and corresponding eigenvalues $\alpha\lambda_1, \ldots, \alpha\lambda_n$.

(ii) $P + I$ has eigenvectors $v^{(1)}, \ldots, v^{(n)}$ and corresponding eigenvalues $\lambda_1 + 1, \ldots, \lambda_n + 1$.

(iii) For all $k \in \mathbb{Z}_+$, P^k has eigenvectors $v^{(1)}, \ldots, v^{(n)}$ and corresponding eigenvalues $\lambda_1^k, \ldots, \lambda_n^k$.

(iv) If P is stochastic, then $|\lambda_i| \leq 1$ for all $i \in [n]$.

Note that (i) and (ii) in Proposition 23 imply that $(P + I)/2$ has eigenvectors $v^{(1)}, \ldots, v^{(n)}$ and corresponding eigenvalues $(\lambda_1 + 1)/2, \ldots, (\lambda_n + 1)/2$.

Proof. (i) Note that $vP = \lambda v$ if and only if $v \cdot \alpha P = \alpha\lambda \cdot v$.

(ii) Note that $v(P + I) = vP + vI = \lambda v + v = (\lambda + 1)v$.

(iii) Note that $vP^k = (vP)P^{k-1} = \lambda vP^{k-1} = \lambda(vP)P^{k-2} = \lambda^2 vP^{k-2} = \ldots = \lambda^k v$.

(iv) Let $i \in [n]$. Let $I = \{j \in [n] : v_j^{(i)} > 0\}$. Then

$$\begin{aligned}
\lambda_i \sum_{j \in I} v_j^{(i)} &= \sum_{j \in I} \sum_{k=1}^{n} v_k^{(i)} P_{k,j} \\
&\leq \sum_{j,k \in I} v_k^{(i)} P_{k,j} && \text{(entries of } v \text{ with coordinates not in } I \text{ are at most 0, } P_{k,j} \geq 0) \\
&= \sum_{k \in I} v_k^{(i)} \sum_{j \in I} P_{k,j} \\
&\leq \sum_{k \in I} v_k^{(i)}. && \left(\sum_{j \in I} P_{k,j} \leq 1 \text{ since } P \text{ is stochastic} \right)
\end{aligned}$$

Therefore, $\lambda_i \leq 1$.

□

Note that if $v^{(1)}, \ldots, v^{(n)}$ form a basis, then any vector w can be expressed as a linear combination of $v^{(1)}, \ldots, v^{(n)}$, i.e., $w = \sum_{i=1}^{n} \alpha_i v^{(i)}$ for some $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, and

$$\|w\|_2 = \sqrt{w \cdot w} = \sqrt{\sum_{i=1}^{n} \alpha_i v^{(i)} \sum_{j=1}^{n} \alpha_j v^{(j)}} = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j v^{(i)} v^{(j)}} = \sqrt{\sum_{i=1}^{n} \alpha_i^2}.$$

Note that the last equality follows from the orthonormality of $v^{(1)}, \ldots, v^{(n)}$.

Mixing times study the following question: How long does it take to reach the stationary distribution?

Definition 24. For $\varepsilon > 0$, the *mixing time* $T(\varepsilon)$ of a Markov chain A with stationary distribution π is the minimum $t \in \mathbb{Z}_+$ such that for all initial distribution $\pi^{(0)}$,

$$\left\| \pi - \pi^{(0)} A^t \right\|_1 < \varepsilon.$$

Theorem 25. *Let P be the transition matrix of the random walk on an undirected, d-regular and unconnected graph with the greatest common divisor of cycle lengths equal to 1. Let π_0 be an initial distribution. Let π be the stationary distribution equal to $(1/n, \ldots, 1/n)$ (so $\pi P = P$). Then*

$$\left\| \pi_0 P^t - \pi \right\|_2 \leq |\lambda_2|^2.$$

Note that the bound $|\lambda_2|^t$ is good (i.e., exponentially decreasing) if $1 - \lambda_2 = \Theta(1)$.