**CS 710: Complexity Theory** Date: Feb. 4th, 2020

# Lecture 5: Random Walk

Instructor: Jin-Yi Cai Scribe: Jialu Bao

## 1 Random Walks

A random walk in graph $G = (V, E)$ starts with a vertex $s \in V$ at time step 0. At each time step $i$, the random walk chooses a neighbor $v$ of the current vertex $u$ with some probability $P_{u,v}$ and jumps to $v$ at the next time step. For simplicity, we work with undirected simple graphs—simple meaning there is no multi-edge and no loop—and assume that the walk picks neighbors with equal probability.

Formally, let $y_i$ be the vertex that the random walk visits at time step $i$.

$$\begin{aligned} P_{u,v} &:= \Pr\left(y_{i+1} = v \mid y_i = u\right) \\ &= \begin{cases} \frac{1}{deg(u)} & \text{if } (u,v) \in E \\ 0 & \text{Otherwise} \end{cases} \end{aligned}$$

## 2 One dimensional Random Walk

Consider the following one dimemsional random walk on a chain of $n$ vertices starting from node 1 and will getting absorbed at node $n$—the random walk stops immediately after it hits node $n$.

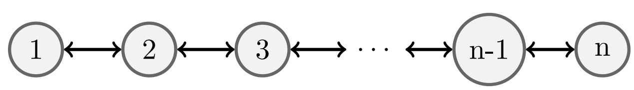


Figure 1: One-dimensional Walk

**Definition 1.** *Define* ***hitting time*** *of a random walk as the expected number of steps it takes from the starting state to the absorbing state.*

**Definition 2.** *Define* $h_{i,j} := \mathrm{E}[$*number of steps the random walk takes to go from node* $i$ *to node* $j$ *in Figure 1*

Then the hitting time of random walks in Figure 1 equals $h_{1,n}$.

Note that although the number of steps between any two states must be integral, the expected value of that number does not have to be integral.

**Lemma 1.** *For the random walk in Figure 1, for any* $i, j, k$ *such that* $1 \le j \le i \le k \le n$,

$$h_{j,k} = h_{j,i} + h_{i,k}$$

*Proof.* By the topology of Figure 1, every path going from node $j$ to node $k$ goes through node $i$. As random walks are memoryless, the number of steps in every path going from node $j$ to node $k$

is the sum of the number of steps it takes from node $j$ to $i$ and the number of steps it takes from node $i$ to $k$. Therefore,

$$\begin{aligned} h_{j,k} &= \mathrm{E}[\textbf{\# of steps to go from } j \textbf{ to } k] \\ &= \mathrm{E}[\textbf{\# of steps to go from } j \textbf{ to } i + \textbf{\# of steps to go from } i \textbf{ to } k] \\ &= \mathrm{E}[\textbf{\# of steps to go from } j \textbf{ to } i] + \mathrm{E}[\textbf{\# of steps to go from } i \textbf{ to } k] \\ &= h_{j,i} + h_{i,k} \end{aligned}$$

The second to the last step follows from the linearity of expectation, and the last step follows from the definition of $h_{j,i}$ and $h_{i,k}$. □

**Corollary 1.** *The hitting time,*

$$\begin{aligned} h_{1,n} &= h_{1,2} + h_{2,n} \\ &= h_{1,2} + h_{2,3} + h_{3,n} \\ &= \ldots \\ &= \sum_{i=1}^{i=n-1} h_{i,i+1} \end{aligned}$$

**Corollary 2.** *For the random walk in Figure 1, for any $i$ such that $2 \leq i \leq n-1$,*

$$h_{i-1,i+1} = h_{i-1,i} + h_{i,i+1}$$

**Lemma 2.** *For $2 \leq i \leq n-1$, $h_{i,i+1} = 1 + h_{i-1,i+1}$, and $h_{12} = 1$.*

*Proof.* Since we can only go node 2 from node 1, there is only one path from node 1 to 2 and that path has 1 step, so $h_{1,2} = 1$.

From node $i$ such that $2 \leq i \leq n-1$, the random walk can either goes to $i-1$ or $i+1$ for the next step, each with probability $\frac{1}{2}$. If it goes directly to $i+1$, then it reaches $i+1$ in 1 step. If it goes to $i-1$, then in expectation it takes another $h_{i-1,i+1}$ steps after that steps to reach node $i+1$. Thus,

$$\textbf{\# of steps to go from } i \textbf{ to } i+1 = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (1 + \textbf{\# of steps to go from } i-1 \textbf{ to } i+1)$$

Thus, by substituting it into the definition of $h_{i,j}$, we have

$$\begin{aligned} h_{i,i+1} &= \mathrm{E}[\textbf{\# of steps to go from } i \textbf{ to } i+1] \\ &= \mathrm{E}[\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (1 + \textbf{\# of steps to go from } i-1 \textbf{ to } i+1)] \\ &= \mathrm{E}[1 + \frac{1}{2}\textbf{\# of steps to go from } i-1 \textbf{ to } i+1)] \\ &= 1 + \frac{1}{2}\mathrm{E}[\textbf{\# of steps to go from } i-1 \textbf{ to } i+1)] \\ &= 1 + \frac{1}{2}h_{i-1,i+1} \end{aligned}$$

□

**Theorem 1.** *The hitting time, $h_{1,n}$, is $(n-1)^2$.*

*Proof.* Substitute $h_{i-1,i+1} = h_{i-1,i} + h_{i,i+1}$ from lemma 1 into $h_{i,i+1} = 1 + \frac{1}{2}h_{i-1,i+1}$, then we have,

$$h_{i,i+1} = 1 + \frac{1}{2}(h_{i-1,i} + h_{i,i+1}) \tag{1}$$

$$\implies h_{i,i+1} = 2 + h_{i-1,i} \tag{2}$$

Since $h_{1,2} = 1$, the recurrence solves to $h_{i,i+1} = 2i - 1$. Substitute it into the formula given by Corollary 1, the hitting time equals

$$h_{1,n} = \sum_{i=1}^{n-1} h_{i,i+1} = (n-1)^2$$

□

## 3 Expected number of times a walk ~~steps~~ visits a node

Now we consider a different problem of random walks, given a undirected graph $G = (V, E)$ and perform the random walk as described in Section 1, then what is the expected number of times a walk visits a node $i$? In the following, we use $t_i$ to denote that number and will try to calculate it.

Because node $n$ is the absorbing node and no other node is absorbing state, the walk will eventually visits $n$ and never steps out after that. Thus, without any computations, we can conclude that $t_n = 1$. Then we consider node $n-1$, every time the walk reaches node $n-1$, with probability $\frac{1}{2}$, it goes to node $n$, gets absorbed and never comes back; with probability $\frac{1}{2}$, it goes to node $n-1$, then either directly or after an indefinite number of steps, it comes back and stopst node $n-1$ again, and repeat the process. Thus, the walk visits a node for only once with probability $\frac{1}{2}$, visits a node twice with probability $\frac{1}{4}$, visits a node three times with probability $\frac{1}{8}$... That is

$$\begin{aligned}
t_{n-1} &= \sum_{k \geq 1} k \cdot \Pr(\text{the walk visits node } n-1 \text{ for } k \text{ times}) \\
&= \frac{1}{2}\sum_{k \geq 1} k \cdot (\frac{1}{2})^{k-1} = \frac{1}{2}\left[\sum_{k \geq 1}(x^k)\right]'_{x=\frac{1}{2}} \\
&= \frac{1}{2}\left[\frac{x}{1-x}\right]'_{x=\frac{1}{2}} = \frac{1}{2}\left[\frac{1}{(1-x)^2}\right]_{x=\frac{1}{2}} \\
&= \frac{1}{2}\left[\frac{1}{(1-\frac{1}{2})^2}\right] = 2
\end{aligned}$$

Alternatively, because the random walk is memoryless, every time it visits node $n-1$, the expected number of upcoming visits to node $n-1$ again stays the same. Thus, if the walk visits node $n-1$ once and then goes to node $n-2$ at the next steps, the expected number of times of additional visits to node $n-1$ is still $t_{n-1}$. Therefore, we have the recurrence $t_{n-1} = \frac{1}{2} \times 1 + \frac{1}{2} \times (1 + t_{n-1})$, which also gives $t_{n-1} = 2$.

How about $t_i$ for the rest of nodes?

**Theorem 2.** *For $3 \le i \le n-2$, $t_i = \frac{1}{2}(t_{i-1} + t_{i+1})$.*

*Proof.* We will first present the intuition and then a more rigorous proof.

Because our graph is just a single chain of nodes, for any node $i$ such that $3 \le i \le n-2$, the walk must passes either node $i-1$ or node $i+1$ en route to other node. Meanwhile, every time the walk visits node $i-1$, it has $\frac{1}{2}$ chance of going to node $i$ the next step; and similarly, every time the walk visits node $i+1$, it has $\frac{1}{2}$ chance of visiting node $i$ the next step. Thus, the expected number of times that the walk visits node $i$ equals $\frac{1}{2}(t_{i-1} + t_{i+1})$.

To prove it formally, we define random variable $W$ to be the number of steps a random walk takes from node 1 to node n, i.e., the length of time that a random walk is alive. We also define $0-1$ random variables $x_{j,i}$ for $i \ge 1, j \ge 0$,

$$x_{j,i} = \begin{cases} 1 & \text{if } W \ge j \text{ and } y_j = i \\ 0 & \text{Otherwise} \end{cases}$$

$W \ge j$ indicates that a random walk is alive at time $j$, which guarantees that $y_j$ is defined. $y_j$ is the node the random walk visited at the $j^{th}$ step. So by definition, $x_{j,i} = 1$ if and only if the walk has not yet ended at the $j^t h$ step and the walk is visiting node $i$ at $j^{th}$ step. Thus,

$$t_i = \mathrm{E}[\text{the number of visits to node } i] = \mathrm{E}[\sum_{j \ge 0} x_{j,i}]$$

Then, by linearity of expectation and the defintion of expected value,

$$\begin{aligned} t_i &= \sum_{j \ge 0} \mathrm{E}[x_{j,i}] \\ &= \sum_{j \ge 0} \Pr[W \ge j \text{ and } y_j = i] \end{aligned}$$

Since the walk starts at node 1 and $i \ge 3$, there is no chance to visit node $i$ in the $0^{th}$ and $1^{st}$ step, so

$$\sum_{j \ge 0} \Pr[W \ge j \text{ and } y_j = i] = \sum_{j \ge 1} \Pr[W \ge j \text{ and } y_j = i] = \sum_{j \ge 2} \Pr[W \ge j \text{ and } y_j = i] \tag{3}$$

Thus,

$$\begin{aligned} t_i &= \sum_{j \ge 0} \Pr[W \ge j \text{ and } y_j = i] = \sum_{j \ge 1} \Pr[W \ge j \text{ and } y_j = i] \\ &= \sum_{j \ge 1} \Pr[W \ge j \text{ and } y_{j-1} = i-1 \text{ and } y_j = i] + \sum_{j \ge 1} \Pr[W \ge j \text{ and } y_{j-1} = i+1 \text{ and } y_j = i] \\ &= \sum_{j \ge 1} \Pr[W \ge j \text{ and } y_{j-1} = i-1] \cdot \Pr[y_j = i \mid y_{j-1} = i-1] + \\ &\qquad \sum_{j \ge 1} \Pr[W \ge j \text{ and } y_{j-1} = i+1] \cdot \Pr[y_j \mid y_{j+1}] \\ &= \sum_{j \ge 1} \Pr[W \ge j \text{ and } y_{j-1} = i-1] \cdot \frac{1}{2} + \sum_{j \ge 1} \Pr[W \ge j \text{ and } y_{j-1} = i+1] \cdot \frac{1}{2} \end{aligned}$$

Since $i \leq n-2$, the random walk had to continue at least one more steps if it visits node $i+1$ or node $i-1$ at $(j-1)^{th}$ step, so

$$\Pr[W \geq j-1 \text{ and } y_{j-1} = i+1] = \Pr[W \geq j \text{ and } y_{j-1} = i+1] \quad (4)$$
$$\Pr[W \geq j-1 \text{ and } y_{j-1} = i-1] = \Pr[W \geq j \text{ and } y_{j-1} = i-1] \quad (5)$$

Thus,

$$\begin{aligned}
t_i &= \sum_{j\geq 1} \Pr[W \geq j \text{ and } y_{j-1} = i-1] \cdot \frac{1}{2} + \sum_{j \geq i} \Pr[W \geq j+1 \text{ and } y_{j+1} = i+1] \cdot \frac{1}{2} \\
&= \sum_{j\geq 1} \Pr[W \geq j-1 \text{ and } y_{j-1} = i-1] \cdot \frac{1}{2} + \sum_{j\geq 1} \Pr[W \geq j-1 \text{ and } y_{j-1} = i+1] \cdot \frac{1}{2} \textbf{ (By 4,5)} \\
&= \sum_{j-1\geq 1} \Pr[W \geq j-1 \text{ and } y_{j-1} = i-1] \cdot \frac{1}{2} + \sum_{j-1\geq 1} \Pr[W \geq j-1 \text{ and } y_{j-1} = i+1] \cdot \frac{1}{2} \textbf{ (By 3)} \\
&= t_{i-1} \cdot \frac{1}{2} + t_{i+1} \cdot \frac{1}{2}
\end{aligned}$$

□

Similarly, $t_1 = \frac{1}{2}t_2 + 1$ because every walk pays one visit to node 1 at the beginning, and after that, it has to go through node 2 once every time it gets to node 1. For node 2, $t_2 = \frac{1}{2}t_3 + t_1$ because every visit to node 2 must come from either node 1 or node 3 at the last step, and from node 1 it always goes to node 2 while from node 3 it goes to node 2 with probability $\frac{1}{2}$. Substituting $t_1 = \frac{1}{2}t_2 + 1$ into $t_2 = \frac{1}{2}t_3 + t_1$, we get $t_2 = t_3 + 2$.

For $3 \leq i \leq n-2$, $t_i = \frac{1}{2}[t_{i-1} + t_{i+1}]$ implies that $t_{i+1} - t_i = t_i - t_{i-1}$. Thus, $t_2 - t_3 = t_3 - t_4 = ... = t_{n-2} - t_{n-1}$. Since $t_2 = t_3 + 2$,

$$t_2 - t_3 = t_3 - t_4 = ... = t_{n-2} - t_{n-1} = 2$$

Since $t_{n-1} = 2$, we can solve that $t_i = 2(n-i)$ for $2 \leq i \leq n-1$. In particular, $t_2 = 2(n-2)$, implying that $t_1 = \frac{1}{2}t_2 + 1 = n-1$.

We can do a sanity check of our calculated result, the sum of $t_i$ for all $i$ should be exactly one more than the hitting time.

$$\begin{aligned}
\sum_{i=1}^{n} t_i &= t_1 + \sum_{i=2}^{n-1} t_i + t_{n-1} + t_n \\
&= (n-1) + \left(\sum_{i=2}^{n-2} 2(n-i)\right) + 2 + 1 \\
&= (n-1) + (n^2 - 3n) + 2 + 1 \\
&= n^2 - 2n + 2 \\
&= (n-1)^2 + 1
\end{aligned}$$

Thus, the sum of our calculated $t_i$ is exactly the hiting time plus one.