CSCI-B609: A Theorist's Toolkit, Fall 2016 20 Sep 2016

Lecture 09: Mixing Time of Random Walks

Lecturer: Yuan Zhou *Scribe: Erfan Sadeqi Azer*

1 1-D Random Walks: a Second Look

Recall that we defined a random walk on a bounded line with set of nodes $\{0, 1, \ldots, n\}$. There is a trap at 0 and a wall after nth node. A drunkard starts at an arbitrary node, say x_0. Then he takes one step at a time, indexed by i and goes form x_{i-1} to x_i with transition probabilities: $x_i = \begin{cases} x_{i-1} - 1 & w.p. 1/2 \\ x_{i-1} + 1 & w.p. 1/2 \end{cases}$, if $x_{i-1} < n$. Otherwise if $x_{i-1} = n$, then $x_i = n - 1$ w.p. 1.

We can consider a parallel universe in order to avoid having the constrain $x_i = n$, and still preserve the properties we would like to have. So, we define a new random walk on a bounded line with set of nodes $\{0, 1, \ldots, n, n+1, \ldots, 2n\}$. There are traps in both 0 and $2n$, and no wall. Starting from x'_0, the transition probabilities are $x'_i = \begin{cases} x'_{i-1} - 1 & w.p. 1/2 \\ x'_{i-1} + 1 & w.p. 1/2 \end{cases}$, for all values of x'_{i-1}. This random walks ends at time step T when either $x'_T = 0$ or $x'_T = 2n$ for the first time.

Notice that given second random walk, we can define $x_i = min\{x'_i, 2n - x'_i\}$ which satisfies the transition probabilities for the first random walk. Specifically, when $x'_{i-1} = n$ next node will be either $n - 1$ or $n + 1$. Both of these nodes correspond to the same node $n - 1$ in the first random walk.

Last session, we proved: $\mathrm{E}[\text{time to fall in the trap}] \leq n^2$. Let us prove asymptotically the same bound using this new random walk.

Lemma 1. *$\Pr[T_{x_0} \geq m] = O(n^2)$.*

Proof. Assume m is even to make sure the drunkard can go back to its starting place. Computations for the case m is odd are similar. Also, **Sterling's** formula $m! \sim \sqrt{2\pi m}(\frac{m}{e})^m$ is used at the end of the following computations.

$$\Pr[T_{x_0} \geq m] \leq \Pr[X'_m \in (0, 2n)] = \Pr[(\sum_{i=1}^{m} b_i) \in (-x_0, 2n - x_0)]$$

$$\leq \Pr[|\sum_{i=1}^{m} b_i| < 2n] = \sum_{j=-2n+1}^{2n-1} \Pr[\sum_{i=1}^{m} b_i = j] = \sum_{j=-n+1}^{n-1} \Pr[\sum_{i=1}^{m} b_i = 2j]$$

$$= \sum_{j=-n+1}^{n-1} \frac{\binom{m}{\frac{m}{2}+j}}{2^m} \leq 2n \frac{\binom{m}{\frac{m}{2}}}{2^m} \sim \frac{2\sqrt{2}n}{\sqrt{\pi m}} = \Theta(\frac{n}{\sqrt{m}}).$$

Now, if we pick a big enough m, like $12n^2$: The value $\frac{2\sqrt{2}n}{\sqrt{\pi m}}$ can be upper-bounded by $1/2$. This proves $\Pr[T_{x_0} \geq 12n^2] \leq \frac{1}{2}$. Similar to the argument from last session, we can use the independence of random transitions in consecutive $12n^2$ intervals to get:

$$\Pr[T_{x_0} \geq t \cdot (12n^2)] \leq \frac{1}{2^t}.$$

This can be used to compute the expectation of T_{x_0} as follows:

$$\mathrm{E}[T_{x_0}] \leq \sum_{t=0}^{+\infty} \Pr[T_{x_0} \geq t \cdot (12n^2)] \cdot 12n^2 \leq 12n^2 \sum_{t=0}^{+\infty} \frac{1}{2^t} = 24n^2.$$

That completes the proof. □

This proof provides more intuition about the distribution of the position of drunkard in the line. At time step j, the drunkard will be at position $2j$ with probability $\frac{\binom{m}{\frac{m}{2}+j}}{2^m}$. Expectedly, x'_0 is the mode of this distribution. Also, the magnitude of the probability at x'_0 is in order of $\Theta(\frac{1}{\sqrt{m}})$. We can get more information about this distribution by looking at the ratio of the probability of being at position $2j$ to the maximum value. Formally, we have:

$$\frac{\binom{m}{\frac{m}{2}+j}}{\binom{m}{\frac{m}{2}}} = \frac{(\frac{m}{2}) \cdot (\frac{m}{2}-1) \dots (\frac{m}{2}-j+1)}{(\frac{m}{2}+1) \cdot (\frac{m}{2}+2) \dots (\frac{m}{2}+j)} = 1 - \frac{1}{m}\Theta(j^2).$$

For values of j bounded by $\sqrt{m}$, this ratio is bounded by a constant, i.e. $\Omega(1)$. It means that the distribution of the position of drunkard in the interval $(-\sqrt{m}, \sqrt{m})$ around initial position is uniform up to constant factors.

In the next section we generalize random walk concepts to spaces more complicated than 1-D line.

2 Random Walk on Graphs

Let $G = (V, E)$ be an undirected connected graph, not necessarily regular. Random walk matrix of G defined as: $W_G = D^{-1/2} A_G D^{-1/2}$, where A_G is the adjacency matrix of G and

$D = \text{diag}(\deg(v_1), \ldots, \deg(v_n))$. Let $\vec{p} = D^{-1/2}\vec{\pi}$, and $L_G = I - W_G$, called *Normalized Laplacian.*

We prove that $\vec{p}$ is the stationary distribution of W_G:

$$W_G \cdot \vec{p} = D^{-1/2} A_G D^{-1/2} D^{-1/2} \vec{\pi} = D^{-1/2}(A_G D^{-1} \vec{\pi}) = D^{-1/2}\vec{\pi} = \vec{p}.$$

This also means $\vec{p}$ is in null space of L_G. Therefore:

$$\vec{p}^T L_G \vec{p} = \sum_{(u,v)\in E} \left(\frac{\vec{p}(u)}{\sqrt{\deg(u)}} - \frac{\vec{p}(v)}{\sqrt{\deg(v)}} \right)^2 = 0.$$

It implies that $\frac{\vec{p}(u)}{\sqrt{\deg(u)}} = \frac{\vec{p}(v)}{\sqrt{\deg(v)}}$, for all $(u, v) \in E$. Thus,

$$\exists c \text{ s.t. } \forall u \in V : \vec{p}(u) = c\sqrt{\deg(u)} \Rightarrow \vec{\pi}(u) = c \cdot \deg(u) = \frac{1}{2|E|} \deg(u).$$

2.1 Mixing time

Recall that if G is connected and non-bipartite, $\lim_{t\to\infty} W_G^t \vec{p} = \vec{\pi}$. We want to study the convergence rate for this limit. First we need to define the measure of "closeness" between $W_G^t \vec{p}$ and $\vec{\pi}$.

Definition 2. Total variation distance *between two distributions* $\vec{p}$ *and* $\vec{q}$ *is defined as follows:*

$$d_{TV}(\vec{p}, \vec{q}) = \frac{1}{2} \sum_a |\vec{p}(a) - \vec{q}(a)| = \frac{1}{2} ||\vec{p} - \vec{q}||_1.$$

Recall that $||\vec{x}||_p$, called "*p-norm*" is defined as $(\sum_a |\vec{x}(a)|^p)^{1/p}$.

Lemma 3. *For any two distributions* $\vec{p}$ *and* $\vec{q}$*, the inequality* $d_{TV}(\vec{p}, \vec{q}) \leq \frac{\sqrt{n}}{2} ||\vec{p} - \vec{q}||_2$ *holds.*

Proof. We use Cauchy–Schwarz inequality

$$d_{TV}(\vec{p}, \vec{q}) = \frac{1}{2} \sum_a |\vec{p}(a) - \vec{q}(a)| \cdot 1 \leq \frac{1}{2} \sqrt{\sum_a (\vec{p}(a) - \vec{q}(a))^2} \cdot \sqrt{\sum_a 1} = \frac{\sqrt{n}}{2} ||\vec{p} - \vec{q}||_2.$$

□

Definition 4. *The* δ-mixing time *of* G *is at most* T *if for any initial distribution* $\vec{p_0}$*:*

$$d_{TV}(W_G^T \vec{p_0}, \vec{\pi}) \leq \delta$$

Let us consider some special cases. For example mixing time for an n-clique is low. We can see that even right after one step the distribution of the random walk is almost the stationary distribution. In contrast, the mixing time is high for a path of length n, where the random walk needs at least n steps to have a chance to to get to the other end of graph and needs n^2 steps in expectation to hit the other side for first time.

One more example that gives better intuition is the graph where there are two partitions. Each partition is highly dense within itself, but there is only one edge between two partitions. The expected number of steps to hit that edge to pass the bridge to the other partition is high. This leads to the concept of conductance.

In the first glance, we might relate this measure to $\frac{1}{\lambda_2(L_G)}$. However $K_{n,n}$ is an counterexample, where the corresponding mixing tie is high, with relatively high $\lambda_2(L_G)$. The following lemma fixes this issue:

Lemma 5. *Let $\lambda = \min\{\lambda_2(L_G), 2 - \lambda_n(L_G)\}$: δ-mixing time of G is $O(\log(n/\delta)/\lambda)$.*

Proof. In order to have $d_{TV}(W_G^t\vec{p_0}, \vec{\pi}) \leq \delta$, we only need $||W_G^t\vec{p_0} - \vec{\pi}||_2^2 \leq (\frac{2\delta}{n})^2$ due to Lemma 4.

Assume G is regular: write $W_G = \sum_{i=1}^n (1-\lambda_i)\psi_i\psi_i^T$. Then,

$$W_G^t\vec{p_0} = \sum_{i=1}^n (1-\lambda_i)^t\psi_i\psi_i^T.$$

Remember that $\psi_1 = \sqrt{n}\vec{\pi}$.

$$= n\vec{\pi}\vec{\pi}^T\vec{p_0} + \sum_{i=2}^n (1-\lambda_i)^t\psi_i\psi_i^T\vec{p_0} = \vec{\pi} + \sum_{i=2}^n (1-\lambda_i)^t\psi_i\psi_i^T\vec{p_0}.$$

$$||W_G^t\vec{p_0} - \vec{\pi}||_2^2 = \sum_{i=2}^n (1-\lambda_i)^{2t}(\psi_i^T\vec{p_0})^2 \leq \sum_{i=2}^n (1-\lambda_i)^{2t} \leq \sum_{i=2}^n (1-\lambda)^{2t} \leq n(1-\lambda)^{2t}.$$

Pick $t = c(\log(n/\delta)/\lambda$. Then, $n(1-\lambda)^{2t} \leq (2\delta/n)^2$.

□

Next session, we will talk about mixing time and related concepts. Specifically, we will formally define conductance to measure how fast a random walk can travel between two arbitrary nodes.