ORIE 6334 Spectral Graph Theory	September 22, 2016
Lecture 11	
Lecturer: David P. Williamson	*Scribe: Pu Yang*

In today's lecture we will focus on discrete time random walks on undirected graphs. Specifically, we will study random walks on an undirected graph $G = (V, E)$, where the time proceeds in unit steps: $t = 1, 2, \ldots$. At each time t, the walk is at some node $i \in V$, and at time $t+1$, the walk will choose one of i's neighbors uniformly at random and move to that neighbor. We will also study another kind of transition where at each time, the walk will stay at the current node i with probabilty $1/2$ and move to a neighbor of i uniformly at random with probability $1/2$. A random walk with such transition probability is called the *lazy random walk*.

Here is a simple example of a random walk on an undirected graph. The number besides each node is the probability the walk is at that node at each time t.

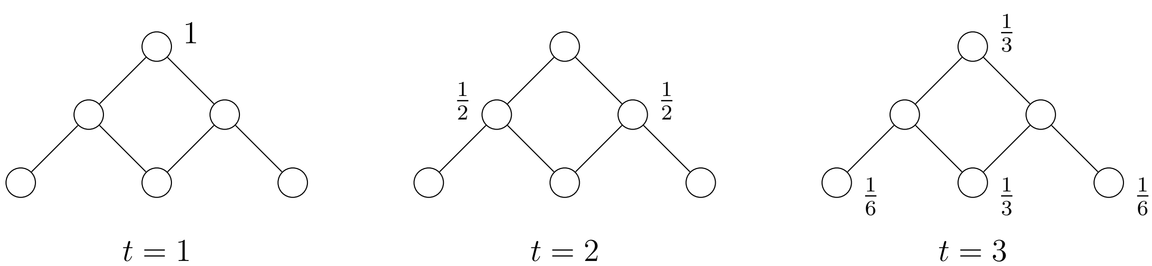

Figure 1: A simple example of a random walk

We are interested in understanding some long term behavior of the walk. For example, under some conditions, a random walk will converge to a probability distribution which is called the stationary distribution. In this lecture, we will use spectral theory to come up with answers to the following questions: what is the stationary distribution? Under what condition will the walk converge? How long does it take to converge? The time the walk takes to converge is called the *mixing time* of the walk.

We will first formally define the random walk. Let p_t be the probability distribution of the position of the walk at time t (that is, $p_t(i)$ is the probability the walk is at node i at time t). Then given the distribution p_t of the walk at time t, the probability the walk is at a node $i \in V$ at time $t + 1$ is

$$p_{t+1}(i) = \sum_{j:(i,j)\in E} p_t(j) \frac{1}{d(j)},$$

[0]This lecture is derived from Lap Chi Lau's lecture notes of Lecture 7 of *CS798: Algorithmic Spectral Graph Theory, Fall 2015* at University of Waterloo, `https://cs.uwaterloo.ca/~lapchi/cs798/notes/L07.pdf`.

where $d(j)$ is the degree of node j.

Let A be the adjacency matrix of G, and let

$$D = \begin{pmatrix} d(1) & & & 0 \\ & \ddots & & \\ & & & \\ 0 & & \ddots & \\ & & & d(n) \end{pmatrix}.$$

Then the above transition relation can be written in matrix form as

$$p_{t+1} = AD^{-1} p_t = (AD^{-1})^{t+1} p_0,$$

where p_0 is the initial distribution of the location of the walk.

We define the stationary distribution as follows.

Definition 1 *A probability distribution π over the set of nodes V of a graph $G = (V, E)$ is a stationary distribution of the random walk if $\pi = (AD^{-1})\pi$.*

Observe that $\pi = \frac{d}{2m}$ where $d = (d(1), d(2), \ldots, d(n))$ is a stationary distribution since

$$AD^{-1} \frac{d}{2m} = A \frac{e}{2m} = \frac{d}{2m}.$$

Here e is the vector of all 1's. Note that this implies the stationary distribution is the eigenvector of AD^{-1} corresponding to the eigenvalue 1.

One question we are interested in is whether the walk will converge to the stationary distribution in the long run. We say such walks are *ergodic*.

Definition 2 *A random walk is* ergodic *if there exists a distribution π such that for all initial distributions p_0, $\lim_{t \to \infty} p_t = \pi$.*

Under what conditions is the walk ergodic? The main theorem we will prove today is the following.

Theorem 1 *A random walk on a graph G is ergodic if and only if G is connected and not bipartite.*

The necessity is straightforward. G must be connected since otherwise two initial distributions which have all mass in two different components respectively will not converge to the same limiting distribution. To see why G cannot be bipartite, let L, R be the bipartition. If the walk starts in L at $t = 0$, then it will always be in L in even steps and in R in odd steps, and the limiting distribution does not exist.

The above analysis is for random walks which always moves to a neighbor at each time. For the lazy random walk, we have that

$$p_{t+1}(i) = \frac{1}{2} p_t(i) + \frac{1}{2} \sum_{j:(i,j) \in E} p_t(j) \frac{1}{d(j)}.$$

Written in matrix form this becomes

$$p_{t+1} = (\frac{1}{2}I + \frac{1}{2}AD^{-1})p_t.$$

We write $W \equiv \frac{1}{2}I + \frac{1}{2}AD^{-1}$, so that $p_{t+1} = W^{t+1}p_0$. It is easy to see $\pi = \frac{d}{2m}$ is also a stationary distribution for the lazy random walk.

For lazy random walks to be ergodic, the requirement for G not being bipartite is no longer necessary since at each time with probability 1/2 the walk will stay in the same part instead of jumping between the two parts alternatively. It is clear that connectivity is still needed. So the theorem we will show is as follows.

Theorem 2 *A lazy random walk on a graph G is ergodic if and only if G is connected.*

We first prove the two theorems on regular graphs (a d-regular graph is a graph such that the degree of all nodes is d), then explain how to extend the result to general non-regular graphs.

For a d-regular graph, $AD^{-1} = \frac{1}{d}A = \mathscr{A}$, where $\mathscr{A}$ is the normalized adjacency matrix of G (and for the lazy random walk we have $W = \frac{1}{2}I + \frac{1}{2}\mathscr{A}$). The stationary distribution is

$$\pi = \frac{d \cdot e}{2m} = \frac{d \cdot e}{nd} = \frac{e}{n},$$

so that each node is equally likely.

Proof of Theorem 1 and Theorem 2 for d-regular graphs: Let $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$ be the eigenvalues of $\mathscr{A}$ and $x_1, x_2, \ldots, x_n$ be corresponding orthonormal eigenvectors. Recall from Lectures 7 and 8 we have the following results:

- $\alpha_1 = 1$, $x_1 = \frac{e}{\sqrt{n}}$.
- If G is connected, $\alpha_2 < 1$.
- $\alpha_n \geq 1$. Moreover, G is not bipartite if and only if $\alpha_n > -1$.

Since x_i's form an orthonormal basis of $\mathbb{R}^n$, for any initial distribution p_0 we can write p_0 as $p_0 = c_1x_1 + \cdots + c_nx_n$ where $c_i = \langle x_i, p_0 \rangle$ for $i = 1, 2, \ldots, n$. Thus,

$$p_t = (AD^{-1})^t p_0 = \mathscr{A}^t p_0 = \mathscr{A}^t \left(\sum_{i=1}^{n} c_i x_i \right) = \sum_{i=1}^{n} c_i \alpha_i^t x_i.$$

Note that only the sufficiency of the condition needs to be shown. Assume G is connected and non-bipartite (for an ordinary random walk). Then from above we have $\alpha_2 < 1$ and $\alpha_n > -1$. Therefore $|\alpha_i| < 1$ hence $\alpha_i^t \to 0$ as $t \to \infty$ for all $i = 2, \ldots, n$ and

$$\lim_{t\to\infty} \mathscr{A}^t P_0 = \lim_{t\to\infty} \sum_{i=1}^{n} c_i \alpha_i^t x_i = \sum_{i=1}^{n} \lim_{t\to\infty} c_i \alpha_i^t x_i = c_1 \alpha_1 x_1.$$

Since $\alpha_1 = 1$ and $x_1 = \frac{e}{\sqrt{n}}$, we have

$$c_1 = \langle p_0, x_1 \rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} p_0(i) = \frac{1}{\sqrt{n}},$$

since p_0 is a probability distribution on the nodes. Hence $c_1\alpha_1 x_1 = \frac{1}{\sqrt{n}} \cdot 1 \cdot \frac{e}{\sqrt{n}} = \frac{e}{n}$, as desired for ordinary random walks.

For the lazy random walk, the proof is similar. Let $w_1 \geq w_2 \geq \cdots w_n$ be eigenvalues of $W = \frac{1}{2}I + \frac{1}{2}\mathscr{A}$. Then we have $w_i = \frac{1}{2} + \frac{1}{2}\alpha_i$. If G is connected, then $1 = \alpha_1 > \alpha_2 \geq \cdots \geq \alpha_n \geq -1$, so $w_1 = 1$ and $0 \leq w_n \leq w_{n-1} \leq \cdots \leq w_2 < 1$. Also W has the same eigenvectors as $\mathscr{A}$ so using the same argument for ordinary random walks we have $\lim_{t\to\infty} W^t p_0 = \frac{e}{n}$. □

Next we look at how long the walk will converge to the stationary distribution on a d-regular graph, i.e., the mixing time. We first define the meaning of "convergence" we will be using here.

Definition 3 *For probability distributions p and q on V, the* total variation distance between p and q is defined as

$$d(p, q) \equiv \sum_{i=1}^{n} |p(i) - q(i)| = \|p - q\|_1.$$

Definition 4 *The (total variation)* mixing time is the smallest t such that $d(p_t, \pi) \leq \frac{1}{4}$.

What is the mixing time for a d-regular graph? As before, assume the graph G is connected and non-bipartite so $\alpha_2 < 1$ and $\alpha_n > -1$. Let $\alpha \equiv \min(1 - \alpha_2, 1 - |\alpha_n|)$ be the *spectral gap* of $\mathscr{A}$. Since

$$p_t = \mathscr{A}^t p_0 = \sum_{i=1}^{n} c_1 \alpha_i x_i = \frac{e}{n} + \sum_{i=2}^{n} c_i \alpha_i x_i,$$

Then we have that

$$\begin{aligned} d(P_t, \pi) &= d(\mathscr{A}^t p_0, \frac{e}{n}) \\ &= \left\| \mathscr{A}^t p_0 - \frac{e}{n} \right\|_1 \\ &= \left\| \sum_{i=2}^{n} c_i \alpha_i^t x_i \right\|_1 \\ &\leq \sqrt{n} \left\| \sum_{i=2}^{n} c_i \alpha_i^t x_i \right\|_2, \end{aligned}$$

where the last inequality is from the Cauchy-Schwartz inequality. Note that

$$\begin{aligned}\left\|\sum_{i=2}^{n} c_i\alpha_i^t x_i\right\|_2^2 &= (c_2\alpha_2^t x_2 + \cdots + c_n\alpha_n^t x_n)^T(c_2\alpha_2^t x_2 + \cdots + c_n\alpha_n^t x_n)\\ &= \sum_{i=2}^{n} c_i^2\alpha_i^{2t},\end{aligned}$$

and from the definition of the spectral gap α we have

$$\begin{aligned}\sum_{i=2}^{n} c_i^2\alpha_i^{2t} &\le (1-\alpha)^{2t}\sum_{i=2}^{n} c_i^2\\ &\le (1-\alpha)^{2t}\sum_{i=1}^{n} c_i^2\\ &\le (1-\alpha)^{2t},\end{aligned}$$

where the last inequality comes from

$$1 \ge \|p_0\|_2^2 = (c_1x_1 + \cdots + c_nx_n)^T(c_1x_1 + \cdots + c_nx_n) = \sum_{i=1}^{n} c_i^2.$$

Therefore

$$d(\mathscr{A}^t p_0, \pi) \le \sqrt{n}(1-\alpha)^t.$$

If we choose $t \ge \frac{\ln n}{\alpha}$, and use the fact that $1 - x \le e^{-x}$ for all x, we have

$$d(\mathscr{A}^t p_0, \pi) \le \sqrt{n}e^{-\alpha t} \le \sqrt{n}\cdot\frac{1}{n} = \frac{1}{\sqrt{n}} \le \frac{1}{4},$$

for n sufficiently large. Therefore the mixing time for random walks is $O(\frac{\ln n}{\alpha})$.

For the lazy random walk, we can say more about the mixing time. The spectral gap for lazy walks is $1 - w_2 = 1 - (\frac{1}{2} + \frac{1}{2}\alpha_2) = \frac{1}{2} - \frac{1}{2}\alpha_2$. Hence the mixing time is $\ln n/(\frac{1}{2} - \frac{1}{2}\alpha_2)$.

Recall the normalized Laplacian matrix $\mathscr{L}$ is defined as $\mathscr{L} = I - \mathscr{A}$. Let $\lambda_1 \le \cdots \le \lambda_n$ be the eigenvalues of $\mathscr{L}$ and by definition $\lambda_i = 1 - \alpha_i$ for all $i = 1, \ldots, n$. Thus, $\alpha_2 = 1 - \lambda_2$ and the mixing time for lazy walks is $\ln n/(\frac{1}{2}\lambda_2)$. By Cheeger's inequality, $\lambda_2 \ge \frac{\phi^2(G)}{2}$, so the mixing time is at most $O(\frac{\ln n}{\phi^2(G)})$.

Finally, we will now extend the above results to non-regular graphs. For a non-regular graph G, $AD^{-1} \ne D^{-1/2}AD^{-1/2} = \mathscr{A}$, and $W \ne \frac{1}{2}I + \frac{1}{2}\mathscr{A}$ for lazy walks.

However, AD^{-1} is similar to $\mathscr{A}$ and W is similar to $\frac{1}{2}I + \frac{1}{2}\mathscr{A}$. Recall a matrix X is *similar to* another matrix Y if there exists a non-singular matrix B such that

$X = BYB^{-1}$. If X and Y are similar, they will have the same spectrum since their characteristic polynomials are the same, shown as follows:

$$\det(\lambda I - X) = \det(\lambda I - BYB^{-1}) = \det(B(\lambda I - Y)B^{-1}) = \det(\lambda I - Y).$$

Since $D^{-1/2}(AD^{-1})D^{1/2} = D^{-1/2}AD^{-1/2} = \mathscr{A}$, AD^{-1} is similar to $\mathscr{A}$, and $W = \frac{1}{2}I + \frac{1}{2}AD^{-1}$ is similar to $\frac{1}{2}I + \frac{1}{2}\mathscr{A}$ because $D^{-1/2}(\frac{1}{2}I + \frac{1}{2}AD^{-1})D^{1/2} = \frac{1}{2}I + \frac{1}{2}\mathscr{A}$.

Therefore, the spectrum for AD^{-1} is still $\alpha_1, \ldots, \alpha_n$ and if G is connected and non-bipartite, we still have $1 = \alpha_1 > \alpha_2 \geq \cdots \geq \alpha_n > -1$. Simialrly, the spectrum of W is still $\frac{1}{2} + \frac{1}{2}\alpha_i$ for $i = 1, \ldots, n$ and if G is connected then $1 = \frac{1}{2} + \frac{1}{2}\alpha_1 > \frac{1}{2} + \frac{1}{2}\alpha_2 \geq \cdots \geq \frac{1}{2} + \frac{1}{2}\alpha_n \geq 0$.

Also, $D^{1/2}x_i$, $i = 1, \ldots, n$ are eigenvectors of AD^{-1}. To see this, note that $D^{-1/2}(AD^{-1}D^{1/2}x_i) = (D^{-1/2}AD^{-1/2})x_i = \mathscr{A}x_i = \alpha_i x_i$ implies $AD^{-1}(D^{1/2}x_i) = \alpha_i(D^{1/2}x_i)$. Since $D^{1/2}$ is full rank so $D^{1/2}x_1, \ldots, D^{1/2}x_n$ form a basis and we can write any intial distribution $p_0 = \sum_{i=1}^n c_i D^{1/2}x_i$, and the remaining proof just follows the argument for regular graphs.