Graphs and Networks Lecture 9

Convergence of Random Walks

Daniel A. Spielman September 30, 2010

9.1 Overview

We begin by reviewing the basics of spectral theory. We then apply this theory to show that lazy random walks do converge to the steady state. In fact, we show that the rate of convergence depends on the gap between the first and second largest eigenvalues of the lazy walk matrix.

An obvious obstruction to convergence of random walks are sets of vertices with very few edges leaving them. We measure this by the *conductance* of the set, and show that the convergence time is at least the reciprocal of the conductance. We finish by stating Cheeger's inequality, which gives a close relation between conductance and the spectral gap. It says that, at least to first order, the only barriers to rapid mixing are sets of low conductance.

I defer the holistic proof of the convergence of random walks to next lecture.

9.2 Review of Spectral Theory

Last lecture, we showed that the distribution of a the ordinary random walk on a graph after t steps is $\boldsymbol{p}_t = \boldsymbol{W}^t \boldsymbol{p}_0$, where $\boldsymbol{W}$ is the walk matrix of the graph. For the lazy random walk, it is given by $\widehat{\boldsymbol{W}}^t \boldsymbol{p}_0$. The important point for us is that it is obtained by mutiplying many times by the same matrix. Spectral theory (the eigenvalues and eigenvectors) is what we use when we want to understand what happens when we multiply by a matrix.

I now recall the basics of the theory. First, recall that $\boldsymbol{v}$ is an eigenvector of a matrix $\boldsymbol{W}$ with eigenvalue λ if

$$\lambda \boldsymbol{v} = \boldsymbol{W} \boldsymbol{v}.$$

The *geometric multiplicity* of the eigenvalue λ is the dimension of the space of vectors $\boldsymbol{v}$ for which this equation holds.

For symmetric matrices, the spectral theory is particularly elegant. While the walk matrices we consider are not usually symmetric, we begin by recalling the theory for the symmetric case.

Theorem 9.2.1. *[Spectral Theory of Symmetric Matrices] For every n-by-n symmetric matrix $\boldsymbol{M}$ there is an orthornormal basis of n eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ and a set of n eigenvalues $\lambda_1, \ldots, \lambda_n$ such that*

$$\lambda_i \boldsymbol{v}_i = \boldsymbol{M} \boldsymbol{v}_i$$

for all i.

Note that some eigenvalues may be repeated in this list. The orthonormality of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ gives us an easy way of expanding every vector in this basis. For every vector $\boldsymbol{x}$

$$\boldsymbol{x} = \sum_{i=1}^{n} (\boldsymbol{v}_i^T \boldsymbol{x}) \boldsymbol{v}_i.$$

Here the terms $\boldsymbol{v}_i^T \boldsymbol{x}$ are scalars, and so are the coefficients of the vectors $\boldsymbol{v}_i$ in the expansion.

Multiplication by $\boldsymbol{M}$ is easily performed by first expanding in the eigenbasis:

$$\boldsymbol{M}\boldsymbol{x} = \boldsymbol{M} \sum_{i=1}^{n} (\boldsymbol{v}_i^T \boldsymbol{x}) \boldsymbol{v}_i = \sum_{i=1}^{n} \boldsymbol{M} (\boldsymbol{v}_i^T \boldsymbol{x}) \boldsymbol{v}_i = \sum_{i=1}^{n} (\boldsymbol{v}_i^T \boldsymbol{x}) \lambda_i \boldsymbol{v}_i.$$

Similarly,

$$\boldsymbol{M}^k \boldsymbol{x} = \sum_{i=1}^{n} (\boldsymbol{v}_i^T \boldsymbol{x}) \lambda_i^k \boldsymbol{v}_i.$$

While the walk matrices $\boldsymbol{W}$ are not symmetric, they are *similar* to symmetric matrices. Let $\boldsymbol{D}^{1/2}$ denote the diagonal matrix whose uth diagonal is $\sqrt{d(u)}$ and let $\boldsymbol{D}^{-1/2}$ be the matrix with $1/\sqrt{d(u)}$ on its corresponding diagonal. We have

$$\boldsymbol{D}^{-1/2} \boldsymbol{W} \boldsymbol{D}^{1/2} = \boldsymbol{D}^{-1/2} \left(\boldsymbol{A} \boldsymbol{D}^{-1} \right) \boldsymbol{D}^{1/2} = \boldsymbol{D}^{-1/2} \boldsymbol{A} \boldsymbol{D}^{-1/2},$$

which is symmetric. For the rest of this lecture, we define

$$\boldsymbol{M} = \boldsymbol{D}^{-1/2} \boldsymbol{A} \boldsymbol{D}^{-1/2}$$

and

$$\widehat{\boldsymbol{M}} = (1/2)(\boldsymbol{I} + \boldsymbol{M}) = \boldsymbol{D}^{-1/2} \widehat{\boldsymbol{W}} \boldsymbol{D}^{1/2}.$$

Observe that $\boldsymbol{M}$ and $\boldsymbol{W}$ have the same eigenvalues, and an easy translation between their eigenvectors. For each eigenvector $\boldsymbol{v}$ of $\boldsymbol{M}$, we have

$$\lambda \boldsymbol{v} = \boldsymbol{M} \boldsymbol{v} = \left(D^{-1/2} \boldsymbol{W} D^{1/2} \right) \boldsymbol{v},$$

so

$$\lambda \left(D^{1/2} \boldsymbol{v} \right) = \boldsymbol{W} \left(D^{1/2} \boldsymbol{v} \right),$$

and we see that $D^{1/2} \boldsymbol{v}$ is a right-eigenvector of $\boldsymbol{W}$. This gives the following formula for multiplication by powers of $\boldsymbol{W}$:

$$\boldsymbol{W}^t \boldsymbol{x} = \left(D^{1/2} \boldsymbol{M} \boldsymbol{D}^{-1/2} \right)^t \boldsymbol{x} = \boldsymbol{D}^{1/2} \boldsymbol{M}^t \boldsymbol{D}^{-1/2} \boldsymbol{x} = \sum_i \lambda_i^t \boldsymbol{D}^{1/2} \boldsymbol{v}_i \left(\boldsymbol{v}_i^T \boldsymbol{D}^{-1/2} \boldsymbol{x} \right). \tag{9.1}$$

The key point here is that as t increases, the only terms that are changing are the powers of the eigenvalues. Moreover, every eigenvalue of absolute value less than 1 will have diminishing contribution. This is why the lazy random walk converges to the steady state: we will show that all of its eigenvalues are between 0 and 1 and that the steady-state vector is the only one with eigenvalue 1.

Before I do that, let's do a sanity check. I'd like to observe that we can use (9.1) to show that $\boldsymbol{W}^t \boldsymbol{\pi} = \boldsymbol{\pi}$, a fact that we already know. As $\boldsymbol{\pi}$ is an eigenvector of $\boldsymbol{W}$ of eigenvalue 1, $\boldsymbol{M}$ has a corresponding eigenvector of eigenvalue 1, which we will call $\boldsymbol{v}_1$ and which is given by

$$\boldsymbol{v}_1 = \frac{\boldsymbol{D}^{-1/2}\boldsymbol{\pi}}{\left\| \boldsymbol{D}^{-1/2}\boldsymbol{\pi} \right\|}.$$

We have to divide by the norm because we require $\boldsymbol{v}_1$ to be a unit vector. Let's see what that norm is. Recall that $\boldsymbol{\pi}(a) = d(a)/2m$, so $(\boldsymbol{D}^{-1/2}\boldsymbol{\pi})(a) = \sqrt{d(a)}/2m$. Thus,

$$\left\| \boldsymbol{D}^{-1/2}\boldsymbol{\pi} \right\| = \frac{1}{2m}\sqrt{\sum_a \sqrt{d(a)}^2} = \frac{1}{2m}\sqrt{\sum_a d(a)} = \frac{1}{\sqrt{2m}}$$

As the basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is orthnormal and $\boldsymbol{D}^{-1/2}\boldsymbol{\pi}$ lies in the same direction as $\boldsymbol{v}_1$, we know that

$$\boldsymbol{v}_i^T \boldsymbol{D}^{-1/2}\boldsymbol{\pi} = 0$$

for every $i \geq 2$ and

$$\boldsymbol{v}_1^T \boldsymbol{D}^{-1/2}\boldsymbol{\pi} = \left\| \boldsymbol{D}^{-1/2}\boldsymbol{\pi} \right\| = \frac{1}{\sqrt{2m}}.$$

So, when we apply equation 9.1, we get

$$\boldsymbol{W}^t \boldsymbol{\pi} = \boldsymbol{D}^{1/2}\boldsymbol{v}_1 (1/\sqrt{2m}) = \boldsymbol{D}^{1/2}\frac{\boldsymbol{D}^{-1/2}\boldsymbol{\pi}}{1/\sqrt{2m}}(1/\sqrt{2m}) = \boldsymbol{\pi}.$$

9.3 The eigenvalues of the Walk Matrix

So that we can apply this theory, we now prove some elementary facts about the eigenvalues of the walk matrix.

Theorem 9.3.1. *Let $\boldsymbol{W}$ be the walk matrix of a connected graph. Then, all eigenvalues of $\boldsymbol{W}$ lie between* 1 *and* -1*, and the eigenvalue* 1 *has multiplicity* 1.

Proof. Our proof of this will be very similar to the proof from last class that the steady-state distribution is unique. Actually, in that proof we already established that the eigenvalue 1 has multiplicity 1. If you check the proof, you will see that we never used the fact that $\boldsymbol{p}$ was a non-negative vector.

Let $\boldsymbol{v}$ be an eigenvector of $\boldsymbol{W}$ of eigenvalue λ. Let a be a vertex for which

$$|\boldsymbol{v}(a)| / d(a) \geq |\boldsymbol{v}(b)| / d(b),$$

for all b. We have

$$\lambda \boldsymbol{v}(a) = \sum_{(a,b)\in E} \boldsymbol{v}(b)/d(b),$$

and so

$$\begin{aligned}
|\lambda| \, |\boldsymbol{v}(a)| &= \left| \sum_{(a,b)\in E} \boldsymbol{v}(b)/d(b) \right| \\
&\leq \sum_{(a,b)\in E} |\boldsymbol{v}(b)| \, / d(b) \\
&\leq \sum_{(a,b)\in E} |\boldsymbol{v}(a)| \, / d(a) \\
&= |\boldsymbol{v}(a)| \, .
\end{aligned}$$

So, $|\lambda| \leq 1$. □

Corollary 9.3.2. *All eigenvalues of* $\widehat{\boldsymbol{W}}$ *lie between* 0 *and* 1*, and the eigenvalue* 1 *has multiplicity* 1*.*

Proof. As

$$\widehat{\boldsymbol{W}} = (1/2)I + (1/2)\,\boldsymbol{W},$$

$\widehat{\boldsymbol{W}}$ has the same eigenvectors as $\boldsymbol{W}$. Moreover, for every eigenvalue λ of $\boldsymbol{W}$ the matrix $\widehat{\boldsymbol{W}}$ has an eigenvalue of $(1+\lambda)/2$. □

We now know enough to show that a lazy random walk must converge to the steady state. We will now make that statement more quantative.

For the rest of the lecture, we let $\lambda_1, \dots, \lambda_n$ denote the eigenvalues of the walk matrix, with the convention

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n.$$

We now measure how quickly the random walk approaches the steady state.

Theorem 9.3.3. *Consider the lazy random walk on a connected graph. For every initial probability distribution* $\boldsymbol{p}_0$ *and every* $t \geq 0$ *we have*

$$\|\boldsymbol{p}_t - \boldsymbol{\pi}\| \leq \sqrt{\frac{\max_a d(a)}{\min_a d(a)}} \lambda_2^t.$$

If the walk starts at vertex a*, then for every vertex* b *we have*

$$|\boldsymbol{p}_t(b) - \boldsymbol{\pi}(b)| \leq \sqrt{\frac{d(b)}{d(a)}} \lambda_2^t.$$

Proof. Let $\boldsymbol{p}_0$ be any probability distribution on the vertices. If it is concentrated on one vertex then $\|\boldsymbol{p}_0\| = 1$, and you should check that otherwise $\|\boldsymbol{p}_0\| < 1$. We will now write

$$\boldsymbol{D}^{-1/2}\boldsymbol{p}_0 = \sum_i \alpha_i \boldsymbol{v}_i,$$

where

$$\alpha_i = \boldsymbol{v}_i^T \boldsymbol{D}^{-1/2}\boldsymbol{p}_0.$$

As the largest entry of $\boldsymbol{D}^{-1/2}$ is the square root of the minimum degree of a vertex in G, we know

$$\left\|\boldsymbol{D}^{-1/2}\boldsymbol{p}_0\right\| \leq \frac{1}{\min d(a)}.$$

As the $\boldsymbol{v}_i$ form an orthornormal basis, we also have

$$\sum_i \alpha_i^2 = \left\|\boldsymbol{D}^{-1/2}\boldsymbol{p}_0\right\| \leq \frac{1}{\min d(a)}.$$

We are most interested in α_1, which we may compute from the formula for $\boldsymbol{v}_1$.

$$\alpha_1 = \boldsymbol{v}_1^T \boldsymbol{D}^{-1/2}\boldsymbol{p}_0 = \sqrt{2m}\boldsymbol{\pi}_1^T \boldsymbol{D}^{-1/2}\boldsymbol{D}^{-1/2}\boldsymbol{p}_0 = \sqrt{2m}\boldsymbol{\pi}_1^T \boldsymbol{D}^{-1}\boldsymbol{p}_0 = \sqrt{2m}(1/2m)\mathbf{1}^T\boldsymbol{p}_0 = 1/\sqrt{2m},$$

as $\mathbf{1}^T\boldsymbol{p} = 1$ for every probability vector $\boldsymbol{p}$.

Applying equation 9.1 and separating the first term from the rest we find

$$\boldsymbol{p}_t = \boldsymbol{W}^t\boldsymbol{p}_0 = \boldsymbol{D}^{1/2}\boldsymbol{v}_1\alpha_1 + \boldsymbol{D}^{1/2}\sum_{i\geq 2}\lambda_i^t\alpha_i\boldsymbol{v}_i.$$

The first term in this sum is simply $\boldsymbol{\pi}$. We now bound the norm of the second term by first computing

$$\begin{aligned}
\left\|\sum_{i\geq 2}\lambda_i^t\alpha_i\boldsymbol{v}_i\right\|^2 &= \sum_{i\geq 2}\left(\lambda_i^t\alpha_i\right)^2 && \text{(as the } \boldsymbol{v}_i \text{ are an orthonormal basis)}\\
&\leq \sum_{i\geq 2}\left(\lambda_2^t\alpha_i\right)^2 && \text{(as } 0 \leq \lambda_i \leq \lambda_2 \text{ for } i \geq 2\text{)}\\
&= \lambda_2^{2t}\sum_{i\geq 2}{\alpha_i}^2\\
&\leq \lambda_2^{2t}\frac{1}{\min_a d(a)}.
\end{aligned}$$

As the largest entry in $\boldsymbol{D}^{1/2}$ is at most the square root of the largest degree of a vertex in the graph, multiplying by $\boldsymbol{D}^{1/2}$ increases the norm of this vector by at most $\sqrt{\max_a d(a)}$.

We thus have

$$\|\boldsymbol{p}_t - \boldsymbol{\pi}\| \leq \sqrt{\frac{\max_a d(a)}{\min_a d(a)}}\lambda_2^t.$$

To prove the second part of the theorem, we consider what happens if we start the walk at vertex a. In this case we have

$$\left\| \boldsymbol{D}^{-1/2} \boldsymbol{p}_0 \right\| = \frac{1}{\sqrt{d(a)}}.$$

To compute the bth entry of $\boldsymbol{p}_t$, let $\boldsymbol{e}_b$ be the elementary unit vector in direction $\boldsymbol{b}$. We have

$$\boldsymbol{p}_t(b) = \boldsymbol{e}_b^T \boldsymbol{p}_t = \boldsymbol{e}_b^T \boldsymbol{\pi} + \boldsymbol{e}_b^T \boldsymbol{D}^{1/2} \sum_{i \geq 2} \lambda_i^t \alpha_i \boldsymbol{v}_i.$$

We upper-bound this last term by

$$\boldsymbol{e}_b^T \boldsymbol{D}^{1/2} \sum_{i \geq 2} \lambda_i^t \alpha_i \boldsymbol{v}_i = \sqrt{d(b)} \boldsymbol{e}_b^T \sum_{i \geq 2} \lambda_i^t \alpha_i \boldsymbol{v}_i \leq \sqrt{d(b)} \left\| \sum_{i \geq 2} \lambda_i^t \alpha_i \boldsymbol{v}_i \right\| \leq \sqrt{d(b)} \lambda_2^t \frac{1}{\sqrt{d(a)}},$$

as before. □

It often happens that λ_2 is relatively close to 1. In this case, we focus on the gap between λ_2 and 1. That is, we write $\lambda_2 = 1 - \mu$. The important term in Theorem 9.3.3 then becomes

$$\lambda_2^t = (1 - \mu)^t \leq e^{-t\mu}.$$

Thus, we see that convergence starts to happen after $1/\mu$ steps.

9.4 The obstructions to rapid mixing

The main reason a random walk would not converge rapidly is if it started inside a set of vertices that has few edges leaving it. This naturally corresponds to a community or a cluster in the graph. We will measure the quality of a cluster of vertices S by its conductance, which we now define.

We first define the boundary of S, written $\partial(S)$, by

$$\partial(S) \stackrel{\text{def}}{=} \{(u, v) \in E : u \in S, v \notin S\}.$$

We will want to divide this by a measure of S. The natural measure on S is given by $\boldsymbol{\pi}$, but we use $\boldsymbol{d}$ for convenience:

$$\boldsymbol{d}(S) \stackrel{\text{def}}{=} \sum_{a \in S} \boldsymbol{d}(a).$$

For small sets S, we define the conductance of S, written $\phi(S)$ to be

$$\frac{|\partial(S)|}{\boldsymbol{d}(S)},$$

where $|\partial(S)|$ is the number of edges in $\partial(S)$.

For larger sets, we consider the smaller of $\boldsymbol{d}(S)$ and $\boldsymbol{d}(V-S)$. This gives us the definition[1]

$$\phi(S) \stackrel{\text{def}}{=} \frac{|\partial(S)|}{\min\left(\boldsymbol{d}(S), \boldsymbol{d}(V-S)\right)}.$$

If $\phi(S)$ is small, than a random walk that starts behind S will take a long time to move its probability mass outside S. In particular, we can show this for the distribution given by restricting $\boldsymbol{\pi}$ to S. We denote this distribution $\boldsymbol{\pi}_S$, where

$$\boldsymbol{\pi}_S(a) = \begin{cases} d(a)/\boldsymbol{d}(S) & \text{if } a \in S \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to show that if $\boldsymbol{p}_0 = \boldsymbol{\pi}_S$ then after one step at most a $\phi(S)/2$ fraction of the probability mass will escape S. The half comes from the laziness of the random walk. In fact, one can prove the following.

Theorem 9.4.1. *If $\boldsymbol{p}_0 = \boldsymbol{\pi}_S$, then after t steps at most a $t\phi(S)/2$ fraction of the probability mass will escape S.*

Let $\phi_G = \min_{S \subset V} \phi(S)$ measure the least conductance of a set of vertices. This theorem tells us that there are initial distributions for which the random walk will not begin to converge until at least after $1/2\phi_G$ steps.

In fact, there is a tight relation between ϕ_G and λ_2, given by Cheeger's inequality.

Theorem 9.4.2. *[Cheeger's inequality for lazy random walks]*

$$\phi_G \geq \mu \geq \phi_G^2/4.$$

Moreover, if $\boldsymbol{v}_2$ is a right-eigenvector of $\boldsymbol{W}$ corresponding to λ_2, then there is a number x for which the set

$$S_x = \{a \in V : \boldsymbol{v}_2(a)/d(a) \geq x\}$$

satisfies

$$\phi(S_x)^2/4 \leq \mu.$$

The argument from the beginning of this section can be used to show the left-hand side of Cheeger's inequality. The interesting part is the right-hand side. The reason that the spectral gap can differ quadratically from the minimum conductance is that one can embedd many sets of low conductance inside each other. This results in even slower mixing.

The last part of Cheeger's inequality is incredibly useful. It gives us a way to find sets of low conductance from eigenvectors. Next lecture we will see how to do this more directly from random walks.

[1] Some authors prefer to divide by the product $\boldsymbol{d}(S)\boldsymbol{d}(V-S)$.