Graphs and Networks Lecture 9

Spectral Analysis of Random Walks

Daniel A. Spielman September 26, 2013

9.1 Disclaimer

These notes are not necessarily an accurate representation of what happened in class. They are a combination of what I intended to say with what I think I said. They have not been carefully edited.

9.2 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.

9.3 Maximum Density Decreases

When first contemplating diffusion, you might think that the maximum amount of mass at any node decreases over time. But, this is not necessarily true. Consider a graph with 4 vertices, a, b, c and d, and 3 edges: (a,b), (a,c) and (a,d). Let's start with mass $1/3$ at b, c and d and mass 0 at a. After one step of the lazy random walk, the mass at a will be $1/2$.

While the maximum mass at a vertex can increase, the maximum mass at a socket cannot. Recall that the amount of mass at a socket belonging to a node a of degree d is $p(a)/d(a)$.

Theorem 9.3.1. *For every probability vector p,*

$$\max_{a \in V} \frac{p(a)}{d(a)} \geq \max_{b \in V} \frac{(\widehat{W} p)(a)}{d(a)}.$$

Proof. Let

$$\alpha = \max_{a \in V} \frac{p(a)}{d(a)}.$$

Now, for any node $b \in V$, let's see how much mass can wind up at node b after one step of the lazy random walk. The amount of mass at b will be one-half of what was there before, plus one half of what each neighbor sends along. A neighbor c of b sends a $1/2d(c)$ fraction of its mass. So,

$$(\widehat{W}p)(b) = (1/2)p(b) + \sum_{(c,b)\in E} p(c)/(2d(c)) \leq (1/2)\alpha d(b) + \sum_{(c,b)\in E} \alpha/2 = (1/2)\alpha d(b) + d(b)\alpha/2 = \alpha d(b).$$

The theorem follows. □

9.4 Obstructions to Convergence

The obstructions to convergence of diffusion are sets of vertices with many internal edges but few edges leaving. These are the generalization of what we have called "tenticles". I now define and quantify them formally. 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:

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

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 $\partial(S)$.

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

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

That is, $\boldsymbol{\pi}_S$ is the distribution on vertices that chooses a vertex in S with probability proportional to its degree.

To measure mass inside S, we introduce the indicator vector of S:

$$\mathbf{1}_S(a) = \begin{cases} 1 & \text{for } a \in S \\ 0 & \text{for } a \notin S. \end{cases}$$

For every vector $\boldsymbol{p}$,

$$\mathbf{1}_S^T \boldsymbol{p} = \sum_{a\in S} \boldsymbol{p}(a).$$

In particular,

$$\mathbf{1}_S^T \boldsymbol{\pi}_S = 1.$$

It is easy to show that if $\boldsymbol{p}_0 = \boldsymbol{\pi}_S$ then a $\phi(S)/2$ fraction of the probability mass will leave S in the first step: the mass leaves along the edges in the boundary of S. For an edge (a, b) with $a \in S$ and $b \notin S$, the amount of mass that flows from a to b (in the lazy walk) is

$$\boldsymbol{p}_0(a)/2d(a) = 1/2\boldsymbol{d}(S).$$

As there are $|\partial(S)|$ edges in the boundary of S, the amount of mass that leaves is

$$|\partial(S)|\,/2\boldsymbol{d}(S) = \phi(S)/2.$$

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. That is,*

$$\mathbf{1}_S^T \boldsymbol{p}_t \geq 1 - t\phi(S)/2.$$

Proof. We know that

$$\max_a \frac{\boldsymbol{p}_0(a)}{d(a)} = 1/\boldsymbol{d}(S).$$

So, by Theorem 9.3.1, for every other time i and every vertex a,

$$\frac{\boldsymbol{p}_t(a)}{d(a)} \leq 1/\boldsymbol{d}(S).$$

So, for any edge (a, b) with $a \in S$ and $b \notin S$, the amount of mass that can move from a to b along edge (a, b) in step i is at most

$$1/2\boldsymbol{d}(S).$$

Thus, the total amount of mass that can leave along edges on the boundary of S is at most $\phi(S)/2$. □

For a linear-algebraic approach to proving this, see Proposition 2.5 of [ST].

9.5 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 9.5.1. *[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

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

9.6 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 9.6.1. *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

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

for all b. We have

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

and so

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

So, $|\lambda| \leq 1$. □

Corollary 9.6.2. *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 $\lambda_1, \ldots, \lambda_n$ 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.

To state the bound correctly, I introduce a new type of norm. For a symmetric matrix $\boldsymbol{B}$ with positive eigenvalues (such as a diagonal matrix with positive entries), the $\boldsymbol{B}$-norm is given by

$$\left\|\boldsymbol{x}\right\|_{\boldsymbol{B}} = \sqrt{\boldsymbol{x}^T \boldsymbol{B} \boldsymbol{x}} = \left\|\boldsymbol{B}^{1/2}\boldsymbol{x}\right\|.$$

Theorem 9.6.3. *Consider the lazy random walk on a connected graph. For every initial probability distribution $\boldsymbol{p}_0$ and every $t \geq 0$ we have*

$$\left\| \boldsymbol{p}_t - \boldsymbol{\pi} \right\|_{\boldsymbol{D}^{-1}} \leq \lambda_2^t \left\| \boldsymbol{p}_0 \right\|_{\boldsymbol{D}^{-1}}.$$

In particular, if the walk starts at vertex a, then for every vertex b we have

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

Proof. Let $\boldsymbol{p}_0$ be any probability distribution on the vertices. Define

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

We begin by observing that

$$\alpha_1 = \boldsymbol{v}_1^T \boldsymbol{D}^{-1/2} \boldsymbol{p}_0 = \frac{(\boldsymbol{d}^{1/2})^T}{\left\| \boldsymbol{d}^{1/2} \right\|} \boldsymbol{D}^{-1/2} \boldsymbol{p}_0 = \frac{\boldsymbol{1}^T}{\left\| \boldsymbol{d}^{1/2} \right\|} \boldsymbol{p}_0 = \frac{1}{\left\| \boldsymbol{d}^{1/2} \right\|}.$$

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 $\boldsymbol{\pi}$:

$$\boldsymbol{D}^{1/2} \boldsymbol{v}_1 \alpha_1 = \boldsymbol{D}^{1/2} \frac{(\boldsymbol{d}^{1/2})^T}{\left\| \boldsymbol{d}^{1/2} \right\|} \alpha_1 = \frac{\boldsymbol{d}}{\left\| \boldsymbol{d}^{1/2} \right\|} \alpha_1 = \frac{\boldsymbol{d}}{\left\| \boldsymbol{d}^{1/2} \right\|^2} = \frac{\boldsymbol{d}}{\sum_a \boldsymbol{d}(a)} = \boldsymbol{\pi}.$$

So,

$$\boldsymbol{p}_t - \boldsymbol{\pi} = \boldsymbol{D}^{1/2} \sum_{i \geq 2} \lambda_i^t \alpha_i \boldsymbol{v}_i.$$

To bound the norm of this term, we note

$$\left\| \boldsymbol{p}_t - \boldsymbol{\pi} \right\|_{\boldsymbol{D}^{-1}} = \left\| \sum_{i \geq 2} \lambda_i^t \alpha_i \boldsymbol{v}_i \right\|.$$

As the vectors $\boldsymbol{v}_i$ are orthonormal, this equals

$$\left(\sum_{i \geq 2} \lambda_i^{2t} \alpha_i^2 \right)^{1/2}.$$

For $i \geq 2$, $\lambda_i^{2t} \leq \lambda_2^{2t}$. So, this is at most

$$\lambda_2^t \left(\sum_{i \geq 2} \alpha_i^2 \right)^{1/2}.$$

To finish the proof, we note that

$$\left(\sum_{i\geq 2} \alpha_i^2\right)^{1/2} \leq \left(\sum_{i} \alpha_i^2\right)^{1/2} = \left\| \boldsymbol{D}^{-1/2}\boldsymbol{p}_0 \right\| = \|\boldsymbol{p}_0\|_{\boldsymbol{D}^{-1}},$$

where the first equality follows from the fact that the $\boldsymbol{v}_i$ form an orthonormal basis. □

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 9.6.3 then becomes

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

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

9.7 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.

References

[ST] Daniel A. Spielman and Shang-Hua Teng. A local clustering algorithm for massive graphs and its application to nearly-linear time graph partitioning. *SIAM Journal on Computing.* To appear. Available at http://arxiv.org/pdf/0809.3232v1.pdf.