Spectral Graph Theory Lecture 8

Random Walks on Graphs

Daniel A. Spielman September 25, 2009

8.1 Overview

We will examine how the eigenvalues of a graph govern the convergence of a random walk on the graph.

8.2 Random Walks

In this lecture, we will consider random walks on undirected graphs. Let's begin with the definitions. Let $G = (V, E, w)$ be a weighted undirected graph. A random walk on a graph is a process that begins at some vertex, and at each time step moves to another vertex. When the graph is unweighted, the vertex the walk moves to is chosen uniformly at random among the neighbors of the present vertex. When the graph is weighted, it moves to a neighbor with probability proportional to the weight of the corresponding edge. Rather than tracking where some individual random walk goes, we will usually be interested in the probability distribution over vertices after a certain number of steps.

We will let the vector $\boldsymbol{p}_t \in \mathbb{R}^n$ denote the probability distribution at time t. I will sometimes write $\boldsymbol{p}_t \in \mathbb{R}^V$ to emphasise that $\boldsymbol{p}_t$ is a vector indexed by the vertices of the graph, or I may even write $\boldsymbol{p}_t : V \to \mathbb{R}$. I will write $\boldsymbol{p}_t(u)$ to indicate the value of $\boldsymbol{p}_t$ at a vertex u–that is the probability of being at vertex u at time t. A probability vector $\boldsymbol{p}$ should satisfy $\boldsymbol{p}(u) \geq 0$, for all $u \in V$, and

$$\sum_u \boldsymbol{p}(u) = 1.$$

Our initial probability distribution, $\boldsymbol{p}_0$, will typically be concentrated one vertex. That is, there will be some vertex v for which $\boldsymbol{p}_0(v) = 1$. In this case, we say that the walk starts at v.

To derive a $\boldsymbol{p}_{t+1}$ from $\boldsymbol{p}_t$, note that the probability of being at a vertex u at time $t+1$ is the sum over the neighbors v of u of the probability that the walk was at v at time t, times the probability it moved from v to u in time $t+1$. Algebraically, we have

$$\boldsymbol{p}_{t+1}(u) = \sum_{v:(u,v)\in E} \frac{w(u,v)}{\boldsymbol{d}(v)} \boldsymbol{p}_t(v), \tag{8.1}$$

where $\boldsymbol{d}(v) = \sum_u w(u,v)$ is the weighted degree of vertex v.

We will often consider lazy random walks, which are the variant of random walks that stay put with probability 1/2 at each time step, and walk to a random neighbor the other half of the time. These evolve according to the equation

$$\boldsymbol{p}_{t+1}(u) = (1/2)\boldsymbol{p}_t(u) + (1/2) \sum_{v:(u,v)\in E} \frac{w(u,v)}{\boldsymbol{d}(v)} \boldsymbol{p}_t(v). \tag{8.2}$$

8.3 Diffusion

There are a few types of diffusion that people study in a graph, but the most common is closely related to random walks. In a diffusion process, we imagine that we have some substance that can occupy the vertices, such as a gas or fluid. At each time step, some of the substance diffuses out of each vertex. If we say that half the substance stays at a vertex at each time step, and the other half is distributed among its neighboring vertices, then the distribution of the substance will evolve according to equation (8.2). That is, probability mass obeys this diffusion equation.

I remark that often people consider finer time steps in which smaller fractions of the mass leave the vertices. In the limit, this results in continuous random walks. But, that is not a topic for this lecture.

8.4 Matrix form

The right way to understand the behavior of random walks is through linear algebra.

Equation (8.2) is equivalent to:

$$\boldsymbol{p}_{t+1} = (1/2)\left(I + AD^{-1}\right)\boldsymbol{p}_t. \tag{8.3}$$

You can verify this by checking that it is correct for any entry $\boldsymbol{p}_{t+1}(u)$, and you should do this yourself. It will prevent much confusion later.

For the rest of the course, I will let W_G denote the *lazy walk matrix* of the graph G, where

$$W_G \stackrel{\text{def}}{=} (1/2)\left(I + A_G D_G^{-1}\right). \tag{8.4}$$

This is the one assymetric matrix that we will deal with in this course. Fortunately, it is similar to a symmetric matrix we have studied, the normalized adjacency matrix. It is also closely related to the normalized Laplacian. We have

$$W = \frac{1}{2} D^{1/2}(I+M)D^{-1/2} = \frac{1}{2} D^{1/2}\left(D^{-1/2}(A+D)D^{-1/2}\right)D^{-1/2} = I - (1/2)D^{1/2} N D^{-1/2}.$$

So, we know that W is diagonalizable, and that for every eigenvector $\boldsymbol{v}_i$ of M with eigenvalue μ_i, the vector $D^{1/2}\boldsymbol{v}_i$ is a right-eigenvector of W of eigenvalue $1/2 + \mu_i/2$:

$$W\left(D^{1/2}\boldsymbol{v}_i\right) = \frac{1}{2}D^{1/2}(I+M)D^{-1/2}\left(D^{1/2}\boldsymbol{v}_i\right) = \frac{1}{2}D^{1/2}(I+M)\boldsymbol{v}_i = \frac{1}{2}D^{1/2}(1+\mu_i)\boldsymbol{v}_i = \frac{1+\mu_i}{2}D^{1/2}\boldsymbol{v}_i.$$

The key thing to remember in the assymetric case is that the eigenvectors of W are not necessarily orthogonal.

You may be wonder why I have decided to consider only lazy walks, rather than more natural walk given by AD^{-1}. There are two equivalent reasons. The first is that all the eigenvalues of W are between 1 and 0. To see this, recall that the Perron eigenvalue of M is 1, so all the eigenvalues of M are between 1 and -1. For the rest of the semester, we will let the eigenvalues of W be:

$$1 = \omega_1 \geq \omega_2 \geq \cdots \geq \omega_n \geq 0, \qquad \text{where} \qquad \omega_i = (1/2)(1 + \mu_i).$$

Yes, I know that ω is not a greek equivalent of w, but it sure looks like it.

8.5 The stable distribution

Regardless of the starting distribution, the lazy random walk on a connected graph always converges to one distribution: the *stable distribution.* In the stable distribution, every vertex is visited with probability proportional to its weighted degree. We denote the vector encoding this distribution by $\boldsymbol{\pi}$, where

$$\boldsymbol{\pi}(i) = \frac{\boldsymbol{d}(i)}{\sum_j \boldsymbol{d}(j)}.$$

We can see that $\boldsymbol{\pi}$ is a right-eigenvector of W of eigenvalue 1. Recall that the eigenvector of M of eigenvalue 1 is proportional to $\boldsymbol{d}^{1/2}$, and we have

$$\boldsymbol{\pi} = \left(\frac{1}{\sum_j \boldsymbol{d}(j)} \right) D^{1/2} \boldsymbol{d}^{1/2}.$$

This is the other reason that we forced our random walk to be lazy. Without laziness, there can be graphs on which the random walks never converge. For example, consider a non-lazy random walk on a bipartite graph. Every-other step will bring it to the other side of the graph. So, if the walk starts on one side of the graph, its limiting distribution at time t will depend upon the parity of t.

To see that the walk converges to $\boldsymbol{\pi}$, we expand $D^{-1/2}$ times the initial distribution in the eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ of M. Let

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

Note that

$$c_1 = \boldsymbol{v}_1^T (D^{-1/2} \boldsymbol{p}_0) = \frac{(\boldsymbol{d}^{1/2})^T}{\left\| \boldsymbol{d}^{1/2} \right\|} (D^{-1/2} \boldsymbol{p}_0) = \mathbf{1}^T \boldsymbol{p}_0 = \frac{1}{\left\| \boldsymbol{d}^{1/2} \right\|},$$

as $\boldsymbol{p}_0$ is a probability vector. We have

$$\begin{aligned}\boldsymbol{p}_t &= W^t \boldsymbol{p}_0\\ &= (D^{1/2}(I/2 + M/2)D^{-1/2})^t \boldsymbol{p}_0\\ &= (D^{1/2}(I/2 + M/2)^t D^{-1/2})\boldsymbol{p}_0\\ &= D^{1/2}(I/2 + M/2)^t \sum_i c_i \boldsymbol{v}_i\\ &= D^{1/2} \sum_i \omega_i^t c_i \boldsymbol{v}_i\\ &= D^{1/2} c_1 \boldsymbol{v}_1 + D^{1/2} \sum_{i \geq 2} \omega_i^t c_i \boldsymbol{v}_i.\end{aligned}$$

As $\omega_i < 1$ for $i \geq 2$, the right-hand term must go to zero. On the other hand, $\boldsymbol{v}_1 = \boldsymbol{d}^{1/2} / \left\| \boldsymbol{d}^{1/2} \right\|$, so

$$D^{1/2} c_1 \boldsymbol{v}_1 = D^{1/2} \left(\frac{1}{\left\| \boldsymbol{d}^{1/2} \right\|} \right) \frac{\boldsymbol{d}^{1/2}}{\left\| \boldsymbol{d}^{1/2} \right\|} = \frac{\boldsymbol{d}}{\left\| \boldsymbol{d}^{1/2} \right\|^2} = \frac{\boldsymbol{d}}{\sum_j \boldsymbol{d}(j)} = \boldsymbol{\pi}.$$

This is a perfect example of one of the main uses of spectral theory: to understand what happens when we repeatedly apply an operator.

8.6 The Rate of Convergence

The rate of convergence to the stable distribution is dictated by ω_2. There are many ways of saying this. We will do so point-wise. Assume that the random walk starts at some vertex $a \in V$. For every vertex b, we will bound how far $\boldsymbol{p}_t(b)$ can be from $\boldsymbol{\pi}(b)$.

Theorem 8.6.1. *For all a, b and t, if $\boldsymbol{p}_0 = \chi_a$, then*

$$|\boldsymbol{p}_t(b) - \boldsymbol{\pi}(b)| \leq \sqrt{\frac{\boldsymbol{d}(b)}{\boldsymbol{d}(a)}} \omega_2^t.$$

Proof. Observe that

$$\boldsymbol{p}_t(b) = \chi_b^T \boldsymbol{p}_t.$$

From the analysis in the previous section, we know

$$\boldsymbol{p}_t(b) = \boldsymbol{\pi}(b) + \chi_b^T D^{1/2} \sum_{i \geq 2} \omega_i^t c_i \boldsymbol{v}_i.$$

We need merely prove an upper bound on the magnitude of the right-hand term. To this end, recall that

$$c_i = \boldsymbol{v}_i^T D^{-1/2} \chi_a.$$

So,

$$\chi_b^T D^{1/2} \sum_{i \geq 2} \omega_i^t c_i \boldsymbol{v}_i = \sqrt{\frac{\boldsymbol{d}(b)}{\boldsymbol{d}(a)}} \chi_b^T \sum_{i \geq 2} \omega_i^t \boldsymbol{v}_i \boldsymbol{v}_i^T \chi_a.$$

Analyzing the right-hand part of this last expression, we find

$$\chi_b^T \sum_{i \geq 2} \omega_i^t \boldsymbol{v}_i \boldsymbol{v}_i^T \chi_a = \sum_{i \geq 2} \omega_i^t \left(\chi_b^T \boldsymbol{v}_i\right) \left(\boldsymbol{v}_i^T \chi_a\right) \leq \sum_{i \geq 2} \omega_i^t \left|\chi_b^T \boldsymbol{v}_i\right| \left|\boldsymbol{v}_i^T \chi_a\right| \leq \omega_2^t \sum_{i \geq 2} \left|\chi_b^T \boldsymbol{v}_i\right| \left|\boldsymbol{v}_i^T \chi_a\right|.$$

We will finish the proof by showing that

$$\sum_{i \geq 2} \left|\chi_b^T \boldsymbol{v}_i\right| \left|\boldsymbol{v}_i^T \chi_a\right| \leq 1.$$

To see this, let $\boldsymbol{V}$ be the matrix having the eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ in its columns. As i varies, $\chi_b^T \boldsymbol{v}_i$ takes on the values in the bth row of $\boldsymbol{V}$. So, the sum over all i of $(\chi_b^T \boldsymbol{v}_i)(\chi_a^T \boldsymbol{v}_i)$ is the inner product of the ath and bth rows of the matrix obtained by replacing every entry of $\boldsymbol{V}$ by its absolute value. As $\boldsymbol{V}$ is orthnormal, every row of $\boldsymbol{V}$ has norm 1, and this does not change if we take the absolute value of every entry of $\boldsymbol{V}$. So, the inner product of the absolute values of two rows of $\boldsymbol{V}$ is at most the product of their norms, which is 1. In our sum we skip the first entry of each row, but this can only decrease the inner product. □

8.7 Examples

Last lecture we proved that

$$\phi_G \geq \nu_2 \geq \phi_G^2/8. \tag{8.5}$$

We are now going to examine some graphs to see when each inequality is tight. We will also see what is going on with their random walks.

In this discussion I will only worry about asymptotics, and so I will write "∼" instead of "=" to mean "equals up to a constant".

First, consider the path P_n and complete binary tree T_n on the same numbers of vertices. Both of these graphs may be cut approximately in half by the removal of one edge. So, we have

$$\phi_{P_n} \sim \frac{1}{n}, \quad \text{and} \quad \phi_{T_n} \sim \frac{1}{n}.$$

On the other hand, we know that

$$\nu_2(P_n) \sim \frac{1}{n^2} \quad \text{and} \quad \nu_2(T_n) \sim \frac{1}{n^2}.$$

Well, that's almost true. You can prove upper bounds on $\nu_2(P_n)$ and $\nu_2(T_n)$ by using the same test vectors as you used for the ordinary Laplacian. To transfer the lower bound for P_n from the ordinary Laplacian to the normalized Laplacian, you need a problem from today's problem set. For the tree, you first need to prove that $\lambda_2(T_n) \sim 1/n$, which you will also do on today's problem set.

These two examples show that neither side of (8.5) can be improved by more than a constant factor.

To understand the random walk on P_n, think about what happens when the walk starts in the middle. Ignoring the steps on which it stays put, it will either move to the left or the right with probability 1/2. So, the position of the walk after t steps is distributed as the sum of t ± 1 random variables. Recall that the standard deviation of such a sum is $\sqrt{t}$. So, we need to have $\sqrt{t}$ comparable to $n/4$ for there to be a reasonable chance that the walk is on the left or right $n/4$ vertices.

To understand the random walk on T_n, first note that whenever it is at a vertex, it is twice as likely to step towards a leaf as it is to step towards the root. So, if the walk starts at a leaf, there is no way the walk can mix until it reaches the root. The height of the walk is like a sum of ± 1 random variables, except that they are twice as likely to be -1 as they are to be 1, and that they sum never goes below 0. One can show that we need to wait approximately n steps before such a walk will hit the root.

Now, let's consider another one of my favorite graphs, the dumbell. The dumbell graph D_n consists of two complete graphs on n vertices, joined by one edge. So, there are $2n$ vertices in total. The conductance of this graph is

$$\phi_{D_n} \sim \frac{1}{n^2}.$$

Using the test vector that is 1 on one complete graph and -1 on the other, we can show that

$$\nu_2(D_n) \lessapprox 1/n^2.$$

To prove that this bound is almost tight, we use the following lemma.

Lemma 8.7.1. *Let G be an unweighted graph of diameter at most r. Then,*

$$\lambda_2(G) \geq \frac{2}{r(n-1)}.$$

Proof. For every pair of vertices (u, v), let $P(u, v)$ be a path in G of length at most r. We have

$$L_{(u,v)} \preccurlyeq r \cdot L_{P(u,v)} \preccurlyeq r L_G.$$

So,

$$K_n \preccurlyeq r \binom{n}{2} G,$$

and

$$n \leq r \binom{n}{2} \lambda_2(G),$$

from which the lemma follows. □

The diameter of D_n is 3, so we have $\lambda_2(D_n) \geq 2/3(n-1)$. Using a problem from today's problem set, we may conclude that $\nu_2(D_n) \gtrapprox 1/n^2$.

To understand the random walk on this graph, first observe that after 1 step the walk will be well mixed on the vertices in the side on which it starts. Because of this, the chance that it finds the edge going to the other side is only around $2/n^2$. This remains true, as the walk becomes better and better mixed on the side on which it starts. So, we must wait some multiple of n^2 steps before there is a resonable chance that the walk reaches the other side of the graph.

A graph on n vertices on which walks mix as slowly as possible may be found by connecting to cliques on n vertices by a path of length n. The analysis of the random walk is similar, except that we should realize that if we reach the first vertex on the path, the chance that we get to the other end before moving back to the clique at which we started is only $1/n$. So, we must wait around n^3 steps before there is a reasonable chance of getting to the other side. One can prove that ν_2 of this graph is at most $O(1/n^3)$ by considering a test vector that is $n/2$ on one clique, $-n/2$ on the other, and increases by 1 along the path. To prove a lower bound on ν_2, we use Lemma 8.7.1 to prove that λ_2 is at least $\omega(1/n^2)$, and then apply Problem 2 from the problem set to prove that ν_2 is at least $\omega(1/n^3)$.