Graphs and Networks Lecture 3

Random Walks on Graphs

Lecturer: Daniel A. Spielman September 13, 2007

3.1 Random Walks

In this lecture, we will consider random walks on undirected graphs. Let's begin with the definitions. Let $G = (V, E)$ be an 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. The vertex the walk moves to is chosen uniformly at random among the neighbors of the present vertex. 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 $\boldsymbol{p}_t : V \rightarrow \mathbb{R}$. 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

$$\boldsymbol{p}_{t+1}(u) = \sum_{v:(u,v)\in E} \boldsymbol{p}_t(v)/d(v), \tag{3.1}$$

where $d(v)$ is the 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

$$\boldsymbol{p}_{t+1}(u) = (1/2)\boldsymbol{p}_t(u) + (1/2) \sum_{v:(u,v)\in E} \boldsymbol{p}_t(v)/d(v). \tag{3.2}$$

3.2 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 (3.2). 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 class.

3.3 Applications of Diffusion

This will be described in class, and filled in later.

3.4 Matrix form

The right way to understand the behavior of random walks is through linear algebra. First, we will re-write equation (3.2) in matrix form. To do this, we first identify the vertices with the set $\{1, \ldots, n\}$. We then define the adjacency matrix, $\boldsymbol{A}$, of the graph by setting $\boldsymbol{A}(u, v)$ to 1 if $(u, v) \in E$ and 0 otherwise (note that I write $\boldsymbol{A}(u, v)$ to indicate the u, vth entry of $\boldsymbol{A}$). We then define the matrix $\boldsymbol{D}$ to be the diagonal matrix of degrees. That is $\boldsymbol{D}(u, u) = d(u)$, and $\boldsymbol{D}(u, v) = 0$ for $u \neq v$. Note that $\boldsymbol{D}^{-1}$ is also a diagonal matrix, and its uth diagonal entry is $1/d(u)$.

Equation (3.2) is equivalent to:

$$\boldsymbol{p}_{t+1} = (1/2)\left(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{D}^{-1}\right)\boldsymbol{p}_t. \tag{3.3}$$

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 $\boldsymbol{W}$ denote the matrix

$$(1/2)\left(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{D}^{-1}\right), \tag{3.4}$$

and I will call it the lazy walk matrix of the graph G.

3.5 The stable distribution

In a moment, we will prove that, regardless of starting vertex, the distribution of a lazy random walk in a connected graph always converges to the same distribution. We will then relate how quickly it converges to the eigenvalues of W.

First, let's see the distribution. Let $\boldsymbol{\pi}$ be the vector given by

$$\boldsymbol{\pi}(u) = \frac{d(u)}{\sum_{v \in V} d(v)}.$$

It is easy to verify that $\boldsymbol{\pi}$ is a probability vector, and we will now observe that

$$\boldsymbol{W}\boldsymbol{\pi} = \boldsymbol{\pi}.$$

So, if the walk ever reaches the distribution $\boldsymbol{\pi}$, it will remain in that distribution. To see this, it sufficies to ignore the denominator and just show that $W\boldsymbol{d} = \boldsymbol{d}$. To prove this, note that $D^{-1}\boldsymbol{d} = \mathbf{1}$, the all-1's vector, and $\boldsymbol{A}\mathbf{1} = \boldsymbol{d}$, as the u-th coordinate of $\boldsymbol{A}\mathbf{1}$ is

$$\sum_{v:(u,v)\in E} 1 = d(u).$$

So, we see that

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

A linear-algebraic interpretation of the equation $\boldsymbol{W}\boldsymbol{\pi} = \boldsymbol{\pi}$ is that $\boldsymbol{\pi}$ is an eigenvector of $\boldsymbol{W}$ of eigenvalue 1.

In a few moments, we will prove that if G is connected, then $\boldsymbol{\pi}$ is the only eigenvector of eigenvalue 1. But, we'll first stop to review a few facts about eigenvectors and eigenvalues.

3.6 Eigenvectors and Eigenvalues

Recall that $\boldsymbol{v}$ is an eigenvector of a matrix $\boldsymbol{W}$ of eigenvalue μ if

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

Also, recall that the eigenvalues of a matrix are the same whether you multiply from the left or the right, but that the eigenvectors can be different. But, we do know that if $\boldsymbol{W}$ has an eigenvalue of μ, then there exists a vector $\boldsymbol{u}$ such that

$$\boldsymbol{u}\boldsymbol{W} = \mu\boldsymbol{u}$$

as well.

For symmetric matrices, the situation is very nice. All the eigenvalues of symmetric matrices are real, and all their eigenvectors are too. Also recall that an n-by-n symmetic matrix $\boldsymbol{M}$ has an orthonormal basis of eigenvectors. That is, for every such $\boldsymbol{M}$, there exist orthonormal vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ and real numbers $\mu_1, \ldots, \mu_n$ such that $\boldsymbol{M}\boldsymbol{v}_i = \mu\boldsymbol{v}_i$.

Since $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are an orthnormal basis, we know that for every other vector x,

$$x = \sum_i \boldsymbol{v}_i(\boldsymbol{v}_i^T x)$$

We will use this formula to understand multiplication by M. We have that for every vector x

$$\begin{aligned} \boldsymbol{M}x &= \boldsymbol{M}\sum_i \boldsymbol{v}_i(\boldsymbol{v}_i^T x) \\ &= \sum_i \boldsymbol{M}\boldsymbol{v}_i(\boldsymbol{v}_i^T x) \\ &= \sum_i \mu_i \boldsymbol{v}_i(\boldsymbol{v}_i^T x) \\ &= \left(\sum_i \mu_i \boldsymbol{v}_i \boldsymbol{v}_i^T\right) x. \end{aligned}$$

We may write this compactly as

$$\boldsymbol{M} = \boldsymbol{V}\Lambda\boldsymbol{V}^T,$$

where Λ is the diagonal matrix of eigenvalues, and $\boldsymbol{V}$ is the matrix whose columns are $v_1, \ldots, v_n$.

The matrix $\boldsymbol{W}$ is symmetric if and only if the graph G is regular (that is, if every vertex has the same degree). Yet, we can still show that all the eigenvalues and eigenvectors of $\boldsymbol{W}$ are real, because $\boldsymbol{W}$ is *similar* to a symmetric matrix. That is, we can multiply $\boldsymbol{W}$ on the left by one matrix and on the right by its inverse to obtain a symmetric matrix:

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

This matrix will occur sufficiently often in our analysis that I will give it a name, $\boldsymbol{N}$, and call it the *normalized lazy walk matrix.* The normalized lazy walk matrix is symmetric, and has the same eigenvalues as the lazy walk matrix. Let me remind you of the corespondence this similarity transformation provides between the eigenvectors. If $\boldsymbol{N}\boldsymbol{x} = \mu\boldsymbol{x}$, then

$$\begin{aligned} \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{1/2}\boldsymbol{x} &= \mu\boldsymbol{x}, && \text{which implies} \\ \boldsymbol{D}^{1/2}\boldsymbol{D}^{-1/2}\boldsymbol{W}\left(\boldsymbol{D}^{1/2}\boldsymbol{x}\right) &= \boldsymbol{D}^{1/2}\mu\boldsymbol{x}, && \text{which implies} \\ \boldsymbol{W}\left(\boldsymbol{D}^{1/2}\boldsymbol{x}\right) &= \mu\boldsymbol{D}^{1/2}\boldsymbol{x}. \end{aligned}$$

That is, $\boldsymbol{D}^{1/2}\boldsymbol{x}$ is a right-eigenvector of $\boldsymbol{W}$ of eigenvalue μ.

3.7 Uniqueness of the stable distribution

I'll now prove to you that if G is connected then the stable distribution is unique. In our proof, we will exploit the left-eigenvector of $\boldsymbol{W}$ of eigenvalue 1. It turns out to be the all-1's vector, which I write $\mathbf{1}$. This is equivalent to saying that the sum of the entries in each row of $\boldsymbol{W}$ is 1. You should check this for yourself.

We will first prove that all eigenvalues of $\boldsymbol{W}$ must lie between 0 and 1. The main reason for this is that all eigenvalues of $\boldsymbol{A}\boldsymbol{D}^{-1}$ have absolute value at most 1, which is intuitively because the sum of the coefficients in each column of this matrix is 1. Here is a formal proof.

Lemma 3.7.1. *Let $\boldsymbol{x}$ and μ satisfy $\boldsymbol{x}\boldsymbol{A}\boldsymbol{D}^{-1} = \mu\boldsymbol{x}$. Then, $|\mu| \leq 1$.*

Proof. Assume without loss of generality that

$$|\boldsymbol{x}(1)| \geq |\boldsymbol{x}(i)|\,,$$

for all i. Then,

$$\mu\boldsymbol{x}(1) = \frac{1}{d_1} \sum_{(1,j)\in E} \boldsymbol{x}(j),$$

and so

$$\begin{aligned}
|\mu|\,|\boldsymbol{x}(1)| &= \left| \frac{1}{d_1} \sum_{(1,j)\in E} \boldsymbol{x}(j) \right| \\
&\leq \frac{1}{d_1} \sum_{(1,j)\in E} |\boldsymbol{x}(j)| \\
&\leq \frac{1}{d_1} \sum_{(1,j)\in E} |\boldsymbol{x}(1)| \\
&= |\boldsymbol{x}(1)|\,.
\end{aligned}$$

So, $|\mu| \leq 1$. □

Corollary 3.7.2. *All eigenvalues of $\boldsymbol{W}$ lie between 0 and 1.*

Proof. From the formual 3.4 for $\boldsymbol{W}$, we see that the eigenvalues of $\boldsymbol{W}$ may be obtained from the eigenvalues of $\boldsymbol{A}\boldsymbol{D}^{-1/2}$ by adding 1 and then dividing by 2. □