Graphs and Networks Lecture 4

Random Walks on Graphs

Lecturer: Daniel A. Spielman September 19, 2006

4.1 Introduction

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 G with initial distribution p_0 is a distribution on sequences of vertices $(v_0, v_1, v_2, \dots)$ where v_0 is chosen according to distribution p_0, and v_{i+1} is chosen to be a random neighbor of v_i. Typically, we will take the initial distribution to be focused on a particular vertex, v_0, in which case we say that the walk starts at v_0.

We will usually be interested in the distribution of the random walk after some number of steps, typically denoted t. We will now describe this in matrix notation. To this end, we let $p_t \in \mathbb{R}^V$ denote the probability distribution of v_t–that is the probability of being at each vertex after t steps. Let A be the adjacency matrix of the graph, d_v the degree of vertex v, and $D = \mathtt{diag}(d_1, \dots, d_n)$, the diagonal matrix of the degrees. Then, set

$$M \stackrel{\text{def}}{=} D^{-1}A,$$

which I will call the walk matrix of G.

We then have

$$p_t = p_{t-1}M.$$

So that you'll believe me, I'll do a simple example.

You can find it at `http://www.cs.yale.edu/homes/spielman/462/lect4ex.pdf`.

4.2 Stationary Distribution

In previous lectures, we used the stationary distribution of such a walk to provide a ranking. In the undirected case, this is not so useful. The reason is that the solution to

$$r = rM$$

is easy to derive. It comes from taking $r(v)$ proportional to d_v. In particular, if m is the number of edges in the graph, then $r(v) = d_v/2m$ is a solution to this equation, and we will also show that every random converges to this distribution, at least if the graph is non-bipartite. To see that r is a solution to this equation, let d be the vector of $(d_1, \dots, d_n)$, and compute:

$$dD^{-1}A = \mathbf{1}A = d.$$

So, if every page on the web contained a link to every page that referenced it, PageRank would do nothing more than count the number of in-links.

4.3 Convergence

It turns out that if G is connected and non-bipartite, then every random walk eventually converges to the distribution r that was just defined. We will now go over a proof of this (leaving some details to the problem set to be given out next week), and even bound how long it takes for the walk to converge.

Our goal is to evaluate $p_0 M^t$. The way one usually examines multiplication of a vector by a matrix is by consideration of the eigenvalues and eigenvectors of M. But, M is not a symmetric matrix, so it is not immediately clear that it admits a basis of eigenvectors. Fortunately, M is similar to a symmetric matrix.

Consider the matrix $D^{-1/2} \stackrel{\text{def}}{=} \texttt{diag}(1/\sqrt{d_1}, \ldots, 1/\sqrt{d_n})$. Then,

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

So, we know that M has the same eigenvalues as

$$N \stackrel{\text{def}}{=} D^{-1/2} A D^{-1/2},$$

which is symmetric. So M has n real eigenvalues. Similarly, if v is an eigenvector of N with eigenvalue λ, then we can see that $w = v D^{1/2}$ is an eigenvector of M with eigenvalue λ:

$$w D^{-1} A = v D^{1/2} D^{-1} A = v D^{-1/2} A = v \left(D^{-1/2} A D^{-1/2} \right) D^{1/2} = \lambda v D^{1/2} = \lambda w.$$

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of N and let $v_1, \ldots, v_n$ be a basis of eigenvectors, where v_i is an eigenvector of λ_i. Recall that we can then express N by

$$N = \sum_i \lambda_i v_i^T v_i.$$

(warning: my v_is are row vectors).

Since it has been a while since some of you have taken linear algebra, let me veryify this formula by checking it on the eigenvectors. We have

$$v_j N = v_j \sum_i \lambda_i v_i^T v_i = \sum_i \lambda_i \left(v_j v_i^T \right) v_i = \lambda_j v_j,$$

as

$$v_j v_i^T = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

We then have that

$$M^t = (D^{-1/2} N D^{1/2})^t = D^{-1/2} N^t D^{1/2},$$

and so

$$p_0 M^t = \sum_i \lambda_i^t \left(p_0 D^{-1/2} v_i^T \right) v_i D^{1/2}.$$

The reason that this converges to r is that v_1 is $(\sqrt{d_1}, \ldots, \sqrt{d_n})/\sqrt{2m}$, r is a multiple of $v_1 D^{1/2}$, $\lambda_1 = 1$, and if G is connected and not bipartite, then $|\lambda_i| < 1$ for $i \geq 2$ (we will show this in a moment). So, we find

$$p_0 M^t \to p_0 D^{-1/2} v_1^T v_1 D^{1/2} = (1/\sqrt{2m}) p_0 \mathbf{1}^T v_1 D^{1/2} = (1/\sqrt{2m}) v_1 D^{1/2} = r.$$

Lemma 4.3.1. *All eigenvalues of M have absolute value at most* 1.

Proof. We will prove this using right-eigenvectors. Assume

$$D^{-1} A w = \lambda w,$$

and assume without loss of generality that

$$|w(1)| \geq |w(i)|,$$

for all i. Then,

$$\lambda w(1) = \frac{1}{d_1} \sum_{(1,j) \in E} w(j),$$

and so

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

So, $|\lambda| \leq 1$. $\square$

It will be an exercise to show that $|\lambda_i| < 1$ for $i \geq 2$ if the graph is connected and non-bipartite.

The difference between 1 and $\max(\lambda_2, |\lambda_n|)$ is called the *spectral gap*, and is related to how quickly the walk converves to the stationary distribution. Define

$$\mu = 1 - \max_{i \geq 2} |\lambda_i|.$$

Lemma 4.3.2. *If $p_0 = \chi_a$, then for all b and $t \geq 0$,*

$$|p_t(b) - r(b)| \leq \sqrt{\frac{d_b}{d_a}} (1 - \mu)^t.$$

Proof. We have $p_t(b) = \chi_a M^t \chi_b^T$. Recalling that

$$D^{-1/2} v_1^T = \mathbf{1}^T / \sqrt{2m}$$

and

$$v_i D^{1/2} = d/\sqrt{2m},$$

we compute

$$\chi_a M^t \chi_b^T = \sum_i \lambda_i^t \left(\chi_a D^{-1/2} v_i^T v_i D^{1/2} \chi_b \right) = \chi_a (\mathbf{1}^T / \sqrt{2m})(d/\sqrt{2m}) \chi_b + \sum_{i \geq 2} \lambda_i^t \chi_a D^{-1/2} v_i^T v_i D^{1/2} \chi_b.$$

Now,

$$\chi_a (\mathbf{1}^T / \sqrt{2m})(d/\sqrt{2m}) \chi_b = d\chi_b / 2m = r(b).$$

On the other hand,

$$\begin{aligned} \sum_{i \geq 2} \lambda_i^t \chi_a D^{-1/2} v_i^T v_i D^{1/2} \chi_b &\leq (1-\mu)^t \sqrt{\frac{d_b}{d_a}} \sum_{i \geq 2} \left| \lambda_i^t \chi_a v_i^T v_i \chi_b \right| \\ &\leq (1-\mu)^t \sqrt{\frac{d_b}{d_a}}, \end{aligned}$$

where the last step follows from the inequality

$$\sum_{i \geq 2} \left| \chi_a v_i^T v_i \chi_b \right| \leq 1, \tag{4.1}$$

which we will now establish. To establish this inequality, note that $\chi_a v_i = v_i(a)$. So, if we let V be the matrix whose rows are $v_1, \ldots, v_n$, then the right-hand-side of (4.1) is the inner product of the ath and bth column of V, excluding the first row. To see that this is at most one, recall that V is an orthonormal matrix, so each column of V is a unit vector. □

4.4 Lazy Random Walks

As it is a pain to tread bipartite and non-bipartite graphs differently, researchers often consider lazy random walks on graphs. These walks stay put with probability 1/2, and step to a random neighbor with probability 1/2. The matrix corresponding to these walks is given by

$$W \stackrel{\text{def}}{=} \frac{1}{2}(I + M),$$

which is similar to the matrix

$$\frac{1}{2}(I + N).$$

Such a matrix has all eigenvalues between 0 and 1, which is another reason it is convenient to study lazy random walks.