Math 262A: Random Walks

Lecture by
Fan Chung Graham

May 20 and 22, 2003

In the next two lectures, we will be looking at random walks as graphs, and their relation to the eigenvalues of the graph. There is a good book on random walks by David Aldous and Jim Fill called *Reversible Markov Chains and Random Walks on Graphs*; it is available on-line at `http://www.stat.berkeley.edu/users/aldous/RWG/book.html`.

First, some definitions. Given a graph G, a walk w is a sequence of vertices $v_0, v_1, \ldots, v_s$. The vertices need not be distinct, but consecutive vertices must be adjacent, that is, $v_i \sim v_{i+1}$. A lazy walk is a walk where we allow consecutive vertices to be the same, that is, either $v_i = v_{i+1}$ or $v_i \sim v_{i+1}$. We can think of a walk as starting a at some vertex and "walking" along the edges. The study of random walks originated with the study of stochastic processes; vertices were states of the process and edges possible transitions between the states.

Define $P(u, v)$ to be the conditional probability $\Pr[v_{i+1} = v | v_i = u]$. This is the probability of going from vertex u to vertex v along the walk. Obviously, we must go somewhere, so for all u,

$$\sum_v P(u, v) = 1.$$

Since we are dealing with undirected graphs, we will insist that $u \sim v$ implies $P(u, v) \neq 0$ and $P(v, u) \neq 0$. As an example, consider an equal probability walk, where the probability of visiting any neighbor of a vertex is proportional to the degree of the vertex. That is,

$$P(u, v) = \begin{cases} \frac{1}{d_u} & u \sim v \\ 0 & \text{otherwise.} \end{cases}$$

We also have an initial distribution f, which is the probability of starting a a given vertex. Since we must start somewhere, it has the property

$$\sum_v f(v) = 1.$$

Observe that P and f have a natural interpretation as a matrix and (row) vector. The product fP then gives the probability distribution on the vertices after taking one step in the walk. Let us look at this more closely:

$$(fP)(v) = \sum_u f(u)P(u, v)$$

Since $u \not\sim v$ implies $P(u,v) = 0$, this simplifies to

$$(fP)(v) = \sum_{u \sim v} f(u)P(u,v)$$

But this is just the sum, over all neighbors u of v, of the probability of starting at u and making a transition to v. After 2 steps, the distribution is $(fP)P$, and after s steps, the distribution is fP^s.

There is an important definition from the theory of random walks, and that is of an *ergodic* walk. A random walk is ergodic when there is a unique stationary distribution π such that for all starting distributions f,

$$\lim_{s \to \infty} fP^s = \pi.$$

The necessary conditions for a walk to be ergodic are irreducibility and non-periodicity. They may sound like foreign words, but in fact, they have very simple interpretations in our graph-theoretic view of random walks. The former simply says that the graph must be connected, and the latter says that the graph is not bipartite. It turns out that these are also sufficient conditions, but we will show this later.

Another definition we will need is that of a *reversible* random walk, which is an ergodic random walk with the property

$$\pi(u)P(u,v) = \pi(v)P(v,u).$$

A reversible random walk is just a weighted, undirected graph. To see this, pick weight assignment $w_{uv} = c\pi(u)P(u,v) = w_{vu}$, where c is some constant. Conversely, given an undirected, weighted graph, the underlying random walk has transition probabilities $P(u,v) = w_{uv}/d_u$, where $d_u = \sum_v w_{uv}$.

Let us now return to our example of a simple graph with some isolated vertices. The weight of each edge is 1, so $P = D^{-1}A$. Why this is may seem rather opaque, but recall that

$$P(u,v) = \begin{cases} \frac{1}{d_u} & u \sim v \\ 0 & \text{otherwise.} \end{cases}$$

Let us try to calculate $P^s = (D^{-1}A)^s$ efficiently. Unfortunately, P is not symmetric, so we cannot diagonalize it. But notice that

$$\begin{aligned} D^{-1}A &= D^{-\frac{1}{2}}D^{-\frac{1}{2}}AD^{-\frac{1}{2}}D^{\frac{1}{2}} \\ &= D^{-\frac{1}{2}}(I - \mathcal{L})D^{\frac{1}{2}}, \end{aligned}$$

where $\mathcal{L}$ is the normalized combinatorial Laplacian we have seen earlier, and which we know is symmetric. Now $P^s = D^{-\frac{1}{2}}(I - \mathcal{L})^s D^{\frac{1}{2}}$.

So what is the stationary distribution π for a simple graph with no isolated vertices? We claim $\mathrm{Vol}(G)\pi = \vec{1}D = (d_{v_1}, d_{v_2}, \ldots, d_{v_n})$, where $\mathrm{Vol}(G) = \sum_v d_v$. To see this, observe that $\pi P = \pi$ since

$$(\pi P)(v) = \sum_u \pi(u)P(u,v) = \sum_{u \sim v} \frac{d_u}{\mathrm{Vol}(G)} \cdot \frac{1}{d_u} = \frac{d_v}{\mathrm{Vol}(G)} = \pi(v).$$

So

$$\lim_{s \to \infty} \pi P^s = \pi,$$

and π must be the stationary distribution.

How quickly does fP^s we converge to π? One way to measure this is to look at $\|fP^s - \pi\|$, where $\|\cdot\|$ is the ℓ_2-norm, defined as

$$\sqrt{\sum_{i=1}^{n} x_i^2}$$

for an n-vector x. We have

$$\begin{aligned}
\|fP^s - \pi\| &= \|fD^{-\frac{1}{2}}(I-\mathcal{L})D^{\frac{1}{2}} - \frac{\vec{1}D}{\mathrm{Vol}(G)}\| \\
&= \|fD^{-\frac{1}{2}}\big((I-\mathcal{L})D^{\frac{1}{2}} - \phi_0^T\phi_0\big)D^{\frac{1}{2}}\|. \qquad (\star)
\end{aligned}$$

Here, ϕ_0 is the first eigenvector of $\mathcal{L}$. Recall that,

$$\phi_0 = \frac{\vec{1}D^{\frac{1}{2}}}{\sqrt{\mathrm{Vol}(G)}}$$

Now let's check $(\star)$:

$$\begin{aligned}
\langle fD^{-\frac{1}{2}}, \phi_0\rangle\phi_0 D^{\frac{1}{2}} &= \langle fD^{-\frac{1}{2}}, \frac{\vec{1}D^{\frac{1}{2}}}{\sqrt{\mathrm{Vol}(G)}}\rangle \frac{\vec{1}D^{\frac{1}{2}}D^{\frac{1}{2}}}{\sqrt{\mathrm{Vol}(G)}} \\
&= \frac{\langle f, \vec{1}\rangle}{\sqrt{\mathrm{Vol}(G)}} \cdot \frac{\vec{1}D}{\sqrt{\mathrm{Vol}(G)}} \\
&= \frac{1}{\sqrt{\mathrm{Vol}(G)}} \cdot \frac{\vec{1}D}{\sqrt{\mathrm{Vol}(G)}} \\
&= \frac{\vec{1}D}{\mathrm{Vol}(G)}.
\end{aligned}$$

Returning to our calculation at $(\star)$, let $M = (I-\mathcal{L})^s - \phi_0^T\phi_0$. Then

$$\|M\| = \sup_v \frac{\langle Mv, v\rangle}{\langle v, v\rangle} = \max_{i\neq 0} |1-\lambda_i|^s.$$

So

$$\begin{aligned}
\|fP^s - \pi\| &\leq \|fD^{-\frac{1}{2}}\| \cdot \|M\| \cdot \|D^{\frac{1}{2}}\| \\
&\leq \|fD^{-\frac{1}{2}}\| \cdot (\max_{i\neq 0}|1-\lambda_i|^s) \cdot \sqrt{d_{\max}}
\end{aligned}$$

We're almost there. Let's calculate $\|fD^{-\frac{1}{2}}\|$:

$$\|fD^{-\frac{1}{2}}\| = \sqrt{\sum_v \big(f(v)d_v^{-\frac{1}{2}}\big)^2} = \sqrt{\sum_v f^2(v)d_v^{-1}} \leq \Big(\sum_v f(v)\Big)d_{\min}^{-\frac{1}{2}} = d_{\min}^{-\frac{1}{2}}.$$

Thus

$$\|fP^s - \pi\| \leq \sqrt{\frac{d_{\max}}{d_{\min}}}(\max_{i\neq 0}|1-\lambda_i|^s).$$

A side note: from here, we get the sufficient and necessary conditions for convergence:

$$\begin{array}{rl} \lambda_1 \neq 0 & \text{(G is connected)} \\ \lambda_{n-1} \neq 2 & \text{(G is not bipartite)} \end{array}$$

Now let $\lambda' = \min\{\lambda_1, 2 - \lambda_{n-1}\}$. Then

$$\|fP^s - \pi\| \leq \sqrt{\frac{d_{\max}}{d_{\min}}} e^{-s \log(1-\lambda')}.$$

The above extends easily to lazy walks. Let

$$w'_{uv} = \begin{cases} w_{uv} & u \neq v \\ w_{uv} + cd_v & u = v, \end{cases}$$

where c is some constant. Then

$$P'(u,v) = \begin{cases} \frac{w_{uv}}{d_v(1+c)} & u \neq v \\ \frac{w_{uv}+cd_v}{d_v(1+c)} & u = v. \end{cases}$$

So $\lambda'_i = \frac{\lambda_i}{1+c}$. This means that we can converge even if the graph is bipartite! To maximize convergence rate, choose $c = \frac{1}{2}(\lambda_1 + \lambda_{n-1}) - 1$.

It is worth mentioning other kinds of distance measures. One is the relative distance, defined as

$$\Delta(s) = \max_{x,y} \frac{|P^s(y,x) - \pi(x)|}{\pi(x)}.$$

There is also the variational distance, also called the total variation. This is defined as

$$\begin{aligned} \Delta_{\mathrm{TV}}(s) &= \max_{A \subseteq V} \max_y \Big|\sum_{x \in A} P^s(y,x) - \sum_{x \in A} \pi(x)\Big| \\ &= \frac{1}{2} \max_y \sum_{x \in V} |P^s(y,x) - \pi(x)|. \end{aligned}$$

One application of random walks is in the area of sampling for approximation. Consider approximating the volume of a complex combinatorial object. If we put the object in a box and sample points uniformly at random inside the box, the fraction of points inside the object of interest gives us an estimate of its volume as a fraction of the rectangle's volume. With care, we can use a random walk to do the sampling, but we need to know how long to walk. This is exactly what the igenvalues tell us.