Spectral Graph Theory and its Applications September 23, 2004

Lecture 7

Lecturer: Daniel A. Spielman

7.1 Random Walks on Weighted Graphs

We now define random walks on weighted graphs. We will let A denote the adjacency matrix of a weighted graph. We will also the graph to have self-loops, which will correspond to diagonal entries in A. Thus, the only restriction on A is that is be symmetric and non-negative.

When our random walk is at a vertex u, it will go to node v with probability proportional to $a_{u,v}$:

$$m_{u,v} \stackrel{\text{def}}{=} \frac{a_{u,v}}{\sum_w a_{u,w}}.$$

So that $m_{u,v}$ can be the probability of moving from v to w, I am going to have to do something I hate: multiplying by vectors from the right.

In matrix notation, we can form the matrix of probabiliites, M, by setting

$$\begin{aligned} d_u &\stackrel{\text{def}}{=} \sum_w a_{u,w} \\ D &\stackrel{\text{def}}{=} \operatorname{diag}(d_1, \dots, d_n) \\ M &\stackrel{\text{def}}{=} D^{-1}A. \end{aligned}$$

I will call M the *walk matrix* of the weighted graph. We must be careful when dealing with M because it is not symmetric, and so its eigenvectors are not necessarily orthogonal, or might not even exist. However, M is very close to symmetric. If we define the *normalized adjacency matrix*

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

we can see that M and N have the same eigenvalues and related eigenvectors. To make this more precise, let v be an eigenvector of N with eigenvalue λ. Setting $w = vD^{1/2}$, we find

$$\begin{aligned} \lambda v &= vN \\ \lambda v &= vD^{1/2}MD^{-1/2} \\ \lambda w D^{-1/2} &= wD^{-1/2}D^{1/2}MD^{-1/2} \\ \lambda w D^{-1/2} &= wMD^{-1/2} \\ \lambda w &= wM. \end{aligned}$$

The main questions we will ask about random walks are: Do they converge to a steady state? How quickly do they converge? And, what can we learn about A from their convergence? For an example of a walk that doesn't converge, consider the graph consisting of two nodes connected by an edge. A random walk starting at one of the nodes will alternate between the two nodes forever. By slightly modifying A, we can force the walk to converge to a steady state: all we need to do is add a small self-loop at each vertex. In the rest of this lecture, we will consider a larger modification: we will add a self-loop at each vertex large enough to guarantee that the walk stays put with probability 1/2. That is, we want:

$$a_{u,u} \geq \sum_{v \neq u} a_{u,v}, \quad \text{which implies } m_{u,u} \geq 1/2.$$

In this case, one can show that M is positive semi-definite, its largest eigenvalue is 1, and the corresponding left eigenvector is $(d_1, \ldots, d_n)$. So, the walk will eventually settle down to hit node i with probability $d_u / \sum_w d_w$. For future use, we set

$$\sigma \stackrel{\text{def}}{=} \sum_w d_w$$

$$\pi_u \stackrel{\text{def}}{=} d_u / \sigma.$$

Knowledge about the second eigenvalue of M can be used to bound how quickly the walk converges to π. Let $p^0(i)$ denote the initial probability of being at node i, and $p^t(i)$ denote the probability of being at node i after t steps, where

$$p^t \stackrel{\text{def}}{=} p^0 M^t.$$

One can prove (see [Lov96, Theorem 5.1])

Theorem 7.1.1. *Let μ_2 denote the second-largest eigenvalue of a positive semi-definite walk matrix M. For any vertex u, let p^0 be the problitity distribution concentrated at u ($p^0(u) = 1$). Then, after t steps we have for every vertex v,*

$$\left| p^t(v) - \pi(v) \right| \leq \sqrt{\frac{d_v}{d_u}} \mu_2^t.$$

Similarly, if μ_2 is large, one can use the corresponding eigenvector to find an initial distribution that does not converge rapidly (This might be an exercise).

7.2 Conductance

For weighted graphs, and for that matter irregular graphs, there is a more natural notion than the isoperimetric number that I defined a few lectures ago. It is called *conductance*. Note: calling one concept conductance and the other isoperimetric number is my own convention. Usage in the literature is mixed.

For a partition of the vertex set of a graph $(S, \bar{S})$, we define the conductance of the cut to be

$$\Phi(S) \stackrel{\text{def}}{=} \frac{\sum_{u \in S, v \notin S} a_{u,v}}{\min\left(\sum_{w \in S} d_w, \sum_{w \notin S} d_w\right)}.$$

To simplify writing expressions such as this, I will define the volume of a set of vertices S by

$$\mathrm{vol}(S) \stackrel{\text{def}}{=} \sum_{w \in S} d_w,$$

the volume of a set of edges F to be

$$\mathrm{vol}(F) \stackrel{\text{def}}{=} \sum_{(u,v) \in F} a_{u,v},$$

and

$$\partial(S) \stackrel{\text{def}}{=} \left\{(u,v) \in E : u \in S, v \in \bar{S}\right\}.$$

So, we can write

$$\Phi(S) = \frac{\mathrm{vol}(\partial(S))}{\min(\mathrm{vol}(S), \mathrm{vol}(\bar{S}))}.$$

Finally, we define the conductance of a graph by

$$\Phi(G) \stackrel{\text{def}}{=} \min_{S \subset V} \Phi(S).$$

Cheeger's Theorem has a nicer form for conductance and walk matrices (See [Lov96, Theorem 3.5] for a proof):

$$\frac{\Phi^2}{8} \leq 1 - \mu_2 \leq \Phi.$$

Of course, you can also get a Laplacian version by taking the normalized Laplacian:

$$L = 2(I - M) = D^{-1/2}(D - A)D^{-1/2}.$$

There is a strong relationship between $\Phi(G)$ and the rate at which random walks converge. The easy direction comes from letting S be a set such that $\Phi(S) = \Phi(G)$ and $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$. Then, consider the initial distribution

$$p^0(u) = \begin{cases} d_u / \sum_{w \in S} d_w & \text{if } u \in S \\ 0 & \text{otherwise.} \end{cases}$$

In one step, the probability the walk will land in a vertex not in S is

$$\sum_{u \in S, v \notin S} p_1(u) m_{u,v} = \frac{\sum_{u \in S, v \notin S} a_{u,v}}{\sum_{u \in S} d_u} = \Phi(S).$$

One can show that in each successive step, even less probability mass will escape. So, we must wait at least $1/4\Phi(S)$ steps before even a quarter of the probability mass escapes to $\bar{S}$, which should have at least half the probabilty mass under π.

In the next section, we will prove a partial converse to this observation. That is, if $\Phi(G)$ is big, then every random walk must converge quickly.

7.3 The Lovasz-Simonovits Theorem

Most people who examine random walks use some function to determine how close the walk is to convergence. Lovasz and Simonovits [LS90] use a curve instead. To describe the curve, I will first have to introduce some notation. For now, fix some probability distribution, p. We will work with a normalized version of p, given by

$$\rho(u) \stackrel{\text{def}}{=} \frac{p(u)}{d_u}.$$

As the walk converges, $\rho(u)$ approaches $1/S$ for all u.

Random walks, and most processes on graphs, are usually best understood by treating the edges as the most important objects, rather than the vertices. I will now try to do that here. First, I will replace every edge (u, v) by two directed edges, one from u to v, denoted (u, v), and one from v to u, denoted (v, u). Then, I will consider the probability mass that is about to be transported over an edge (u, v), and denote it by

$$p(u, v) \stackrel{\text{def}}{=} p(u)m_{u,v}.$$

We will usually work with a normalized version of this term:

$$\rho(u, v) \stackrel{\text{def}}{=} \frac{p(u, v)}{a_{u,v}} = \frac{p(u)}{d_u}.$$

So, $\rho(u, v)$ only depends upon u.

Now, let $e_1, \dots, e_{2m}$ be an ordering of the directed edges satisfying

$$\rho(e_1) \geq \rho(e_2) \geq \cdots \geq \rho(e_{2m}).$$

We now define some points on the critical curve, $I(x)$. For each $0 \leq k \leq 2m$, we set

$$s_k \stackrel{\text{def}}{=} \sum_{i=1}^{k} a_{e_i}, \text{ and the point}$$

$$I(s_k) \stackrel{\text{def}}{=} \sum_{i=1}^{k} a_{e_i} \left(\rho(e_i)\right).$$

Observe that $s_{2m} = S$ and $I(s_{2m}) = 1$. We now extend I to a function on all of $[0, \sigma]$ by making it piecewise linear between these points. Note that the slope of the curve I between s_k and s_{k+1} is

$$\frac{I(s_{k+1}) - I(s_k)}{s_{k+1} - s_k} = \frac{a_{e_{k+1}} \left(\rho(e_{k+1})\right)}{a_{e_{k+1}}} = \rho(e_{k+1}).$$

Two important conclusions follow.

- As $\rho(e)$ only depends on the start vertex of edge e, and the slope only depends on $\rho(e)$, the curve does not depend on the order in which we place edges with the same start vertex.
- As $\rho(e_i)$ is monotonically decreasing, the slopes are as well. Thus, the curve is concave.

As the walk converges, the curve approaches the line from $(0,0)$ to $(\sigma, 1)$. We will now show that the curve for each time step of a walk lies under the curve for the previous step. Our notation will be to superscript all terms by t, the time step. Note that at each time step we may have a different ordering of the vertices.

Theorem 7.3.1. *For every initial distrbution p^0, all t, and every $x \in [0, \sigma]$,*

$$I^t(x) \leq I^{t-1}(x).$$

Before proving this theorem, we make one simple claim about I:

Claim 7.3.2. *For every $c_1, \ldots, c_{2m}$ such that $c_i \leq a_{e_i}$,*

$$\sum_{i=1}^{2m} c_i \left(\rho(e_i)\right) \leq I\left(\sum_{i=1}^{2m} c_i\right).$$

Proof sketch. This should be obvious: since the terms $(\rho(e_i))$ are monotonically decreasing, one maximizes the sum by maxing out the coefficients of the leading terms, as much as possible. Stated differently, if $c_1 < a_{e_1}$, then increasing c_1 and decreasing some other c_i to preserve $\sum c_i$ will increase the sum. Once you max out c_1, proceed with c_2, and so on. □

Proof of Theorem 7.3.1. Order the edges so that

$$\rho(u_1, v_1) \geq \rho(u_2, v_2) \geq \cdots \geq \rho(u_{2m}, v_{2m}).$$

It suffices to prove the theorem in the case where $x = s_k^t$ for some k so that $(u_1, v_1), \ldots, (u_k, v_k)$ are exactly the set of edges entering some set of vertices, $W = \{u_1, \ldots, u_k\}$. We then have

$$\begin{aligned}
I^t(s_k) &= \sum_{i=1}^{k} a_{(u_i,v_i)} \rho^t(u_i, v_i) \\
&= \sum_{i=1}^{k} p^t(u_i, v_i) \\
&= \sum_{i=1}^{k} p^t(u_i) \\
&= \sum_{i=1}^{k} p^{t-1}(v_i, u_i), \text{ as mass out equals mass in,} \\
&= \sum_{i=1}^{k} a_{(v_i,u_i)} \rho^{t-1}(v_i, u_i), \\
&\leq I^{t-1}\left(\sum_{i=1}^{k} a_{(v_i,u_i)}\right) \quad \text{by Claim 7.3.2} \\
&= I^{t-1}(s_k),
\end{aligned}$$

□

That was easy, so we will push it a little further: we will prove that the curve I^t has to lie below I^{t-1} by an amount depending on $\Phi(G)$.

Theorem 7.3.3. *For every initial distrbution p^0, all t, and every $x \in [0, \sigma/2]$,*

$$I^t(x) \leq \frac{1}{2}\left(I^{t-1}(x - 2\Phi x)) + I^{t-1}(x + 2\Phi x))\right)$$

and for $x \in [\sigma/2, \sigma]$,

$$I^t(x) \leq \frac{1}{2}\left(I^{t-1}(x - 2\Phi(\sigma - x))) + I^{t-1}(x + 2\Phi(\sigma - x)))\right).$$

This theorem tells us that we can draw chords below the curve I^{t-1}, below which I^t must lie. If you examine the proof, you will find that it only depends on the conductance of the level sets under p^t. So, if the walk stagnates, then you know that one of the level sets has poor conductance.

Before proving Theorem 7.3.3, we will show how it can be applied.

Theorem 7.3.4. *For every initial probability distribution, p^0, every $x \in [0, \sigma]$ and every time t,*

$$I^t(x) \leq \min\left(\sqrt{x}, \sqrt{\sigma - x}\right)\left(1 - \frac{1}{2}\Phi^2\right)^t . + x/\sigma.$$

In particular, for every set of vertices W,

$$\left|\sum_{w \in W} p^t(w) - \pi(w)\right| \leq \left(\sqrt{x}, \sqrt{\sigma - x}\right)\left(1 - \frac{1}{2}\Phi^2\right)^t,$$

where $x = \sum_{w \in W} d_w$.

Proof of Theorem 7.3.4. Consider the curve

$$R^0(x) = \min\left(\sqrt{x}, \sqrt{\sigma - x}\right) + x/\sigma.$$

It is easy to show that

$$I^0(x) \leq R^0(x), \text{for all } x \in [0, \sigma].$$

While we can not necessarily reason about what happens to the curves I^t when we draw the chords indicated by Theorem 7.3.3, we can reason about the chords under R^0. If we set

$$R^t(x) = \frac{1}{2}\left(R^{t-1}(x - 2\Phi x) + R^{t-1}(x + 2\Phi x)\right),$$

for $x \in [0, \sigma/2]$, and

$$R^t(x) = \frac{1}{2}\left(R^{t-1}(x - 2\Phi(\sigma - x)) + R^{t-1}(x + 2\Phi(\sigma - x))\right),$$

for $x \in [\sigma/2, \sigma]$, then an elementary calculation reveals that

$$R^t(x) \leq \min\left(\sqrt{x}, \sqrt{\sigma - x}\right)\left(1 - \frac{1}{2}\Phi^2\right)^t + x/\sigma.$$

As all the curves are concave, we have

$$I^t(x) \le R^t(x),$$

which proves the theorem. □

Before I prove Theorem 7.3.3, I want to conjecture that a better proof exists. Please try to find one!

Proof of Theorem 7.3.3. We will only consider the case $x \in [0, \sigma/2]$, and again observe that it suffices to prove the theorem in the case where $x = s_k$, for some k. Moreover, we may assume that the edges $(u_1, v_1), \ldots, (u_k, v_k)$ consist of all edges leaving some vertex set $W = \{u_1, \ldots, u_k\}$.

Applying the same derivation, we find

$$\sum_{i=1}^{k} a_{(u_i,v_i)} \rho^t(u_i, v_i) = \sum_{i=1}^{k} a_{(v_i,u_i)} \rho^{t-1}(v_i, u_i).$$

At this point, we stop and divide the edges $\{(v_i, u_i)\}_{i=1}^k$ into two classes. Class W_1 will consist of all edges (v_i, u_i) where $v_i \in W$ and $v_i \neq u_i$; that is, the set of internal edges excluding self-loops. Class W_2 will consist of all other edges: the self-loops (w, w) for $w \in W$ and incoming edges (v_i, u_i) for $v \notin W$ and $u \in W$. We obtain the sum

$$\sum_{(u,v) \in W_1} a_{(u,v)} \rho^{t-1}(u, v) + \sum_{(u,v) \in W_2} a_{(u,v)} \rho^{t-1}(u, v).$$

We will show momentarily that

$$\sum_{(u,v) \in W_1} a_{(u,v)} \rho^{t-1}(u, v) \le (1/2) I^{t-1}(x - 2\Phi x), \tag{7.1}$$

and

$$\sum_{(u,v) \in W_2} a_{(u,v)} \rho^{t-1}(u, v) \le (1/2) I^{t-1}(x + 2\Phi x), \tag{7.2}$$

which will complete the proof.

To prove (7.1), observe the sum of the weights of the internal, non-self-loop edges is at most $x/2 - \Phi x$. So, by Claim 7.3.2, we immediately have

$$\sum_{(v,u) \in W_1} a_{(v,u)} \rho^{t-1}((v, u)) \le I^{t-1}(x/2 - \Phi x).$$

To prove the stronger bound required by (7.1), note that we have been very loose by maxing out some coefficients, and letting others be zero. If we instead set $c_{(v,u)} = a_{(v,u)}/2$ and $c_{(v,v)} = \sum_{u:(v,u) \in W} a_{(v,u)}/2$, we have

$$c_{(v,u)} \le a_{(v,u)}/2 \tag{7.3}$$

for all (v, u) and

$$\sum_{(v,u)} c_{(v,u)} = \sum_{(v,u) \in W_1} a_{(v,u)} \leq x/2 - \Phi x,$$

so

$$\begin{aligned} \sum_{(v,u) \in W_1} a_{(v,u)} \rho^{t-1}(v, u) &= \sum_{(v,u)} c_{(v,u)} \rho^{t-1}(v, u) \\ &= (1/2) \sum_{(v,u)} 2c_{(v,u)} \rho^{t-1}(v, u) \\ &\leq (1/2) I^{t-1}(x - 2\Phi x), \end{aligned}$$

by (7.3) and Claim 7.3.2. The proof of (7.2) is similar. □

For some examination of how this proof technique can be used to find cuts around a vertex, see [ST03, Section 3].

References

[Lov96] Laszlo Lovasz. Random walks on graphs: a survey. In T. Szonyi D. Miklos, V. T. Sos, editor, *Combinatorics, Paul Erdös is Eighty, Vol. 2*, pages 353–398. Janos Bolyai Mathematical Society, Budapest, 1996.

[LS90] L. Lovasz and M. Simonovits. The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. In IEEE, editor, *Proceedings: 31st Annual Symposium on Foundations of Computer Science: October 22–24, 1990, St. Louis, Missouri*, volume 1, pages 346–354, 1109 Spring Street, Suite 300, Silver Spring, MD 20910, USA, 1990. IEEE Computer Society Press.

[ST03] Daniel A. Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. Technical report, 2003. Available `http://arxiv.org/abs/cs.DS/0310051`.