Graphs and Networks Lecture 10

Holistic Convergence of Random Walks

Daniel A. Spielman October 5, 2010

10.1 Overview

There are two things that I want to do in this lecture. The first is the holistic proof of the convergence of random walks. This will use the plot of the cumulative distribution we developed in Lecture 8. This proof was developed by Lovàsz and Simonovits [LS90]. The name “holistic” is my fault.

I give this non-spectral proof for three reasons. First, it’s just cool. Second, its very different from the spectral proof, and so is likely to extend to different situations. Third, it has algorithmic applications that the spectral proof does not.

10.2 The Holistic Approach to Convergence

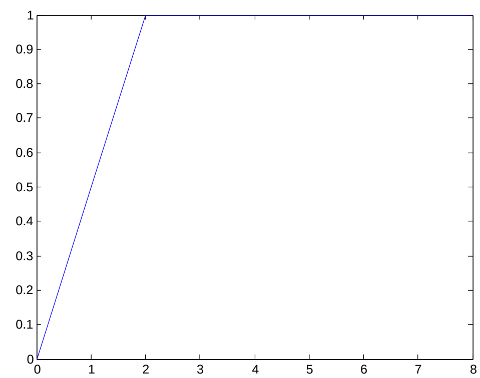
Recall from Lecture 8 that we can view a random walk as actually living on the sockets in a graph instead of on the vertices. For this lecture, we will exploit this view by transforming every edge into two directed edges (going in opposite directions). We now view the walk as living on the edges. As soon as it enters a vertex, it choose one of the edges leaving that vertex at random.

But, in this lecture we want to consider lazy random walks. These stay put with probability one-half. We model this by putting self-loops at every vertex. If a vertex has d out-going edges, then it should get d self-loops as well. These loops are directed edges from the node to itself. If our original graph had m edges, then this directed graph has $4m$ edges: the edges first double to make the directed edges, and they double again to make the self-loops. Every vertex in the original that had degree d now has $2d$ in-coming and out-going edges, d of which are the same as they are self-loops. Recall that the stable distribution occupies each vertex with probability proportional to its degree. So, it becomes uniform on the directed edges.

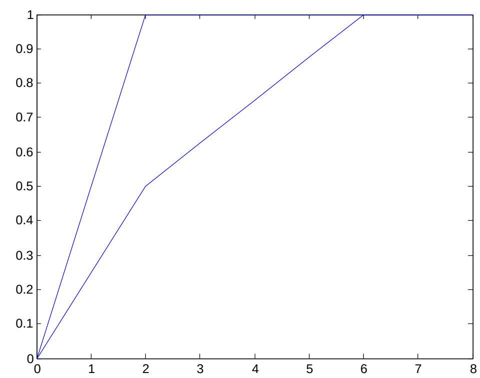
Let $\boldsymbol{p}_t$ be the probability distribution of the lazy random walk after t steps. We will let $\boldsymbol{q}_t$ denote the induced distribution on the directed edges. That is, for a directed edge (u, v), $\boldsymbol{q}_t(u, v) = \boldsymbol{p}_t(u)/d(u)$. We now consider the cumulative distribution function of $\boldsymbol{q}_t$, which we call C^t. We defined $C^t(k)$ to be the sum of the largest k values of $\boldsymbol{q}_t$.

Let’s see an example using an insanely simple graph: the path with 3 vertices and two edges. Note that this results in 8 directed edges. We run the walk for a few steps, plot the function C, and list the values on the edges.

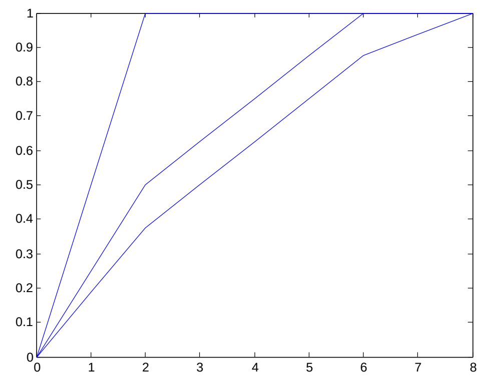
```
>> A = diag([1 1],1);
>> A = A + A'

A =

     0     1     0
     1     0     1
     0     1     0

>> deg = sum(A);
>> Di = diag(1./deg);
>> Wl = (1/2)*(eye(3) + A*Di)

Wl =

    0.5000    0.2500         0
    0.5000    0.5000    0.5000
         0    0.2500    0.5000

>> [ai,aj] = find(A);
>> sock = sparse([1:4],ai,1./deg(ai));
>> sock = [sock;sock]/2;
>> p0 = [1;0;0];
>> sock*p0

ans =

         0
    0.5000
         0
         0
         0
    0.5000
         0
         0

>> C = cumsum(sort(sock*p0,'descend'));
>> plot([0:8],[0;C])
>> hold on;
```



```
>> p1 = Wl*p0

p1 =

    0.5000
    0.5000
         0

>> C = cumsum(sort(sock*p1,'descend'))

C =

    0.2500
    0.5000
    0.6250
    0.7500
    0.8750
    1.0000
    1.0000
    1.0000

>> plot([0:8],[0;C])
```



```
>> p2 = W1*p1

p2 =

    0.3750
    0.5000
    0.1250

>> C = cumsum(sort(sock*p2,'descend'))

C =

    0.1875
    0.3750
    0.5000
    0.6250
    0.7500
    0.8750
    0.9375
    1.0000

>> plot([0:8],[0;C])
```



A few things should become obvious from examining these plots. First, I am drawing the plot as a piecewise linear function between the integer points. But, it is piecewise linear over a longer range. It is always flat accross the odd integers, because every probability value on a directed edge occurs at least twice. In fact, for a node of (original) degree d, each of its $2d$ outgoing directed edges will carry the same value (note that there are d self-loops and d edges leaving). Imagine that the vertices have been numbered so that

$$\boldsymbol{p}_t(1)/d(1) \geq \boldsymbol{p}_t(2)/d(2) \geq \cdots \geq \boldsymbol{p}_t(n)/d(n).$$

We will then have

$$\begin{aligned} C^t(d(1)) &= \boldsymbol{p}_t(1), \\ C^t(d(1)+d(2)) &= \boldsymbol{p}_t(1)+\boldsymbol{p}_t(2) \qquad \text{and} \\ C^t(d(1)+d(2)+\cdots+d(k)) &= \boldsymbol{p}_t(1)+\boldsymbol{p}_t(2)+\cdots+\boldsymbol{p}_t(k). \end{aligned}$$

The function C^t will be piecewise linear between these points. We call these points the *extreme points* as they are the only points at which the function can be non-linear. Let's give them names. Set

$$x_k^t = d(1)+d(2)+\cdots+d(k).$$

Note that to make this precise we need to re-number the nodes for each t. That is, x_k^t is the sum of the degrees of the k vertices for which the value of $\boldsymbol{p}_t(v)/d(v)$ is largest.

Here are a few other useful properties of the curve C^t.

Lemma 10.2.1. *Extend $C^t(x)$ to all real $x \in [0, 4m]$ by making it piecewise linear between integral points.*

a. The funcation $C^t(x)$ is concave.

b. For every $x \in [0, 4m]$, every s such that $x \pm s \subset [0, 4m]$ and every $r < s$,

$$\frac{1}{2}\left(C^t(x+s) + C^t(x-s)\right) \leq \frac{1}{2}\left(C^t(x+r) + C^t(x-r)\right).$$

c. For every set of directed edges F that does not contain any self-loops,

$$\boldsymbol{q}_t(F) \leq \frac{1}{2} C^t(2\,|F|).$$

Proof. Parts a and b will be on problem set 2, so I won't prove them. But, I will give the geometric explanation of the expression in part b. The line segment between the points

$$\left(x - s, C^t(x-s)\right) \quad \text{and} \left(x+s, C^t(x+s)\right).$$

Is a chord lying under the curve C^t. It contains exactly the points

$$\left(x - s + 2s\alpha, (1-\alpha)C^t(x-s) + \alpha C^t(x+s)\right),$$

for α between 0 and 1. Setting $\alpha = 1/2$ gives the point with ordinate x.

Part c is almost obvious. From the definition of C^t, it is clear that

$$\boldsymbol{q}_t(F) \leq C^t(|F|).$$

To prove part c, observe that for every edge in F there is a self loop will the same probability under $\boldsymbol{q}_t$. □

We will eventually prove convergence by proving that for each critical point x_k^t, $C^t(x_k^t)$ lies below a chord accross the curve C^{t-1}, and that the width of this chord increases with the conductance of the graph. But, we first just show that the curve at time t lies beneath the curve at time $t-1$.

Lemma 10.2.2. *For all t and all integers $0 \leq k \leq n$,*

$$C^t(x_k^t) \leq C^{t-1}(x_k^t)$$

Proof. Observe that $C^t(x_k^t)$ is the sum of the probabilities at time t on the edges leaving some set of k vertices. So, it is also the sum of the probabilities that the walk was at those vertices at time t, and thus the sum of the probabilities at time $t-1$ on the edges pointing in to those vertices. There are exactly x_k^t such edges. As $C^{t-1}(x_k^t)$ is the sum of the largest probabilities on that many edges at time $t-1$, this sum on the edges entering those k vertices is at most this large. □

As C^{t-1} is concave, C^t is piecewise linear and C^t lies below C^{t-1} at all its exteme points, we may conclude that C^t lies beneath C^{t-1}.

We now prove a stronger inequality.

Theorem 10.2.3. *Let G be a graph with conductance at least ϕ. Then, for every initial probability distribution $\boldsymbol{p}_0$, every $t \geq 0$ and every extreme point x_k^t,*

a. if $x_k^t \leq 2m$ then

$$C^t(x_k^t) \leq \frac{1}{2}\left(C^{t-1}\left(x_k^t - \phi x_k^t\right) + C^{t-1}\left(x_k^t + \phi x_k^t\right)\right),$$

b. if $x_k^t \geq 2m$ then

$$C^t(x_k^t) \leq \frac{1}{2}\left(C^{t-1}\left(x_k^t - \phi(4m - x_k^t)\right) + C^{t-1}\left(x_k^t + \phi(4m - x_k^t)\right)\right).$$

This theorem says that each extreme point on the curve C^t lies beneath a chord drawn accross the curve C^{t-1}. The width of the chord depends on how close x_k^t is to either endpoint. On the right-hand side the length of the chord is proportional to x_k^t. When x_k^t is past the half-way point, the length of the chord is proportional to the distance from x_k^t to the right side: $4m - x_k^t$.

Proof. As in the proof of Lemma 10.2.2, we first observe that $C^t(x_k^t)$ is the sum of $\boldsymbol{p}_t$ over some set of k vertices. Let's call this set S. We now partition the directed edges attached to S. Let S^{self} be all the self-loops attached to S. Of the edges that are not self-loops, we let S^{out} denote the set of directed edges leaving vertices in S and we let S^{in} denote the set of directed edges entering vertices in S. Note that S^{in} and S^{out} can overlap. Their intersection consists of exactly those edges that go from one vertex of S to another.

We have

$$\boldsymbol{p}_t(S) = \boldsymbol{q}_{t-1}(S^{self}) + \boldsymbol{q}_{t-1}(S^{in}).$$

As each edge in S^{out} may be paired with an edgein S^{self} that carries the same probability,

$$\boldsymbol{q}^{t-1}(S^{self}) = \boldsymbol{q}^{t-1}(S^{out}),$$

so

$$\begin{aligned}
\boldsymbol{p}_t(S) &= \boldsymbol{q}_{t-1}(S^{out}) + \boldsymbol{q}_{t-1}(S^{in}) \\
&= \boldsymbol{q}_{t-1}(S^{out} \cup S^{in}) + \boldsymbol{q}_{t-1}(S^{out} \cap S^{in}).
\end{aligned}$$

Now, the number of edges in $S^{out} \cap S^{in}$ is exactly equal to the number of edges attached to vertices in S minus the number that leave. So,

$$\left|S^{out} \cap S^{in}\right| = d(S) - |\partial(S)|.$$

Similarly,

$$\left|S^{out} \cup S^{in}\right| = d(S) + |\partial(S)|.$$

So, by part c of Lemma 10.2.1,

$$\boldsymbol{q}_{t-1}(S^{out} \cup S^{in}) + \boldsymbol{q}_{t-1}(S^{out} \cap S^{in}) \leq \frac{1}{2}\left(C^{t-1}(d(S) + |\partial(S)|) + C^{t-1}(d(S) - |\partial(S)|)\right).$$

For the rest of the proof, let's assume that $d(S) = x_k^t \leq 2m$. The other case is similar. In this case,

$$|\partial(S)| \geq \phi d(S),$$

so part b of Lemma 10.2.1 implies

$$\begin{aligned}\frac{1}{2}\left(C^{t-1}(d(S) + |\partial(S)|) + C^{t-1}(d(S) - |\partial(S)|)\right) &\leq \frac{1}{2}\left(C^{t-1}(d(S) + \phi d(S)) + C^{t-1}(d(S) - \phi d(S))\right) \\ &= \frac{1}{2}\left(C^{t-1}(x_k^t + \phi x_k^t) + C^{t-1}(x_k^t - \phi x_k^t)\right).\end{aligned}$$

□

Question We have only proved this for the extreme points. Does it hold for all x?

Theorem 10.2.3 tells us that when the graph has high conductance, the extreme points of the curve C^t must lie well beneath the curve C^{t-1}. It remains to use this fact to prove a concrete bound on how quickly C^t must converge to a straight line. We do this by establishing that each C^t lies beneath some conrete curve that we can understand well. That is, we will show that C^0 lies beneath some initial curve. We then show that C^t lies beneath the curve that we get by placing chords accross this initial curve t times, and we analyze how this curve behaves when we do that. We will call these curves U^t. We define

$$U^t(x) = x/4m + \min\left(\sqrt{x}, \sqrt{4m - x}\right)\left(1 - \frac{1}{8}\phi^2\right)^t.$$

As t grows, these curves quickly approach the straight line.

We will prove two lemmas about these curves.

Lemma 10.2.4. *For every* $x \in (0, 2m]$,

$$U^t(x) \leq \frac{1}{2}\left(U^{t-1}(x - \phi x) + U^{t-1}(x + \phi x)\right),$$

and for every $x \in [2m, 4m)$,

$$\frac{1}{2}\left(U^{t-1}(x - \phi(4m - x)) + U^{t-1}(x + \phi(4m - x))\right). \leq U^t(x).$$

Proof. This proof follows by considering the Taylor series for $\sqrt{1+x}$:

$$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \cdots,$$

from which we learn

$$\sqrt{1+x} \leq 1 + \frac{1}{2}x - \frac{1}{8}x^2.$$

We apply this to show that

$$\sqrt{k-\phi k}+\sqrt{k+\phi k}=\sqrt{k}\left(\sqrt{1-\phi}+\sqrt{1+\phi}\right)\leq\sqrt{k}\left(1-\frac{\phi}{2}-\frac{\phi^2}{8}+1+\frac{\phi}{2}-\frac{\phi^2}{8}\right)=\sqrt{k}\left(2-\frac{2\phi^2}{8}\right).$$

□

Lemma 10.2.5. *For every $t \geq 0$ and every $x \in [0, 4m]$,*

$$C^t(x) \leq U^t(x).$$

Proof. We prove this by induction on t. The base case of $t = 0$ is simple, so we skip it. To handle the induction, assume that for every x

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

For every extreme point $x_k^t \leq 2m$, we may apply Theorem 10.2.3 and Lemma 10.2.4 to show

$$\begin{aligned} C^t(x_k^t) &\leq \frac{1}{2}\left(C^{t-1}(x_k^t - \phi x_k^t) + C^{t-1}(x_k^t + \phi x_k^t)\right) \\ &\leq \frac{1}{2}\left(U^{t-1}(x_k^t - \phi x_k^t) + U^{t-1}(x_k^t + \phi x_k^t)\right) \\ &\leq U^t(x_k^t). \end{aligned}$$

□

Here's the implication of this lemma for convergence of the random walk.

Theorem 10.2.6. *For every initial probability distribution and every set of vertices S,*

$$\boldsymbol{p}_t(S) - \boldsymbol{\pi}(S) \leq \sqrt{d(S)}\left(1 - \frac{1}{8}\phi^2\right)^t \leq \sqrt{d(S)}\exp\left(-\frac{1}{8}t\phi^2\right).$$

10.3 Finding Sets of Small Conductance

I would now like to observe that this theorem gives us another approach to finding sets of small conductance. Last lecture, we saw Cheeger's inequality which said that we can find such sets by examining eigenvectors. We now know that we can find them by studying random walks.

If you look at this proof, you will see that we actually employed a weaker quantity than the conductance of the graph. We only needed a lower bound on the conductance of the sets S. That appeared in the proof. If each of these sets had high conductance, then we obtained fast convergence.

On the other hand, we know that if we start the random walk behind a set of small conductance, then it will converge slowly. That means that one of the sets S encoutered during the analysis must have low conductance as well. Let's make that more concrete. For each t and k, let S_k^t be the set of k vertices u maximizing the quantity $\boldsymbol{p}_t(u)/d(u)$. Break ties arbitrarily. If each of these sets S_k^t has high conductance then the walk converges quickly. So, if the walk converges slowly, then one of these sets S_k^t has low conductance. Actually, many do.

Remark Given $\boldsymbol{p}_t$, you can find the k for which the set S_k^t has least conductance in time $O(m)$. You will probably need to do this if you take the experimental route in this problem set.

By simulating the random walk, we can identify these sets, and then check if each has low conductance. For example, let's say that you wanted to find a set of low conductance near some particular vertex v. You could try to do this by starting a random walk at v, and examining the sets S_k^t that arise.

We can say something formal about this. First, recall from last lecture that if S is a set of conductance ϕ and if $\boldsymbol{p}_0 = \boldsymbol{\pi}_S$ is the initial distribution, then

$$\boldsymbol{p}_t(S) \geq 1 - t\phi.$$

We could also express this by letting χ_S be the characteristic vector of the set S. We could then say

$$\chi_S^T \boldsymbol{p}_t \geq 1 - t\phi \quad \text{and} \quad \chi_{V-S}^T \boldsymbol{p}_t \leq t\phi.$$

What if we instead start from one vertex of S, chosen according to $\boldsymbol{\pi}_S$?

Proposition 10.3.1. *Let v be a vertex chosen from S with distribution $\boldsymbol{\pi}_S$. Then, with probability at least* $1/2$*,*

$$\chi_{V-S}^T \widehat{\boldsymbol{W}}^t \chi_v \leq 2t\phi.$$

Proof. This follows from Markov's inequality, as

$$\mathbf{E}_v \left[\chi_{V-S}^T \widehat{\boldsymbol{W}}^t \chi_v \right] = \chi_{V-S}^T \widehat{\boldsymbol{W}}^t \boldsymbol{\pi}_S.$$

□

So, we know that if we start the walk from most vertices of S, then most of its mass stays inside S. Let's see what this says about the curve C^t. For concreteness, let's consider the case when

$$\boldsymbol{\pi}(S) \leq 1/4 \quad \text{and} \quad t = \phi/4.$$

We then know that with probability at least $1/2$ over the choice of v,

$$\boldsymbol{p}_t(S) \geq 1 - 2t\phi = 1/2.$$

Question Can you say anything better than this?

Now, let θ be the lowest conductance among the sets S_k^t that we find during the walk. By Theorem 10.2.6, we have

$$\begin{aligned} 1/4 &\leq \boldsymbol{p}_t(S) - \boldsymbol{\pi}(S) \\ &\leq \sqrt{d(S)} \exp\left(-\frac{1}{8} t\theta^2\right) \\ &\leq \sqrt{m/2} \exp\left(-\frac{1}{8} t\theta^2\right). \end{aligned}$$

Taking logs and rearranging terms, this gives

$$\theta \leq \sqrt{8 \ln 2\sqrt{2m}/t} = \sqrt{32 \ln 2\sqrt{2m}}\sqrt{\phi}.$$

So, we find a set whose conductance is a little more than the square root of the conductance of ϕ.

With a little more work, one can show that there is a set S_k^t that satisfies a similar guarantee and lies mostly inside S. So, starting from a random vertex inside a set of small conductance, we can find a set of small conductance lying mostly inside that set.

You are probably now asking whether we can find that set. One obstacle is that S might contain very small sets of low conductance within itself, and we might find one of these instead. Other obstacles come from computational hardness. It turns out to be NP-hard to find sets of minimum conductance. It is also computationall hard to find sets of approximate minimum conductance.

But, it is still a very reasonable to improve upon this result. OK, there are even some improvements (which I'll eventually work into the notes). But, so far none improve on this $\sqrt{\phi}$ term. I do not yet know a really good reason that we should not be able to find a small set of conductance at most $O(\phi \log n)$. (although some think this could be hard too, need a reference)

10.4 Thoughts after lecture

There might be a cleaner way to do all of this. Perhaps we should view a lazy random walk as alternating between vertices and edges. From a vertex, it moves to a random attached edge. From an edge, it moves to a random attached vertex.

References

[LS90] L. Lovàsz 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.