Graphs and Networks Lecture 10

Convergence of Random Walks and Conductance - Draft

Daniel A. Spielman October 1, 2013

10.1 Overview

I present a bound on the rate of convergence of random walks in graphs that depends upon the conductance. This proof was developed by Lovàsz and Simonovits [LS90]. The name "holistic" is my fault.

First, it's just cool. Second, it is 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 Conductance of a Graph

Recall that we defined the conductance of a set S to be

$$\phi(S) = \frac{|\partial(S)|}{d(S)}.$$

We now wish to write a definition of the *conductance of a graph* that gives a lower bound on the conductance of sets of vertices in the graph. It does not make sense to consider sets S for which $d(S) > d(V)/2$, as then the complementary set has larger conductance. So, we define the conductance of a graph G to be

$$\phi_G = \min_{S : d(S) \leq d(V)/2} \phi(S) = \min_{S \subset V} \max\left(\phi(S), \phi(V - S)\right).$$

I remark that it is sometimes more convenient to handle the balance between large and small sets by removing the min, but computing

$$m \frac{|\partial(S)|}{d(S) d(V - S)}.$$

It never differs too much from

$$\max\left(\phi(S), \phi(V - S)\right).$$

10.3 The Holistic Approach to Convergence

At the end of Lecture 8 I gave a "holistic" way of understanding the convergence of a random walk. I will explain it again, with a slight change of notation.

We began by identifying a walk on vertices with a walk on sockets. For an edge (u,v), I will call (u,v) the socket of that edge where it attaches to u, and (v,u) the socket where it attaches to v. If $\boldsymbol{p}$ is a probability vector on the vertices, I will map it to a probability vector on the sockets by distributing the mass at a vertex evenly among its sockets. I call the resulting vector on sockets $\boldsymbol{q}$. That is

$$\boldsymbol{q}(u,v) = \boldsymbol{p}(u)/d(u), \quad \text{for all } v \text{ so that } (u,v) \in E.$$

I then want to consider the sum of the largest k values in $\boldsymbol{q}$, for varying k. To make this precise, I define a function C_k that maps a vector $\boldsymbol{q}$ to the sum of its largest k values. That is,

$$C_k(\boldsymbol{q}) = \max_{|S|=k} \mathbf{1}_S^T \boldsymbol{q}.$$

If $\boldsymbol{x}$ is a vector such that $\boldsymbol{x}(1) \geq \boldsymbol{x}(2) \cdots \geq \boldsymbol{x}(n)$, then

$$C_k(\boldsymbol{x}) = \boldsymbol{x}(1) + \cdots + \boldsymbol{x}(k).$$

We now measure the convergence of the walk by looking at a plot of k against $C_k(\boldsymbol{q})$. We make it piece-wise linear between integer points. When the walk converges, we have $\boldsymbol{q}(u,v) = 1/2m$ for all sockets (u,v), where m is the number of edges in the graph and so $2m$ is the number of sockets. For this vector, $C_k(\boldsymbol{q}) = k/2m$, and the graph is a straight line. For all other vectors, the graph is convex. For all probability vectors (non-negative vectors that sum to one), the graph goes from $(0,0)$ to $(2m,1)$.

We will show that the curves obtained at one step of the walk lies beneath the curve from the previous step, and that the curves approach the straight line faster when the graph has high conductance.

10.4 Simplification

To simplify this lecture, I am going to assume in all the proofs that the graph is d-regular for a constant d. I will state the main theorem in full generality at the end of lecture.

10.5 The Curve Goes Down

In Lecture 9 we proved that the maximum mass at a socket does not increase during a walk. In this section we prove a generalization of this fact.

I begin by pointing out something about diffusion (and random walks) that should be obvious: the total amount of mass is conserved. That is, for every vector $\boldsymbol{x}$

$$\mathbf{1}^T \boldsymbol{W} \boldsymbol{x} = \mathbf{1}^T \boldsymbol{x}.$$

This is because $\mathbf{1}$ is always a left-eigenvector of a walk matrix. The same holds for a lazy walk matrix.

Theorem 10.5.1. *Let G be d-regular. Let $\boldsymbol{p}$ be a probability vector. Then, for all $0 \leq k \leq n$,*

$$C_k(\widehat{\boldsymbol{W}}\boldsymbol{p}) \leq C_k(\boldsymbol{p}).$$

I remark that for general graphs we have

$$C_k(\boldsymbol{q}_{t+1}) \leq C_k(\boldsymbol{q}_t),$$

where $\boldsymbol{q}_t$ is the distribution on sockets at the th step of a walk.

We begin by proving this in the case that $\boldsymbol{p}$ is uniform[1] on a subset of vertices.

Lemma 10.5.2. *Let G be d-regular. Let S be a subset of the vertices of G, and let $s = |S|$. Then, for all $0 \leq k \leq n$,*

$$C_k(\widehat{\boldsymbol{W}}\mathbf{1}_S) \leq \min(k, s).$$

In particular,

$$C_k(\widehat{\boldsymbol{W}}\mathbf{1}_S) \leq C_k(\mathbf{1}_S) = \min(k, s).$$

Proof. Every entry of $\mathbf{1}_S$ is 1. As G is d-regular, Theorem 9.3.1 (Maximum Density Decreases) tells us that every entry of $\widehat{\boldsymbol{W}}\mathbf{1}_S$ is at most 1. So,

$$C_k(\widehat{\boldsymbol{W}}\mathbf{1}_S) \leq k.$$

On the other hand, the conservation of mass tells us that

$$C_k(\widehat{\boldsymbol{W}}\mathbf{1}_S) = \max_{|K|=k} \mathbf{1}_K^T \widehat{\boldsymbol{W}}\mathbf{1}_S \leq \mathbf{1}^T \widehat{\boldsymbol{W}}\mathbf{1}_S = s,$$

where the inequality exploits the fact that $\widehat{\boldsymbol{W}}\mathbf{1}_S$ is a non-negative vector. □

We will prove Theorem 10.5.1 by writing a probability vector as a non-negative combination of characteristic vectors of sets.

Let $\boldsymbol{p}$ be a non-negative vector. By renumbering the vertices, we may assume without loss of generality that

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

Now, let S_i be the set $\{1, \ldots, i\}$. For $1 \leq i < n$, set

$$\delta_i = \boldsymbol{p}(i) - \boldsymbol{p}(i+1).$$

Also set

$$\delta_n = \boldsymbol{p}(n).$$

We may now write

$$\boldsymbol{p} = \sum_{i=1}^{n} \delta_i \mathbf{1}_{S_i}.$$

[1] In the general case, we use $\boldsymbol{\pi}_S$.

Claim 10.5.3. *For every* $0 \le k \le n$,

$$C_k(\boldsymbol{p}) = \sum_{i=1}^{n} \delta_i C_k(\mathbf{1}_{S_i}).$$

Proof. We have $p(j) = \sum_{i=j}^{n} \delta_j$. So,

$$\begin{aligned} C_k(\boldsymbol{p}) &= \sum_{j=1}^{k} \boldsymbol{p}(j) \\ &= \sum_{j=1}^{k} \sum_{i=j}^{n} \delta_i \\ &= \sum_{i=1}^{n} \delta_i \sum_{j \le \min(i,k)} 1 \\ &= \sum_{i=1}^{n} \delta_i \min(i,k) \\ &= \sum_{i=1}^{n} \delta_i C_k(\mathbf{1}_{S_i}). \end{aligned}$$

□

Proof of Theorem 10.5.1 . We have

$$C_k(\widehat{\boldsymbol{W}}\boldsymbol{p}) = C_k\left(\widehat{\boldsymbol{W}} \sum_{i=1}^{n} \delta_i \mathbf{1}_{S_i}\right) \le \sum_{i=1}^{n} \delta_i C_k(\widehat{\boldsymbol{W}}\mathbf{1}_{S_i}) \le \sum_{i=1}^{n} \delta_i C_k(\mathbf{1}_{S_i}) = C_k(\boldsymbol{p}),$$

where the second-to-last inequality follows from the fact that

$$C_k(p+q) \le C_k(p) + C_k(q),$$

the last inequality follows from Lemma 10.5.2 and the last equality follows from Claim 10.5.3. □

10.6 The Stronger Inequality

While we've proved that C_k decreases with every step of the walk, we have not yet established how quickly it decreases. In this section, we will show that it decreases quickly if the graph has large conductance.

The main lemma we will prove is:

Lemma 10.6.1. *Let G be d-regular. Let S be a subset of the vertices of G. Let K be a set of vertices so that*

$$C_k(\widehat{\boldsymbol{W}}\mathbf{1}_S) = \mathbf{1}_K^T \widehat{\boldsymbol{W}}\mathbf{1}_S.$$

Set $x = \phi(K)k$. Then,

$$C_k(\widehat{\boldsymbol{W}}\mathbf{1}_S) \leq \frac{1}{2}\left(C_{k-x}(\mathbf{1}_S) + C_{k+x}(\mathbf{1}_S)\right). \tag{10.1}$$

This implies that the curve of values of $C_k(\widehat{\boldsymbol{W}}\boldsymbol{p})$ lies beneath a number of chords drawn below the values of $C_k(\boldsymbol{p})$.

Theorem 10.6.2. *Let G be d-regular, and assume that $\phi_G \geq \phi$. Let $\boldsymbol{p}$ be a probability vector. Then, for all $0 \leq k \leq n$,*

$$C_k(\widehat{\boldsymbol{W}}\boldsymbol{p}) \leq \frac{1}{2}\left(C_{k-x}(\boldsymbol{p}) + C_{k+x}(\boldsymbol{p})\right),$$

where $x = \phi \min(k, n-k)$.

Proof. For every set K of size k, $|\partial(K)| / \min(k, n-k) \geq \phi$. So, $\phi(K)k = x$.

$$\begin{aligned}
C_k(\widehat{\boldsymbol{W}}\boldsymbol{p}) = C_k\left(\widehat{\boldsymbol{W}}\sum_{i=1}^{n}\delta_i\mathbf{1}_{S_i}\right) &\leq \sum_{i=1}^{n}\delta_i C_k(\widehat{\boldsymbol{W}}\mathbf{1}_{S_i}) \\
&\leq \sum_{i=1}^{n}\delta_i\frac{1}{2}\left(C_{k-x}(\mathbf{1}_{S_i}) + C_{k+x}(\mathbf{1}_{S_i})\right) = \frac{1}{2}\left(C_{k-x}(\boldsymbol{p}) + C_{k+x}(\boldsymbol{p})\right),
\end{aligned}$$

where the last inequality follows from Lemma 10.6.1 and the last equality follows from Claim 10.5.3. □

We again begin our proof by analyzing the case of characteristic vectors of sets of vertices.

Lemma 10.6.3. *Let G be d-regular. Let S and K be subset of the vertices of G, and let $s = |S|$ and $k = |K|$. Then,*

$$\mathbf{1}_K^T(\widehat{\boldsymbol{W}}\mathbf{1}_S) \leq s/2 + k/2 - \phi(K)k/2. \tag{10.2}$$

Proof. We first consider the case in which $K = S$. If every edge from S landed in S, that is if $\phi(S) = 0$, then we would get

$$\mathbf{1}_S^T\widehat{\boldsymbol{W}}\mathbf{1}_S = s,$$

and every edge would contribute $1/2d$ to the sum. Thus, the amount of mass from K that does not enter K equals the amount that goes over the boundary edges. This is

$$(1/2d)\,|\partial(K)| = (1/2d)\phi(K)d(K) = (1/2d)\phi(K)dk = \phi(K)k/2.$$

So, (10.2) holds in this case.

We now consider the case in which $k = s$, but $K \neq S$. Let t be the number of vertices of K that are not in S. Each of these vertices has degree d and so can receive an amount of mass from S equal to

$$(1/2d)d = 1/2.$$

On the other hand, for each vertex of K that is outside of S there must be another vertex of S that is not in K. The set K will fail to receive the half of the mass that stays at this vertex. So,

$$\mathbf{1}_K^T(\widehat{\boldsymbol{W}}\mathbf{1}_S) \leq s - \phi(K)k/2 + t/2 - t/2 = \leq s - \phi(K)k/2.$$

We now extend this argument to sets K of sizes different from S. If K is bigger than S, then we view it as $K_0 \cup K_1$ where $|K_0| = |S|$ and $K_1 \subseteq V - S$. We know from the first part that K_0 can receive at most $s - \phi(K)k/2$ mass from S. Each vertex in K_1 can receive at most an additional $(1/2)$ mass from S. So,

$$\mathbf{1}_K^T(\widehat{\boldsymbol{W}}\mathbf{1}_S) \leq\leq s - \phi(K)k/2 + (k-s)/2 = s/2 + k/2 - \phi(K)k/2.$$

On the other hand, if $k < s$ then we write $K = K_0 - K_1$, where $|K_0| \leq |S|$ and $K_1 \subseteq S$. This gives $|K_1| \geq s - k$. Every vertex in K_1 corresponds to a half-unit of mass from S that does not land in K. So,

$$\mathbf{1}_K^T(\widehat{\boldsymbol{W}}\mathbf{1}_S) \leq\leq s - \phi(K)k/2 - (s-k)/2 = s/2 + k/2 - \phi(K)k/2.$$

□

Extend the definition of $C_k(\boldsymbol{p})$ to non-integer k by making it linear between integers.

Proof of Lemma 10.6.1. From Lemma 10.6.3, we know

$$C_k(\widehat{\boldsymbol{W}}\mathbf{1}_S) \leq s/2 + k/2 - \phi(K)k/2 = s/2 + k/2 - x/2.$$

We have

$$C_{k-x}(\mathbf{1}_S) = \min(k-x, s)$$

and

$$C_{k+x}(\mathbf{1}_S) = \min(k+x, s).$$

So,

$$\frac{1}{2}\left(C_{k-x}(\mathbf{1}_S) + C_{k+x}(\mathbf{1}_S)\right) = \frac{1}{2}\left(\min(k-x, s) + \min(k+x, s)\right).$$

For $k - x \leq s \leq k + x$, this gives

$$= \frac{1}{2}(k - x + s) = k/2 + s/2 - x/2.$$

For $s < k - x$, this gives

$$\frac{1}{2}(2s) = s,$$

which is an upper bound on

$$C_k(\widehat{\boldsymbol{W}}\mathbf{1}_S)$$

by conservation of mass. For $s > k + x$ we obtain

$$\frac{1}{2}((k-x) + (k+x)) = k,$$

which is an upper bound on $C_k(\widehat{\boldsymbol{W}}\mathbf{1}_S)$ because each entry of $\widehat{\boldsymbol{W}}\mathbf{1}_S$ is at most 1. □

Theorem 10.6.2 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

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

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

We will prove two lemmas about these curves.

Lemma 10.6.4. *For every $x \in (0, n/2]$,*

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

and for every $x \in [n/2, n)$,

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

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 10.6.5. *For every $t \geq 0$ and every $x \in [0, n]$,*

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

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 $x_k^t \leq n/2$, we may apply Theorem **??** and Lemma 10.6.4 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 10.6.6. *For every initial probability distribution and every set of vertices S,*

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

10.7 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 10.7.1. *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

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

□

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 10.6.6, 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

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

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)

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.