CS 761 : Randomized Algorithms , Spring 2023, Waterloo

Lecture 14: Mixing Time

We analyze the mixing time of random walks through graph conductance using the Lovász-Simonovits curve.

This also gives us a "local" algorithm for graph partitioning.

Random Walks and Eigenvalues

We start with a discussion of the algebraic approach to analyze mixing time.

In the following, we assume the graph is undirected and d-regular.

Recall that if the graph is bipartite, then a random walk may not converge to its stationary distribution.

To avoid this problem, we consider lazy random walks, where we stay at the same vertex with probability $1/2$.

This won't change the mixing time much, but makes the random walk aperiodic.

So, let $W = \frac{1}{2}I + \frac{1}{2d}A$ be the transition matrix of the random walk, where A is the adjacency matrix.

Since W is a symmetric matrix (as the graph is undirected), it is known that all eigenvalues of W are real.

Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of W.

It can be shown that $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n \geq 0$ if G is connected.

Also, it is known that there is an orthonormal basis of eigenvectors $v_1, v_2, \ldots, v_n$ (i.e. $\langle v_i, v_j \rangle = 0$ if $i \neq j$, $\langle v_i, v_i \rangle = 1$ and any vector $x \in \mathbb{R}^n$ can be written as a linear combination of $v_1, \ldots, v_n$).

Let p_i be the probability distribution after i steps of random walk. Then $p_t = W^t p_0$.

Write $p_0 = c_1 v_1 + \ldots + c_n v_n$ as a linear combination of the eigenvectors.

Then $p_t = W^t p_0 = c_1 \lambda_1^t v_1 + \ldots + c_n \lambda_n^t v_n$, because v_i is an eigenvector with eigenvalue λ_i.

When $t \to \infty$, since $1 > \lambda_2 \geq \ldots \geq \lambda_n \geq 0$, $\lambda_i^t \to 0$ for $2 \leq i \leq n$, and thus $p_t \to c_1 v_1$.

Note that $v_1 = (\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$ and $c_1 = \langle p_0, v_1 \rangle = \frac{1}{\sqrt{n}}$, so we have $c_1 v_1 = (\frac{1}{n}, \ldots, \frac{1}{n})$, the stationary distribution when the graph is regular.

This proves that the random walk will converge to the stationary distribution.

Eigenvalue gap

Now, suppose $\lambda_2 \leq 1-\varepsilon$ for some constant $\varepsilon > 0$. Let π be the stationary distribution.

Then $\|p_t - \pi\|_2^2 = \|\sum_{i\geq 2} c_i \lambda_i^t v_i\|_2^2 = \sum_{i \geq 2} \lambda_i^{2t} c_i^2 \|v_i\|_2^2 \leq \|p_0\|_2^2 (1-\varepsilon)^{2t} \leq (1-\varepsilon)^{2t}$.

So, when $t = \Omega(\log n)$, then $\|p_t - \pi\|_2 \leq 1/\text{poly}(n)$, and the two distributions are very close.

This proves that when there is a large gap between the first and the second eigenvalue, then the mixing time is only $\Theta(\log n)$, logarithmic in the size of the graph.

Cheeger's inequality

When is there a large gap between the first and the second eigenvalue?

Cheeger's inequality shows that it happens if and only if the graph is an expander graph.

Again, assume the graph is d-regular.

Define the conductance of a set S as $\phi(S) = |\delta(S)|/(d \cdot |S|)$, and $\phi(G) = \min_{S:|S|\le n/2} \phi(S)$.

Then $0 \le \phi(G) \le 1$, and $\phi(G)$ is large (bounded away from zero) if and only if G is an expander.

Cheeger's inequality : $\frac{1}{2}(1-\lambda_2) \le \phi(G) \le \sqrt{2(1-\lambda_2)}$

So, λ_2 is bounded away from one iff $\phi(G)$ is bounded away from zero.

Roughly speaking, random walks have small mixing time iff the graph is an expander.

So, to prove upper bound on mixing time, one can prove a lower bound on the graph conductance.

Graph Partitioning

Finding a set of small conductance, called a sparse cut, is an important algorithmic problem that is useful in different areas of computer science, e.g. image segmentation, clustering, community detection in social networks.

The proof of Cheeger's inequality actually provides an efficient algorithm to find a sparse cut with conductance $O(\sqrt{1-\lambda_2})$, but we won't prove Cheeger's inequality in this course (see CS 466).

Instead, we will prove a direct connection between conductance and mixing time, and this would imply a graph partitioning algorithm with similar guarantee as Cheeger's inequality.

One advantage of this approach is that it gives a local graph partitioning algorithm that (sometimes) does not need to read the whole graph.

This is useful in applications when we have a massive graph, e.g. the social network.

From now on, we work again on general undirected graphs, not necessarily regular.

Lovász Simonovits Curve

In the stationary distribution, the probability at a vertex v is $d(v)/2m$, and so the probability at an edge is $1/m$. It is uniformly distributed on the edges.

For the analysis, it is more convenient to think of the graph as a directed graph, where each undirected edge is replaced by two opposite directed edges (i.e. replace uv by $u \to v$ and $v \to u$).

Since we consider lazy random walks, we also add one self-loop for each outgoing edge.

So there are totally $4m$ directed edges in the graph, and for each vertex v it has $d(v)$ outgoing edges, $d(v)$ incoming edges, and $d(v)$ self-loops.

Let p_t be the probability distribution of the lazy random walk after t steps.

We let q_t denote the induced distribution on the directed edges, i.e. $q_t(u,v) = p_t(u)/2d(u)$.

We consider the culmulative distribution function of q_t, which we call C^t.

We define $C^t(k)$ to be the sum of the largest k values of q_t.

Let $V = \{1, 2, \dots, n\}$.

If we order the vertices so that $p_t(1)/2d(1) \geq p_t(2)/2d(2) \geq \dots \geq p_t(n)/2d(n)$, then

$C^t(2d(1)) = p_t(1)$, $C^t(2d(1)+2d(2)) = p_t(1)+p_t(2)$, and $C^t(2d(1)+\dots+2d(k)) = p_t(1)+\dots+p_t(k)$.

The function will be piecewise linear between these points. We call these points the extreme points.

Let $x_k^t = 2d(1)+\dots+2d(k)$ be the location of the k-th extreme point.

Note that we need to re-order the vertices for each t.

Here are some basic properties of the curve C^t.

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

Then ① the function $C^t(x)$ is concave.

② for every x with $x \pm s \in [0, 4m]$ and every $r < s$, we have

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

③ for every subset F of directed edges with no self-loops, we have $q_t(F) \leq \frac{1}{2} C^t(2|F|)$.

Proof We prove ②, and note that ① follows immediately by setting $r = 0$.

② is equivalent to $C^t(x+s) - C^t(x+r) \leq C^t(x-r) - C^t(x-s)$, which follows from the definition as the s-r edges appear later have probabilities at most the s-r edges earlier, as we order the edges in non-increasing probabilities.

③ is also easy as the self-loops corresponding to F have the same probabilities, and so $C^t(2|F|) \geq 2q_t(F)$. □

Curve to Line

In the stationary distribution every edge has the same probability, and so the curve becomes a line.

The following theorem says that the curve converges to the straight line faster if the graph conductance is bigger.

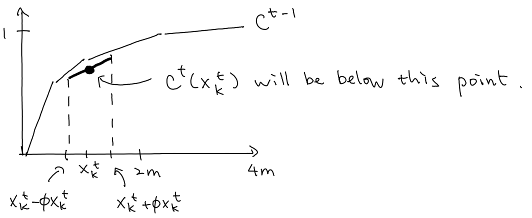
Theorem (Lovász-Simonovits) Let G be a graph with conductance at least ϕ.

Then, for any initial probability distribution p_0, every $t \geq 0$ and every x,

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

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

Pictorically, the theorem shows that



Intuitively, if the "chord" drawn is longer (when the conductance is bigger), then the new point will be lower, and thus the curve will drop faster.

When $x \leq 2m$, the length of the chord is proportional to the conductance and x; when $x \geq 2m$ the length is proportional to ϕ and $4m - x$, that is, the smaller side of the cut.

Proof. We will just prove the theorem for an extreme point x, and the rest follows from concavity.

Observe that $C^t(x_k^t)$ is the sum of the probabilities of the k largest vertices.

Let this set of k vertices be S.

Let S^{self} be all the self-loops attached to S. For those edges that are not self-loops, let S^{out} be the set of edges with their tails in S, and let S^{in} be the set of edges with their heads in S. Note that S^{in} and S^{out} can overlap; their intersection is the set of edges with both endpoints in S.

Note that the sum of the probabilities in S at time t is equal to the sum of the probabilities of the edges coming in S at time $t-1$.

That is, $p_t(S) = q_{t-1}(S^{self}) + q_{t-1}(S^{in})$.

As the probability at each self-loop is equal to its corresponding outgoing edge, we have

$$q_{t-1}(S^{self}) = q_{t-1}(S^{out}).$$

So, $p_t(S) = q_{t-1}(S^{out}) + q_{t-1}(S^{in}) = q_{t-1}(S^{out} \cap S^{in}) + q_{t-1}(S^{out} \cup S^{in})$.

Observe that $|S^{in} \cap S^{out}| = vol(S) - |\delta(S)|$, and $|S^{in} \cup S^{out}| = vol(S) + |\delta(S)|$.

So, by ③ of the previous lemma, we have

$$p_t(S) \leq \frac{1}{2}\left(C^{t-1}(2vol(S) + 2|\delta(S)|) + C^{t-1}(2vol(S) - 2|\delta(S)|)\right),$$

where $vol(S)$ is defined as $\sum_{v \in S} d(v)$.

Assume $x \leq 2m$. The other case is similar.

Then $|\delta(S)| \geq \phi\, vol(S)$. So, by ② of the previous lemma,

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

Since $x_k^t = 2\mathrm{vol}(S)$, this implies $p_t(S) \le \frac{1}{2}\left(C^{t-1}(x_k^t + \phi x_k^t) + C^{t-1}(x_k^t - \phi x_k^t)\right)$. $\square$

Upper Bound

We prove that the curve C^t is always below the curve U^t, defined by

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

It is clear that $C^0 \le U^0$.

The proof will follow by induction once we show that

- for every $x \in (0, 2m]$, $\frac{1}{2}\left(U^{t-1}(x-\phi x) + U^{t-1}(x+\phi x)\right) \le U^t(x)$
- for every $x \in [2m, 4m)$, $\frac{1}{2}\left(U^{t-1}(x-\phi(4m-x)) + U^{t-1}(x+\phi(4m-x))\right) \le U^t(x)$,

because e.g. $C^t(x) \le \frac{1}{2}\left(C^{t-1}(x-\phi x) + C^{t-1}(x+\phi x)\right) \le \frac{1}{2}\left(U^{t-1}(x-\phi x) + U^{t-1}(x+\phi x)\right) \le U^t(x)$.

For $x \in (0, 2m]$, $\frac{1}{2}\left(U^{t-1}(x-\phi x) + U^{t-1}(x+\phi x)\right) \le \frac{1}{2}\left(\sqrt{x-\phi x} + \sqrt{x+\phi x}\right)\left(1-\frac{1}{8}\phi^2\right)^{t-1} = \frac{1}{2}\sqrt{x}\left(\sqrt{1-\phi} + \sqrt{1+\phi}\right)\left(1-\frac{1}{8}\phi^2\right)^{t-1}$

Note that by Taylor series for $\sqrt{1+\phi}$ at $\phi = 0$, we have $\sqrt{1+\phi} = 1 + \frac{1}{2}\phi - \frac{1}{8}\phi^2 + \frac{1}{16}\phi^3 - \dots$,

and so $\frac{1}{2}\left(U^{t-1}(x-\phi x) + U^{t-1}(x+\phi x)\right) \le \frac{1}{2}\sqrt{x}\left(1 - \frac{1}{2}\phi - \frac{1}{8}\phi^2 - \dots + 1 + \frac{1}{2}\phi - \frac{1}{8}\phi^2 + \dots\right)\left(1-\frac{1}{8}\phi^2\right)^{t-1} \le \sqrt{x}\left(1-\frac{1}{8}\phi^2\right)^t = U^t(x)$.

By subtracting C^t from U^t, we have the following bound on the convergence rate of the random walks.

Theorem $p_t(S) - \pi(S) \le \sqrt{\mathrm{vol}(S)}\left(1-\frac{1}{8}\phi^2\right)^t$.

So, if ϕ is a constant, then after $\Theta(\log n)$ steps, then p_t is very close to the stationary distribution.

Sparse Cut by Random Walks

We show how to apply the Lovász-Simonovits theorem to find a sparse cut.

Suppose S is a set of conductance ϕ.

Then, it is not difficult to show that there is a vertex v in S such that if the random walk starts at v, then $p_t(S) \ge 1 - t\phi$. That is, at each step $\le \phi$ probability mass leaves S.

Let $\pi(S) = \sum_{v \in S} \pi(v) \le \frac{1}{2}$.

Let's run the random walk at vertex v for $t = \frac{1}{4\phi}$ steps.

Then $p_t(S) \ge 1 - \left(\frac{1}{4\phi}\right)\phi \ge 1 - \frac{1}{4} \ge 3/4$.

Notice that in the proof of the Lovász-Simonovtis theorem, we only use the fact that the conductance of S_k^i is at least θ, where S_k^i denotes the set of k highest probability vertices at time step i.

So, if $\min \phi(S_k^i) = \theta$ for $1 \le i \le \frac{1}{4\phi}$ and for all k, then by the theorem we have

$$\begin{aligned} 1/4 \le p_t(S) - \pi(S) &\le \sqrt{\mathrm{vol}(S)}\left(1-\tfrac{1}{8}\theta^2\right)^t \\ &\le \sqrt{2m}\, e^{-\frac{t\theta^2}{8}}, \end{aligned}$$

and this implies that $e^{t\theta^2/8} \le 4\sqrt{2m}$, hence $\theta \le \sqrt{8\ln(4\sqrt{2m})/t} = O(\sqrt{\phi \ln m})$.

and this implies that $e^{t\theta^2/8} \le 4\sqrt{2m}$, hence $\theta \le \sqrt{8\ln(4\sqrt{2m})/t} = O(\sqrt{\phi \ln m})$.

It means that one of the sets in S_k^i must have conductance at most $O(\sqrt{\phi}\sqrt{\ln m})$.

We can check all sets S_k^i for $i \le 1/4\phi$ in polynomial time, thereby obtaining an approximation algorithm for sparse cut almost as good as the spectral partitioning algorithm.

Local Graph Partitioning Algorithm

It is possible to keep exploring a small part of the graph by truncated random walk.

That is, whenever the probability at a vertex v is at most $\varepsilon \cdot d(v)$, then we set the probability to be zero.

Note that the number of vertices with positive probability is at most $1/\varepsilon$.

By choosing an appropriate ε, we can keep the set small and the error small.

Moreover, we can have some control over the output size using this approach.

The following is a more precise statement.

Theorem For an undirected graph $G=(V,E)$ and a set $U \subseteq V$, given $\phi \geq \phi(U)$ and $k \geq \mathrm{vol}(U)$, there exists initial vertices such that the truncated random walk algorithm can find a set S with $\phi(S) \le O(\sqrt{\phi/\varepsilon})$ and $\mathrm{vol}(S) \le O(k^{1+\varepsilon})$ for any $\varepsilon > 2/k$. The runtime of the algorithm is $\tilde{O}(k^{1+2\varepsilon}\phi^{-2})$.

References

- Lovász, Simonovits. The mixing time of Markov chains, an isoperimetric inequality, and computing the volume.
- Spielman, Teng. A local clustering algorithm for massive graphs and its applications to nearly linear time graph partitioning.
- Kwok, Lau. Finding small sparse cuts by random walk.