CS 270 Combinatorial Algorithms and Data Structures, Spring 2015

Lecture 24: Small-set expansion

Today we random walks to find small sparse cuts. This will lead to a "local" graph partitioning algorithm with performance similar to the spectral partitioning algorithm. Also, the analysis will allow us to prove that if there are many small eigenvalues in the Laplacian matrix, then there is always a small sparse cut.

Small-set expansion

We are interested in finding a small sparse cut, i.e. a set S with $\phi(S)$ small and $|S|$ small.

Let $\phi_\delta(G) := \min_{S: |S| \leq \delta |V|} \phi_\delta(S)$ be the δ-small-set expansion of G.

This problem is natural and has applications say in finding a small community in a social network.

Often the graph is very big. and it would be useful to have an algorithm with running time only depends on the output size (more precisely, depends on $|S|$ and $\text{polylog}(|V|)$), so that its running time is sublinear when $|S|$ is small. We call such algorithms "local" algorithms.

The spectral partitioning algorithm can be implemented in near-linear time, and today we will see a random walk algorithm with similar performance guarantee and can be implemented locally.

The analysis will show some close connection of spectral partitioning and random walks.

Another motivation to study this problem is from the small-set expansion conjecture by Raghavendra and Steurer, which roughly says that the performance ratio of any polynomial time algorithm that approximates $\phi_\delta(G)$ must depend on δ.

More formally, the conjecture states that for any ε, there exists δ such that it is NP-hard to distinguish the following two cases:

1. There exists $S \subseteq V$ with $\phi(S) \leq \varepsilon$ and $|S| \leq \delta n$.
2. $\phi(S) \geq 1-\varepsilon$ for every set $S \subseteq V$ with $|S| \leq \delta n$.

If this conjecture is true, then the unique games conjecture (which we will define next) is true, and this would imply optimal inapproximability results for a few problems, including the 0.878-approximation for maxcut and the 2-approximation for minimum vertex cover.

If the conjecture is false, then probably the unique games conjecture is false (no formal proof for this direction), and hopefully the techniques in disproving it would lead to improved approximation algorithms.

The uniques game conjecture can be stated as follows.

Consider the problem of linear equations over two variables modulo R, i.e. $x_3 + x_5 \equiv 2 \pmod R$, etc.

The conjecture by Khot states that for any ε there exists R s.t. it is NP-hard to distinguish the following:

① There exists a solution satisfying $(1-\varepsilon)$-fraction of equations.

② No solutions satisfying more than ε-fration of equations.

Actually, the result that we are going to discuss about eigenvalues and small-set expansion can be used to design a subexponential time algorithm to distinguish between these two cases.

Random walk vectors

By the analysis of Cheeger's inequality, we know that if we are given a vector $x \in \mathbb{R}^n$, then we can find a sparse cut $S \subseteq \text{supp}(x) := \{ i \mid x_i > 0 \}$ with $\phi(S) \leq \sqrt{2R(x)}$ where $R(x) = \frac{x^T \mathcal{L} x}{x^T x} = \frac{\sum_{i \sim j} (x_i - x_j)^2}{d \sum_i x_i^2}$.

Again, we assume the graph is d-regular throughout.

Algorithm outline

The algorithm is very simple. So let me state it informally first, without specifying the parameters.

Let $W := \frac{1}{2} I + \frac{1}{2} \mathcal{A}$ be the lazy random walk matrix.

① For each vertex $i \in V$, compute $W^t \chi_i$ for some appropriate t, where $\chi_i \in \mathbb{R}^n$ has one in i-th position and zero otherwise.

② "Truncate" $W^t \chi_i$ to a vector with "small" support.

③ Apply cheeger rounding to the truncated vectors to obtain a small sparse cut.

Analysis outline

For ①, we will prove that the vectors $W^t \chi_i$ would have small Rayleigh quotient, for all $i \in V$.

For ②, we will prove that if there is a small sparse cut S, there exists some vertex $i \in S$ such that $\|W^t \chi_i\|_2$ is large, and that will imply that it can be truncated into a vector of small support, without increasing the Rayleigh quotient ~~too~~ much.

For ③, once we have a vector with small Rayleigh quotient and small support, then cheeger rounding would produce a small sparse cut, and this part is straightforward.

For ② , there are two approaches to show that $\|W^t x_i\|_2$ is large.

One approach is to argue about the "staying" or "escaping" probability of a set S, and this will give us a bicriteria approximation algorithm for the small-set expansion problem.

Another approach is to bound $\|W^t x_i\|_2$ using eigenvalues, and this gives us a sufficient (spectral) condition for the existence of a small sparse cut.

<u>**Rayleigh quotient**</u>

Now we carry out the analysis of the first step, to show that the Rayleigh quotient of $W^t x_i$ is small when t is large enough.

This should not be surprising, because we know that $W^t x_i \to \pi = \frac{\vec{1}}{n}$ (when G is d-regular), and so the Rayleigh quotient tends to zero when $t \to \infty$.

What is important is the precise convergence rate, as in the second step we cannot afford to set t too large, and this is the tension for the correct choice of t.

The analysis is similar to the analysis of the power method, which is a way to compute the largest eigenvector of a matrix.

<u>Lemma 1</u> $R(W^t x_i) \leq 2 - 2\|W^t x_i\|_2^{\frac{1}{t}}$, where $R(x) = \frac{x^T \mathcal{L} x}{x^T x}$.

<u>proof</u> Let the eigenvalues of W be $1 = \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq 0$.

Note that $W = \frac{1}{2} I + \frac{1}{2} \mathcal{A} = I - \frac{1}{2}(I - \mathcal{A}) = I - \frac{1}{2}\mathcal{L}$.

Therefore, $R(W^t x_i) := \frac{(x_i W^t)\mathcal{L}(W^t x_i)}{\|W^t x_i\|_2^2} = 2 - 2\frac{(x_i W^t) W (W^t x_i)}{\|W^t x_i\|_2^2}$.

Write $x_i = \sum_{i=1}^n c_i v_i$, where $v_1, \ldots, v_n$ are the eigenvectors of W with eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_n$.

Then, $W^t x_i = \sum_{i=1}^n c_i \alpha_i^t v_i$, and $\|W^t x_i\|^2 = \sum_{i=1}^n c_i^2 \alpha_i^{2t}$.

Hence, $\frac{(x_i W^t) W (W^t x_i)}{\|W^t x_i\|^2} = \frac{\sum_{i=1}^n c_i^2 \alpha_i^{2t+1}}{\sum_{i=1}^n c_i^2 \alpha_i^{2t}}$.

Now, we want to apply the power means inequality, which states that if $\sum_{i=1}^n w_i = 1$ and $w_i \geq 0 \ \forall i$, then $\left(\sum_{i=1}^n w_i y_i^p\right)^{\frac{1}{p}} \geq \left(\sum_{i=1}^n w_i y_i^q\right)^{\frac{1}{q}}$ for $p \geq q$.

Notice that $c_i^2 \geq 0 \ \forall i$ and $\sum_{i=1}^n c_i^2 = \underset{\uparrow \text{ orthonormal basis}}{=} \langle \sum_{i=1}^n c_i v_i, \sum_{i=1}^n c_i v_i \rangle = \langle x_i, x_i \rangle = 1$.

So, we can apply the power means inequality by setting $w_i = c_i^2$ and $y_i = \alpha_i$ to get
$$\left(\sum_{i=1}^n c_i^2 \alpha_i^{2t+1}\right)^{\frac{1}{2t+1}} \geq \left(\sum_{i=1}^n c_i^2 \alpha_i^{2t}\right)^{\frac{1}{2t}}.$$

This implies that $$\frac{\sum_{i=1}^n c_i^2 \alpha_i^{2t+1}}{\sum_{i=1}^n c_i^2 \alpha_i^{2t}} \geq \left(\sum_{i=1}^n c_i^2 \alpha_i^{2t}\right)^{\frac{1}{2t}} = \left(\|W^t x_i\|_2^2\right)^{\frac{1}{2t}} = \|W^t x_i\|_2^{\frac{1}{t}}.$$

Therefore, we have $R(W^t x_i) \leq 2 - 2\|W^t x_i\|_2^{\frac{1}{t}}$. □

To get a feeling what it gives us, first observe that $\|W^t x_i\|_2 \geq \frac{1}{\sqrt{n}}$, which is minimized when $W^t x_i = \frac{\vec{1}}{n}$.

So, $R(W^t x_i) \leq 2\left(1 - \frac{1}{\sqrt{n}}^{\frac{1}{t}}\right) = 2\left(1 - e^{-\frac{1}{2}\log n/t}\right) \approx \log n/t$ (since $e^{-x} \approx 1 - x$).

Therefore, if we set $t = \frac{\log n}{\lambda_2}$, then $R(W^t x_i) \leq \lambda_2$,

and if we set $t = \frac{1}{\lambda_2}$, then $R(W^t x_i) \leq \lambda_2 \log n$.

We will use these choices of t later on.

Large 2-norm and small support vectors

Ideally, we can directly obtain a vector of small support, but the support size of a vector is a discrete property and is not easy to be reasoned analytically.

Instead, we consider a "relaxation" of the discrete property, and it turns out that there is a simple property that is easy to handle and is close to having a small support (for our purpose).

For any vector $x \in \mathbb{R}^n$ with at most δn nonzeros, we have $\|x\|_1 \leq \sqrt{\delta n}\,\|x\|_2$ by Cauchy-Schwarz.

So, for a vector with small support, its 2-norm must be large and this is the quantity we work with.

In our problem, we consider random walk vectors of the form $W^t x_i$, which is a probability distribution.

So, if $W^t x_i$ has at most δn nonzeros, then $1 = \|W^t x_i\|_1 \leq \sqrt{\delta n}\,\|W^t x_i\|_2$ and thus $\|W^t x_i\|_2^2 \geq \frac{1}{\delta n}$.

The good news is that if $W^t x_i$ has small Rayleigh quotient and large 2-norm, there is a simple operation to turn it into a vector of small Rayleigh quotient and small support.

<u>Lemma 2</u> Let $x \in \mathbb{R}^n$ be a non-negative vector with $\|x\|_1^2 \leq \delta n \|x\|_2^2$.

Then there exists a vector $y \in \mathbb{R}^n$ with $|\text{supp}(y)| = O(\delta n)$ and $R(y) = O(R(x))$.

<u>proof</u> The proof is by a simple truncation argument.

By scaling, we can assume that $\|x\|_2^2 = \delta n$ and $\|x\|_1 \leq \delta n$.

Let $y \in \mathbb{R}^n$ be the vector with $y_i = \max\{x_i - \frac{1}{4}, 0\}$.

Then, it is clear that $|\text{supp}(y)| \leq 4\delta n$, as otherwise $\|x\|_1 > \delta n$.

We just need to compare $R(y) = \frac{y^T \mathcal{L} y}{y^T y} = \frac{\sum_{i\sim j}(y_i - y_j)^2}{d \sum_i y_i^2}$ with $R(x) = \frac{\sum_{i\sim j}(x_i - x_j)^2}{d \sum_i x_i^2}$.

First, notice that for each $ij \in E$, we have $(y_i - y_j)^2 \leq (x_i - x_j)^2$, as truncation won't make an edge longer.

So, the numerator of y is not larger than the numerator of x, and it remains to compare the denominators.

Note that $y_i^2 \geq x_i^2 - \frac{1}{2}x_i$, and so $\sum_i y_i^2 \geq \sum_i x_i^2 - \frac{1}{2}\sum_i x_i = \delta n - \frac{1}{2}\delta n = \delta n/2 = \|x\|_2^2/2$

Combining, we have $R(y) = \frac{\sum_{i\sim j}(y_i - y_j)^2}{d \sum_i y_i^2} \leq \frac{\sum_{i\sim j}(x_i - x_j)^2}{d \sum_i x_i^2/2} = 2R(x)$, and we are done. $\square$

With this truncation lemma, it suffices for us to find a vector with small Rayleigh quotient and large 2-norm, and then we can truncate it to obtain a vector with small Rayleigh quotient and small support, and then we can just apply Cheeger's rounding in L22 to finish the proof.

Henceforth, we focus on bounding the 2-norm of a random walk vector.

Higher eigenvalues

Let the eigenvalues of W be $1 = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n \geq 0$.

Suppose there are many large eigenvalues, say $\alpha_k \geq 1 - \varepsilon$ for a small ε and a large r, which corresponds to the case that $\lambda_k \leq 2\varepsilon$ for the normalized Laplacian matrix $\mathcal{L}$.

Consider W^t, which has eigenvalues $1 = \alpha_1^t \geq \alpha_2^t \geq \ldots \geq \alpha_n^t \geq 0$.

Since trace = sum of eigenvalues, we have $\sum_{i=1}^n \chi_i^T W^t \chi_i = \sum_{i=1}^n W_{i,i}^t = \mathrm{Tr}(W^t) = \sum_{i=1}^n \alpha_i^t \geq k(1-\varepsilon)^t$.

It follows that there exists i with $\|W^{t/2}\chi_i\|_2^2 \geq \frac{k}{n}(1-\varepsilon)^t$.

Suppose $k = n^{2\beta}$ for some $0 < \beta < \frac{1}{2}$.

By setting $t = \frac{\beta \ln n}{\varepsilon}$, we have $\|W^{t/2}\chi_i\|_2^2 \geq \frac{k}{n}(1-\varepsilon)^t = \frac{n^{2\beta}}{n}(1-\varepsilon)^{\frac{\beta \ln n}{\varepsilon}} \approx \frac{n^\beta}{n}$.

By Lemma 1, we have
$$\begin{aligned} R(W^{t/2}\chi_i) &\leq 2 - 2\|W^{t/2}\chi_i\|_2^{\frac{2}{t}} \leq 2 - 2\left(n^{\beta-1}\right)^{\frac{1}{t}} = 2 - 2e^{(\beta-1)\frac{\ln n}{t}} \\ &= 2\left(1 - e^{(\beta-1)(\frac{\varepsilon}{\beta})}\right) \\ &\approx 2(1-\beta)\left(\frac{\varepsilon}{\beta}\right) \end{aligned}$$

So, we have $R(W^{t/2}\chi_i) = O(\frac{\varepsilon}{\beta})$ and $\|W^{t/2}\chi_i\| \geq \frac{1}{n^{-\beta}n}$.

By Lemma 2, we obtain a vector y with $R(y) = O(\frac{\varepsilon}{\beta})$ and $\mathrm{supp}(y) = O(n^{1-\beta})$ by using $\delta = n^{-\beta}$.

By Cheeger rounding, we get a set $S \subseteq V$ with $\phi(S) = O(\sqrt{R(y)}) = O\left(\sqrt{\frac{\varepsilon}{\beta}}\right)$ and $|S| = O(n^{1-\beta})$.

This proves the result by Arora, Barak and Steurer about higher eigenvalues and small-set expansion.

<u>Theorem</u> For $k = n^{2\beta}$, there is a polynomial time algorithm to find a set S with

Theorem For $K = n^{2\beta}$, there is a polynomial time algorithm to find a set S with $\phi(S) = O\left(\sqrt{\frac{\lambda_K}{\beta}}\right)$ and $|S| = O(n^{1-\beta})$, where λ_K is the K-th smallest eigenvalue of $\mathcal{L}$.

They use this result to decompose the graph recursively until all the pieces have few small eigenvalues (as otherwise the process can be continued).

In each such piece, they show that one can do exhaustive search on the eigenspace to find a good solution (since the dimension is low, the exhaustive search can be faster), and combining the two steps gives a subexponential time algorithm for unique games.

Approximation algorithm

The above section gives a sufficient condition for the existence of a small sparse cut.

In this section, we present a bicriteria approximation algorithm for finding a small sparse cut.

The approach is similar, but we need a different way to bound $\|W^t x_i\|_2$.

The idea is that if S is a small sparse cut, then when we start a random walk from a vertex $i \in S$, the walk will stay within S with a reasonable probability, and so the entries in $W^t x_i$ corresponding to the vertices in S will have large values, and thus $\|W^t x_i\|_2$ large.

So, let's try to analyze the probability that the random walk stay within S for t steps.

Claim Let $p_0 = \frac{\chi_S}{|S|}$ and $p_i = W^i p_0$. Then $\sum_{v \in S} p_t(v) \geq 1 - t \cdot \phi(S)$.

proof We prove it by a simple inductive argument.

We lower bound $\sum_{v \in S} p_t(v)$ by the probability that the random walk stays within S in <u>all</u> t steps.

Equivalently, we upper bound the probability that the random walk go outside of S in any of these t steps.

We start with p_0, the uniform distribution in S, where each vertex in S has probability $\frac{1}{|S|}$.

Since the graph is d-regular, each edge going out of S will carry $\frac{1}{d|S|}$ probability out of S.

So, the total probability escaping out of S is $|\delta(S)| \cdot \frac{1}{d|S|} = \phi(S)$ in the first step.

We would like to argue that the total escaping probability at each step is at most $\phi(S)$, and thus the total escaping probability is at most $t \cdot \phi(S)$, and thus the staying probability is at least $1 - t \cdot \phi(S)$, and this would imply the claim.

To finish the proof, we just need to observe the invariant that the probability at each vertex in S at each time step is at most $\frac{1}{|S|}$, and thus the same calculation holds.

The observation follows from the equation $p_{i+1}(v) = \frac{1}{2} p_i(v) + \frac{1}{2d} \sum_{\substack{u: uv \\ u \in S}} p_i(u) \le \frac{1}{|S|}$. □

<u>Corollary</u> There exists a vertex $v \in S$ such that if $p_0 = \chi_v$ then $\sum_{i \in S} p_t(i) \ge 1 - t \cdot \phi(S)$.

<u>proof</u> We use the fact that $\frac{\chi_S}{|S|}$ is a convex combination of $\chi_i : i \in S$.

Let $p_{t,i} = W^t \chi_i$ and $p_{t,\pi} = W^t \left(\frac{\chi_S}{|S|}\right)$. Note that $\frac{1}{|S|} \sum_{i \in S} W^t \chi_i = W^t \left(\frac{\chi_S}{|S|}\right)$.

So, $\frac{1}{|S|} \sum_{i \in S} \sum_{j \in S} p_{t,i}(j) = \sum_{j \in S} p_{t,\pi}(j) \ge 1 - t \cdot \phi(S)$ by the claim.

Therefore, there exists a vertex v with $\sum_{j \in S} p_{t,v}(j) \ge 1 - t \cdot \phi(S)$. □

<u>Corollary</u> There exists $S' \subseteq S$ with $|S'| \ge |S|/2$ such that if $p_0 = \chi_v$ for $v \in S'$, then $\sum_{j \in S} p_t(j) \ge 1 - 2t\,\phi(S)$.

<u>proof</u> by Markov's inequality (exercise).

Now, we can bound the 2-norm of the random walk vectors.

<u>Lemma</u> There exists $i \in S$ with $\|W^t \chi_i\|_2^2 \ge \frac{1}{|S|} (1 - t \cdot \phi(S))^2$.

<u>proof</u> Choose the vertex $i \in S$ that is guaranteed by the first corollary.

Then $\|W^t \chi_i\|_2^2 \ge \sum_{j \in S} (W^t \chi_i)(j)^2 \ge \frac{1}{|S|} \left(\sum_{j \in S} (W^t \chi_i)(j) \right)^2 \ge \frac{1}{|S|} (1 - t \cdot \phi(S))^2$. □

<u>Putting together</u>

We are ready to complete the analysis.

Set $t = \frac{1}{2\phi(S)}$.

Then, by Lemma 3, there exists i with $\|W^t \chi_i\|_2^2 \ge \frac{1}{4|S|}$.

By Lemma 1, $R(W^t \chi_i) \le 2(1 - \|W^t \chi_i\|_2^{\frac{2}{t}}) \le 2\left(1 - \frac{1}{2\sqrt{|S|}}^{2\phi(S)}\right) = 2\left(1 - e^{-\ln(2\sqrt{|S|}) \cdot 2\phi(S)}\right) = O(\phi(S) \ln(|S|))$.

By Lemma 2, there exists y with $R(y) = O(\phi(S) \ln(|S|))$ and $\text{supp}(y) = O(|S|)$.

By Cheeger rounding in L22, we find a set S' with $\phi(S') = \sqrt{\phi(S) \ln(|S|)}$ and $|S'| = O(|S|)$.

We prove the following bicriteria result.

<u>Theorem</u> If there is a set S^* with $\phi(S^*) = \phi$ and $|S^*| = \delta n$, then we can find in polynomial time a set S with $\phi(S) = O\left(\sqrt{\phi \log |S^*|}\right)$ and $|S| = O(\delta n)$.

Notice that there is an extra $\sqrt{\log|S|}$ factor in the approximation ratio, compared to spectral partitioning.

In HW3, you are asked to remove this $\sqrt{\log|S|}$ factor in some situation.

If you can remove the $\sqrt{\log|S^*|}$ factor and $|S| = O(|S^*|)$, then most likely you have disproved the small-set expansion conjecture!

Local algorithms

One advantage of the random walk algorithm is that it can be implemented locally without exploring the whole graph.

The idea is that we can truncate the random walk vector in every step, by setting very small entries to zero.

By doing so, one can still prove that the resulting vector is a good approximation of the original vector, and the same analysis will go through.

By truncation, we can assume the vectors are of small support, and one can show that the ~~total~~ running time is $O(d \cdot |S| \cdot \text{polylog}(|S|) / \phi^2(S))$, which is sublinear if d and $|S|$ are small.

The details are straightforward but tedious and are omitted.

There are other local graph partitioning algorithms using pagerank vectors and evolving sets, and they seem to work well in pratical applications.

References

You are referred to the paper "subexponential algorithms for unique games and related problems" by Arora, Barak, and Steurer for the original proof and its application in solving unique games.

Our presentation follows the paper " Improved Cheeger's inequality and analysis of local graph partitioning by vertex expansion and expansion profile".