CS 798 - Algorithmic Spectral Graph Theory, Fall 2015, Waterloo

Lecture 8: Local graph partitioning

We show that random walks can be used to find a small non-expanding set, with performance guarantee similar to that of the spectral partitioning algorithm, while the running time could be sublinear to the graph size.

Small sparse cuts

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

Given a target size δn (where δ could be a small constant, or could depend on n say $\frac{1}{\sqrt{n}}$) and a vertex v, we would like to find a set S with $|S| \leq \delta n$ which contains v and $\phi(S)$ small.

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, but there is no control over the size of the output set.

We will show a random walk algorithm with similar performance guarantee as spectral partitioning, with some control over the size of the output set, and may run in sublinear time.

Intuition

Suppose we want to distinguish between the following two cases:

① There is a set S with $\phi(S) \leq \varepsilon_1$ and $|S| \leq \delta n$.

② Every set S with $|S| \leq 2\delta n$ has $\phi(S) \geq \varepsilon_2 >> \varepsilon_1$.

In the first case, if we start a random walk from a vertex in S, then we expect that the random walk will stay in S with high probability, while in the second case, we expect the random walk to mix quickly, at least for sets up to size $2\delta n$.

To distinguish the two cases, we compute $W^t x_i$ for an appropriately chosen t, where W is the laze walk matrix and $x_i \in \mathbb{R}^n$ is the vector with one in i-th position and zero otherwise.

lazy walk matrix and $x_i \in \mathbb{R}^n$ is the vector with one in i-th position and zero otherwise.

We look at the sum of the δn largest entries in $W^t x_i$, call this sum $C_{\delta n}$.

In the first case, if $i \in S$, then we expect that $C_{\delta n}$ is close to one, as most probabilities stayed in S.

In the second case, for every vertex i in the graph, we expect that $C_{\delta n}$ is at most $\frac{1}{2}$, because the probability would have spread evenly in at least $2\delta n$ vertices.

Spielman and Tang [ST04] designed the first local graph partitioning algorithm using random walks, and their proof is based on the work of Lovász and Simonovits, who developed a combinatorial approach to study mixing time using (small-set) expansion, which can make the above intuition rigorous.

Lovász and Simonovits method is interesting and can be used to give an alternative proof of Cheeger's inequality. We will discuss a bit more in class, but refer to the project page for references.

Spectral approach

We will present a more spectral approach for local graph partitioning, closer to what we have seen so far.

The idea is based on the work of Arora, Barak and Steurer [ABS10], which we will study next time.

We assume the graph is d-regular.

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 $S \subseteq supp(x) := \{ i \mid x_i \neq 0 \}$ with $\phi(S) \leq \sqrt{2R(x)}$ where $R(x) = \frac{x^T \mathcal{L} x}{x^T x} = \frac{\sum_{i \sim j} (x_i - x_j)^2}{d \sum_i x_i^2}$.

So, if we can find a vector x with $|supp(x)| \leq \delta n$ and $R(x)$ small, then we can use it to find a small sparse cut.

We call a vector x with $|supp(x)| \leq \delta n$ a combinatorially δ-sparse vector.

This combinatorial condition is not easy to work with directly.

One idea in [ABS10] is to relax this condition, so that it is easier to work with and has essentially the same effect.

By Cauchy-Schwarz, a combinatorially δ-sparse vector x satisfies $\|x\|_1 \leq \sqrt{\delta n}\, \|x\|_2$.

We call a vector x analytically δ-sparse if $\|x\|_1 \leq \sqrt{\delta n}\, \|x\|_2$.

It will turn out that if we find an analytically sparse vector with small Rayleigh quotient, then we can find a combinatorially sparse vector vector with small Rayleigh quotient.

And we will see that it is much easier to reason about analytical sparsity.

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 $W^t x_i$ for some appropriate t.

② "Truncate" $W^t x_i$ 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 $W^t x_i$ would have small Rayleigh quotient, for all $i \in V$.

The analysis in this step is very similar to the analysis of the power method.

For ②, we will prove that if there is a small sparse cut S, there exists some vertex $i \in S$ such that $W^t x_i$ is analytically sparse. Furthermore, an analytically δ-sparse vector can be truncated to a combinatorially $O(\delta)$-sparse vector with similar Rayleigh quotient.

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 should be clear by now.

Power method

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.

Lemma 1 $\quad R(W^t x_i) \le 2 - 2\|W^t x_i\|^{\frac{1}{t}}$, where $R(x) = \frac{x^T \mathcal{L} x}{\quad}$.

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

~~proof~~ Let the eigenvalues of W be $1 = \alpha_1 \ge \alpha_2 \ge \dots \ge \alpha_n \ge 0$.

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 $v_1, \dots, v_n$ are the eigenvectors of W with eigenvalues $\alpha_1, \alpha_2, \dots, \alpha_n$.

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 $w_i \ge 0 \ \forall i$, then $\left(\sum_{i=1}^n w_i y_i^p\right)^{\frac{1}{p}} \ge \left(\sum_{i=1}^n w_i y_i^q\right)^{\frac{1}{q}}$ for $p \ge q$.

Notice that $c_i^2 \ge 0 \ \forall i$ and $\sum_{i=1}^n c_i^2 = \langle \sum_{i=1}^n c_i v_i, \sum_{i=1}^n c_i v_i \rangle = \langle x_i, x_i \rangle = 1$.

$\uparrow$ orthonormal basis

So, we can apply the power means inequality by setting $w_i = c_i^2$ and $y_i = \alpha_i$ to get

$$\left(\sum_{i=1}^n c_i^2 \alpha_i^{2t+1}\right)^{\frac{1}{2t+1}} \ge \left(\sum_{i=1}^n c_i^2 \alpha_i^{2t}\right)^{\frac{1}{2t}}.$$

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

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

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

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

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

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

We will eventually choose $\lambda = \phi(G)$ and apply the second bound, and potentially in the homework you will apply the first bound.

Combinatorially sparse vectors from analytically sparse vectors

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

So, if $W^t x_i$ is δ-analytically sparse, then $1 = \|W^t x_i\|_1 \le \sqrt{\delta n}\,\|W^t x_i\|_2$ and thus $\|W^t x_i\|_2^2 \ge \frac{1}{\delta n}$.

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 combinatorially sparse vector with small Rayleigh quotient.

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

Then there exists a vector $y \in \mathbb{R}^n$ with $|supp(y)| = 4\delta n$ and $R(y) = 2R(x)$.

proof The proof is by a simple truncation argument.

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

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

Then, it is clear that $|supp(y)| \le 4\delta n$, 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 $(y_i - y_j)^2 \le (x_i - x_j)^2$, 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 $y_i^2 \ge x_i^2 - \frac{1}{2} x_i$, and so $\sum_i y_i^2 \ge \sum_i x_i^2 - \frac{1}{2} \sum_i x_i = \delta n - \frac{1}{2}\delta n = \delta n / 2 = \|x\|_2^2 / 2$

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

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 to finish the proof.

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

Analytically sparse vector from staying probability

The idea is that if S is a small sparse cut, then when we start a random walk from 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 $\|W^t x_i\|$ 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 $\sum_{v \in S} p_t(v) \ge 1 - t \cdot \phi(S)$.

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 all t steps.

Equivalently, we upper bound the probability that the random walk go outside of S in any of these 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 $p_{i+1}(v) = \frac{1}{2} p_i(v) + \frac{1}{2d} \sum_{\substack{u: uv \\ u \in S}} p_i(u) \leq \frac{1}{|S|}$. □

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

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

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

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

<u>Proof</u> The average escaping probability is $t \cdot \phi(S)$. So, there are at most half of the vertices with escaping probability at least $2t\,\phi(S)$, hence the corollary.

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

<u>Lemma 3</u> There exists $S' \subseteq S$ with $|S'| \geq |S|/2$ such that for $i \in S'$, then $\| W^t \chi_i \|_2^2 \geq \frac{1}{|S|} (1 - 2t\,\phi(S))^2$

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

Then $\| W^t \chi_i \|_2^2 \geq \sum_{j \in S} (W^t \chi_i)(j)^2 \geq \frac{1}{|S|} \left(\sum_{j \in S} (W^t \chi_i)(j) \right)^2$ by Cauchy Schwarz

$$\geq \frac{1}{|S|} (1 - 2t\,\phi(S))^2. \quad \square$$

Approximation algorithm

We are ready to complete the analysis.

Set $t = \frac{1}{4\phi(S)}$.

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

By Lemma 1, $R(W^t x_i) \leq 2(1 - \|W^t x_i\|_2^{\frac{2}{t}}) \leq 2(1 - \frac{1}{2\sqrt{|S|}}^{4\phi(S)}) = 2(1 - e^{-\ln(2\sqrt{|S|}) \cdot 4\phi(S)}) = O(\phi(S)\ln(|S|))$.

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

By Cheeger rounding in L03, we find a set S' with $\phi(S') = \sqrt{\phi(S)\ln(|S|)}$ and $|S'| = O(|S|)$.

We prove the following bicriteria approximation result.

Theorem 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 $\phi(S) = O(\sqrt{\phi \log|S^*|})$ and $|S| = O(\delta n)$.

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 $O(d \cdot |S| \cdot \mathrm{polylog}(|S|) / \phi^2(S))$, which is sublinear if d and $|S|$ are small.

The details are straightforward but tedious and are omitted (see [KLL16]).

There are other local graph partitioning algorithms using pagerank vectors and evolving sets, and they seem to work well in pratical applications. We will explain a bit in class if there is time.

Recently, it is shown that these local graph partitioning algorithms also perform better (like spectral partitioning) when λ_k or ϕ_k is large, or when the robust vertex expansion is large.

References

[ABS10] Subexponential algorithms for Unique Games and related problems, by Arora, Barak, Steurer, 2010.

[ST04] A local clustering algorithm for massive graphs and its application to near linear time graph partitioning, by Spielman and Teng, 2008.

[ACL06] Local graph partitioning using PageRank vectors, by Andersen, Chung, Lang, 2006.

[AP08] Finding sparse cuts locally using evolving sets, by Andersen and Peres, 2008.

[KLL16] Improved Cheeger's inequality and analysis of local graph partitioning using vertex expansion and expansion profile, by Kwok, Lau, Lee, 2016.