Notes 13: Local graph partitioning

1. Small sparse cut

Given an undirected graph G with positive edge weights, consider the problem of finding a small sparse cut: a vertex set S with small conductance $\varphi(S)$ and has small size:

$$\operatorname{argmin}\left\{\varphi(S) \mid S \subseteq V, |S| \leqslant \delta n\right\}$$

This is sometimes motivated by finding a small community in a social network.

The spectral partitioning algorithm of Cheeger–Alon–Milman can find a set of small conductance, but the set may be large (containing up to half of the vertices).

We will study an algorithm with the following guarantee: If a graph G has a subset S with small conductance, then the algorithm will find a a subset T with $|T| \leqslant 16|S|$ and $\varphi(T) \leqslant O\left(\sqrt{\varphi(S)\log|S|}\right)$.

Compared with Cheeger–Alon–Milman, we gain in the guarantee that T is small, but we pay an extra $\sqrt{\log|S|}$ factor in conductance.

2. Analytic sparsity

For simplicity we consider only d-regular graphs, and further assume d is normalized to be 1. The proof of Cheeger–Alon–Milman inequality shows that given any $x \in \mathbb{R}^V$, we can find a sparse cut $T \subseteq \operatorname{supp}(x) = \{i \in V \mid x_i \neq 0\}$ and $\varphi(T) \leqslant \sqrt{2R(x)}$, where $R(x) = x^\top \mathcal{L} x / x^\top x$.

If we can solve the problem of minimizing Rayleigh quotient over vector $x \in \mathbb{R}^V$ of small support,

$$\operatorname{argmin}\left\{R(x) \mid x \in \mathbb{R}^V, |\operatorname{supp}(x)| \leqslant \delta n\right\},$$

then sweep cut algorithm of Cheeger–Alon–Milman outputs a desired subset T from x. But the combinatorial sparsity condition $|\operatorname{supp}(x)| \leqslant \delta n$ is difficult to work with.

The idea is to relax the combinatorial sparsity condition to the analytic sparsity condition

$$\|x\|_1^2 \leqslant \delta n \|x\|_2^2 .$$

This condition is satisfied whenever $|\operatorname{supp}(x)| \leqslant \delta n$ (by Cauchy–Schwarz). Also, if x is the probability vector of a distribution μ, then $\|x\|_1^2 = 1$, and

$$\|x\|_2^2 = \sum_{i \in V} \mu(i)^2 = \mathop{\mathbb{P}}_{i\sim\mu,\, j\sim\mu}[i = j]$$

is the collision probability of μ (the probability for two independent samples from μ to coincide). In particular, if x is the probability vector of the uniform distribution over a subset $S \subseteq V$, then $\|x\|_2^2 = \sum_{i\in S} 1/|S|^2 = 1/|S|$. Therefore the ratio $\|x\|_1^2/\|x\|_2^2$ is a robust way to measure the size of the support of a distribution.

Turns out any analytically sparse vector with small Rayleigh quotient can be "rounded" into a combinatorially sparse vector with small Rayleigh quotient.

3. Algorithm outline

At a high level, the algorithm is as follows:

(1) For every vertex i, run lazy random walk from i for t steps for some t depending on $\varphi(S)$
(2) Truncate t-step lazy walk probability vector π_t into a vector with small support
(3) Apply Cheeger–Alon–Milman sweep cut to this vector and output a small sparse cut

Why do we expect this algorithm to work? If the random walk starts at a vertex $i \in S$, since $\varphi(S)$ is small, most of the probability mass of $\pi_t^\top = \mathbb{1}_i^\top W^t$ will stay inside S. Here W is the transition matrix of the lazy random walk, and $\mathbb{1}_i$ is the indicator vector for vertex i (the probability vector for the initial distribution of starting the random walk at i). After some time t, the lazy random walk should have become close to the "stationary distribution" in S. Therefore Cheeger–Alon–Milman thresholding should reveal S.

To analyze step (1), we will show that π_t has small Rayleigh quotient, provided the collision probability $\|\pi_t\|_2^2$ is not too small (due to having substantial mass in S).

To analyze step (2), we will show that if there is a small sparse cut S, then π_t will be analytically sparse for some starting vertex $i \in S$. Further, an analytically sparse vector can be truncated to a combinatorially sparse vector with similar Rayleigh quotient.

To analyze step (3), we apply a lemma in proving Cheeger–Alon–Milman inequality.

4. Collision probability and Rayleigh quotient

We keep track of how the collision probabilty $\|\pi_t\|_2^2$ changes over time.

- Initially, $\|\pi_0\|_2^2 = \|\mathbb{1}_i\|_2^2 = 1$.
- $\|\pi_{t+1}\|_2^2 = \|W\pi_t\|_2^2 \leqslant \|\pi_t\|_2^2$, as W has all eigenvalues bounded by 1 in magnitude. So collision probability $\|\pi_t\|_2^2$ can only decrease over time.
- $\|\pi_t\|_2^2 \to \|\mathbb{1}/n\|_2^2 = 1/n$ as t grows.

In fact, the ratio $\|\pi_{t+1}\|_2^2/\|\pi_t\|_2^2$ is nondecreasing in t, so $\|\pi_t\|_2^2$ converges to $\|\mathbb{1}/n\|_2^2$ more and more slowly over time. This is proved in the following claim.

Claim 4.1. $\dfrac{\|\pi_{t+1}\|_2^2}{\|\pi_t\|_2^2} \leqslant \dfrac{\|\pi_{t+2}\|_2^2}{\|\pi_{t+1}\|_2^2}$.

Proof. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of W and $v_1, \dots, v_n$ be its orthonormal eigenvectors. Using the eigen-expansion $\pi_t = \sum_{1\leqslant \ell \leqslant n} c_\ell \lambda_\ell^t v_\ell$ of π_t, we have $\|\pi_t\|_2^2 = \sum_{1\leqslant \ell\leqslant n} c_\ell^2 \lambda_\ell^{2t}$. The desired inequality is $\|\pi_{t+1}\|_2^4 \leqslant \|\pi_{t+2}\|_2^2 \|\pi_t\|_2^2$, and it becomes

$$\left(\sum_{1\leqslant \ell\leqslant n} c_\ell^2 \lambda_\ell^{2t+2}\right)^2 \leqslant \left(\sum_{1\leqslant \ell\leqslant n} c_\ell^2 \lambda_\ell^{2t+4}\right)\left(\sum_{1\leqslant \ell\leqslant n} c_\ell^2 \lambda_\ell^{2t}\right) ,$$

which follows by Cauchy–Schwarz. □

What happens when $\|\pi_t\|_2^2$ decreases slowly? $\|\pi_{t+1}\|_2^2/\|\pi_t\|_2^2$ will be close to 1, or equivalently $1 - (\|\pi_{t+1}\|_2^2/\|\pi_t\|_2^2)$ is close to 0. We can express

$$1 - \frac{\|\pi_{t+1}\|_2^2}{\|\pi_t\|_2^2} = 1 - \frac{\|W\pi_t\|_2^2}{\|\pi_t\|_2^2} = \frac{\pi_t^\top (I - W^\top W)\pi_t}{\pi_t^\top \pi_t} = \frac{\pi_t^\top \mathcal{L}' \pi_t}{\pi_t^\top \pi_t}$$

as the Rayleigh quotient $R_{\mathcal{L}'}(\pi_t)$ for the matrix $\mathcal{L}' = I - W^2$. Turns out $\mathcal{L}'$ is the normalized Laplacian of some graph H! This graph H is the two-step lazy random walk, where every step in H corresponds to two consecutive steps in W. More precisely, H also has vertex set V, and every edge (i,k) in H corresponds to a length-2 path $(i,j),(j,k)$ in the lazy random walk W. The weight w_{ik} of (i,k) in H is $w_{ij}w_{jk}$, the product of weights of the two edges in the path in W. H has normalized adjacency matrix W^2. We won't prove these claims about H since our proof does not depend on them, and will leave them as easy exercises.

This means when $\|\pi_t\|_2^2$ decreases slowly at time t, the probability vector π_t corresponds two small Rayleigh quotient (hence a sparse cut, by Cheeger–Alon–Milman) in the two-step lazy random walk graph H.

We can translate small Rayleigh quotient $R_{\mathcal{L}'}(\pi_t)$ (for the two-step walk) into small Rayleigh quotient $R(\pi_t)$ (for the original lazy walk) using the following claim:

Claim 4.2. *For any $x \in \mathbb{R}^V$ and lazy random walk transition W, $x^\top W^2 x \leqslant x^\top W x$. Therefore*

$$R_{\mathcal{L}'}(x) = \frac{x^\top (I - W^2)x}{x^\top x} \geqslant \frac{x^\top (I-W)x}{x^\top x} = R(x) .$$

Proof. $W = I^{-1}W$ coincides with the normalized adjacency matrix of G, since G is assumed to be 1-regular, so the degree matrix is I.

Since W is lazy, $W = \frac{1}{2}I + \frac{1}{2}W'$, where W' is the transition/normalized adjacency matrix of the non-lazy random walk on G. Then $W - W^2 = \frac{1}{4}I - \frac{1}{4}(W')^2 = \frac{1}{4}\mathcal{L}_{W'} \succcurlyeq 0$. □

The above claim is the only place we require the random walk to be lazy.

We get the following upperbound on Rayleigh quotient $R(\pi_{t-1})$ if we can lower bound the collision probability $\|\pi_t\|_2^2$.

Proposition 4.3. $R(\pi_{t-1}) \leqslant 1 - \|\pi_t\|_2^{2/t}$.

Proof. Since $\|\pi_0\|_2^2 = 1$,

$$\|\pi_t\|_2^2 = \frac{\|\pi_t\|_2^2}{\|\pi_0\|_2^2} = \frac{\|\pi_t\|_2^2}{\|\pi_{t-1}\|_2^2} \frac{\|\pi_{t-1}\|_2^2}{\|\pi_{t-2}\|_2^2} \cdots \frac{\|\pi_1\|_2^2}{\|\pi_0\|_2^2} \leqslant \left(\frac{\|\pi_t\|_2^2}{\|\pi_{t-1}\|_2^2}\right)^t ,$$

where the inequality is Claim 4.1. This inequality and Claim 4.2 implies

$$R(\pi_{t-1}) \leqslant R_{\mathcal{L}'}(\pi_{t-1}) = 1 - \frac{\pi_{t-1}^\top W^\top W \pi_{t-1}}{\pi_{t-1}^\top \pi_{t-1}} = 1 - \frac{\|\pi_t\|_2^2}{\|\pi_{t-1}\|_2^2} \leqslant 1 - \|\pi_t\|_2^{2/t} .$$ □

5. Truncating analytically sparse vector

Lemma 5.1. *Suppose $x \in \mathbb{R}_{\geqslant 0}^V$ satisfies $\|x\|_1^2 \leqslant s\|x\|_2^2$. Then it can be truncated into a vector $y \in \mathbb{R}_{\geqslant 0}^V$ with $|\mathrm{supp}(y)| \leqslant 4s$ and $R(y) \leqslant 2R(x)$.*

Proof. By scaling, assume $\|x\|_2^2 = s$ and $\|x\|_1 \leqslant s$.

Let $y \in \mathbb{R}_{\geqslant 0}^V$ be the vector $y_i = \max\{x_i - 1/4, 0\}$.

Then $s \geqslant \|x\|_1 \geqslant \sum_{i \in \mathrm{supp}(y)} x_i \geqslant |\mathrm{supp}(y)| \frac{1}{4}$, because every $i \in \mathrm{supp}(y)$ contributes $x_i \geqslant 1/4$ to $\|x\|_1$. Hence $|\mathrm{supp}(y)| \leqslant 4s$.

We will compare $R(y)$ and $R(x)$, where $R(x) = \dfrac{x^\top \mathcal{L} x}{x^\top x} = \dfrac{\sum_{(i,j)\in E}(x_i - x_j)^2}{d\sum_{i\in V} x_i^2}$.

For the numerator, $(y_i - y_j)^2 \leqslant (x_i - x_j)^2$ because truncation can only reduce the difference. Hence $y^\top \mathcal{L} y^\top \leqslant x^\top \mathcal{L} x$.

For the denominator, we have $y_i^2 \geqslant x_i^2 - \frac{1}{2} x_i$, so

$$\sum_{i\in V} y_i^2 \geqslant \sum_{i\in V} x_i^2 - \frac{1}{2}\sum_{i\in V} x_i \geqslant s - \frac{1}{2}s = \frac{s}{2} = \frac{1}{2}\sum_{i\in V} x_i^2 .$$

Hence $y^\top y \geqslant x^\top x/2$.

Therefore $R(y) = y^\top \mathcal{L} y / y^\top y \leqslant x^\top \mathcal{L} x/(x^\top x/2) = 2R(x)$. □

6. Analytically sparse vector from small sparse cut

Given a probability π over V, we write $\pi(S) = \sum_{i\in S} \pi(i)$ to denote its total probability in $S \subseteq V$.

Claim 6.1. *If initial distribution $\mu_0 = \mathbb{1}_S/|S|$ is uniform over subset S, and $\mu_t = W^t \mu_0$, then $\mu_t(S) \geqslant 1 - t\varphi(S)$.*

Proof. We lowerbound $\mu_t(S)$ by the probability the random walk stays inside S for all t steps. We will upperbound the probability it leaves S in any of the t steps.

Every vertex i in the initial distribution μ_0 carries $\mu_0(i) = 1/|S|$ probability. Since the graph is d-regular, an edge going out of S carries $\frac{w_{ij}}{d|S|}$ probability out of S. Total probability escaping out of S in the first step is $\displaystyle\sum_{i\in S, j\in \overline{S}} \frac{w_{ij}}{d|S|} = \varphi(S)$.

We can finish the proof if the escape probability for every step is at most $\varphi(S)$. This is true by repeating the above calculations (changing "=" to "⩽"), and observing every vertex i at any time t carries probability $\mu_t(i)$ at most $1/|S|$.

Why is $\mu_i(t) \leqslant 1/|S|$ for any i and any t? This is true for initially $t = 0$ for all vertices i. For future time steps, $\mu_i(t+1)$ is a weighted average of $\mu_j(t)$ over neighbors j of i, so it remains true for time $t+1$. □

Corollary 6.2. *There is a starting point $i \in S$ such that if $\pi_0^{(i)} = \mathbb{1}_i$ and $\pi_t^{(i)} = W^t \pi_0^{(i)}$, then $\pi_t^{(i)}(S) \geqslant 1 - t\varphi(S)$.*

Proof. The uniform distribution μ_0 over S is the average, over a uniformly random $i \in S$, of initial distributions $\mathbb{1}_i$ starting from a single vertex i in S, because $\mu_0 = \frac{\mathbb{1}_S}{|S|} = \mathbb{E}_{i\sim\mu_0}[\mathbb{1}_i]$.

Now $\mu_t(S)$ is the same averaging of $\pi_t^{(i)}(S)$, because

$$\mu_t(S) = (W^t \mu_0)(S) = \left(W^t \mathop{\mathbb{E}}_{i\sim\mu_0}[\mathbb{1}_i]\right)(S) = \mathop{\mathbb{E}}_{i\sim\mu_0}[W^t \mathbb{1}_i](S) = \mathop{\mathbb{E}}_{i\sim\mu_0}[\pi_t^{(i)}(S)] .$$

The key observation here is that the t-step lazy random walk W^t is a linear operator, so taking average first and then t-step walk is the same as taking t-step walk first

Some vertex i in S must achieve staying probability $\pi_t^{(i)}(S)$ at least the average $\mu_t(S)$. □

Lemma 6.3. *For any distribution π, its collision probability $\|\pi\|_2^2 \geqslant \pi(S)^2/|S|$.*

Proof. Expand $\|\pi\|_2^2$ and apply Cauchy–Schwarz,

$$\|\pi\|_2^2 \geqslant \sum_{j\in S} \pi(j)^2 \geqslant \frac{1}{|S|}\left(\sum_{j\in S}\pi(j)\right)^2 = \frac{1}{|S|}\pi(S)^2 .$$

This Cauchy–Schwarz inequality implies that the distribution over S with the smallest collision probability is the uniform distribution, and has collision probability $1/|S|$. □

7. Algorithm

We know the graph contains a small subset S with conductance $\varphi(S)$. Corollary 6.2 implies that if we are lucky to choose $i \in S$ as the starting point of our random walk, then even after $t+1 = 1/2\varphi(S)$ steps, there is still $\pi_{t+1}(S) \geqslant 1/2$ probability mass of staying in S.

Lemma 6.3 then implies the collision probability $\|\pi_{t+1}\|_2^2 \geqslant 1/4|S|$.

Proposition 4.3 gives the following upperbound on Rayleigh quotient:

$$R(\pi_t) \leqslant 1 - \|\pi_{t+1}\|_2^{2/(t+1)} \leqslant 1 - \frac{1}{(4|S|)^{2\varphi(S)}} = 1 - \exp(-2\varphi(S)\ln(4|S|)) = O(\varphi(S)\ln|S|) ,$$

where the last equality is due to $1 - e^{-x} = O(x)$ for small x near 0.

π_t is analytically sparse and has sparsity ratio $\|\pi_t\|_1^2/\|\pi_t\|_2^2 = 1/\|\pi_t\|_2^2 \leqslant 1/\|\pi_{t+1}\|_2^2 \leqslant 4|S|$.

Lemma 5.1 truncates π_t to some nonnegative vector y with $|\mathrm{supp}(y)| \leqslant 16|S|$ and $R(y) = O(\varphi(S)\ln|S|)$.

Cheeger–Alon–Milman outputs a super-level set $T = \{i \in V \mid y_i > r\}$ of y with $|T| \leqslant 16|S|$ and $\varphi(T) \leqslant \sqrt{2R(y)} = O\left(\sqrt{\varphi(S)\ln|S|}\right)$.

8. Small-set expansion

The above conductance guarantee has an extra $\sqrt{\log|S|}$ factor. Is there an efficient approximation algorithm whose approximation factor is independent of the size of S?

Such an algorithm, if exists, will solve the Small-Set-Expansion problem, defined as follows:

Small-Set-Expansion

Parameters:	conductance bound ε and size bound δ
Input:	regular undirected graph G
Goal:	decide between the following two cases:
(Yes)	Some $S \subseteq V$ with $|S| \leqslant \delta n$ satisfies $\varphi(S) \leqslant \varepsilon$
(No)	All $S \subseteq V$ with $|S| \leqslant 16\delta n$ satisfies $\varphi(S) \geqslant 1-\varepsilon$

You may think of the problem as asking if a graph has a hidden small "community" (subset with small conductance). And it only asks for deciding between two extreme cases of conductance: either some small subset has conductance very close to 0, or all small subsets have conductance very close to 1.

A conjecture known as Small-Set-Expansion Hypothesis says that Small-Set-Expansion is hard to solve.

Conjecture 8.1 (Raghavendra and Steurer 2010)**.** *For every $\varepsilon > 0$, there is $\delta > 0$ such that Small-Set-Expansion with parameters ε and δ is NP-hard.*

In particular, if Small-Set-Expansion Hypothesis holds and $P \neq NP$, then no efficient algorithm can avoid the dependence on $|S|$.

Small-Set-Expansion Hypothesis also implies the Unique-Games Conjecture, a central open problem in approxmiation algorithms that we will not define here. The latter conjecture says that certain constraint satisfaction problem called Unique-Games is NP-hard to approxmiate.

If Unique-Games Conjecture holds and $P \neq NP$, then a simple SDP algorithm will be the best approximation algorithm for many problems. A consequence is that Goemans–Williamson rounding algorithm for MaxCut (with approximation factor $0.878\ldots$) will be optimal.