CS 229r Spectral Graph Theory in Computer Science, Lecture 9 Oct 1, 2020

Lecture 9

Instructor: Salil Vadhan *Scribes: Ben Edelman, Alex Lin*

A question was asked last time: Is the upper bound side of Cheeger's inequality, $\phi(G) \le \sqrt{2\nu_2}$, always tight for Cayley graphs corresponding to abelian groups, as it is for the cycle graph? The answer is no. For the Boolean hypercube graph, the lower bound $\nu_2/2 \le \phi(G)$ is the tight side.

1 Recap and proof of Cheeger's inequality

Definition 1 (Conductance)**.** *The conductance of a set S of vertices is the ratio of the weight of the boundary of S to the (degree-weighted) weight of S.*

$$\phi(S) := \frac{w(\delta S)}{d(S)} = \frac{\sum_{a\in S, b\notin S} w_{a,b}}{\sum_{a\in S} d(a)}$$

The conductance of a graph G is the minimum conductance over all subsets of vertices with at most half the total weight.

$$\phi(G) := \min_{S: d(S)\le d(V)/2} \phi(S)$$

Theorem 2 (Cheeger's inequality)**.** *Let G be an undirected weighted graph. Then*

$$\frac{\nu_2}{2} \le \phi(G) \le \sqrt{2\nu_2}$$

"Cheeger's inequality" typically only refers to the upper bound $\phi(G) \le \sqrt{2\nu_2}$.

In Problem Set 1, we proved that in any connected unweighted graph, λ_2 can't be negligibly small. Cheeger's inequality gives us another way to see why this is true, though the resulting bound isn't optimal. Suppose for simplicity that G is regular. Then $\phi(G) = \Omega(1/(nd))$ because every set $S \subseteq V$ has at least one edge on its boundary. Applying Cheeger's, we see that $\nu_2 = \Omega(1/(n^2d^2))$.

1.1 Proof of Cheeger's inequality

We are given a vector $y \perp \vec{d}$ such that its generalized Rayleigh quotient is bounded above by some constant ρ:

$$\frac{y^T L y}{y^T D y} \le \rho$$

For example, we could have $y = D^{1/2}\psi_2$ and $\rho = \nu_2 \le 2\phi(G)$.

Our goal is to "round" y to obtain a vertex set S (i.e., an indicator vector $\vec{1}_S$) with conductance $\le \sqrt{2\rho}$:

$$\frac{w(\delta S)}{\min\{d(S), d(V-S)\}} \le \sqrt{2\rho} \overset{\text{if } \rho=\nu_2}{\le} 2\sqrt{\phi(G)}$$

The first step is to sort the coordinates of y so that $y(1) \le \cdots \le y(n)$, and then to center y, letting

$$z := y - s\vec{1}$$

for a choice of s such that

$$\sum_{a: z(a)<0} d(a) \le \frac{d(V)}{2} \quad \text{and} \sum_{a: z(a)>0} d(a) \le \frac{d(V)}{2}$$

Note that the generalized Rayleigh quotient of z is also bounded above by ρ.

$$\frac{z^T L z}{z^T D z} \leq \frac{y^T L y}{y^T D y} \leq \rho$$

We assume without loss of generality that $z(1)^2 + z(n)^2 = 1$. This just corresponds to a scaling of y at the beginning of the argument.

We then choose some threshold τ and take $S_\tau = \{a : z(a) \leq \tau\}$. In fact, following Trevisan's proof, we will carefully define a distribution on thresholds τ such that

$$\mathop{\mathbb{E}}_t[w(\delta S_\tau)] \leq \sqrt{2\rho} \cdot \mathop{\mathbb{E}}_\tau[\min\{d(S_\tau), d(V - S_\tau)\}]$$

A common approximation algorithm technique is to:

1. Write the optimization objective as an NP-hard integer program.
2. Solve a relaxation of the program.
3. Round the solution to an integer solution using a randomized procedure.

Though the original proof of Cheeger's inequality did not use a randomized threshold, we follow Trevisan in doing so because it emphasizes the fact that the proof gives a relaxation-based approximation algorithm for finding a vertex set with low conductance.

The distribution on τ is given by the following probability density function:

$$f(t) = \begin{cases} 2|t| \text{ if } t \in [z(1), z(n)] \\ 0 \text{ otherwise} \end{cases}$$

The following two claims show why this arbitrary-seeming distribution is useful:

Claim 3. $\mathbb{E}_\tau[\min\{d(S_\tau), d(V - S_\tau)\}] = z^T D z$

Proof.

$$\begin{aligned} \mathop{\mathbb{E}}_\tau[\min\{d(S_\tau), d(V - S_\tau)\}] &= \sum_a d(a) \cdot \Pr[a \text{ is in ``smaller'' of } S_\tau \text{ and } V - S_\tau] \\ &= \sum_a d(a) \cdot \Pr[\tau \text{ is between } z(a) \text{ and } 0] \qquad \text{(centering)} \\ &= \sum_a d(a) \cdot z(a)^2 \\ &= z^T D z \end{aligned}$$

□

Claim 4. $\mathbb{E}_\tau[w(\delta S_\tau)] \leq \sqrt{2}\sqrt{z^t L z} \cdot \sqrt{z^T D z}$

Proof.

$$\begin{aligned} \mathop{\mathbb{E}}_\tau[w(\delta S_\tau)] &= \sum_{(a,b) \in E} w_{a,b} \cdot \Pr[\tau \text{ is between } z(a) \text{ and } z(b)] && (1) \\ &\leq \sum_{(a,b) \in E} w_{a,b} \cdot |z(b) - z(a)| \cdot (|z(b)| + |z(a)|) && (2) \\ \text{(Cauchy-Schwarz)} &\leq \sqrt{\sum_{(a,b) \in E} w_{a,b} \cdot (z(b) - z(a))^2} \cdot \sqrt{\sum_{(a,b) \in E} w_{a,b} \cdot (|z(b)| + |z(a)|)^2} && (3) \\ &\leq \sqrt{z^T L z} \cdot \sqrt{2 z^T D z} && (4) \end{aligned}$$

Step (2) follows because if $z(a)$ and $z(b)$ have the same sign, the probability τ lies between them is exactly $|z(b)-z(a)|\cdot(|z(b)|+|z(a)|)$ by definition of the distribution, and if they have opposite signs the probability will only be smaller. □

In the preceding proof we used the ubiquitous Cauchy-Schwarz inequality, which is worth becoming familiar with.

Lemma 5 (Cauchy-Schwarz Inequality). *For every two vectors $u, v \in \mathbb{R}^n$,*

$$|\langle u, v\rangle| \leq \|u\|\|v\|$$

i.e. for all $u_1, \ldots, u_n,\ v_1, \ldots, v_n \in \mathbb{R}$

$$\left(\sum_{i=1}^{n} u_i v_i\right)^2 \leq \left(\sum_{i=1}^{n} u_i^2\right) \cdot \left(\sum_{i=1}^{n} v_i^2\right)$$

2 Higher-order Cheeger's inequality

In this section, we assume that $G = (V, E)$ is an undirected graph. We begin our analysis by recapping a simple fact.

Fact 6 (Multiplicity of eigenvalue 0). *Let ν_k be the k-th smallest eigenvalue of the normalized Laplacian N for graph G. Then, $\nu_k = 0$ if and only if G has at least k connected components.*

There are several ways to prove this fact. We give two possible solutions here.

Proof. If G has exactly k connected components, then we can write its normalized Laplacian N as a partition k disjoint blocks along the diagonal. Specifically,

$$N = \begin{bmatrix} N_1 & 0 & 0 & \ldots & 0 \\ 0 & N_2 & 0 & \ldots & 0 \\ 0 & 0 & N_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & N_k \end{bmatrix},$$

where $N_1, \ldots, N_k$ are the normalized Laplacians for the the k connected subgraphs of G. Therefore, the eigenvalues of N are the union (with multiplicity) of the eigenvalues of $N_1, \ldots, N_k$. We know that each N_i has zero as an eigenvalue with the all-ones vector as the corresponding eigenvector. Thus, N has the zero eigenvalue with multiplicity k. □

Proof. Observe that the i-th eigenvalue ν_i of the normalized Laplacian N is zero if and only if the i-th eigenvalue λ_i of the (unnormalized Laplacian) L is also zero. To see this, observe the following relationship between the Rayleigh quotient of N and the generalized Rayleigh quotient of L:

For any vector x, let $y := D^{-1/2}x$. Then, we have

$$\frac{x^T N x}{x^T x} = \frac{x^T D^{-1/2} L D^{-1/2} x}{x^T D^{-1/2} D D^{-1/2} x} = \frac{y^T L y}{y^T D y}.$$

From this relationship, we can see that $x^T N x = 0$ if and only if $y^T L y = 0$, which implies that

$$\nu_1 = \nu_2 = \ldots = \nu_k = 0 \iff \lambda_1 = \lambda_2 = \ldots = \lambda_k = 0.$$

Next, we know that $\lambda_k = 0$ if and only if the dimension of the kernel of L is at least k, since the k orthogonal eigenvectors corresponding to $\lambda_1, \ldots, \lambda_k$ span a k-dimensional subspace within the kernel of L. Indeed, we know that the kernel of L specifically contains $1_{S_1}, 1_{S_2}, \ldots, 1_{S_k}$, where each 1_{S_i} is a vector that has 1's for a particular set of vertices S_i and 0's everywhere else. This final fact is true if and only if G has at least k connected components. □

Higher-order Cheeger's inequality can be thought of as a "softer" version of Fact 3. Intuitively, it shows that $\nu_k \approx 0$ if and only if G is "close" to being partitioned into k sub-components with few edges going between them. Put another way, $\nu_k \approx 0$ if and only if there exist disjoint subsets $S_1, S_2, \ldots, S_k \subset V$ with small conductance $\phi(S_1), \phi(S_2), \ldots, \phi(S_k)$. We can think of $S_1, \ldots, S_k$ as *clustering* vertices of the graph into k separate components.

We formalize the notion of the higher-order conductance of a graph G with the following definition:

Definition 7 (Higher-order conductance of a graph)**.**

$$\phi_k(G) := \min_{\substack{S_1,\ldots,S_k \subset V \\ \forall i,j, S_i \cap S_j = \emptyset}} \max\{\phi(S_1), \ldots, \phi(S_k)\}$$

Theorem 8 (Higher-order Cheeger's inequality)**.** *Let G be an undirected weighted graph. Then, we can found $\phi_k(G)$ as follows:*

$$\frac{\nu_k}{2} \leq \phi_k(G) \leq \mathit{poly}(k) \cdot \sqrt{\nu_k}$$

2.1 Lower bound

We start by proving the lower bound, which is the easier side:

$$\frac{\nu_k}{2} \leq \phi_k(G).$$

Proof. Let $\rho := \phi_k(G)$. This implies that there exists k disjoint sets $S_1, \ldots, S_k$ such that, for each $i = 1, \ldots, k$

$$\phi(S_i) = \frac{1_{S_i}^T L 1_{S_i}}{1_{S_i}^T D 1_{S_i}} \leq \rho,$$

where 1_{S_i} is a vector that has 1's for vertices $a \in S_i$ and 0's otherwise.

Now, let L be the Laplacian of G and N be the normalized Laplacian. Define $S := \mathrm{span}(1_{S_1}, \ldots, 1_{S_k})$, which is a k-dimensional subspace since the component vectors are mutually orthogonal. Thus,

$$\nu_k = \min_{\substack{T \subseteq \mathbb{R}^n \\ \dim(T)=k}} \max_{x \in T} \frac{x^T N x}{x^T x} = \min_{\substack{T \subseteq \mathbb{R}^n \\ \dim(T)=k}} \max_{y \in T} \frac{y^T L y}{y^T D y} \leq \max_{y \in S} \frac{y^T L y}{y^T D y}.$$

The final quantity of the right-hand side can be upper bounded by 2ρ, which yields our desired result. To see why this last part of true, consider the following intuition: We know from above that for the basis vectors of S (i.e. $1_{S_1}, \ldots, 1_{S_k}$), the generalized Rayleigh quotient is at most ρ. Any vector $y \in S$ can be expressed as a linear combination of these k basis vectors, so we know that a very loose bound on the maximum Rayleigh quotient is $k\rho$, by the triangle inequality. However, if we re-express the numerator of the Rayleigh quotient as the quadratic form of the Laplacian, we observe that at most two basis vectors can factor into any term, which gives us the tighter 2ρ bound. □

2.2 Upper bound

The full proof is rather involved, but we will now describe the proof idea for the upper bound:

$$\phi_k(G) \leq \mathrm{poly}(k) \cdot \sqrt{\nu_k}$$

For simplicity we assume G is d-regular and unweighted.

Let $F_1, \ldots, F_k : V \to \mathbb{R}$ be k orthonormal vectors such that

$$F_i^T N F_i = \frac{F_i^T L F_i}{d} \leq \rho$$

Let $F : V \to \mathbb{R}^k$ be

$$F(a) = \begin{pmatrix} F_1(a) \\ \vdots \\ F_k(a) \end{pmatrix}$$

as in Hall's graph drawing method. We know

$$\begin{aligned} \frac{1}{d} \sum_{(a,b)\in E} \|F(a) - F(b)\|^2 &= \frac{1}{d} \sum_{(a,b)\in E} \sum_{i=1}^{k} (F_i(a) - F_i(b))^2 \\ &= \frac{1}{d} \sum_{i=1}^{k} \sum_{(a,b)\in E} (F_i(a) - F_i(b))^2 \\ &= \sum_{i=1}^{k} \frac{F_i^T L F_i}{d} \\ &\le k\rho \end{aligned}$$

Also,

$$\sum_{a\in V} \|F(a)\|^2 = k \tag{5}$$

Our goal, by analogy to the non-generalized setting, is to "round" the F_i's and obtain disjoint vertex sets $S_1, \ldots, S_k$, all with small conductance. We can't just round each F_i separately because of the disjointness condition.

It may be difficult to directly partition V into $S_1, \ldots, S_k$, so we take a two-step approach. First, we assign the vertices of the graph to m disjoint sets $T_1, \ldots, T_m$, where $m >> k$. Then, we reduce the number of sets to k. Note that the union of $T_1, \ldots, T_m$ may not contain all points $F(a)$ for $a \in V$. Nonetheless, we would like the $T_1, \ldots, T_m$ to satisfy the following three properties:

(A) $\sum_{a\in\bigcup T_i} ||F(a)||^2 \ge k \cdot (1 - O(1/k))$

(B) $\forall i, \sum_{a\in T_i} ||F(a)||^2 \le 1 + O(1/k)$

(C) $\forall i \ne j, a \in T_i, b \in T_j, \left\| \frac{F(a)}{||F(a)||} - \frac{F(b)}{||F(b)||} \right\| \ge \frac{1}{\text{poly}(k)}$.

Intuitively, the Property (A) (in conjunction with Equation (5)) shows that the $T_1, \ldots, T_m$ cover the vast majority of vertex points $F(a)$ for $a \in V$. Property (B) shows that each set T_i on its own does not have too much mass. And finally, Property (C) indicates that the sets should be far from one another, which is the first step towards partitioning the graph. Specifically, the distance between any two (normalized) points is at least $1/\text{poly}(k)$. This normalized distance can also be equated to $\sqrt{2 - 2\cos\theta}$, where θ is the angle between $F(a)$ and $F(b)$.

How do we obtain $T_1, \ldots, T_m$ to satsify these three properties? Trevisan provides a strategy, which we briefly sketch below. A pictorial representation by Prof. Vadhan is given in Figure 1.

1. Place a grid over $\mathbb{R}^k$, where the side length of each square is $O(1/k)$ (and randomly shift it to ensure that each point gets covered with some average probability).
2. Within each grid square, remove a small portion of each side with length $O(1/k^2)$. This leaves a smaller square inside the original square.
3. Within each smaller square, look at the sectors that intersect with the unit circle and form cones from the origin to that sector.

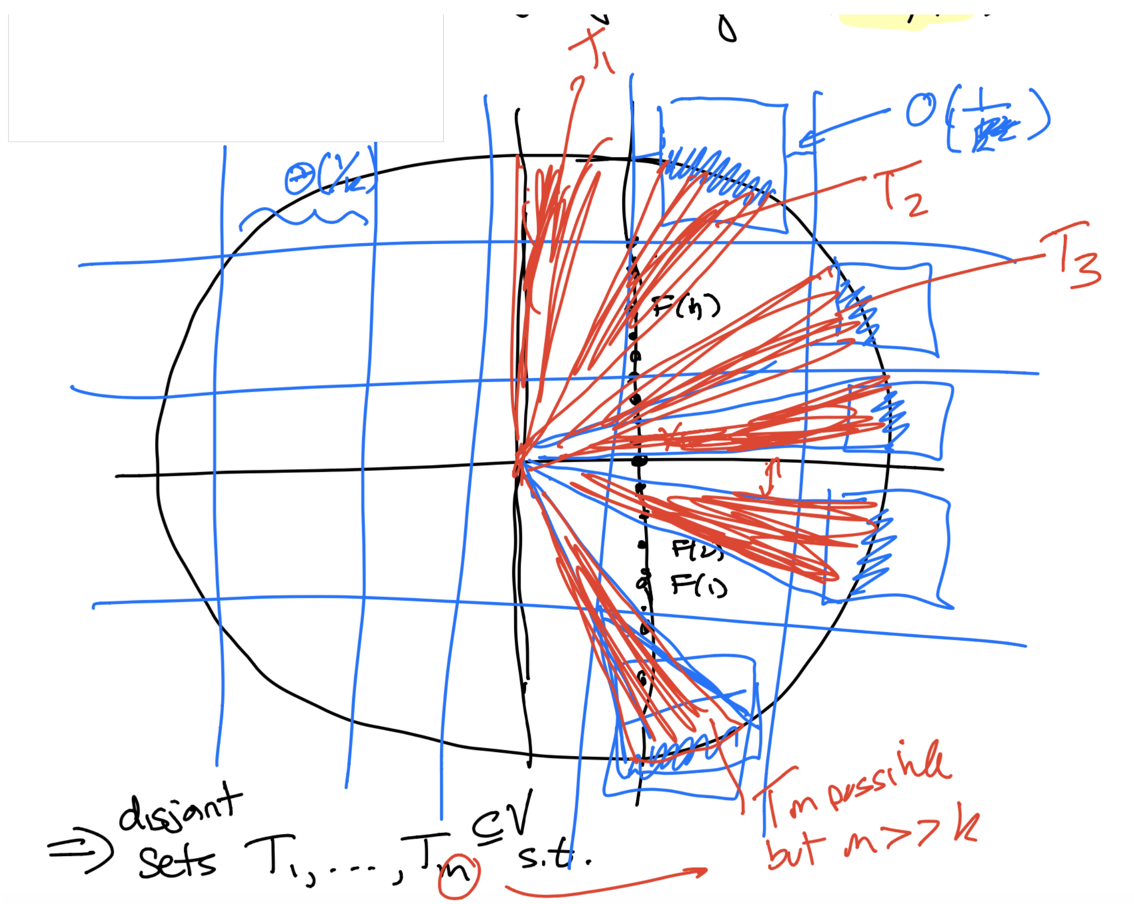

Figure 1: A pictorial representation by Prof. Vadhan of how to construct $T_1, \ldots, T_m$ in two dimensions (i.e. for $k = 2$).

4. Each cone outlines an area that corresponds to one of the sets T_i of points within $\mathbb{R}^k$.

We can show that using the aforementioned strategy, most points have a good chance of being in one of the $T_1, \ldots, T_m$, which satisfies Property (A). Property (B) is satisfied because each cone is not so large. Finally, Property (C) is satisfied because of the intentionally constructed space between the cones.

Now that we have $T_1, \ldots, T_m$, we start merging them to reduce the number of disjoint sets. In particular, any sets that are "too small" in mass get merged to yield sets $A_1, \ldots, A_k$ that satisfy the following two properties:

(D) $\forall i, \sum_{a \in A_i} ||F(a)||^2 \geq \frac{1}{2}$

(E) $\forall i \neq j, a \in A_i, b \in A_j, \left\| \frac{F(a)}{||F(a)||} - \frac{F(b)}{||F(b)||} \right\| \geq \frac{1}{\text{poly}(k)}$.

Property (D) asserts that each A_i has sufficient mass. Using Property (B) from before, we can ensure that Property (D) is satisfied for at least k disjoint sets after merging. Property (E) follows due to Property (C) from before.

Next, for each set A_i, we identify a non-negative vector y_i such that its generalized Rayleigh quotient is not too large, i.e.

$$\frac{y_i^T N y_i}{y_i^T y_i} \leq \text{poly}(k) \cdot \frac{F^T N F}{F^T F} = \text{poly}(k) \cdot \rho.$$

We can do this by defining y_i in the following manner:

$$y_i(a) = \begin{cases} ||F(a)||, & \text{if } a \in A_i \\ 0, & \text{if } F(a) \text{ is far away from the cone(s) corresponding to } A_i \\ \text{interpolate between } ||F(a)|| \text{ and } 0 & \text{otherwise.} \end{cases}$$

Finally, we can apply Cheeger-like rounding to obtain disjoint set $S_1, \ldots, S_k$ with small conductance, i.e.

$$\phi(S_i) \leq \sqrt{\frac{2y_i^T N y_i}{y_i^T y_i}} \leq \text{poly}(k) \cdot \sqrt{\rho}.$$