University of Michigan–Ann Arbor

Department of Electrical Engineering and Computer Science

EECS 498 004 **Advanced Graph Algorithms**, Fall 2021

Lecture 7: Cheeger's inequality

September 21, 2021

Instructor: Thatchaphol Saranurak

Scribe: Jingyi Gao

1. Recap

So far, we've seen several characterization of expanders. Intuitively, expanders are graphs that are robust against any adversarial deletions. From the perspective of cut, when we compare all graphs with the same degree profile, the size of every cut of expander is almost maximum among them. On the other hand, regarding flows, for all graphs with a degree profile no greater than the expander, they are embeddable into expander with small congestion.
All of the three characterization of expander are combinatorial. Today, we are going to introduce an algebraic way to characterize expanders based on **eigenvalue**s. This invokes an area called **spectral graph theory.** Roughly speaking, in this area, we associate a graph G with some matrix M and relate eigenvalues of M to combinatorial properties of G.
A very nice resource is the textbook by Spielman [1].

2. Review of Linear Algebra

We start by reviewing some linear algebra facts.

2.1. Theorem. *Let $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ be a symmetric real matrix. Then there exists $\lambda_1, ..., \lambda_n \in \mathbb{R}$ and orthonormal vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in \mathbb{R}^n$ such that*

- *For all i, $\boldsymbol{M}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$.*
- *In fact, $\boldsymbol{M} = \sum_i \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\top$*

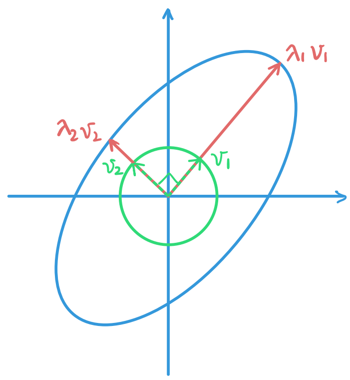
We call $\lambda_1, ..., \lambda_n$ the eigenvalues *of $\boldsymbol{M}$ and each $\boldsymbol{v}_i$ is called the* eigenvector *corresponding to λ_i.*

2.2. *Remark.*
- A useful way to think of the relation between $\boldsymbol{M}$ and its eigenvectors and eigenvalues is that any real symmetric matrix $\boldsymbol{M}$ corresponds to an ellipsoid. Consider the image of $\boldsymbol{M}C$, where $C = \{x \in \mathbb{R}^n : |x| = 1\}$. Let $x \in C$ and we can write x in terms of basis

[1] http://cs-www.cs.yale.edu/homes/spielman/sagt/sagt.pdf

$\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ (an orthonormal basis consists of all eigenvectors of $\boldsymbol{M}$,) $x = x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n$, where $\sum_{i=1}^n x_i^2 = 1$. Then $y = \boldsymbol{M}x = \sum_{i=1}^n x_i\lambda_i\boldsymbol{v}_i$. This is to say that the component of the image point $y = \boldsymbol{M}x$ has components $y_i = x_i\lambda_i$ in terms of the basis $\boldsymbol{v}_i$. Hence the image of $\boldsymbol{M}x$ is on the ellipsoid $\left\{\sum_{i=1}^n y_i\boldsymbol{v}_i : \sum_{i=1}^n \left(\frac{y_i}{\lambda_i}\right)^2 = 1\right\} = \boldsymbol{M}C$ with axes along the vectors $\boldsymbol{v}_i, i = 1, \cdots, n$ with length of the axes being $\lambda_i, i = 1, \cdots, n$, respectively.

- In short, any $\boldsymbol{M} \in \mathbb{SR}^{n\times n}$ corresponds to an ellipsoid, which is the image of unit circle under mapping $\boldsymbol{M}$. Each eigenvector $\boldsymbol{v}_i$ is the $i-$th axis of the ellipsoid and each eigenvalue λ_i describes how much we stretch the unit circle along the $i-$th axis. A picture for $n = 2$ is as below.



2.3. Theorem (Variational Characterization of Eigenvalues)**.** *Let $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ be a symmetric real matrix with eigenvalues $\lambda_1 \le \ldots \le \lambda_n$ and eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. Then*

- $\lambda_1 = \min_{\boldsymbol{x}\neq \boldsymbol{0}} \frac{\boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}}$
- $\lambda_n = \max_{\boldsymbol{x}\neq \boldsymbol{0}} \frac{\boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}}$
- $\lambda_2 = \min_{\boldsymbol{x}\perp \boldsymbol{v}_1, \boldsymbol{x}\neq \boldsymbol{0}} \frac{\boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}} = \min_{2\text{-}dim\ \mathcal{V}} \max_{\boldsymbol{x}\in\mathcal{V}\setminus\{\boldsymbol{0}\}} \frac{\boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}}$
- *In general, we have*

$$\lambda_k = \min_{k\text{-}dim\ \mathcal{V}} \max_{\boldsymbol{x}\in\mathcal{V}-\{\boldsymbol{0}\}} \frac{\boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}} = \min_{\boldsymbol{x}\perp \mathrm{span}\{\boldsymbol{v}_1,\ldots\boldsymbol{v}_{k-1}\}, \boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}}$$

We call $R_M(\boldsymbol{x}) = \frac{\boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}}$ the Rayleigh quotient of x w.r.t. $\boldsymbol{M}$.

3. The Laplacian of Graphs

Given an undirected graph $G = (V, E, w)$ with non-negative weights $w : E \longrightarrow \mathbb{R}_{\ge 0}$, the *adjacency matrix* $\boldsymbol{A}$ is such that $\boldsymbol{A}_{u,v} = w_{u,v}$ for all $(u, v) \in E$. Note that since the graph is undirected, the matrix is symmetric. The *degree matrix* $\boldsymbol{D}$ is a diagonal matrix such that $\boldsymbol{D}_{uu} = \boldsymbol{d}(u) = \sum_v w_{uv}$ for all $u \in V$, where $\boldsymbol{d}(u) = \deg_G(u)$ is the short hand for degree, and $\boldsymbol{d}(S) = \mathrm{vol}_G(S)$ is the volume of

the vertex set S.
The **Laplacian matrix** is defined as the difference of these two matrices: $\boldsymbol{L}_G = \boldsymbol{D} - \boldsymbol{A}$.[2]

3.1. Fact. *$\boldsymbol{L}_G$ is a real symmetric matrix. Therefore, all the eigenvalues of $\boldsymbol{L}_G$ are real.*[3]

Let $\boldsymbol{L}_{(u,v)}$ be the Laplacian of the subgraph of G that contains only one unweighted edge (u,v). Then in this case, since

$$D = \begin{array}{cc} \begin{array}{ccccccccccc} & \cdots & & u & & \cdots & & v & & \cdots & \end{array} & \\ \begin{pmatrix} 0 & & & & & \cdots & & & & & 0 \\ \vdots & & & & & & & & & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & & & & & & & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & & & & & & \vdots \\ 0 & & & & & \cdots & & & & & 0 \end{pmatrix} & \begin{array}{c} \vdots \\ \\ u \\ \\ \vdots \\ \\ v \\ \\ \vdots \end{array} \end{array}$$

$$A = \begin{array}{cc} \begin{array}{ccccccccccc} & \cdots & & u & & \cdots & & v & & \cdots & \end{array} & \\ \begin{pmatrix} 0 & & & & & \cdots & & & & & 0 \\ \vdots & & & & & & & & & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & & & & & & & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & & & & & & & & \vdots \\ 0 & & & & & \cdots & & & & & 0 \end{pmatrix} & \begin{array}{c} \vdots \\ \\ u \\ \\ \vdots \\ \\ v \\ \\ \vdots \end{array} \end{array}$$

[2]Note that L_G does not change when we add self-loops.

[3]In fact, it is quite easy to see that L_G is positive semidefinite (this can be show by endowing the graph an arbitrary direction, then $\boldsymbol{L}_G$ is the incidence matrix multiply by its transpose) and hence all its eigenvalues are non-negative.

the Laplacian of a single-edge subgraph

$$
L_{(u,v)} = D - A = \begin{array}{cc}
\begin{array}{ccccccccccc} & \cdots & & u & & \cdots & & v & & \cdots & \end{array} & \\
\begin{pmatrix}
0 & & & & & \cdots & & & & & 0 \\
\vdots & & & & & & & & & & \vdots \\
0 & \cdots & 0 & 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & & & & & & & & & & \vdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & \cdots & 0 & -1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
\vdots & & & & & & & & & & \vdots \\
0 & & & & & \cdots & & & & & 0
\end{pmatrix} &
\begin{array}{c} \vdots \\ \\ u \\ \\ \vdots \\ \\ v \\ \\ \vdots \end{array}
\end{array}
$$

which looks very simple.

By decomposing the whole graph into edges, the Laplacian of the graph L_G is the sum of these simple matrices:

3.2. Proposition. $L_G = \sum_{e \in E} w_e L_e$.

Since the Laplacian is a linear combination of Laplacian of a single-edge subgraph, $L_{(u,v)}$ is important and we should take a closer look. Observe that for $L_{(u,v)}$, we have $\forall x \in \mathbb{R}^V$

$$x^\top L_{(u,v)} x = x_u^2 - 2x_u x_v + x_v^2 = (x_u - x_u)^2$$

whence

$$\mathbf{1}_S^\top L_{(u,v)} \mathbf{1}_S = \begin{cases} 1 & \text{if } |\{u, v\} \cap S| = 1 \text{ (i.e. } \{u, v\} \text{ is cut)} \\ 0 & \text{otherwise} \end{cases}$$

Since $x^\top L_G x = \sum_{e \in E} w_e (x^\top L_e x)$, we can conclude that

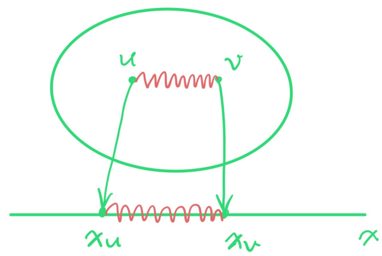
3.3. Proposition. *$\forall x \in \mathbb{R}^V$, we have*

$$x^\top L_G x = \sum_{(u,v) \in E} w_{uv} (x_u - x_v)^2.$$

In particular, if $x = \mathbf{1}_S$, we have

$$x^\top L_G x = w_G(\partial_G S).$$

3.4. *Note.* A intuitive interpretation of this proposition is that we can think of in the original graph, all edges (u, v) are springs with spring constant $w_{(u,v)}$. We use x to keep track of each edge (u, v), in the way that $(x_u - x_v)^2$ scaled by the "spring constant," namely the edge weight, is the energy of the spring corresponds to (u, v). $x^T L_G x$ stands for the total spring energy in the graph G.



3.5. Proposition. *Let $\lambda_1(\boldsymbol{L}_G) \le \cdots \le \lambda_n(\boldsymbol{L}_G)$ be the eigenvalues of $\boldsymbol{L}_G$. We have*

1. *$\lambda_1(\boldsymbol{L}_G) = 0$ with a corresponding eigenvector $\boldsymbol{v}_1 = \mathbf{1}$.*
2. *$\lambda_k(\boldsymbol{L}_G) = 0$ if and only if G contains at least k connected components.*

Bizonyítás.

1. By definition,

$$\lambda_1(\boldsymbol{L}_G) = \min_{\boldsymbol{x} \neq \mathbf{0}} R_{L_G}(\boldsymbol{x}) = \min_{\boldsymbol{x} \neq \mathbf{0}} \frac{\sum_{(u,v)\in E} w_{uv}(x_u - x_v)^2}{\boldsymbol{x}^\top \boldsymbol{x}}.$$

Since $(x_u - x_v)^2 \ge 0$, we have $\lambda_1(\boldsymbol{L}_G) \ge 0$. On the other hand, if we let $\boldsymbol{x} = \mathbf{1}$, then $R_{\boldsymbol{L}_G}(\mathbf{1}) = 0$. Therefore, as the minimum over all possible $\mathbb{R}_{\boldsymbol{L}_G}(x)$, $\lambda_1(\boldsymbol{L}_G) \le R_{\boldsymbol{L}_G}(\mathbf{1}) = 0$. Hence $\lambda_1(\boldsymbol{L}_G) = 0$. To find its corresponding eigenvector, as previously mentioned on page 3, the Laplacian of the subgraph with a single unweighted edge has exactly 2 non-zero rows containing exactly one 1 and one -1 each. Hence when we multiply this matrix by an all one vector, we will get zero vector. Therefore, $\boldsymbol{L}_G\mathbf{1} = \sum_{e\in E} w_e \boldsymbol{L}_e \mathbf{1} = \mathbf{0}$. Hence the eigenvector corresponding to the eigenvalue 0 is $\boldsymbol{v}_1 = \mathbf{1}$.

Another way to find $\boldsymbol{v}_1 = \mathbf{1}$ is by eyeballing $\boldsymbol{L}_G\mathbf{1} = (\boldsymbol{D} - \boldsymbol{A})\mathbf{1} = \boldsymbol{D}\mathbf{1} - \boldsymbol{A}\mathbf{1}$. The i^{th} row of $\boldsymbol{D}\mathbf{1}$ is the degree of u, and the i^{th} row of $\boldsymbol{A}\mathbf{1}$ is the sum of the weight of all edges adjacent to u, which is also the degree of u. Therefore for any arbitrary i, the i^{th} row of $(\boldsymbol{D} - \boldsymbol{A})\mathbf{1}$ is 0, whence $(\boldsymbol{D} - \boldsymbol{A})\mathbf{1} = \mathbf{0}$.

2. For $\lambda_k(\boldsymbol{L}_G)$, recall that

$$\lambda_k(\boldsymbol{L}_G) = \min_{k\text{-dim } \mathcal{V}} \max_{\boldsymbol{x}\in\mathcal{V}-\{\mathbf{0}\}} R_{L_G}(\boldsymbol{x})$$

" $\Longleftarrow$ " Suppose there are k connected components $C_1, \ldots, C_k$.

Let $\mathcal{V} = \operatorname{span}\{1_{C_1}, \ldots 1_{C_k}\}$. Note that $\dim(\mathcal{V}) = k$. For any $\boldsymbol{x} \in \mathcal{V}$ and for each component C_i, entries of $\boldsymbol{x}$ correspond to vertices in C_i are the same constant since by definition of $\mathcal{V}$, we can represent any $\boldsymbol{x} \in \mathcal{V}$ by using basis $\boldsymbol{x} = \sum_{i=1}^k \mathbf{1}_{C_i}$. Therefore $\sum_{(u,v)\in E(C_i)} w_{uv}(x_u - x_v)^2 = 0$. Since there's no edges between each connected components $R_{\boldsymbol{L}_G}(\boldsymbol{x}) = \sum_{i=1}^k \sum_{(u,v)\in E(C_i)} w_{uv}(x_u - x_v)^2 = 0$, and $\lambda_k(\boldsymbol{L}_G) \le 0$. Combining with the fact that $\lambda_k(\boldsymbol{L}_G)$ is non-negative, it is 0

" $\Longrightarrow$ " Suppose $\lambda_k(\boldsymbol{L}_G) = 0$.

Let $\mathcal{V}$ be a k-dimensional space where $\max_{\boldsymbol{x}\in\mathcal{V}-\{\mathbf{0}\}} R_{L_G}(\boldsymbol{x}) = 0$. For any $\boldsymbol{x} \in \mathcal{V}$ and for each component C_i, entries of $\boldsymbol{x}$ in C_i must be constant. Otherwise, $R_{L_G}(\boldsymbol{x}) > 0$. Therefore $\mathcal{V} \subseteq \operatorname{span}\{1_{C_1}, \ldots 1_{C_z}\}$ where z counts the connected components in G. So $k = \dim(\mathcal{V}) \le z$, which in English means that the number of connected components of G is at least k.

□

3.6. *Remark.* Basically, $\lambda_1(\boldsymbol{L}_G)$ is not informative at all since it is always zero. But $\lambda_2(\boldsymbol{L}_G) > 0$ iff G is connected.

In other words, if $\lambda_2 = 0$, then the graph is not connected. A intuitive question to ask is that what if λ_2 is very closed to 0, would that means that the graph is very closed to be disconnected? And the answer is yes. We'll see the proof in the following section.

4. The Normalized Laplacian of Graphs

4.1. Definition. Excluding isolated vertices(namely $\boldsymbol{D}$ is of full rank,) the **normalized Laplacian** matrix is

$$\boldsymbol{N}_G = \boldsymbol{D}^{-1/2}\boldsymbol{L}_G\boldsymbol{D}^{-1/2} = I - \boldsymbol{D}^{-1/2}\boldsymbol{A}_G\boldsymbol{D}^{-1/2}.$$

Basically a scaled Laplacian.

4.2. Exercise. We have

1. $\lambda_1(\boldsymbol{N}_G) = 0$ with a corresponding eigenvector $\boldsymbol{v}_1 = \boldsymbol{d}^{1/2}$, where $\boldsymbol{d} = (\deg(u))_{u\in V(G)}$.
2. $\lambda_k(\boldsymbol{N}_G) = 0$ if and only if G contains at least k connected component.

4.3. Lemma.

$$\lambda_2(\boldsymbol{N}_G) = \min_{\boldsymbol{x}\perp\boldsymbol{d}} \frac{\boldsymbol{x}^\top\boldsymbol{L}_G\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{D}\boldsymbol{x}}.$$

Bizonyítás.

Firstly, by definition

$$\lambda_2(\boldsymbol{N}_G) = \min_{\boldsymbol{v}\perp\boldsymbol{d}^{1/2}} \frac{\boldsymbol{v}^\top\boldsymbol{N}_G\boldsymbol{v}}{\boldsymbol{v}^\top\boldsymbol{v}}$$

since the first eigenvector of $\boldsymbol{N}$ is $\boldsymbol{d}^{1/2}$. Now, define $\boldsymbol{x} = \boldsymbol{D}^{-1/2}\boldsymbol{v}$, so $\boldsymbol{v} = \boldsymbol{D}^{1/2}\boldsymbol{x}$. We have $\frac{\boldsymbol{v}^\top\boldsymbol{N}_G\boldsymbol{v}}{\boldsymbol{v}^\top\boldsymbol{v}} = \frac{\boldsymbol{v}^\top\boldsymbol{D}^{-1/2}\boldsymbol{L}_G\boldsymbol{D}^{-1/2}\boldsymbol{v}}{\boldsymbol{v}^\top\boldsymbol{v}} = \frac{\boldsymbol{x}\boldsymbol{L}_G\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{D}\boldsymbol{x}}$. Also, $\boldsymbol{v}\perp\boldsymbol{d}^{1/2}$ iff $\boldsymbol{x}\perp\boldsymbol{d}$. Because $0 = \boldsymbol{v}^\top\boldsymbol{d}^{1/2} = \boldsymbol{x}^\top\boldsymbol{D}^{1/2}\boldsymbol{d}^{1/2} = \boldsymbol{x}^\top\boldsymbol{d}$. Since the mapping $\boldsymbol{v} \mapsto \boldsymbol{D}^{-1/2}\boldsymbol{v} = \boldsymbol{x}$ is a bijection, we have $\lambda_2(\boldsymbol{N}_G) = \min_{\boldsymbol{x}\perp\boldsymbol{d}} \frac{\boldsymbol{x}^\top\boldsymbol{L}_G\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{D}\boldsymbol{x}}$ as desired. □

4.4. *Note.* The above lemma is the reason why we look at $\boldsymbol{N}_G$-the second eigenvalue of normalized Laplacian is closely related to the conductance since they are minimizing the same objective but over different domains. Indeed, we can see the similarity between $\lambda_2(\boldsymbol{N}_G)$ and $\Phi(G)$:

$$\lambda_2(\boldsymbol{N}_G) = \min_{\boldsymbol{x}\perp\boldsymbol{d}} \frac{\boldsymbol{x}^\top\boldsymbol{L}_G\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{D}\boldsymbol{x}}$$

and

$$\Phi(G) = \min_{\boldsymbol{x}=\boldsymbol{1}_S:0<\boldsymbol{d}(S)\leq\boldsymbol{d}(V)/2} \frac{\boldsymbol{x}^\top\boldsymbol{L}_G\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{D}\boldsymbol{x}}$$

where $\boldsymbol{x} = \boldsymbol{1}_S$, $\boldsymbol{x}^\top\boldsymbol{L}_G\boldsymbol{x} = w_G(\partial_G S)$ and $\boldsymbol{x}^\top\boldsymbol{D}\boldsymbol{x} = \boldsymbol{d}(S)$.

5. Cheeger's Inequality: Statement

It turns out that $\lambda_2(\boldsymbol{N}_G)$ does not only tell us if G is connected or not. It actually measures how much G is well-connected.

5.1. Theorem (Cheeger's Inequality)**.** *We have*

$$\frac{\lambda_2(\boldsymbol{N}_G)}{2} \leq \Phi(G) \leq \sqrt{2\lambda_2(\boldsymbol{N}_G)}.$$

Furthermore, there is an algorithm that find a cut S where $\Phi_G(S) \leq \sqrt{2\lambda_2(\boldsymbol{N}_G)} \leq 2\sqrt{\Phi(G)}$.

5.2. *Note.*

1. What this means - G **is expander if and only if the second eigenvalue of $\boldsymbol{N}_G$ is big.**
2. So when $\Phi(G) = \Omega(1)$, this gives $O(1)$-approximation algorithm for $\Phi(G)$.
3. But the approximation can be very bad when $\Phi(G)$ is small. For example, when $\Phi(G) = O(\frac{1}{n})$, the algorithm return a cut where $\Phi_G(S) \leq O\left(\sqrt{\frac{1}{n}}\right)$ which can be much larger than $O(\frac{1}{n})$.

5.3. Question. *Is there a $O(1)$-approximation algorithm $\Phi(G)$ for any graph G? If so, this would refute the small-set expansion conjecture.*

- Best known approximation: $O(\sqrt{\log n})$ via ARV.
  - We saw $O(\log n)$-approx algorithm (which is tight because of the flow-cut gap).
- The Cheeger's Inequality is tight.
  - the second inequality: for cycle of size n, $\Phi(G) = \Theta(1/n)$ and $\lambda_2(\boldsymbol{N}_G) = O(1/n^2)$
  - the first inequality: for d-dimensional hypercube on $n = 2^d$ nodes, $\Phi(G) = 1/d$ and $\lambda_2(\boldsymbol{N}_G) = 2/d$
  - See proofs: https://lucatrevisan.github.io/teaching/expanders2016/lecture15.pdf

6. Easy Direction: Sparse Cut implies Small Eigenvalue

6.1. Lemma. $\lambda_2(\boldsymbol{N}_G)/2 \leq \Phi(G)$

Bizonyítás. Consider any $S \subset V$ where wlog $0 < \boldsymbol{d}(S) \leq \boldsymbol{d}(V)/2$. Define $\boldsymbol{y}$ based on S,

$$\boldsymbol{y} = \boldsymbol{1}_S - \sigma\boldsymbol{1}$$

where $\sigma = \boldsymbol{d}(S)/\boldsymbol{d}(V)$ is chosen so that $\boldsymbol{y} \perp \boldsymbol{d}$. We have

$$\lambda_2(\boldsymbol{N}_G) \leq \frac{\boldsymbol{y}^\top \boldsymbol{L}_G \boldsymbol{y}}{\boldsymbol{y}^\top \boldsymbol{D} \boldsymbol{y}}.$$

Note that $\boldsymbol{y}^\top \boldsymbol{L}_G \boldsymbol{y} = w_G(\partial_G S)$, since $\boldsymbol{y}^\top \boldsymbol{L}_G \boldsymbol{y} = \sum_{(u,v)\in E} w_{uv}(y_u - y_v)^2$ is translation invariant whence is the same as $\mathbf{1}_S \boldsymbol{L}_G \mathbf{1}_S = w_G(\partial_G S)$.
Next, we compute $\boldsymbol{y}^\top \boldsymbol{D} \boldsymbol{y}$

$$
\begin{aligned}
\boldsymbol{y}^\top \boldsymbol{D} \boldsymbol{y} &= \sum_{u\in S} \boldsymbol{d}(u)(1-\sigma)^2 + \sum_{u\notin S} \boldsymbol{d}(u)\sigma^2 \\
&= \boldsymbol{d}(S)(1-\sigma)^2 + \boldsymbol{d}(V \setminus S)\sigma^2 \\
&= \boldsymbol{d}(S) - 2\sigma \boldsymbol{d}(S) + \sigma^2 \boldsymbol{d}(V) \\
&= \boldsymbol{d}(S)(1-\sigma) \\
&= \boldsymbol{d}(S)\frac{\boldsymbol{d}(V\setminus S)}{\boldsymbol{d}(V)} \geq \frac{\boldsymbol{d}(S)}{2}
\end{aligned}
$$

So we have

$$\lambda_2(\boldsymbol{N}_G) \leq 2 \cdot \frac{w_G(\partial_G S)}{\boldsymbol{d}(S)}$$

for all S where $0 < \boldsymbol{d}(S) \leq \boldsymbol{d}(V)/2$. Therefore,

$$\lambda_2(\boldsymbol{N}_G) \leq 2 \cdot \Phi(G).$$

□

7. Harder direction: Small Eigenvalue implies Sparse Cut

7.1. Theorem. *Given any $x \perp \boldsymbol{d}$, we can find a cut S in G where*

$$\Phi_G(S) \leq \sqrt{2\frac{\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{D} \boldsymbol{x}}}$$

7.2. Corollary. $\Phi(G) \leq \sqrt{2\lambda_2(\boldsymbol{N}_G)}$.

Sketch of proof of Theorem 7.1. We have $\lambda_2(\boldsymbol{N}_G) = \min_{x\perp \boldsymbol{d}} \frac{\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{D} \boldsymbol{x}}$.

- Remark that it will be necessary that the theorem works for any $x \perp \boldsymbol{d}$
  - because we will show how to find $x \perp \boldsymbol{d}$ where $\frac{\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{D} \boldsymbol{x}}$ is *almost* minimum in the next lecture.
- Given a vector $\boldsymbol{x}$, there are three main steps for constructing S.
  1. Firstly, we define a *centered* vector $\boldsymbol{y}$ where $\frac{\boldsymbol{y}^\top \boldsymbol{L}_G \boldsymbol{y}}{\boldsymbol{y}^\top \boldsymbol{D} \boldsymbol{y}} \leq \frac{\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{D} \boldsymbol{x}}$,
  2. then, we further define *non-negative* vector $\boldsymbol{z}$ where $\frac{\boldsymbol{z}^\top \boldsymbol{L}_G \boldsymbol{z}}{\boldsymbol{z}^\top \boldsymbol{D} \boldsymbol{z}} \leq \frac{\boldsymbol{y}^\top \boldsymbol{L}_G \boldsymbol{y}}{\boldsymbol{y}^\top \boldsymbol{D} \boldsymbol{y}}$
  3. $\boldsymbol{z}$ will have a nice enough structure such that it determines a *threshold cut* S where $\Phi_G(S) \leq \sqrt{2\frac{\boldsymbol{z}^\top \boldsymbol{L}_G \boldsymbol{z}}{\boldsymbol{z}^\top \boldsymbol{D} \boldsymbol{z}}}$

□

We proceed to elaborate on the three steps in the following sections.

7.1. Centering

Without loss of generality assume that

$$x_1 \le x_2 \le \cdots \le x_n.$$

7.3. Definition. Say y is *centered w.r.t.* $\boldsymbol{d}$ if

$$\sum_{u:y_u<0} \boldsymbol{d}(u) \le \boldsymbol{d}(V)/2 \quad \text{and} \quad \sum_{u:y_u>0} \boldsymbol{d}(u) \le \boldsymbol{d}(V)/2$$

Namely the total volume of both vertices with strictly less then zero value in $\boldsymbol{y}$ and vertices with strictly greater then zero value in $\boldsymbol{y}$ are less than a half.

Proof of the first step of Theorem 7.1. Given an arbitrary vector $\boldsymbol{x}$, we can make it into a centered vector by translation.

Let j be the minimum index where $\sum_{u=1}^{j} \boldsymbol{d}(u) \ge \boldsymbol{d}(V)/2$. Set

$$\boldsymbol{y} = \boldsymbol{x} - x_j \mathbf{1}$$

So we have $y_j = 0$, $\sum_{u:y_u<0} \boldsymbol{d}(u) \le \boldsymbol{d}(V)/2$ and $\sum_{u:y_u>0} \boldsymbol{d}(u) \le \boldsymbol{d}(V)/2$. Namely, $\boldsymbol{y}$ is centered w.r.t. $\boldsymbol{d}$. Also observe that

$$\frac{\boldsymbol{y}^\top \boldsymbol{L}_G \boldsymbol{y}}{\boldsymbol{y}^\top \boldsymbol{D} \boldsymbol{y}} \le \frac{\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{D} \boldsymbol{x}}$$

because (a) the numerator is not changing at all since $\boldsymbol{y}^\top \boldsymbol{L}_G \boldsymbol{y} = (\boldsymbol{x} - x_j\mathbf{1})^\top \boldsymbol{L}_G (\boldsymbol{x} - x_j\mathbf{1}) = \boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}$, as $\mathbf{1} \in \ker(\boldsymbol{L}_G)$;
(b) and the denominator can only possibly be larger, i.e. $\boldsymbol{y}^\top \boldsymbol{D} \boldsymbol{y} \ge \boldsymbol{x}^\top \boldsymbol{D} \boldsymbol{x}$. To see this, let $\boldsymbol{v}_s = \boldsymbol{x} + s\mathbf{1}$, where s is a translating parameter. The minimum of $\boldsymbol{v}_s^\top \boldsymbol{D} \boldsymbol{v}_s$ is achieved at s for which $\boldsymbol{v}_s^\top \boldsymbol{d} = 0$. This is because if we take the derivative of this value $\frac{\partial(\boldsymbol{v}_s^\top \boldsymbol{D} \boldsymbol{v}_s)}{\partial s} = \mathbf{1}^\top \boldsymbol{D} \boldsymbol{v}_s + \boldsymbol{v}_s^\top \boldsymbol{D} \mathbf{1} = 2\boldsymbol{d}^\top \boldsymbol{v}_s$, we know that when $\boldsymbol{d}^\top \boldsymbol{v}_s$ reaches 0 the value reaches its minimum. But notice that $\boldsymbol{x}$ is orthogonal to $\boldsymbol{d}$, where we have $\boldsymbol{x}^\top \boldsymbol{d} = 0$ and hence we must have $\boldsymbol{y}^\top \boldsymbol{D} \boldsymbol{y} \ge \boldsymbol{x}^\top \boldsymbol{D} \boldsymbol{x}$.

□

7.2. Non-negative

We next proceed to show the second step of the proof. Given the vector $\boldsymbol{y}$ above, we decompose it into two vectors, one comprising the positive entries and the other comprising the negative entries.

$$\boldsymbol{y} = \boldsymbol{y}^+ - \boldsymbol{y}^-$$

where

$$y_j^+ = \begin{cases} y_j & y_j > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and } y_j^- = \begin{cases} -y_j & y_j < 0 \\ 0 & \text{otherwise} \end{cases}$$

Note that $\boldsymbol{y}^+$ and $\boldsymbol{y}^-$ has several neat properties

- have disjoint non-zero entries,

- are non-negative, and
- are small w.r.t. $\boldsymbol{d}$. More precisely, $\sum_{u:y_u^+>0} \boldsymbol{d}(u) \leq \boldsymbol{d}(V)/2$ and $\sum_{u:y_u^->0} \boldsymbol{d}(u) \leq \boldsymbol{d}(V)/2$.

Suppose the following claim is true.

7.4. Claim.

$$\min\{\frac{(\boldsymbol{y}^+)^\top \boldsymbol{L}_G \boldsymbol{y}^+}{(\boldsymbol{y}^+)^\top \boldsymbol{D} \boldsymbol{y}^+}, \frac{(\boldsymbol{y}^-)^\top \boldsymbol{L}_G \boldsymbol{y}^-}{(\boldsymbol{y}^-)^\top \boldsymbol{D} \boldsymbol{y}^-}\} \leq \frac{\boldsymbol{y}^\top \boldsymbol{L}_G \boldsymbol{y}}{\boldsymbol{y}^\top \boldsymbol{D} \boldsymbol{y}}$$

Then we can simply take one of $\boldsymbol{y}^+$ and $\boldsymbol{y}^-$ that makes the inequality holds to be $\boldsymbol{z}$ where

$$\frac{\boldsymbol{z}^\top \boldsymbol{L}_G \boldsymbol{z}}{\boldsymbol{z}^\top \boldsymbol{D} \boldsymbol{z}} \leq \frac{\boldsymbol{y}^\top \boldsymbol{L}_G \boldsymbol{y}}{\boldsymbol{y}^\top \boldsymbol{D} \boldsymbol{y}}$$

with $\boldsymbol{z}$ being non-negative and $\sum_{u:z_u>0} \boldsymbol{d}(u) \leq \boldsymbol{d}(V)/2$. It remains to prove the claim.

Proof of claim 7.4. We first show $\forall uv \in E, ((y_u^+ - y_v^+) - (y_u^- - y_v^-))^2 \geq \left((y_u^+ - y_v^+)^2 + (y_u^- - y_v^-)^2\right)$.
We will prove this by considering the contribution of every edge. Consider any edge (u, v). If both u and v such that $y_u, y_v \geq 0$ or $y_u, y_v \leq 0$, then (u, v) contributes the exact same to both sides of inequality.
Otherwise, if $y_u > 0 > y_v$, then (u, v) contributes $(y_u^+ + y_v^-)^2$ to the left, but $(y_u^+)^2 + (y_v^-)^2$ to the right, where we know the inequality $(y_u^+ + y_v^-)^2 \geq (y_u^+)^2 + (y_v^-)^2$ always holds.
Therefore, we have

$$\begin{aligned}
\frac{\boldsymbol{y}^\top \boldsymbol{L}_G \boldsymbol{y}}{\boldsymbol{y}^\top \boldsymbol{D} \boldsymbol{y}} &= \frac{\sum_{uv\in E} w_{uv}(y_u - y_v)^2}{\boldsymbol{y}^\top \boldsymbol{D} \boldsymbol{y}} \\
&\text{by decomposing and rearranging} \\
&= \frac{\sum_{uv\in E} w_{uv} \left((y_u^+ - y_v^+) - (y_u^- - y_v^-)\right)^2}{(\boldsymbol{y}^+)^\top \boldsymbol{D} \boldsymbol{y}^+ + (\boldsymbol{y}^-)^\top \boldsymbol{D} \boldsymbol{y}^- - 2((\boldsymbol{y}^-)^\top \boldsymbol{D} \boldsymbol{y}^+)} \\
&\text{according to what we've just shown} \\
&\geq \frac{\sum_{uv\in E} w_{uv} \left((y_u^+ - y_v^+)^2 + (y_u^- - y_v^-)^2\right)}{(\boldsymbol{y}^+)^\top \boldsymbol{D} \boldsymbol{y}^+ + (\boldsymbol{y}^-)^\top \boldsymbol{D} \boldsymbol{y}^-} \\
&\text{since } \forall a_1, a_2, b_1, b_2, \frac{a_1 + a_2}{b_1 + b_2} \geq \min\left(\frac{a_1}{b_1}, \frac{a_2}{b_2}\right) \\
&\geq \min\left\{\frac{\sum_{uv\in E} w_{uv}(y_u^+ - y_v^+)^2}{(\boldsymbol{y}^+)^\top \boldsymbol{D} \boldsymbol{y}^+}, \frac{\sum_{uv\in E} w_{uv}(y_u^- - y_v^-)^2}{(\boldsymbol{y}^-)^\top \boldsymbol{D} \boldsymbol{y}^-}\right\} \\
&= \min\left\{\frac{(\boldsymbol{y}^+)^\top \boldsymbol{L}_G \boldsymbol{y}^+}{(\boldsymbol{y}^+)^\top \boldsymbol{D} \boldsymbol{y}^+}, \frac{(\boldsymbol{y}^-)^\top \boldsymbol{L}_G \boldsymbol{y}^-}{(\boldsymbol{y}^-)^\top \boldsymbol{D} \boldsymbol{y}^-}\right\}
\end{aligned}$$

□

7.3. Threshold Cuts

It suffices to show the following lemma,

7.5. Lemma. *Given a non-negative* $\boldsymbol{z}$ *as above, there is a number* τ *where* $S_\tau = \{u \mid \boldsymbol{z}_u \geq \tau\}$ *such that*

$$\Phi_G(S_\tau) \leq \sqrt{2\frac{\boldsymbol{z}^\top \boldsymbol{L}_G \boldsymbol{z}}{\boldsymbol{z}^\top \boldsymbol{D} \boldsymbol{z}}}$$

7.6. *Note.* Visually, when we set all entries of $\boldsymbol{y}$ on a real axis, if $\boldsymbol{z} = \boldsymbol{y}^-$, then the cut S_τ would be a cut includes all vertices correspond to points including everything from the left most on the axis all the way through the smallest u such that $\boldsymbol{z}_u \geq \tau$. This is because when we defined $\boldsymbol{y}^-$ we flipped the sign of each entry contained in this part in $\boldsymbol{y}$;
Conversely, when $\boldsymbol{z} = \boldsymbol{y}^+$, then the cut S_τ would includes all vertices correspond to points including everything from the right most on the axis all the way through the smallest u such that $\boldsymbol{z}_u \geq \tau$.

To show the Lemma, w.l.o.g we may assume $\max_u \boldsymbol{z}_u = 1$ since scaling does not change the ratio $\frac{\boldsymbol{z}^\top \boldsymbol{L}_G \boldsymbol{z}}{\boldsymbol{z}^\top \boldsymbol{D} \boldsymbol{z}}$. We are going to use a probabilistic argument, where we first let $\tau > 0$ be sampled such that τ^2 **is uniformly distributed in** $[0, 1]$. Namely, $\Pr[\tau \leq \alpha] = \Pr[\tau^2 \leq \alpha^2] = \alpha^2$ for all $\alpha \in [0, 1]$.[4] Let $S_\tau = \{u \mid \boldsymbol{z}_u \geq \tau\}$, then the key claim is that

7.7. Claim.

$$\frac{\mathbb{E}_\tau w_G(\partial_G S_\tau)}{\mathbb{E}_\tau \boldsymbol{d}(S_\tau)} \leq \sqrt{2\frac{\boldsymbol{z}^\top \boldsymbol{L}_G \boldsymbol{z}}{\boldsymbol{z}^\top \boldsymbol{D} \boldsymbol{z}}}$$

Given the claim, even though

$$\mathbb{E}_\tau \left(\frac{w_G(\partial_G S_\tau)}{\boldsymbol{d}(S_\tau)} \right) \neq \frac{\mathbb{E}_\tau w_G(\partial_G S_\tau)}{\mathbb{E}_\tau \boldsymbol{d}(S_\tau)},$$

we are done since there must exist a τ s.t.

$$\min_\tau \frac{w_G(\partial_G S_\tau)}{\boldsymbol{d}(S_\tau)} \leq \frac{\mathbb{E}_\tau w_G(\partial_G S_\tau)}{\mathbb{E}_\tau \boldsymbol{d}(S_\tau)}$$

since $\frac{\sum_{i=1}^k a_i}{\sum_{i=1}^k b_i} \geq \min_i \frac{a_i}{b_i}$. Since we have $\boldsymbol{d}(S_\tau) \leq \sum_{u: \boldsymbol{z}_u > 0} \boldsymbol{d}(u) \leq \boldsymbol{d}(V)/2$,

$$\Phi_G(S_\tau) = \frac{w_G(\partial_G S_\tau)}{\boldsymbol{d}(S_\tau)}$$

which implies

$$\Phi_G(S_\tau) \leq \sqrt{2\frac{\boldsymbol{z}^\top \boldsymbol{L}_G \boldsymbol{z}}{\boldsymbol{z}^\top \boldsymbol{D} \boldsymbol{z}}}$$

as desired.

[4] For the existential argument, we consider the probability density function $f(x) = 2x$ because $\int_{x=0}^{\alpha} 2x = x^2$.

Proof of claim 7.7. To show $\frac{\mathbb{E}_\tau w_G(\partial_G S_\tau)}{\mathbb{E}_\tau \boldsymbol{d}(S_\tau)} \le \sqrt{2\frac{\boldsymbol{z}^\top \boldsymbol{L}_G \boldsymbol{z}}{\boldsymbol{z}^\top \boldsymbol{D}\boldsymbol{z}}}$, for the denominator,

$$\begin{aligned}\mathbb{E}_\tau \boldsymbol{d}(S_\tau) &= \sum_u \boldsymbol{d}(u)\Pr[u \in S_\tau] \\ &= \sum_u \boldsymbol{d}(u)\Pr[\tau \le z_u] \\ &= \sum_u \boldsymbol{d}(u) z_u^2 = \boldsymbol{z}^\top \boldsymbol{D}\boldsymbol{z}\end{aligned}$$

For the numerator,

$$\begin{aligned}\mathbb{E}_\tau w_G(\partial_G S_\tau) = &\sum_{uv} w_{uv}\Pr[\{u, v\} \text{ is cut}] \\ &\text{since } \tau \text{ is the exact value that separate the vertices within the cut and vertices that are not in the cut} \\ &= \sum_{uv} w_{uv}\Pr[z_u < \tau \le z_v] \qquad \text{assuming } z_v > z_u \\ &\text{this probability is exactly } \Pr[\tau \le z_v] - \Pr[\tau \le z_u] \text{ i.e.} \\ &= \sum_{uv} w_{uv}(z_v^2 - z_u^2) \\ &= \sum_{uv} w_{uv}(z_v - z_u)(z_v + z_y)\end{aligned}$$

By Cauchy-Schwarz, we have

$$\begin{aligned}\sum_{uv} w_{uv}(z_v - z_u)(z_v + z_y) &\le \sqrt{\sum_{uv} w_{uv}(z_v - z_u)^2} \cdot \sqrt{\sum_{uv} w_{uv}(z_v + z_u)^2} \\ &\le \sqrt{\boldsymbol{z}^\top \boldsymbol{L}_G \boldsymbol{z}} \cdot \sqrt{2\boldsymbol{z}^\top \boldsymbol{D}\boldsymbol{z}}\end{aligned}$$

since

$$\begin{aligned}\sum_{uv} w_{uv}(z_v + z_u)^2 &\le \sum_{uv} w_{uv}(2z_v^2 + 2z_u^2) \\ &= \sum_u 2\deg(u) \cdot z_u^2 \\ &= \sum_u 2\boldsymbol{d}(u) \cdot z_u^2 = 2\boldsymbol{z}^\top \boldsymbol{D}\boldsymbol{z}.\end{aligned}$$

[5]

□

8. Conclusion: Fiedler's Algorithm

The proof above is in fact algorithmic - we can find a sparse cut S where $\Phi_G(S) \le \sqrt{2\lambda_2(\boldsymbol{N}_G)} \le 2\sqrt{\Phi(G)}$. Specifically, the algorithm is as follows:

[5] Note that it is this last step where we need that $\boldsymbol{d}(u) \ge \deg(u)$. So it would not work if we want to prove a generalized Cheeger's inequality for $\Phi(G, \boldsymbol{d})$ for arbitrary vector $\boldsymbol{d}$

1. Compute the second eigenvector $\boldsymbol{v}_2$ of $\boldsymbol{N}_G$. We have $\boldsymbol{v}_2 \perp \boldsymbol{d}^{1/2}$ and $\frac{\boldsymbol{v}_2^\top \boldsymbol{N}_G \boldsymbol{v}_2}{\boldsymbol{v}_2^\top \boldsymbol{v}_2} = \lambda_2(\boldsymbol{N}_G)$.
2. Compute $\boldsymbol{x} = \boldsymbol{D}^{-1/2}\boldsymbol{v}_2$. We have $x \perp \boldsymbol{d}$ and $\frac{\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{D} \boldsymbol{x}} = \lambda_2(\boldsymbol{N}_G)$.
3. Sort indices such that $x_1 \le \cdots \le x_n$.
4. Define $S'_i = \{1, \dots, i\}$ for all $i < n$, where $\forall j \in [i]$, j represents the vertex in G corresponding to entry x_j.
5. Return S^* with minimum $\Phi_G(S'_i)$ among all $i \in \{1, \dots, n-1\}$.

8.1. Lemma (Correctness)**.** *Consider the cut S_τ from the proof above. We have $(S_\tau, V \setminus S_\tau) = (S'_i, V \setminus S'_i)$ for some i. So $\Phi_G(S^*) \le \Phi_G(S_\tau) \le \sqrt{2\lambda_2(\boldsymbol{N}_G)}$.*

8.2. Exercise. If instead of $\boldsymbol{v}_2$, we are given a vector $\boldsymbol{v} \perp \boldsymbol{d}^{1/2}$ where $\frac{\boldsymbol{v}^\top \boldsymbol{N}_G \boldsymbol{v}}{\boldsymbol{v}^\top \boldsymbol{v}} \le \lambda_2(\boldsymbol{N}_G) + \epsilon$. We would have $\Phi_G(S^*) \le \sqrt{2\lambda_2(\boldsymbol{N}_G) + 2\epsilon}$

Solution. Let $\boldsymbol{x} = \boldsymbol{D}^{-1/2}\boldsymbol{v}$, then $\boldsymbol{x} \perp \boldsymbol{d}$. Theorem 7.1 shows that as long as $\boldsymbol{x} \perp \boldsymbol{d}$, the S^* found by Fiedler's algorithm satisfies $\Phi_G(S^*) \le \sqrt{2\frac{\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{D} \boldsymbol{x}}}$. Since $\frac{\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{D} \boldsymbol{x}} = \frac{\boldsymbol{v}^\top \boldsymbol{N}_G \boldsymbol{v}}{\boldsymbol{v}^\top \boldsymbol{v}} \le \lambda_2(\boldsymbol{N}_G) + \epsilon, \implies \Phi_G(S^*) \le \sqrt{2\lambda_2(\boldsymbol{N}_G) + 2\epsilon}$. □

8.3. Lemma (Time)**.** *Given the second eigenvector $\boldsymbol{v}_2$, we can compute S^* in $O(m + n \log n)$ time.*

- Modulo the time for computing $\boldsymbol{v}_2$, the algorithm is very fast and gives a good approximation when $\Phi(G)$ is big.
- In the next lecture, we will show how to obtain the vector which is almost as good as $\boldsymbol{v}_2$ in $\tilde{O}(m/\Phi(G))$ or even $\tilde{O}(m)$ time.

8.4. Question (Possible project)**.** *I think that using ideas from Section 7 of https://arxiv.org/pdf/1910.07950.pdf, we should be able to improve the above algorithm/analysis in two ways*

1. *The analysis based on "median cut" will be more intuitive.*
2. *The algorithm will only need to "query" only $O(\log^2 n)$ cuts of the form S'_i instead of $n-1$ cuts.*

The analysis in the paper is for PageRank but it can possibly work for Cheeger's cut too.

9. Extensions of Cheeger's Inequality

- Extension to Directed graphs [6]
- Improvement when vertex expansion is high[7]
- What about quantitative implications of λ_k for $k > 2$?
  - $\lambda_k(\boldsymbol{N}_G) = 0$ iff G has at least k connected components
  - What if $\lambda_k(\boldsymbol{N}_G) \approx 0$? We can also say that G can be clustered into k parts by deleting few edges.

[6] https://core.ac.uk/download/pdf/206608976.pdf
[7] https://arxiv.org/pdf/1504.00686.pdf