1. Ratio Cut Objectives

<u>High-level goal:</u> Given $G=(V,E)$, find $S \subseteq V$ such that $|S| \approx |\bar{S}|$ and $\mathrm{cut}(S)$ is small.

<u>More specific objectives</u>

- Balanced graph partitioning

$$\begin{aligned} \min \quad & \mathrm{cut}(S) \\ \text{s.t.} \quad & |S| = |\bar{S}| \end{aligned}$$

(assumes $|V|$ is even, can be relaxed)

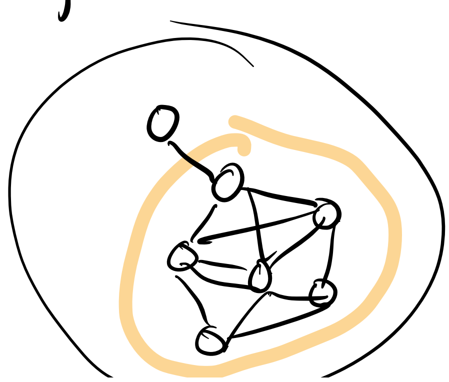
- sparsest cut

$$\min_{S \subseteq V} \boxed{\frac{\mathrm{cut}(S)}{|S|} + \boxed{\frac{\mathrm{cut}(\bar{S})}{|\bar{S}|}}}$$

$\mathrm{cut}(S) = \mathrm{cut}(\bar{S})$

Q: why not just minimize $\min_{S \subseteq V} \boxed{\frac{\mathrm{cut}(S)}{|S|}}$?

V



Q: Why do some people say $\boxed{\dfrac{cut(S)}{|S|\,|\bar{S}|}}$ is the sparsest cut objective?

$$\frac{cut(S)}{|S|} + \frac{cut(S)}{|\bar{S}|} = \frac{|\bar{S}|\,cut(S)}{|S|\,|\bar{S}|} + \frac{|S|\,cut(S)}{|S|\,|\bar{S}|}$$

$$= |V| \cdot \frac{cut(S)}{|S| \cdot |\bar{S}|}$$

· Expansion

$$\min \frac{cut(S)}{|S|} \quad \text{s.t. } |S| \le \frac{|V|}{2}$$

I.E

$$\min \frac{cut(S)}{\min\{|S|, |\bar{S}|\}} \quad \text{s.t. } S \subseteq V$$

For any $S \subseteq V$, $|S| \le \frac{|V|}{2}$

$$\boxed{\frac{cut(S)}{|S|}} \le \boxed{\frac{cut(S)}{|S|} + \boxed{\frac{cut(S)}{|\bar{S}|}}} \le 2\,\boxed{\frac{cut(S)}{|S|}}$$

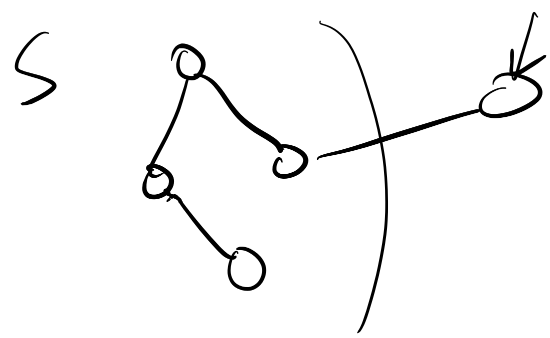
These differ only by a factor of two, so an approximation algorithm for one of them will be an approx. alg. for the other.

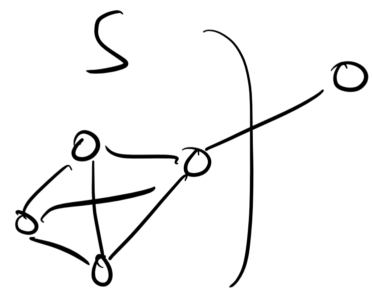
(Degree-weighted versions)

$|S| \longrightarrow vol(S)$

- Normalized cut

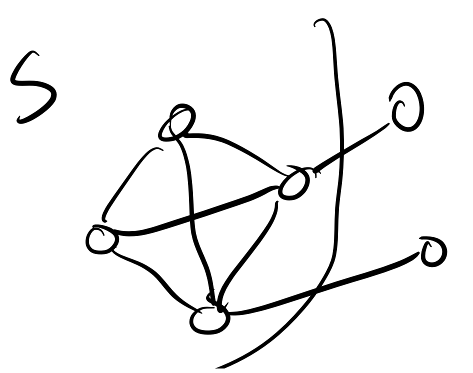
$$\min \quad \frac{cut(S)}{vol(S)} + \frac{cut(S)}{vol(\bar{S})}$$





- Conductance

$$\min_{S \subseteq V} \boxed{\frac{cut(S)}{\min\{vol(S),\, vol(\bar{S})\}}}$$



$$vol(S) \le \frac{vol(V)}{2}$$

Def: A d-regular graph is a graph where every node has degree d.

Observation: minimizing conductance is equivalent to minimizing expansion in degree regular graphs, because

$$vol(S) = d \cdot |S|$$

2. Spectral Graph Theory and the Laplacian

As seen previously, the Laplacian matrix is given by

$$L = D - A$$

where D is the degree matrix and A is the adjacency.

($G = (V, E)$ is unweighted and undirected)

- $L = BB^T$
- symmetric $\Rightarrow$ real eigenvalues
- positive semidefinite (non-negative eigenvalues)
- $L\underline{e} = \underline{0}$ where e is the all ones vector

$$L\underline{e} = D\underline{e} - A\underline{e} = \underline{d} - \underline{d} = \underline{0}$$

$\Rightarrow 0$ is always the smallest eigenvalue.

- $e_S^T L e_S = \text{cut}(S)$

Let $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ be the eigenvalues of L counting multiplicities.

__Proposition__: If $G = (V, E)$ has K connected components, then the geometric multiplicity of $\lambda_1 = 0$ is at least K.

Proof: Let the connected components be denoted by $S_1, S_2, \dots, S_K$ and for $i = 1, 2, \ldots, k$ define vector $\underline{x}_i$ by

$$\underline{x}_i(v) = \begin{cases} 1 & \text{if } v \in S_i \\ 0 & \text{otherwise} \end{cases}$$

Observe that $\underline{x}_i^T \underline{x}_j = 0$ if $i \neq j$ and

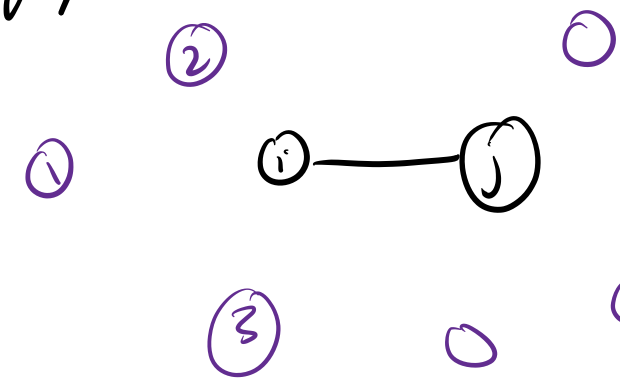
$$L\underline{x}_i = 0 \quad \text{for each} \quad i \in \{1, 2, \ldots, k\}$$

$$\text{null}(L - \lambda_1 I) = \text{null}(L).$$

Finding connected components (the most basic notion of clustering) can be encoded using eigenvectors.

2.1 The Laplacian Quadratic form

Let L_{ij} be the Laplacian for the graph with one edge $(i,j) \in E$



$$L_{ij} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$L_{ij} = \begin{array}{cc} & \begin{matrix} 1\ 2 \cdots & i & \cdots & j & \cdots\ n \end{matrix} \\ \begin{matrix} 1, 2, \ldots \\ i \\ \vdots \\ j \\ \vdots\ n \end{matrix} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ & & 0 & & 0 \\ & -1 & & 1 & \\ & & & & \end{bmatrix} \end{array}$$

Take a vector $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$

Then $$L_{ij}x = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ x_i - x_j \\ \vdots \\ x_j - x_i \\ 0 \\ 0 \end{bmatrix}$$

and $$x^T L_{ij} x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \\ \\ x_i - x_j \\ \\ x_j - x_i \\ \\ \\ \end{bmatrix}$$

$$= x_i(x_i - x_j) - x_j(x_i - x_j)$$

$$= (x_i - x_j)^2$$

For an arbitrary graph $G = (V, E)$,
I have a matrix L_{ij} for each
$(i,j) \in E$ and the Laplacian L of
G is

$$L = \sum_{(i,j) \in E} L_{ij}$$

For an arbitrary $\underline{x} \in \mathbb{R}^n$

$$\boxed{\underline{x}^T L \underline{x}} = \sum_{(i,j) \in E} \underline{x}^T L_{ij} \underline{x} = \boxed{\sum_{(i,j) \in E} (x_i - x_j)^2}$$

Can be used to quickly show

$$\boxed{e_S^T L e_S = cut(S)}$$

3. More Linear Algebra

Lemma 1: $M \in \mathbb{R}^{n \times n}$ is a symmetric matrix, and let $x_1, x_2, \ldots, x_k$ ($k < n$) be orthogonal eigenvector. Then there exists $x_{k+1} \in \mathbb{R}^n$ that is orthogonal to $\{x_1, x_2, \ldots, x_k\}$ and is an eigenvector of M as well.

Proof: see Michael Mahoney's lecture notes.

Theorem 2: Let $M \in \mathbb{R}^{n \times n}$ be symmetric, with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then there are n orthonormal vectors $x_1, x_2, \ldots, x_n$ such that $Mx_i = \lambda_i x_i$. This means:

$$\begin{array}{ccccc} \lambda_i & \leq \lambda_{i+1} & \leq \cdots & \leq \lambda_j \\ x_i & x_{i+1} & \cdots & x_j \end{array}$$

$$MX = X\Lambda \iff M = X \Lambda X^T$$

where $X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$

and $\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$

$$MX = \begin{bmatrix} Mx_1 & Mx_2 & \cdots & Mx_n \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \lambda_2 x_2 & \cdots & \lambda_2 x_n \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}}\ \underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_n \end{bmatrix}}$$

$$= X\Lambda$$

$$M\underbrace{XX^{-1}}_{I} = X\Lambda X^{-1}$$

$$\boxed{M = X\Lambda X^T}$$

Because $X^{-1} = X^T$

$$\begin{bmatrix} x_1^T \\ x_2^T \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} x_1^T x_1 & x_1^T x_2 \\ x_2^T x_1 & x_2^T x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Recall: if λ is an eigenvalue of A

$\gamma(\lambda)$ = # times λ is a root of $\det(xI - A)$

algebraic

$\delta(\lambda)$ = dimension of nullspace$(\lambda I - A)$

geometric

Standard fact: $\gamma(\lambda) \geq \delta(\lambda)$

Observation: For a symmetric real-valued matrix A, $\gamma(\lambda) = \delta(\lambda)$ for every eigenvalue λ.

Proof: Assume that $\lambda = \lambda_i = \lambda_{i+1} = \cdots = \lambda_j$ for $j > i \geq 1$ for a symmetric real-valued matrix. Then there are orthonormal eigenvectors $x_i, x_{i+1}, \ldots, x_j$

Thm 2

associated with eigenvalue λ,
so nullspace$(\lambda I - \underline{A})$ has dimension at least $(j-i+1)$, but then it must have dimension exactly $(j-i+1)$ since geometric multiplicity is always $\leq$ algebraic multiplicity. ∎

So for L, algebraic and geometric multiplicities always coincide.

Variational Characterization of Eigenvalues

Finding extreme eigenvalues can be cast as an optimization problem.

Let $M \in \mathbb{R}^{n \times n}$ be symmetric with eigenvalues $\lambda_1 \le \lambda_2 \le \dots \le \lambda_n$

If $\boxed{Mx_1 = \lambda_1 x_1} \Rightarrow \boxed{x_1^T M x_1 = \lambda_1}$

if $x_1, x_2, \dots, x_n$ is a set of orthonormal eigenvectors.

Solving the following optimization problem produces the smallest eigenpair $(\underline{x}_1, \lambda_1)$

$$(\star) \quad \begin{aligned} \min \quad & x^T M x \\ \text{s.t.} \quad & x \neq 0 \quad \|x\|_2 = 1 \end{aligned}$$

How do I find the next smallest eigenpair (x_2, λ_2)? Solve

$$(*) \quad \begin{aligned} \min \quad & x^T M x \\ \text{s.t.} \quad & \|x\|_2 = 1 \\ & x^T x_1 = 0 \end{aligned}$$

produces second smallest eigenpair (x_2, λ_2)

where x_1 is the minimizer for problem $(*)$.

Lemma: If x is the solution to $(*)$, then $Mx = \lambda_2 x$.

Proof: Let $\lambda = x^T M x \Rightarrow Mx = \lambda x$

So x is an eigenvector for M

Furthermore:

(i) $\lambda \geq \lambda_1$

(ii) There is no eigenvalue $\hat{\lambda}$ such that $\lambda_1 \le \hat{\lambda} < \lambda$

To see (i), if this were not true, then λ would be the solution to $(\ast)$, rather than λ_1. To see (ii) if such a $\hat{\lambda}$ existed, it would be the solution to problem $(\ast)$. ∎

<u>Theorem 3</u>: Let $M \in \mathbb{R}^{n \times n}$ be symmetric with eigenvalues $\lambda_1, \dots, \lambda_k$ and corresponding ortho~~normal~~gonal eigenvectors $x_1, \dots, x_k$ $(k<n)$. Then

$$\boxed{\lambda_{k+1} = \min_{\substack{x \neq 0 \\ x \perp x_i \text{ for } i=1,2,\dots,k}} \frac{x^T M x}{x^T x}}$$

Def: $R(M, x) = \dfrac{x^T M x}{x^T x}$ is called the Raleigh quotient for M and x.

Key Example:

Let L be the Laplacian of $G = (V, E)$ (undirected, unweighted).

$$\lambda_2 = \min_{\substack{x \neq 0 \\ x^T e = 0}} \frac{x^T L x}{x^T x}$$

λ_2 is called the "algebraic connectivity" of G, or the Fiedler values. The vector x_2 such that $L x_2 = \lambda_2 x_2$ is the Fiedler vector.

Lemma 4: The graph G is connected if and only if $\lambda_2 > 0$.

Proof: (1) G is disconnected $\Rightarrow \lambda_2 = 0$

If G is disconnected, there is some $S \subseteq V$ that is disconnected from $\bar{S}$, and e_S and $e_{\bar{S}}$ are orthogonal eigenvectors for eigenvalue 0.

(2) Assume the graph is connected, we will show $\lambda_2 > 0$.

If
$$\boxed{x^T L x = \sum_{(i,j) \in E} (x_i - x_j)^2 = 0}$$

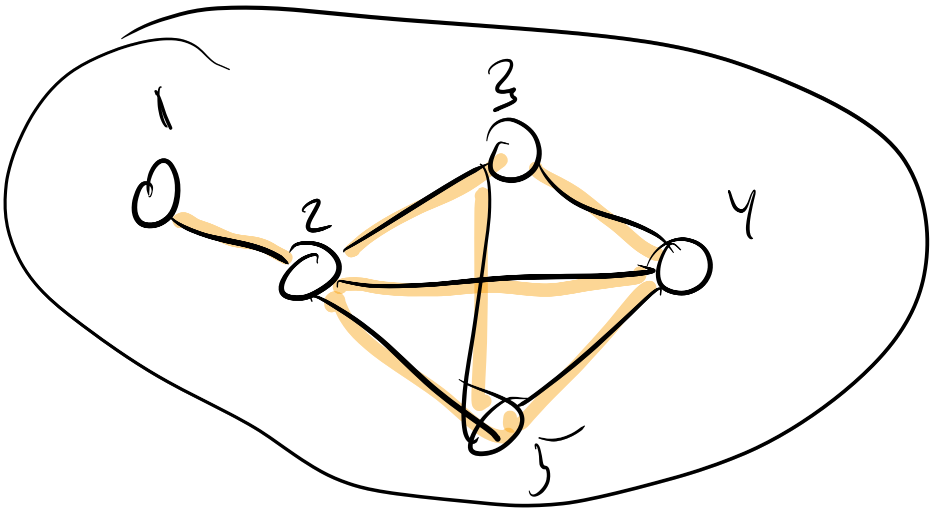
Then $x_i = x_j$ for all $(i,j) \in E$. But this means that $\underline{x} = c \cdot \underline{e}$ where $c > 0$ is a constant, for every $x \in \mathbb{R}^n$ satisfying $x^T L x = 0$.

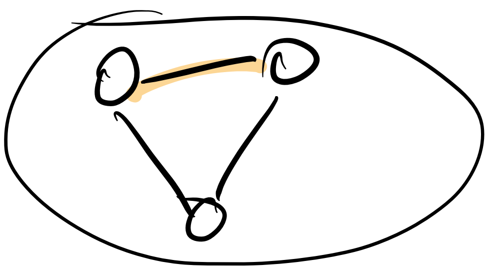
Thus it is impossible to find a vector $x_2 \in \mathbb{R}^n$ such that $x_2^T e = 0$ and

$x_2^T L x_2 = 0$ and hence

$$\lambda_2 = \min_{\substack{x \neq 0 \\ x^T e = 0}} \frac{x^T L x}{x^T x} > 0.$$

$\blacksquare$





Def: The Fiedler cut $S \subseteq V$ is

$$\boxed{S = \{ v \in V : \underline{x_2}(v) > 0 \}}$$

Lemma: The second smallest eigenvalue λ_2 of L is a relaxation of the sparsest cut problem

$$\boxed{\min_{S \subseteq V} \varphi(S) = \frac{cut(S)}{|S|} + \frac{cut(S)}{|\bar{S}|}}$$

in the sense that $\lambda_2 \leq \varphi(S)$ for all $S \subseteq V$.

Proof: Recall that

$$\lambda_2 = \min_{\substack{x \neq 0 \\ x^T e = 0}} \frac{x^T L x}{x^T x}$$

We will show that the minimum sparsest cut problem is equivalent to

$$\min_{S \subseteq V} \frac{y^T L y}{y^T y}$$

where
$$y(i) = \begin{cases} \frac{1}{|S|} & \text{if } i \in S \\ -\frac{1}{|\bar{S}|} & \text{if } i \in \bar{S} \end{cases}$$

For a vector y of this form, $\boxed{y^T e = 0}$

To show this, define $x_S(i) = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{if } i \in \bar{S} \end{cases}$

and define $x_{\bar{S}}$ similarly, then for $S \subseteq V$

$$y_S = \frac{x_S}{|S|} - \frac{x_{\bar{S}}}{|\bar{S}|}$$

Then
$$y_S^T L y_S = \sum_{(i,j) \in E} \left(y_S(i) - y_S(j) \right)^2$$

$$= \begin{cases} \frac{1}{|S|} + \frac{1}{|\bar{S}|} & \text{if } ij \in \partial S \\ 0 & \text{otherwise} \end{cases}$$

$$= \underline{\mathrm{cut}(S) \left(\frac{1}{|S|} + \frac{1}{|\bar{S}|} \right)^2}$$

Meanwhile,

$$y_S^T y_S = \left(\frac{x_S^T}{|S|} - \frac{x_{\bar{S}}^T}{|\bar{S}|}\right)\left(\frac{x_S}{|S|} - \frac{x_{\bar{S}}}{|\bar{S}|}\right)$$

$$= \frac{x_S^T x_S}{|S|^2} + \frac{x_{\bar{S}}^T x_{\bar{S}}}{|\bar{S}|^2}$$

$$= \frac{|S|}{|S|^2} + \frac{|\bar{S}|}{|\bar{S}|^2} = \frac{1}{|S|} + \frac{1}{|\bar{S}|}$$

So $$\frac{y_S^T L y_S}{y_S^T y_S} = \frac{\text{cut}(S)\left(\frac{1}{|S|} + \frac{1}{|\bar{S}|}\right)^2}{\left(\frac{1}{|S|} + \frac{1}{|\bar{S}|}\right)} = \varphi(S)$$

Also:

$$y_S^T e = \sum_{i \in S} \frac{1}{|S|} + \sum_{i \in \bar{S}} -\frac{1}{|\bar{S}|}$$

$$= \frac{|S|}{|S|} - \frac{|\bar{S}|}{|\bar{S}|} = 1 - 1 = 0.$$

Hence $\lambda_2 \leq \varphi(S) \quad \forall \; S \subseteq V$.

$\blacksquare$

Spectral Clustering

We will now focus more on the conductance objective

$$\phi(S) = \frac{\text{cut}(S)}{\min\{\text{vol}(S), \text{vol}(\bar{S})\}}$$

We will show how to approximately minimize this using spectral clustering.

Normalized Laplacian

(Assume each node has positive degree)

Def: The normalized Laplacian $\mathcal{L}$ for a simple graph $G=(V,E)$ is

$$\mathcal{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = D^{-\frac{1}{2}}(D-A)D^{-\frac{1}{2}}$$

$$= D^{-\frac{1}{2}} D D^{-\frac{1}{2}} - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

$$\mathcal{L} = I - \underbrace{D^{-\frac{1}{2}} A D^{-\frac{1}{2}}}_{\text{normalized adjacency matrix}}$$

<u>Observation</u>: define $\underline{d}^{1/2} = D^{1/2}\underline{e}$

where $\underline{e} = [1, 1, \dots, 1]^T$. Then

$$\mathcal{L}\underline{d}^{\frac{1}{2}} = 0$$

<u>Proof</u>: $\mathcal{L}\underline{d}^{1/2} = \left(I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right) d^{\frac{1}{2}}$

$$= d^{1/2} - D^{-\frac{1}{2}} A \underbrace{D^{-\frac{1}{2}} d^{1/2}}$$

$$= d^{1/2} - D^{-\frac{1}{2}} \underbrace{A\underline{e}}$$

$$= d^{1/2} - \underline{D^{-\frac{1}{2}}}\,\underline{d}$$

$$= d^{1/2} - d^{1/2} = \underline{0}$$

$\Rightarrow$ 0 is the smallest eigenvalue of $\mathcal{L}$ always and it has eigenvector $d^{1/2}$.

The second smallest eigenvalue is given by

$$\boxed{\lambda_2 = \min_{\substack{x \neq 0 \\ \boxed{x^T D^{1/2} e = 0}}} \frac{x^T \mathcal{L} x}{x^T x}}$$

$d^{1/2} = D^{1/2} e$

If we do a change of variables $\boxed{y = D^{-\frac{1}{2}} x} \Leftrightarrow x = D^{1/2} y$, we get:

$$\lambda_2 = \min_{\substack{y \neq 0 \\ y^T d = 0}} \frac{x^T D^{-\frac{1}{2}} L D^{-\frac{1}{2}} x}{x^T x} \longrightarrow y^T D y$$

$$\boxed{\lambda_2 = \min_{\substack{y \neq 0 \\ y^T d = 0}} \frac{y^T L y}{y^T D y}}$$

$$= \min_{\sum_{i=1}^{n} y_i d_i = 0} \frac{\sum_{(i,j)\in E} (y_i - y_j)^2}{\sum_{i=1}^{n} d_i y_i^2} = \min_{y^T d} \underline{\underline{R(y)}}$$

<u>Theorem (Cheeger's Inequality)</u>

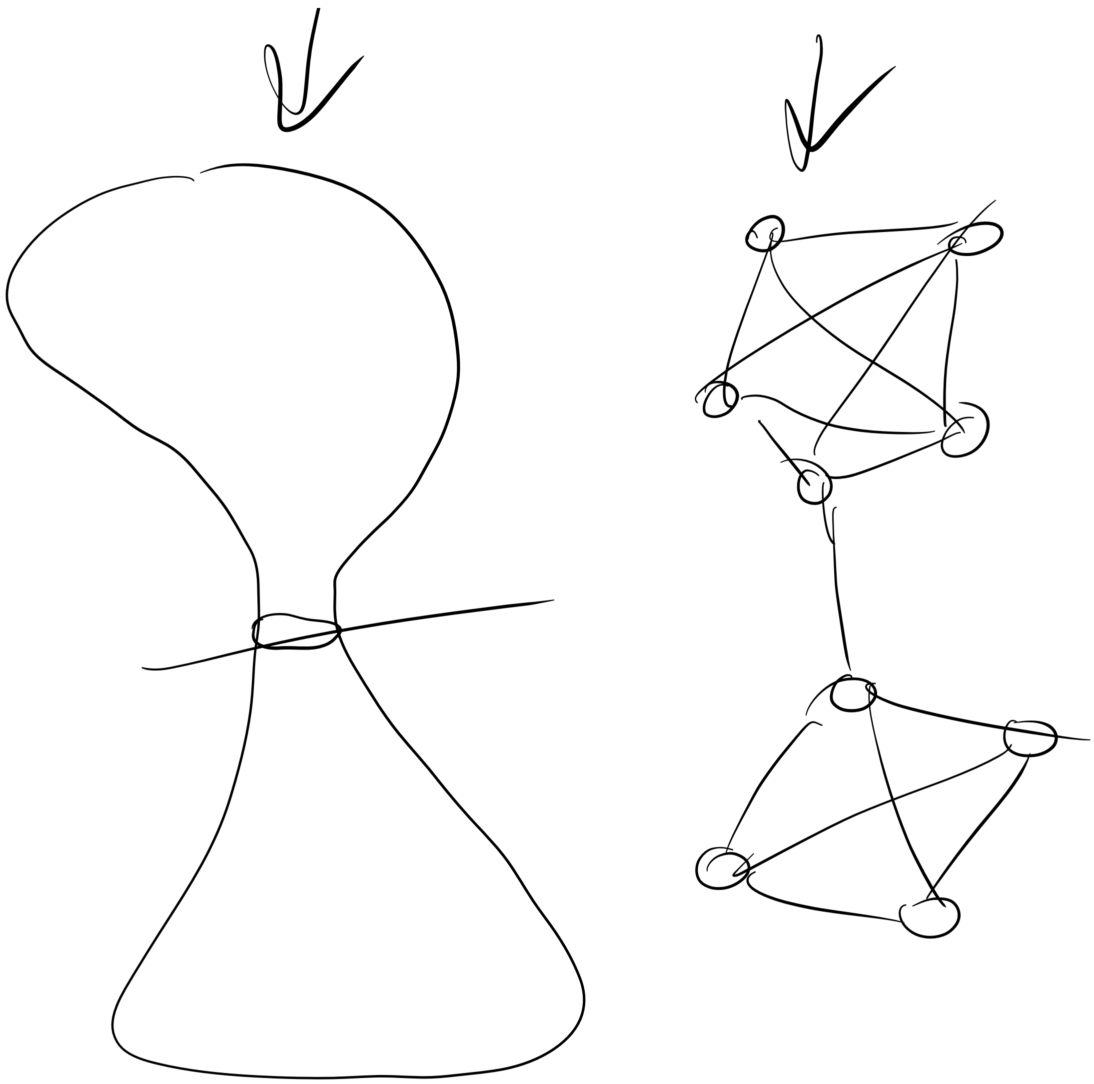
Let λ_2 be the second smallest eigenvalue of $\mathcal{L} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ for a connected graph. Then

$$\frac{\lambda_2}{2} \leq \boxed{\phi(S^*)} \leq \boxed{\sqrt{2\lambda_2}}$$

where S^* is <u>the minimum conductance set.</u>

$$S^* = \operatorname*{argmin}_{S \subseteq V} \frac{\mathrm{cut}(S)}{\min\{\mathrm{vol}(S), \mathrm{vol}(\bar{S})\}}$$

/



Proof of $\frac{\lambda_2}{2} \le \phi(S^*)$: is HW.

This follows using the strategy outlined for showing that

$$\boxed{\lambda_2(L)} \le \boxed{\min_{S \subseteq V} \frac{cut(S)}{|S|} + \frac{cut(S)}{|\bar{S}|}}$$

$$\le \boxed{\min_{\substack{S \subseteq V \\ |S| \le \frac{|V|}{2}}} \frac{2\,\mathrm{cut}(S)}{|S|}}$$

The harder direction is showing that $\phi(S^*) \le \sqrt{2\lambda_2}$

We will prove this by explicitly constructing a set S such that

$$\phi(S^*) \le \phi(S) \le \sqrt{2\lambda_2}$$

We will use the following spectral clustering algorithm

Alg: Spectral Clustering

Input: Graph $G=(V,E)$ (simple)

Output: A set $S \subseteq V$ with $\phi(S) \leq \sqrt{2\lambda_2}$

- Find $$y = \underset{y^T d = 0}{\operatorname{argmin}} \frac{y^T L y}{y^T D y}$$

- Sort $V = \{v_1, v_2, \ldots, v_n\}$ so that

$$y_{v_1} \leq y_{v_2} \leq \cdots \leq y_{v_n}$$

- For $i = 1, 2, \ldots, n-1$

$$S_i = \{v_1, v_2, \ldots, v_i\}$$

- output S_j with smallest conductance.

Lemma 1: For $y \in \mathbb{R}^n$ with $-1 \le y_i \le 1$ define

$$\text{supp}(y) = \{ i : y(i) \neq 0 \}$$

There exists $S \subseteq \text{supp}(y)$ such that

$$\boxed{\frac{\text{cut}(S)}{\text{vol}(S)}} \le \sqrt{2R(y)}$$

If we choose y as in the spectral clustering algorithm, then

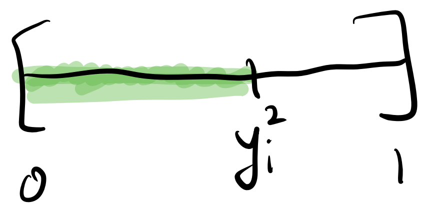
$$\sqrt{2R(y)} = \sqrt{2\lambda_2}.$$

Proof: Generate $t \in (0,1)$ uniformly at random and set

$$S_t = \{ i : y_i^2 \geq t \}$$

$$E[\mathrm{val}(S_t)] = \sum_{i=1}^{n} \underbrace{\Pr[\text{node } i \text{ is in } S_t]}\, d_i$$

$$= \sum_{i=1}^{n} \Pr[y_i^2 \geq t]\, d_i = \sum_{i=1}^{n} y_i^2 d_i$$



Assume WLOG that $(i,j) \in E$ means $y_i^2 \leq y_j^2$.

$$E[\mathrm{cut}(S_t)] = \sum_{(i,j) \in E} \Pr[ij \in \partial S]$$

$$= \sum_{(i,j)\in E} \Pr\left[y_j^2 \geq t \ \& \ y_i^2 < t\right]$$

$$= \sum_{(i,j)\in E} \left(y_j^2 - y_i^2\right)$$

$$= \sum_{ij\in E} (y_j - y_i)(y_j + y_i)$$

$$\leq \sqrt{\sum_{ij\in E} (y_j - y_i)^2 \sum_{ij\in E} (y_j + y_i)^2}$$

since $a^T b \leq \|a\|_2 \|b\|_2$

Cauchy-Schwartz

$$\leq \sqrt{\sum_{ij} (y_j - y_i)^2} \cdot \sqrt{2 \sum_{ij\in E} y_j^2 + y_i^2}$$

Since $(a+b)^2 \leq 2a^2 + 2b^2$

$$= \sqrt{\sum_{ij\in E} (y_j - y_i)^2} \sqrt{2 \sum_{i=1}^{n} d_i y_i^2}$$

$$= \sqrt{2 R(y)} \cdot \sum_{i=1}^{n} d_i y_i^2$$

$$R(y) = \frac{\sum (y_i - y_j)^2}{\sum d_j y_i^2}$$

$$\Rightarrow \sqrt{\sum (y_i - y_j)^2} = \sqrt{R(y)} \cdot \sqrt{\sum_{i=1} d_i y_i^2}$$

$$= \underbrace{\sqrt{2R(y)} \cdot E[\mathrm{vol}(S_t)]}$$

So

$$E\left[\mathrm{cut}(S_t) - \sqrt{2R(y)}\ \mathrm{vol}(S_t)\right] \leq 0$$

For some choice of t

$$\mathrm{cut}(S_t) \leq \sqrt{2R(y)}\ \mathrm{vol}(S_t)$$

$$\Rightarrow \quad \frac{\mathrm{cut}(S_t)}{\mathrm{vol}(S_t)} \leq \sqrt{2R(y)}$$