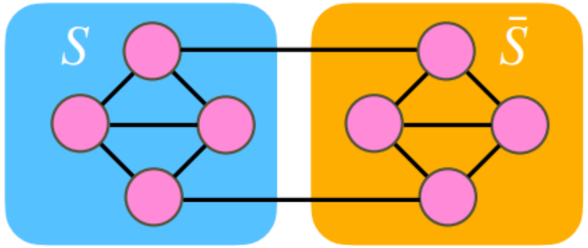
Advanced Algorithms

Spectral methods and algorithms

Jingcheng Liu

Recap



$$\varphi(S) := \frac{|E(S,\bar{S})|}{d|S|}$$

We saw a spectral partitioning algorithm on day 1:
To find a sparse cut with small **conductance** in a d-regular graph, we

1. Compute the second largest eigenvector $\boldsymbol{x} \in \mathbb{R}^n$ of the adjacency matrix.
2. Sort the vertices so that $x_1 \geq x_2 \geq \cdots \geq x_n$.
3. Let $S_i := \begin{cases} \{1, \dots, i\} & \text{if } i \leq \frac{n}{2} \\ \{i+1, \dots, n\} & \text{otherwise} \end{cases}$, and output $S_i = \operatorname{argmin}_{1 \leq i \leq n} \varphi(S_i)$.

Theorem: $\exists i, \varphi(S_i) \leq 2\sqrt{\varphi(G)}$

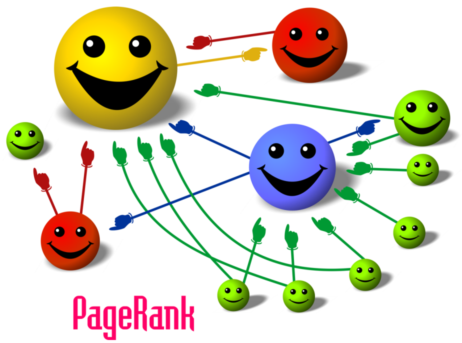
Why eigenvectors?

Overview

Analysis of the spectral partitioning algorithm

- Introduction to spectral graph theory
    - Connectedness
    - Bipartiteness (2-coloring)
- Cheeger’s inequality on d-regular graphs
    - Easy direction: a sparse cut implies λ_2 is small
    - Hard direction: a small λ_2 means we can find a sparse cut from v_2
    - Improvements of Cheeger’s
    - Generalizations of Cheeger’s

Spectral graph theory



Spectral theory

eigenvalues + eigenvectors + related linear algebra

Graph structures

- Connectedness
- Coloring / Clustering
- Mixing of random walks
- Expander graphs

In Theoretical CS

- Pagerank
- Sparsification
- Solving linear systems
- Counting / Sampling
- Expander codes
- Hardness of approximation
- Derandomization
- Max flow and more

And Beyond

- Image segmentation
- Electrical networks
- Reliable / Efficient networks
- Epidemic modelling
- Economic networks

Graphs as matrices

Eigenvalues and eigenvectors

$$Av = \lambda v$$

- λ : **eigenvalue**
- v : **eigenvector**
- **characteristic polynomial** of A: $\det(A - xI)$
- $\det(A - xI) = 0$ gives all the eigenvalues
- **multiplicity** of λ:
  - Geometric: **dimension** of the eigenspace corresponding to λ
  - Algebraic: how many times λ appears as a **root**
  - For ***diagonalizable matrices***, they are the same

Undirected graph $G = (V, E)$ has **adjacency matrix**

$$A_{u,v} = 1 \text{ iff } uv \in E$$

A is an $n \times n$ real symmetric matrix:

- It has **real eigenvalues** $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$
- there is an **orthonormal basis** of **eigenvectors** $v_1, v_2, \ldots, v_n$ such that

$$Av_i = \alpha_i v_i, \forall i$$

$$v_i^{\mathsf{T}} v_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise} \end{cases}$$

Adjacency matrix is NOT only a data structure
Its algebraic properties as a matrix are useful too:
rank, determinant, eigenspaces, …

Complexity of Linear algebra

All the following can be solved in $\tilde{O}(n^{\omega})$ arithmetic operations:

- Matrix multiplication
- Matrix inverse
- Determinant
- Characteristic polynomial
- Solving linear equations $Ax = b$
- Singular value decomposition
- Eigen-decomposition of symmetric matrices

In fact, almost linear time (in theory) for matrices that we will care about..

Spectrum of the adjacency matrix

Let $G = (V, E)$ be an undirected graph, α_1 be the largest eigenvalue of the adjacency matrix $A(G)$

Claim: $d_{\text{avg}} \leq \alpha_1 \leq d_{\max}$

Proof of the upperbound:

Let v be the eigenvector corresponding to α_1, so that $Av = \alpha_1 v$

Without loss of generality we can assume that $\max_i v_i > 0$

Choose an index j so that $v_j = \max_i v_i$

Then $Av = \alpha_1 v$ in the j-th row means that

$$\alpha_1 v_j = \sum_i A_{ji}\, v_i \leq d_{\max} \cdot v_j$$

Here the inequality follows from $v_j = \max_i v_i$, and there are at most $d_{\max}$ neighbors of j

Since $v_j > 0$, $\alpha_1 v_j \leq d_{\max} \cdot v_j \Rightarrow \ \alpha_1 \leq d_{\max}$

Remark. This argument can be adapted to prove: for a connected G, $\alpha_1 = d_{\max}$ iff G is regular

Spectrum of Bipartite graphs

Spectrum also tells us something about graph coloring

We start with 2-colorability (Bipartiteness)

Let $G = (V, E)$ be an undirected graph, and $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$ be its eigenvalues

Claim: The spectrum of $A(G)$ is symmetric about 0 (i.e., $\alpha_i = -\alpha_{n-i+1}$) iff G is bipartite

Spectrum of Bipartite graphs

$$A(G) = \begin{array}{c c} & \begin{array}{cc} U & V \end{array} \\ \begin{array}{c} U \\ V \end{array} & \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \end{array}$$

Lemma: Let G be bipartite, and α be an eigenvalue of $A(G)$ with multiplicity k, then $-\alpha$ is also an eigenvalue of $A(G)$ with multiplicity k

Proof: If $\alpha = 0$, the lemma is vacuously true. So we assume $\alpha \neq 0$.

Let $\binom{x}{y}$ be an eigenvector of A corresponding to α: $\binom{By}{B^T x} = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \binom{x}{y} = \alpha \binom{x}{y}$

So $B^T x = \alpha y, By = \alpha x$. On the other hand, $A \binom{x}{-y} = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \binom{x}{-y} = \binom{-By}{B^T x} = \binom{-\alpha x}{\alpha y} = -\alpha \binom{x}{-y}$

This means $-\alpha$ is also an eigenvalue of A

Finally, notice that the multiplicity of α being $k \Leftrightarrow$ there exists k linearly independent eigenvectors corresponding to α

Apply the above argument to every one of those, we get that the multiplicity of $-\alpha$ is also k

Spectrum of Bipartite graphs

$$A(G) = \begin{array}{c c} & \begin{array}{cc} U & V \end{array} \\ \begin{array}{c} U \\ V \end{array} & \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \end{array}$$

Lemma: If the spectrum of $A(G)$ is symmetric about 0 (i.e., $\alpha_i = -\alpha_{n-i+1}$), then G is bipartite

Proof: Note that for every odd integer k, $\sum_i \alpha_i^k = 0$

Since the eigenvalues of A^k is $\alpha_1^k, \alpha_2^k, \dots, \alpha_n^k$, thus for every odd integer k,

$$\text{trace}(A^k) = \sum_i \alpha_i^k = 0$$

On the other hand, $\text{trace}(A^k)$ has a combinatorial meaning:

$$(A^k)_{i,j} = \text{the number of } k\text{—walks going from } i \text{ to } j$$

Since $\text{trace}(A^k) = \sum_i (A^k)_{i,i} = 0$, and $(A^k)_{i,i} \geq 0$, so we must have $(A^k)_{i,i} = 0$

This means: for every odd integer k, there is no cycle of length k. Thus, all cycles are of even length.

Side note: Graph Coloring

For k-coloring, we do not expect a spectral characterization (why?)

For an approximation, the chromatic number $\chi(G)$ satisfies

$$\left\lceil \frac{\alpha_1}{-\alpha_n} \right\rceil + 1 \leq \chi(G) \leq \lfloor \alpha_1 \rfloor + 1$$

The upperbound is known as Wilf's Theorem, and the lowerbound as Hoffman's bound

See Dan Spielman's Spectral Graph Theory book for a proof

Many matrices associated with a graph

- **Adjacency matrix** $A(G)$

$$A_{u,v} = 1 \text{ iff } uv \in E$$

- **Laplacian matrix**: let $D(G)$ be the diagonal degree matrix

$$L(G) \coloneqq D(G) - A(G)$$

Later in class:

- **Normalized Laplacian matrix**: $\mathcal{L}(G) \coloneqq D^{-\frac{1}{2}} L(G)\, D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$
- Random walk matrix
  - Consider $\overrightarrow{p_{t+1}} = \overrightarrow{p_t}(D^{-1}A)$
  - The **transition matrix** $P \coloneqq D^{-1}A$

Laplacian matrix $L(G) \coloneqq D(G) - A(G)$

For regular graphs, $L(G) = dI - A(G)$, eigenspace is roughly the same as $A(G)$

This is not true for irregular graphs, and the difference is important

$$L_{ij} = \begin{cases} \deg(i), & \text{if } i = j \\ -1, & \text{if } ij \in E \\ 0, & \text{otherwise} \end{cases}$$

Consider the Laplacian on a single edge $e = (u, v)$, $L_e = b_e b_e^\top$

$$L_e = \begin{array}{cc} \begin{array}{ccccc} & u & & v & \end{array} & \\ \begin{pmatrix} & \vdots & & \vdots & \\ \cdots & 1 & \cdots & -1 & \cdots \\ & \vdots & & \vdots & \\ \cdots & -1 & \cdots & 1 & \cdots \\ & \vdots & & \vdots & \end{pmatrix} & \begin{array}{c} \\ u \\ \\ v \\ \\ \end{array} \end{array}$$

Decomposition of Laplacian

$$L_e = \begin{array}{cc} \begin{array}{ccccc} & u & & v & \end{array} & \\ \begin{pmatrix} & \vdots & & \vdots & \\ \cdots & 1 & \cdots & -1 & \cdots \\ & \vdots & & \vdots & \\ \cdots & -1 & \cdots & 1 & \cdots \\ & \vdots & & \vdots & \end{pmatrix} & \begin{array}{c} \\ u \\ \\ v \\ \\ \end{array} \end{array}$$

$$L(G) := D(G) - A(G) = \sum_{e \in E(G)} L_e = \sum_{e \in E(G)} b_e b_e^\top$$

Theorem: $\vec{1}$ is an eigenvector of L with eigenvalue 0

Proof: Notice that each row of L sum up to 0, so $L\vec{1} = 0$

Theorem: The smallest eigenvalue of L is 0

Proof: Note that for every x,

$$x^\top L\, x = \sum_e x^\top b_e b_e^\top\, x = \sum_e (x_u - x_v)^2 \geq 0$$

Thus L is a positive semi-definite (PSD) matrix, with all eigenvalues non-negative.
We also saw that 0 is an eigenvalue, this concludes the proof.

PSD often simply written as $L \succcurlyeq 0$

λ_2 of the Laplacian

$$L(G) := D(G) - A(G) = \sum_{e \in E(G)} b_e b_e^\top$$

Theorem: The second smallest eigenvalue of $L(G)$ is 0 iff G is disconnected

Proof: Suppose that is G disconnected, with components $G = G_1 \uplus G_2$

$$L(G) = \begin{array}{cc} & \begin{array}{cc} V_1 & V_2 \end{array} \\ \begin{array}{c} V_1 \\ V_2 \end{array} & \begin{bmatrix} L(G_1) & 0 \\ 0 & L(G_2) \end{bmatrix} \end{array}$$

$\vec{1}_{G_1}, \vec{1}_{G_2}$ are eigenvectors with eigenvalue 0, and are linearly independent

Conversely, if G is connected, and let $x \neq 0$ be any vector such that $L\,x = 0$

$$x^\top L\,x = \sum_e (x_u - x_v)^2 = 0 \;\Rightarrow\; \forall uv \in E, x_u = x_v$$

Since G is connected, $\forall uv \in E, x_u = x_v \Rightarrow\; \forall u \in V, v \in V, x_u = x_v \Rightarrow\; x = c\vec{1}$

This argument can be adapted to prove:
$\lambda_k(L) = 0$ iff G has k connected components

Spectrum of the Laplacian

$$L(G) := D(G) - A(G) = \sum_{e \in E(G)} b_e b_e^\top$$

Denote eigenvalues of the Laplacian by $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$

Corollary: $\lambda_2(L) > 0$ iff G is connected

Robust generalizations:

$\lambda_2(L)$ is small $\iff$ G is "almost disconnected"

$\lambda_k(L)$ is small $\iff$ G is "close to having k disconnected components"

$\alpha_1 \approx -\alpha_n$ $\iff$ G has an "almost bipartite component"

Intuition behind spectral algorithms for finding
- sparse cuts
- k-way cuts
- Maximum cuts

Recap: Graph conductance

We first define what it means to be “almost disconnected”

The **conductance** of a set $S \subseteq V$ is defined as $\varphi(S) := \frac{|E(S,\bar{S})|}{\mathrm{vol}(S)}$, where $\mathrm{vol}(S) := \sum_{v \in S} \deg(v)$

When the graph is d-regular, $\varphi(S) := \frac{|E(S,\bar{S})|}{d|S|}$

Note: the **expansion** of a set S is defined as $\frac{|E(S,\bar{S})|}{|S|}$

For d-regular graphs, they’re basically the same.

The conductance of a graph G is defined as $\varphi(G) := \min_{S:\mathrm{vol}(S) \leq m} \varphi(S)$

Note that $0 \leq \varphi(G) \leq 1$

Recap: Expander graphs and sparse cuts

A graph G with constant $\phi(G)$ (e.g. $\phi(G) = 0.1$) is called an **expander graph**

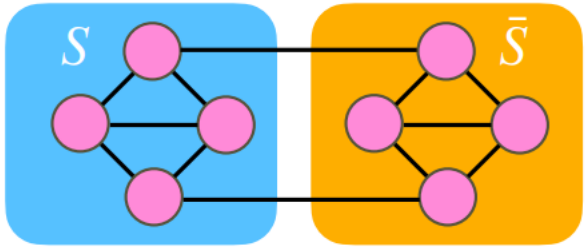
A set S with small $\phi(S)$ is called a **sparse cut**

Both concepts are very useful

Finding a sparse cut is useful in designing divide-and-conquer algorithms, and have applications in

- image segmentation
- data clustering
- community detection
- VLSI-design
- ….

Recap: Spectral partitioning



$$\varphi(S) := \frac{|E(S,\bar{S})|}{d|S|}$$

To find a sparse cut with small **conductance** in a **general graph**, we

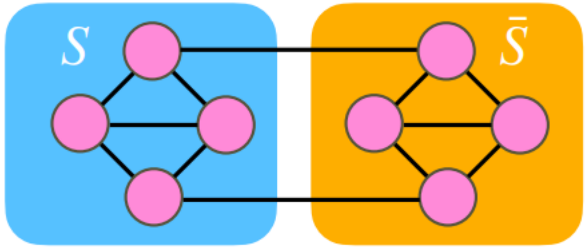
1. Compute the second smallest eigenvector $\boldsymbol{x} \in \mathbb{R}^n$ of the **normalized Laplacian** $\mathcal{L}(G) := D^{-\frac{1}{2}} L(G)\, D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$

2. Sort the vertices so that $x_1 \geq x_2 \geq \cdots \geq x_n$.

3. Let $S_i := \begin{cases} \{1, \dots, i\} & \text{if } i \leq \frac{n}{2} \\ \{i+1, \dots, n\} & \text{otherwise} \end{cases}$, and output $S_i = \operatorname{argmin}_{1 \leq i \leq n} \varphi(S_i)$.

Intuition: $\varphi(G) \approx \lambda_2(\mathcal{L})$

Recap: Spectral partitioning



$$\varphi(S) := \frac{|E(S,\bar{S})|}{d|S|}$$

To find a sparse cut with small **conductance** in a d-**regular graph**, we

1. Compute the second smallest eigenvector $\boldsymbol{x} \in \mathbb{R}^n$ of the **normalized Laplacian** $\frac{L}{d}$
2. Sort the vertices so that $x_1 \geq x_2 \geq \cdots \geq x_n$.
3. Let $S_i := \begin{cases} \{1, \dots, i\} & \text{if } i \leq \frac{n}{2} \\ \{i+1, \dots, n\} & \text{otherwise} \end{cases}$, and output $S_i = \underset{1 \leq i \leq n}{\operatorname{argmin}}\, \varphi(S_i)$.

Intuition: $\varphi(G) \approx \lambda_2(\mathcal{L}) = \lambda_2(L/d)$

Courant-Fischer Theorem

Theorem: For a real symmetric matrix A, the maximum eigenvalue

$$\lambda_n(A) = max_{x \neq 0} \frac{x^\top A\, x}{x^\top x} \longrightarrow$$

Rayleigh quotient

$$R_A(x) = \frac{x^\top A\, x}{x^\top x}$$

Proof: Since equality can be attained. It suffices to show $\frac{x^\top A\, x}{x^\top x} \leq \lambda_n(A)$

Let $v_1, v_2, .., v_n$ be an orthonormal basis of A

$$x^\top A x = (a_1 v_1 + \cdots + a_n v_n)^\top A (a_1 v_1 + \cdots + a_n v_n)$$
$$= \lambda_1 a_1^2 + \cdots + \lambda_n a_n^2 \leq \lambda_n (a_1^2 + \cdots + a_n^2)$$

$$x^\top x = (a_1 v_1 + \cdots + a_n v_n)^\top (a_1 v_1 + \cdots + a_n v_n) = a_1^2 + \cdots + a_n^2$$

Thus, we have $\frac{x^\top A\, x}{x^\top x} \leq \lambda_n$

Courant-Fischer Theorem

For a real symmetric matrix A, the maximum eigenvalue:

$$\lambda_n(A) = max_{x\neq 0} \frac{x^\top A\, x}{x^\top x}$$

The smallest eigenvalue:

$$\lambda_1(A) = min_{x\neq 0} \frac{x^\top A\, x}{x^\top x}$$

More generally,

$$\lambda_k(A) = min_{x\neq 0, x^\top v_i = 0, \forall i \in \{1,\ldots,k-1\}} \frac{x^\top A\, x}{x^\top x}$$

$$\lambda_k(A) = max_{x\neq 0, x^\top v_i = 0, \forall i \in \{k+1,\ldots,n\}} \frac{x^\top A\, x}{x^\top x}$$

Cheeger's inequality

In a d**-regular graph,** the **normalized Laplacian** $\mathcal{L} = \frac{L}{d}$

Cheeger's Inequality [Cheeger 70, Alon-Milman 85]

$$\frac{\lambda_2(\mathcal{L})}{2} \leq \varphi(G) \leq \sqrt{2\lambda_2(\mathcal{L})}$$

The first inequality is called the easy direction, and the second is called the hard direction
We start with some **intuition** in the case when G is a d-regular graph.

For the easy direction: think of λ_2 as a "relaxation" of the graph conductance problem.

$$\varphi(G) = \min_{x \in \{0,1\}^n, |x| \leq \frac{n}{2}} \frac{\sum_{ij \in E} (x_i - x_j)^2}{d \sum_{i \in V} x_i^2} \qquad \text{and} \qquad \lambda_2(\mathcal{L}) = \min_{x \perp 1} \frac{\sum_{ij \in E} (x_i - x_j)^2}{d \sum_{i \in V} x_i^2}.$$

Question: What does the second eigenvector $\boldsymbol{x}$ look like when the graph G is disconnected, i.e., $G = G_1 \uplus G_2$?

Easy direction

Think of λ_2 as a “relaxation” of the graph conductance problem.

$$\varphi(G) \approx \min_{x \perp 1\,:\, x \text{ is binary}} \frac{\sum_{ij \in E} \left(x_i - x_j\right)^2}{d \sum_{i \in V} x_i^2} \qquad \text{and} \qquad \lambda_2(\mathcal{L}) = \min_{x \perp 1} \frac{\sum_{ij \in E} \left(x_i - x_j\right)^2}{d \sum_{i \in V} x_i^2}.$$

Proof: Given a set S with $\varphi(S) = \varphi(G)$, we try to find $x \perp 1$ with $R_{\mathcal{L}}(x) \leq 2\varphi(S)$

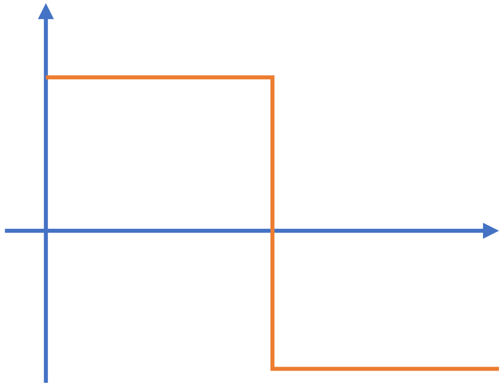
- Consider $x_i = \begin{cases} \frac{1}{|S|}, & \text{if } i \in S \\ -\frac{1}{n-|S|}, & \text{otherwise} \end{cases}$
- Then $\lambda_2(\mathcal{L}) \leq R_{\mathcal{L}}(x) = \frac{\sum_{ij \in E}\left(x_i - x_j\right)^2}{d \sum_{i \in V} x_i^2} = \frac{E(S,\bar{S})}{d}\left(\frac{1}{|S|} + \frac{1}{n-|S|}\right) \leq 2\varphi(S)$

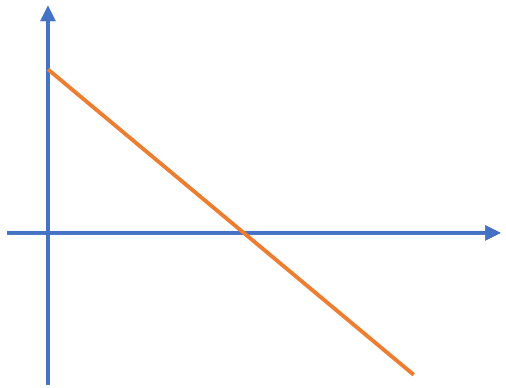
Hard direction

In the hard direction, we are given the second eigenvector x, which has small Rayleigh quotient, and we need to find a binary vector x'

$$\varphi(G) \approx \min_{x \perp 1 \,:\, x \text{ is binary}} \frac{\sum_{ij \in E} (x_i - x_j)^2}{d \sum_{i \in V} x_i^2} \qquad \text{and} \qquad \lambda_2 = \min_{x \perp 1} \frac{\sum_{ij \in E} (x_i - x_j)^2}{d \sum_{i \in V} x_i^2}$$

To gain some intuition, consider sorting x





Hard direction

WLOG, assume the number of positive entries in x is at most the number of negative entries.

Zero out the negative entries of x to obtain y.

Working with y, we ensure that the output set S satisfies $|S| \leq n/2$.

Lemma. $R(y) \leq R(x) = \lambda_2$

Proof: Consider a row i with $y_i > 0$. We have

$$(Ly)_i = \deg(i) - \sum_{j\sim i} y_j \leq \deg(i) - \sum_{j\sim i} x_j = (Lx)_i = \lambda_2 x_i$$

Then $y^\top L\, y = \sum_i y_i (Ly)_i \leq \sum_{i:x_i>0} \lambda_2 x_i^2 = \lambda_2 \sum_i y_i^2$

Hard direction

$\mathrm{supp}(y) \coloneqq \{i \mid y(i) \neq 0\}$.

Claim. Given any y, there exists a subset $S \subseteq \mathrm{supp}(y)$ such that

$$\varphi(S) \leq \sqrt{2R(y)} \leq \sqrt{2\lambda_2}$$

Proof plan. By scaling, assume that $\max_i y(i) = 1$.

For $0 < t \leq 1$, we consider a "threshold set" $S_t \coloneqq \{i \mid y(i)^2 \geq t\}$.

We want to prove that there exists a t such that $\varphi(S_t) \leq \sqrt{2R(y)}$.

The **idea** is to choose t uniformly randomly from $(0,1)$!

We will show that $\frac{\mathbb{E}_t[|E(S_t,\bar{S}_t)|]}{\mathbb{E}_t[d|S_t|]} \leq \sqrt{2R(y)}$.

This would imply that there exists t such that $\frac{|E(S_t,\bar{S}_t)|}{d|S_t|} \leq \sqrt{2R(y)}$, as desired.

Hard direction

$\mathrm{supp}(y) \coloneqq \{i \mid y(i) \neq 0\}.$

Claim. Given any y, there exists a subset $S \subseteq \mathrm{supp}(y)$ such that

$$\varphi(S) \leq \sqrt{2R(y)} \leq \sqrt{2\lambda_2}$$

Proof. Choose a random "threshold set" $S_t \coloneqq \{i \mid y(i)^2 \geq t\}$

We will show that $\frac{\mathbb{E}_t[|E(S_t,\bar{S}_t)|]}{\mathbb{E}_t[d|S_t|]} \leq \sqrt{2R(y)}$

Let's first calculate

$$\mathbb{E}_t[d|S_t|] = d\sum_i \Pr[i \in S_t] = d\sum_i \Pr[t \leq y(i)^2] = d\sum_i y(i)^2$$

Hard direction

Cauchy-Schwarz inequality:
$$\langle u, v\rangle \le \sqrt{\langle u, u\rangle} \cdot \sqrt{\langle v, v\rangle}$$

Proof (cont.) Choose a random "threshold set" $S_t \coloneqq \{i \mid y(i)^2 \ge t\}$.

It remains to show that $\mathbb{E}_t[|E(S_t, \bar{S}_t)|] \le \sqrt{2R(y)} \cdot \left(d \sum_{i\in V} y_i^2\right)$, or equivalently, we want to show

$$\mathbb{E}_t[|E(S_t, \bar{S}_t)|] \le \sqrt{2 \sum_{ij\in E} (y_i - y_j)^2 \cdot d \sum_{i\in V} y_i^2}$$

$$\mathbb{E}_t[|E(S_t, \bar{S}_t)|] = \sum_{ij\in E} \Pr[i \in S_t, j \notin S_t] + \Pr[i \notin S_t, j \in S_t] = \sum_{ij\in E} |y_i^2 - y_j^2| = \sum_{ij\in E} |y_i - y_j| \cdot (y_i + y_j)$$

Apply Cauchy-Schwarz inequality:

$$\sum_{ij\in E} |y_i - y_j| \cdot (y_i + y_j) \le \sqrt{\sum_{ij\in E} |y_i - y_j|^2} \cdot \sqrt{\sum_{ij\in E} (y_i + y_j)^2} \le \sqrt{\sum_{ij\in E} |y_i - y_j|^2} \sqrt{2 \sum_{ij\in E} y_i^2 + y_j^2} = \sqrt{\sum_{ij\in E} |y_i - y_j|^2} \sqrt{2d \sum_{i\in V} y_i^2}$$

Combined, this concludes the proof of $\frac{\mathbb{E}_t[|E(S_t,\bar{S}_t)|]}{\mathbb{E}_t[d|S_t|]} \le \sqrt{2R(y)}$. Then we notice that this implies

$$\mathbb{E}_t\left[|E(S_t, \bar{S}_t)| - \sqrt{2R(y)}d|S_t|\right] \le 0$$

This means that there must be a choice of t, $|E(S_t, \bar{S}_t)| - \sqrt{2R(y)}d|S_t| \le 0 \Rightarrow \varphi(S_t) \le \sqrt{2R(y)}$

Hard direction summary

Easy direction is to show that $\lambda_2 = \min_{x \in \mathbb{R}^n} \frac{\sum_{ij \in E} (x_i - x_j)^2}{d \sum_{i \in V} x_i^2}$ is a “relaxation” of $\varphi(G) \approx \min_{x:x \text{ "binary"}} \frac{\sum_{ij \in E} (x_i - x_j)^2}{d \sum_{i \in V} x_i^2}$.

For the hard direction, given an optimizer x for λ_2, we want to produce a set S with $\varphi(S) \leq \sqrt{2\lambda_2}$.

The idea is simply to try all “threshold” sets of x.

We truncate the vector x to obtain y to ensure that the output set is of size at most $n/2$.

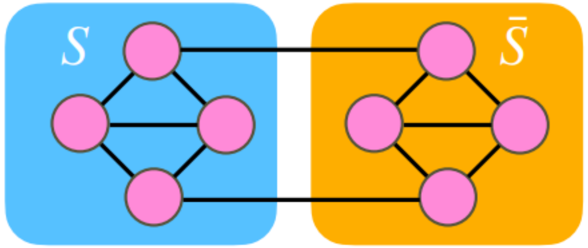
In the analysis, we choose a <u>random</u> threshold for y and prove that $\frac{\mathbb{E}_t[|E(S_t, \bar{S}_t)|]}{\mathbb{E}_t[d|S_t|]} \leq \sqrt{2R(y)}$.

In general, this is called a “rounding” algorithm, where we turn a “fractional” solution to an integral solution.

This is the most common way to design approximation algorithms for NP-hard optimization problems.

Today we see an example of “randomized rounding”, a useful technique in rounding algorithms.

Summary: Spectral partitioning



$\varphi(S) := \frac{|E(S,\bar{S})|}{d|S|}$

To find a sparse cut with small **conductance** in a **general graph**, we

1. Compute the second smallest eigenvector $\boldsymbol{x} \in \mathbb{R}^n$ of the **normalized Laplacian** $\mathcal{L}(G) := D^{-\frac{1}{2}} L(G)\, D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$
2. Sort the vertices so that $x_1 \geq x_2 \geq \cdots \geq x_n$.
3. Let $S_i := \begin{cases} \{1, \dots, i\} & \text{if } i \leq \frac{n}{2} \\ \{i+1, \dots, n\} & \text{otherwise} \end{cases}$, and output $S_i = \operatorname{argmin}_{1 \leq i \leq n} \varphi(S_i)$.

Theorem: $\exists i, \varphi(S_i) \leq 2\sqrt{\varphi(G)}$

Aside: Spectral Partitioning in Planar Graphs

Planar graph separator theorem: The removal of $O(\sqrt{n})$ vertices partitions the planar graph into disjoint subgraphs, each of which has at most $2n/3$ vertices

Theorem (Spielman-Teng'07). For bounded degree planar graphs, a recursive spectral partitioning finds a separator of size $O(\sqrt{n})$

Recent Generalizations

Previously, spectral graph theory is mostly about the second eigenvalue.

In the past decade, there are a few interesting generalizations of Cheeger's inequality using other eigenvalues!

We will discuss some of them.

Last Eigenvalue

$\mathcal{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ is the **normalized adjacency matrix**

Homework. $\alpha_1(A) = -\alpha_n(A)$ iff G is bipartite, and $\alpha_n(\mathcal{A}) = -1$ iff G is bipartite.

Let α_n be the smallest eigenvalue of $I + \mathcal{A}$. Then homework implies that $\alpha_n = 0$ iff G is bipartite.

Exercise. $\alpha_n = \min_{x \in \mathbb{R}^n} \frac{\sum_{ij \in E} (x_i + x_j)^2}{d \sum_{i \in V} x_i^2}$ for d-regular graphs.

Define $\beta(G) = \min_{x \in \{-1,0,+1\}^n} \frac{\sum_{ij \in E} |x_i + x_j|}{d \sum_{i \in V} |x_i|}$. (By the way, we can also think of $\varphi(G)$ this way.)

This is called the bipartiteness ratio of G.

Theorem. [Trevisan 09] $\frac{1}{2}\alpha_n \le \beta(G) \le \sqrt{2\alpha_n}$.

Proof idea. Pick a random $t \in [0,1]$.

A vertex i gets -1 if $x_i^2 \ge t$ and $x_i \le 0$, gets $+1$ if $x_i^2 \ge t$ and $x_i \ge 0$, and gets 0 otherwise.

This result is used to design a spectral algorithm for approximating maximum cut of a graph.

k-th Eigenvalue

Homework. $\lambda_k(\mathcal{L}) = 0$ if and only if G has at least k components.

There are two interesting ways to generalize this statement.

- If λ_k is small, then there is a sparse cut S with $|S| \lesssim \frac{n}{k}$.

- If λ_k is small, then there are k disjoint sparse cuts.

The second result is more general, but the first result is quantitatively stronger.

[Arora,Barak,Steurer,10] proved that when k is large enough, there is a set S with $\varphi(S) \lesssim \sqrt{\lambda_k}$ and $|S| \approx \frac{n}{k}$.

The proof uses ideas about random walks.

The algorithm is used for solving “unique games”.

Small set expansion, local graph partitioning

Note that the Cheeger rounding works with any vector with small Rayleigh quotient

One could try to run a random walk to find such vectors, this will be efficient in both time and space

Further, if we only care about finding a small sparse cut (e.g., a small community), the algorithm has a running time that only depends on the output size

The question of finding small set expansion is closely related to the Unique Games problem

Higher Order Cheeger's Inequality

Theorem. [Lee, Oveis-Gharan, Trevisan 12] [Louis, Raghavendra, Tetali, Vempala 12]

$$\frac{\lambda_k}{2} \leq \varphi_k(G) \leq O(k^2 \cdot \text{polylog}(k))\sqrt{\lambda_k}, \qquad \text{where } \varphi_k(G) := \min_{\text{disjoint } S_1,\ldots,S_k} \max_{1 \leq i \leq k} \phi(S_i).$$

Furthermore, $\varphi_k(G) \leq O\left(\sqrt{\ln k}\right) \cdot \sqrt{\lambda_{1.01k}}$.

The proof is by a spectral embedding, where each vertex is mapped to a point in $\mathbb{R}^k$ using the k eigenvectors.

The vectors are orthonormal, so the points are "well spread out".

The algorithm by [LRTV] is very simple:

(1) Generate k random directions. (2) Put each point to its closest direction. (3) Run Cheeger on each direction.

The algorithm by [LOT] is similar to a clustering heuristic that was proposed in machine learning.

Improved Cheeger's Inequality

Theorem. [Kwok, Lau, Lee, Oveis-Gharan, Trevisan 13] For any $k \geq 2$,

$$\frac{\lambda_2}{2} \leq \varphi(G) \leq O\left(\frac{k\lambda_2}{\sqrt{\lambda_k}}\right).$$

Cheeger's inequality is when $k = 2$.

Performance achieved by the same spectral partitioning algorithm.

Constant factor approximation when λ_k is large for a small k,

which happens in image segmentation when there are only few outstanding objects in an image.

Tight up to a constant factor for any k.

The proof is by showing that if λ_k is large, then the second eigenvector looks like a k-step function.

See Chapter 5.4 of Lap Chi Lau's book for more discussions.

What next

Random walks on undirected graphs

- Fundamental theorem of Markov chains
- Spectral analysis
- Mixing time
- Random sampling
- Electrical networks