Graph Signal Processing – Part I: Graphs, Graph Spectra, and Spectral Clustering

Qincheng Lu

Ph.D. Student Supervised by Prof. Xiao-Wen CHANG

McGill
School of Computer Science

October 19, 2020

Examples

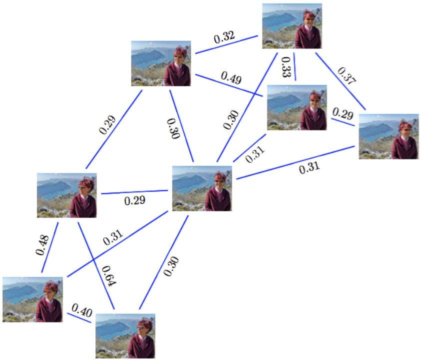


Figure: Images graph

Figure: Minnesota roadmap graph

Graph and Graph Signal

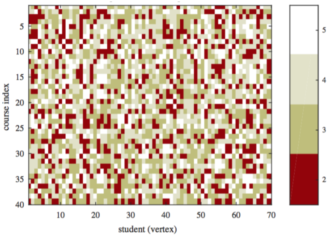


Figure: Marks per student and per course

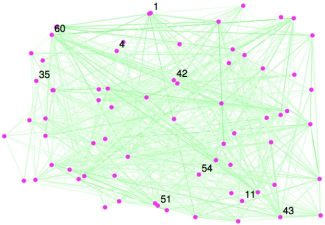


Figure: 2D map with random position

Definition: Graph $\mathcal{G} = \{\mathcal{V}, \mathcal{B}\}$, $\mathcal{B} \subset \mathcal{V} \times \mathcal{V}$
Graph Signal $f \rightarrow \mathbb{R}^N$

Classical Discrete Signal Processing

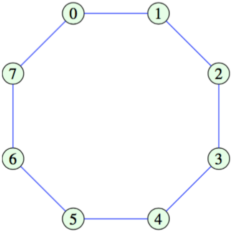


Figure: Time series data

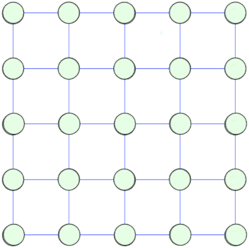


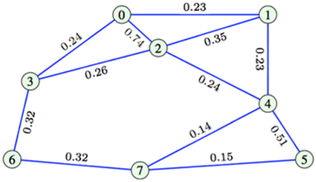
Figure: Digital image data

Adjacency Matrix

1. Adjacency matrix $\mathbf{A}$ for N vertices is an $N \times N$ matrix.
$$A_{mn} \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } (m,n) \in \mathcal{B} \\ 0, & \text{if } (m,n) \notin \mathcal{B} \end{cases}$$
2. For undirected graph, if $(n,m) \in \mathcal{B}$ then also $(m,n) \in \mathcal{B}$, $\mathbf{A} = \mathbf{A}^\intercal$.

$$\mathcal{V} = \{0,1,2,3,4,5,6,7\}$$
$$\mathcal{B} \subset \{0,1,2,3,4,5,6,7\} \times \{0,1,2,3,4,5,6,7\}$$
$$\mathcal{B} = \{(0,1),(1,2),(2,0),(2,3),(2,4),(2,7),(3,0),$$
$$(4,1),(4,2),(4,5),(5,7),(6,3),(6,7),(7,2),(7,6)\}.$$

$$\mathbf{A}_{\text{un}} = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ \\ \end{array} \begin{array}{c} \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \\ \begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \end{array}$$



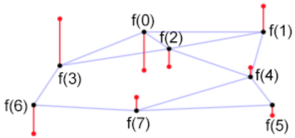
$$\mathbf{W} = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ \\ \end{array} \begin{array}{c} \begin{bmatrix} 0 & 0.23 & 0.74 & 0.24 & 0 & 0 & 0 & 0 \\ 0.23 & 0 & 0.35 & 0 & 0.23 & 0 & 0 & 0 \\ 0.74 & 0.35 & 0 & 0.26 & 0.24 & 0 & 0 & 0 \\ 0.24 & 0 & 0.26 & 0 & 0 & 0 & 0.32 & 0 \\ 0 & 0.23 & 0.24 & 0 & 0 & 0.51 & 0 & 0.14 \\ 0 & 0 & 0 & 0 & 0.51 & 0 & 0 & 0.15 \\ 0 & 0 & 0 & 0.32 & 0 & 0 & 0 & 0.32 \\ 0 & 0 & 0 & 0 & 0.14 & 0.15 & 0.32 & 0 \end{bmatrix} \\ \begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \end{array}$$

3. Adjacency matrix $\mathbf{A}$ is a special case of the weight matrix $\mathbf{W}$

Laplacian Matrix

1. Degree matrix $\mathbf{D}$ is a diagonal matrix, where $D_{mm} \overset{\text{def}}{=} \sum_{n=0}^{N-1} \mathbf{W}_{mn}$
2. Laplacian matrix is defined as $\mathbf{L} \overset{\text{def}}{=} \mathbf{D} - \mathbf{W}$

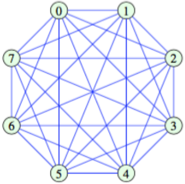
   For undirected graph
   1. Symmetric and positive semidefinite
   2. Off-diagonal entries are non-positive for non-negative weights
   3. Rows sum up to zero
   4. Eigenvalues are non-negative real numbers
   5. Eigenvectors are real and orthogonal
3. Notion of "smoothness":
   $\mathbf{f}^\intercal\mathbf{L}\mathbf{f} = \frac{1}{2}\sum_{i,j=0}^{N-1} \mathbf{W}_{ij}(f(i)-f(j))^2 = \frac{1}{2}\sum_{i,j} \mathbf{W}_{ij}(f(i)^2 - 2f(i)f(j) + f(j)^2) = \sum_i y(i)^2\mathbf{DD}_{ii} + \sum_j y(j)^2\mathbf{DD}_{jj} - 2\sum_{i,j} y(i)y(j)\mathbf{W}_{ij}$

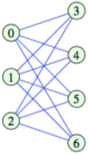


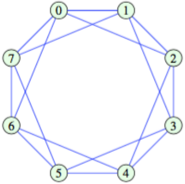
$$\mathbf{L} = \begin{bmatrix} 1.21 & -0.23 & -0.74 & -0.24 & 0 & 0 & 0 & 0 \\ -0.23 & 0.81 & -0.35 & 0 & -0.23 & 0 & 0 & 0 \\ -0.74 & -0.35 & 1.59 & -0.26 & -0.24 & 0 & 0 & 0 \\ -0.24 & 0 & -0.26 & 0.82 & 0 & 0 & -0.32 & 0 \\ 0 & -0.23 & -0.24 & 0 & 1.12 & -0.51 & 0 & -0.14 \\ 0 & 0 & 0 & 0 & -0.51 & 0.66 & 0 & -0.15 \\ 0 & 0 & 0 & -0.32 & 0 & 0 & 0.64 & -0.32 \\ 0 & 0 & 0 & 0 & -0.14 & -0.15 & -0.32 & 0.61 \end{bmatrix}$$

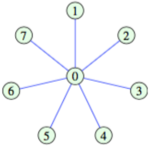
4. Normalized laplacian matrix:
   $\mathbf{L}_N \overset{\text{def}}{=} \mathbf{D}^{-1/2}(\mathbf{D}-\mathbf{W})\mathbf{D}^{-1/2} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$

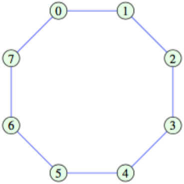
Frequently Used Graph Topologies


(a) Complete graph

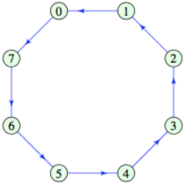

(b) Bipartite graph


(c) Regular graph


(d) Star graph


(e) Circular graph

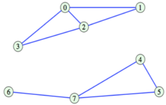

(f) Path graph


(g) Directed circular graph


(h) Directed path graph

Disconnected Graph

1. Adjacency matrix and Laplacian matrix are of block-diagonal form.
2. The multiplicity of the zero eigenvalue of the Laplacian = the number of disjoint components.



$$\mathbf{A} = \begin{array}{cc} \begin{array}{c}0\\1\\2\\3\\4\\5\\6\\7\end{array} & \left[\begin{array}{cccc|cccc} 0&1&1&1&0&0&0&0\\ 1&0&1&0&0&0&0&0\\ 1&1&0&1&0&0&0&0\\ 1&0&1&0&0&0&0&0\\ \hline 0&0&0&0&0&1&0&1\\ 0&0&0&0&1&0&0&1\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&1&1&1&0 \end{array}\right] \\ & \begin{array}{cccccccc}0&1&2&3&4&5&6&7\end{array} \end{array} \qquad \mathbf{L} = \left[\begin{array}{cccc|cccc} 3&-1&-1&-1&0&0&0&0\\ -1&2&-1&0&0&0&0&0\\ -1&-1&3&-1&0&0&0&0\\ -1&0&-1&2&0&0&0&0\\ \hline 0&0&0&0&2&-1&0&-1\\ 0&0&0&0&-1&2&0&-1\\ 0&0&0&0&0&0&1&-1\\ 0&0&0&0&-1&-1&-1&3 \end{array}\right]$$

1. $\mathbf{v_0} = (1, ..., 1, 0, ..., 0)^\intercal$, $\mathbf{v_1} = (0, ..., 0, 1, ..., 1)^\intercal$ are two eigenvectors of eigenvalue $\lambda_0 = 0$ for $\mathbf{L}$

Find a Partition into Two Sets of Vertices $\mathcal{E}$, $\mathcal{H}$

Minimum k-cuts Problem, $k=2$

- Consider an undirected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{B}\}$ with set of edge weights $\mathbf{W}$
- We want to find $\mathcal{E}$ and $\mathcal{H}$ ($\mathcal{E} \subset \mathcal{V}$, $\mathcal{H} \subset \mathcal{V}$, $\mathcal{E} \cup \mathcal{H} = \mathcal{V}$ and $\mathcal{E} \cap \mathcal{H} = \emptyset$)
- Such that cut $\text{Cut}(\mathcal{E}, \mathcal{H}) = \sum_{m \in \mathcal{E}, n \in \mathcal{H}} \mathbf{W}_{mn}$ is minimized.

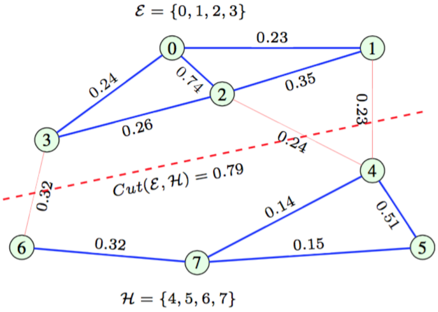


Figure: A cut for a weighted undirected graph

Minimum 2-cuts Problem

- Combinatorial problem: Brute force approach on N vertices takes

$$\binom{N}{1} + \binom{N}{2} + ... + \binom{N}{N/2-1} + \binom{N}{N/2}/2 = O(2^N)$$

- Express partition $(\mathcal{E}, \mathcal{H})$ as a vector $\mathbf{x}$: $\mathbf{x}_i \stackrel{\text{def}}{=} \begin{cases} +1, & \text{if } i \in \mathcal{E} \\ -1, & \text{if } i \in \mathcal{H} \end{cases}$

$$\begin{aligned} \mathbf{x}^\intercal \mathbf{L} \mathbf{x} &= \frac{1}{2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \mathbf{W}_{mn} \left(\mathbf{x}(n) - \mathbf{x}(m)\right)^2 \\ &= 4 \sum_{m \in \mathcal{E}, n \in \mathcal{H}} \mathbf{W}_{mn} \\ &= 4 \text{Cut}(\mathcal{E}, \mathcal{H}) \end{aligned}$$

- Fiedler vector $\mathbf{x} = \arg\min_{\mathbf{y} \in \mathbb{R}^N} \mathbf{y}^\intercal \mathbf{L} \mathbf{y}$

Minimum Normalized 2-cuts Problem

- Normalized (ratio) cut $\mathrm{CutN}(\mathcal{E},\mathcal{H}) = (\frac{1}{N_{\mathcal{E}}} + \frac{1}{N_{\mathcal{H}}}) \sum_{m\in\mathcal{E}, n\in\mathcal{H}} \mathbf{W}_{mn}$
- where $N_{\mathcal{E}}$ and $N_{\mathcal{H}}$ are the respective numbers of vertices in the sets $\mathcal{E}$ and $\mathcal{H}$.
- The normalized indicator $\mathbf{x}$: $\mathbf{x}_i \stackrel{\text{def}}{=} \begin{cases} +1/(N_{\mathcal{E}} e_x), & \text{if } i \in \mathcal{E} \\ -1/(N_{\mathcal{H}} e_x), & \text{if } i \in \mathcal{H} \end{cases}$, $||\mathbf{x}||_2^2 = 1$, $e_x^2 = \frac{1}{N_{\mathcal{E}}} + \frac{1}{N_{\mathcal{H}}}$
- The indicator $\mathbf{x}$ is orthogonal to the eigenvector of $\mathbf{L}$ for $\lambda_0 = 0$

$$\begin{aligned}
\frac{\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}}{\mathbf{x}^\mathsf{T}\mathbf{x}} &= \frac{1}{||\mathbf{x}||_2^2} \frac{1}{2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \mathbf{W}_{mn} \left(\mathbf{x}(n) - \mathbf{x}(m)\right)^2 \\
&= \frac{1}{||\mathbf{x}||_2^2} \frac{1}{e_x^2} \left(\frac{1}{N_{\mathcal{E}}} + \frac{1}{N_{\mathcal{H}}}\right)^2 \sum_{m\in\mathcal{E}, n\in\mathcal{H}} \mathbf{W}_{mn} \\
&= \mathrm{CutN}(\mathcal{E},\mathcal{H})
\end{aligned}$$

Minimum Normalized 2-cuts Problem (Continue)

- $\frac{\mathbf{x}^\intercal \mathbf{L}\mathbf{x}}{\mathbf{x}^\intercal \mathbf{x}} = \mathsf{CutN}(\mathcal{E}, \mathcal{H})$, with indicator $\mathbf{x}$ normalized to unit energy
- $\min\{\mathbf{x}^\intercal \mathbf{L}\mathbf{x}\}$ subject to $\mathbf{x}^\intercal \mathbf{x} = 1$
- $\mathcal{L}(\mathbf{x}) = \mathbf{x}^\intercal \mathbf{L}\mathbf{x} - \lambda(\mathbf{x}^\intercal \mathbf{x} - 1) \Rightarrow \partial \mathcal{L}(\mathbf{x})/\partial \mathbf{x}^\intercal = \mathbf{0} \Rightarrow \mathbf{L}\mathbf{x} = \lambda \mathbf{x}$
- $\mathbf{x}$ is an eigenvector of $\mathcal{L}$
- $\min\{\mathbf{x}^\intercal \mathbf{L}\mathbf{x}\} = \min\{\lambda \mathbf{x}^\intercal \mathbf{x}\} = \min\{\lambda\}$

Section 2

Eigenvalue Decomposition of A

1. $\mathbf{Ax}$ is output after a movement of graph signal $\mathbf{x}$ along walks of length one.
2. The output signal from a system on a graph

$$\mathbf{y} = h_0\mathbf{A}^0\mathbf{x} + h_1\mathbf{A}^1\mathbf{x} + \ldots + h_{M-1}\mathbf{A}^{M-1}\mathbf{x} = \sum_{m=0}^{M-1} h_m\mathbf{A}^m\mathbf{x}$$

3. Given $\mathbf{A} = \mathbf{U\Lambda U}^{-1}$, $\mathbf{A}^m = \mathbf{U\Lambda}^m\mathbf{U}^{-1}$
4. Characteristic polynomial of $\mathbf{A}$

$$\begin{aligned} P(\lambda) = \mathrm{det}|\mathbf{A} - \lambda\mathbf{I}| &= \lambda^N + c_1\lambda^{N-1} + c2\lambda^{N-2} + \ldots + c_N \\ &= (\lambda - \mu_1)^{p_1}(\lambda - \mu_2)^{p_2} \cdots (\lambda - \mu_{N_m})^{p_{N_m}} \\ p_1 + p_2 + \ldots + p_{N_m} &= N, N_m \leq N \end{aligned}$$

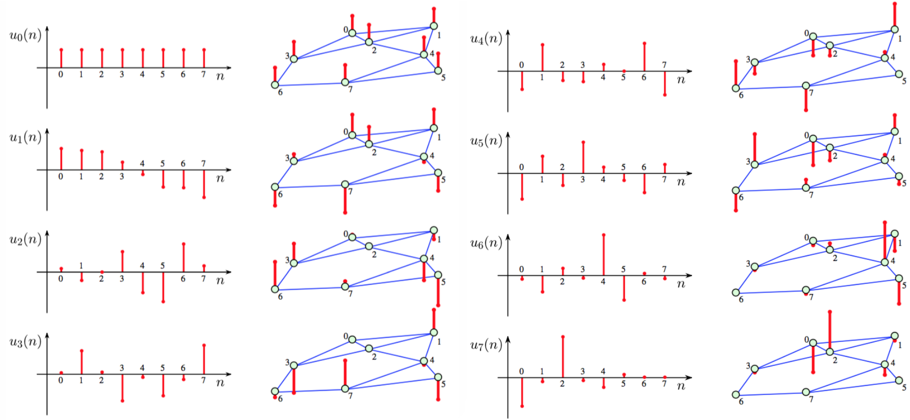
5. The minimal polynomial of $\mathbf{A}$

$$P_{\mathsf{min}}(\lambda) = (\lambda - \mu_1)(\lambda - \mu_2)\cdots(\lambda - \mu_{N_m})$$

Eigenvalue Decomposition of L

1. The set of the eigenvalues of the graph Laplacian $\mathbf{L}$ is called graph spectrum
2. Eigenvalues are usually sorted increasingly: $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{N-1}$
3. If $\lambda_1 \neq 0$, λ_1 is called algebraic connectivity
4. The smoothness of an eigenvector $\mathbf{u}_k$ is $\mathbf{u}_k^\intercal \mathbf{L} \mathbf{u}_k = \lambda_k$

The Smoothness of an Eigenvector



Eigenvalue Decomposition of $\mathbf{L}_N$

The normalized Laplacian $\mathbf{L}_N = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ has eigenvalues $\lambda_i^{(N)}$

$$\mathbf{L}_N(u,v) = \begin{cases} 1, & \text{if } u = v \text{ and } \mathbf{D}_{vv} \neq 0 \\ -\frac{1}{\sqrt{\mathbf{D}_{vv}\mathbf{D}_{uu}}}, & \text{if } u \text{ and } v \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases}$$

- The sum of $\lambda_i^{(N)}$ is equal to N if there are no isolated vertices
- $0 = \lambda_0^{(N)} \leq \lambda_1^{(N)} \leq ... \leq \lambda_{N-1}^{(N)} \leq 2$
- $\lambda_{N-1}^{(N)} = 2$ if and only if the graph is a bipartite graph
- Rayleigh quotient of $\mathbf{L}_N$, where $\mathbf{g} = \mathbf{D}^{1/2}\mathbf{f}$

$$R(\mathbf{L}_N, \mathbf{g}) = \frac{\mathbf{g}^\intercal \mathbf{L}_N \mathbf{g}}{\mathbf{g}^\intercal \mathbf{g}} = \frac{\mathbf{f}^\intercal \mathbf{L}\mathbf{f}}{(\mathbf{D}^{1/2}\mathbf{f})^\intercal(\mathbf{D}^{1/2}\mathbf{f})} = \frac{\sum_{u\sim v}(f(u)-f(v))^2}{\sum_u f(u)^2\mathbf{D}_{uu}}$$

- $\mathbf{D}^{1/2}\mathbb{1}$ is an eigenvector of $\mathbf{L}_N$ with eigenvalue 0

$$\mathbf{f}^\intercal\mathbb{1} = 0 \Rightarrow \mathbf{g}^\intercal\mathbf{D}^{1/2}\mathbb{1} = 0;\ \mathbf{L}\mathbb{1} = \mathbf{0} \Rightarrow \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}(\mathbf{D}^{1/2}\mathbb{1}) = 0$$

- $\inf\limits_{\mathbf{g}\perp\mathbf{D}^{1/2}\mathbb{1}, ||\mathbf{g}||_2^2=1} R(\mathbf{L}_N) = \lambda_1^{(N)}$

Upper Bound of $\lambda^{(N)}$

- $(f(x)-f(y))^2 \leq 2(f(x)^2+f(y)^2) \Rightarrow \frac{\sum_{x\sim y}(f(x)-f(y))^2}{\sum_x f(x)^2 \mathbf{D}_{xx}} \leq 2$
- The equality holds for $f(x) = -f(y)$ for every edge (x,y) in $\mathcal{G} \Rightarrow$ bipartite

  Another approach:
- Let $\mathcal{E}$ and $\mathcal{H}$ be two disjoint sets of bipartite $\mathcal{G}$

$$
\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{W}_{\mathcal{E}\mathcal{H}} \\ \mathbf{W}_{\mathcal{E}\mathcal{H}}^{\intercal} & \mathbf{0} \end{bmatrix}, \; \mathbf{L}_N = \begin{bmatrix} \mathbf{I} & \mathbf{L}_{\mathcal{E}\mathcal{H}} \\ \mathbf{L}_{\mathcal{E}\mathcal{H}}^{\intercal} & \mathbf{I} \end{bmatrix}
$$

$$
\mathbf{u} = \begin{bmatrix} \mathbf{u}_{\mathcal{E}} \\ \mathbf{u}_{\mathcal{H}} \end{bmatrix}, \; \mathbf{L}_N \mathbf{u} = \begin{bmatrix} \mathbf{u}_{\mathcal{E}} + \mathbf{L}_{\mathcal{E}\mathcal{H}} \mathbf{u}_{\mathcal{H}} \\ \mathbf{u}_{\mathcal{H}} + \mathbf{L}_{\mathcal{E}\mathcal{H}}^{\intercal} \mathbf{u}_{\mathcal{E}} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{u}_{\mathcal{E}} \\ \mathbf{u}_{\mathcal{H}} \end{bmatrix}
$$

$$
\Rightarrow \begin{cases} \mathbf{L}_{\mathcal{E}\mathcal{H}} \mathbf{u}_{\mathcal{H}} = (\lambda - 1)\mathbf{u}_{\mathcal{E}} \\ \mathbf{L}_{\mathcal{E}\mathcal{H}}^{\intercal} \mathbf{u}_{\mathcal{E}} = (\lambda - 1)\mathbf{u}_{\mathcal{H}} \end{cases} \Rightarrow \begin{cases} \mathbf{u}_{\mathcal{E}} - \mathbf{L}_{\mathcal{E}\mathcal{H}} \mathbf{u}_{\mathcal{H}} = (2 - \lambda)\mathbf{u}_{\mathcal{E}} \\ \mathbf{L}_{\mathcal{E}\mathcal{H}}^{\intercal} \mathbf{u}_{\mathcal{E}} - \mathbf{u}_{\mathcal{H}} = (\lambda - 2)\mathbf{u}_{\mathcal{H}} \end{cases}
$$

$$
\mathbf{L}_N \begin{bmatrix} \mathbf{u}_{\mathcal{E}} \\ -\mathbf{u}_{\mathcal{H}} \end{bmatrix} = (2-\lambda) \begin{bmatrix} \mathbf{u}_{\mathcal{E}} \\ -\mathbf{u}_{\mathcal{H}} \end{bmatrix}
$$

- For bipartite $\mathcal{G}$, $\lambda_{\text{min}} = 0 \Rightarrow \lambda_{\text{max}} = 2$

Lower Bound of the Rayleigh Quotient of $\mathbf{L}_N$

- Let $\mathbf{u}_i^{(N)}$, $0 \le i \le N-1$ be the eigenvectors of $\mathbf{L}_N$.
- $\mathbf{u}_0^{(N)}$, ..., $\mathbf{u}_{N-1}^{(N)}$ are orthonormal, $\mathbf{u}_0^{(N)} = \mathbf{D}^{1/2}\mathbb{1}$
- Let $\mathbf{g} \perp \mathbf{D}^{1/2}\mathbb{1}$, $\mathbf{g} = \alpha_1 \mathbf{u}_1 + ... + \alpha_{N-1}\mathbf{u}_{N-1}$

$$\mathbf{g}^\intercal \mathbf{L}_N \mathbf{g} = \alpha_1^2 \lambda_1^{(N)} + ... + \alpha_{N-1}^2 \lambda_{N-1}^{(N)}$$
$$\mathbf{g}^\intercal \mathbf{g} = \alpha_1^2 + ... + \alpha_{N-1}^2 = 1$$

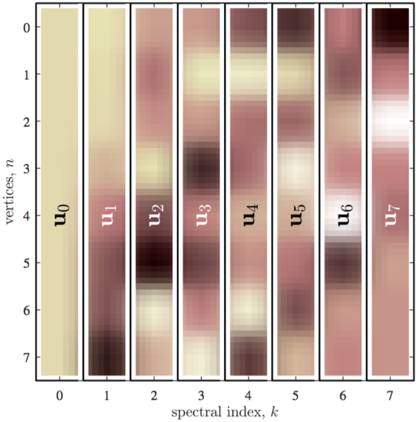
- $R(\mathbf{L}_N, \mathbf{g}) = \frac{\mathbf{g}^\intercal \mathbf{L}_N \mathbf{g}}{\mathbf{g}^\intercal \mathbf{g}}$
- $\lambda_1^{(N)} \le R(\mathbf{L}_N, \mathbf{g}) \le \lambda_{N-1}^{(N)} \le 2$, for $\mathbf{g} \perp \mathbf{D}^{1/2}\mathbb{1}$

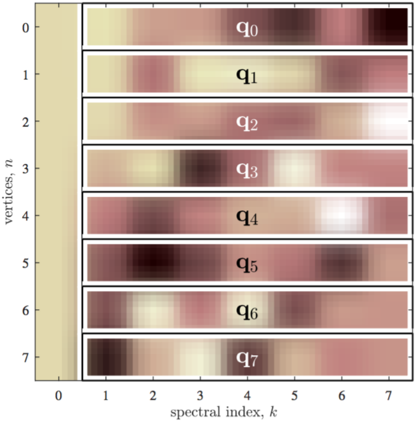
Introduction

1. Clustering based on graph topology
2. Spectral (eigenvector-based) methods for graph clustering

Spectral Space

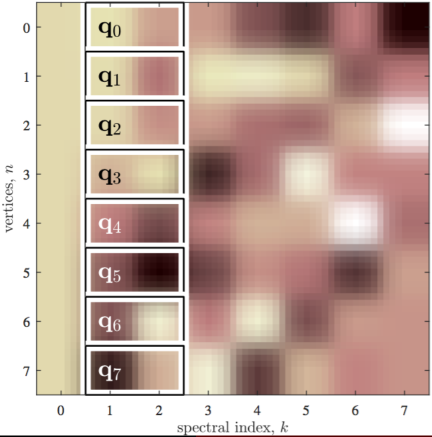
1. Spectral space for a graph with N vertices: an N dimensional space whose basis are the orthogonal eigenvector of the graph Laplacian.
2. Eigenmaps: For every vertex n, we define the $(N-1)$-dimensional spectral vector as $\mathbf{q}_n = [u_1(n), ..., u_{N-1}(n)]$ (omit $\mathbf{u}_0$)

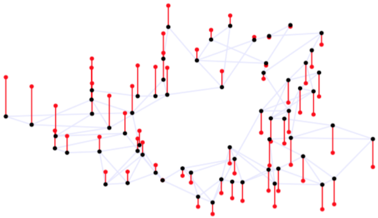




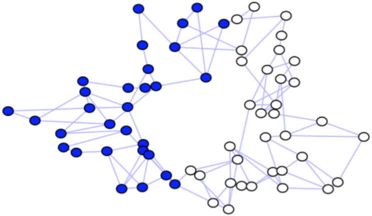
Spectral Space (Continue)

1. Spectrally similar if $d_{mn} <$ threshold, $d_{mn} \stackrel{\text{def}}{=} ||\mathbf{q}_m - \mathbf{q}_n||_2$
2. Spectral dimensionality reduction: restrict the definition of spectral similarity to only a few lower-order eigenvectors, e.g. $\mathbf{q}_n = [u_1(n), u_2(n)]$
3. Spectral scalar $\mathbf{q}_n = [u_1(n)]$
4. Simplest way to do spectral clustering is assigning vertices to two sets according to the sign of spectral scalar.

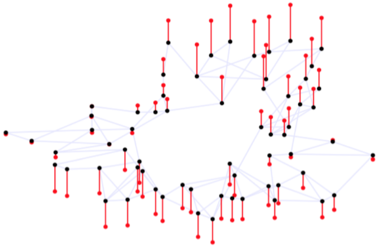




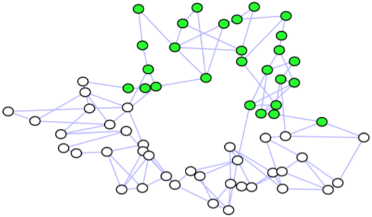
(a) Eigenvector, $u_1(n)$



(b) Clustering based on the eigenvector, $u_1(n)$

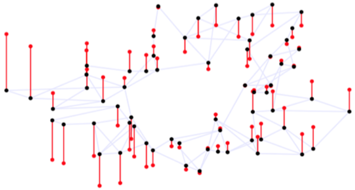


(c) Eigenvector, $u_2(n)$

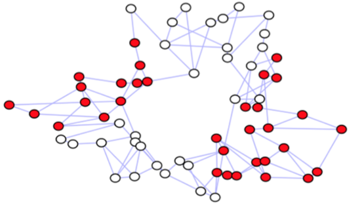


(d) Clustering based on the eigenvector, $u_2(n)$

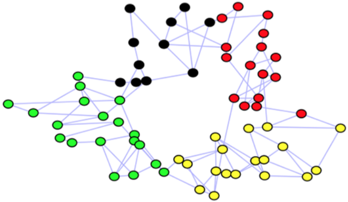
Spectral Vertex Clustering (Continue)



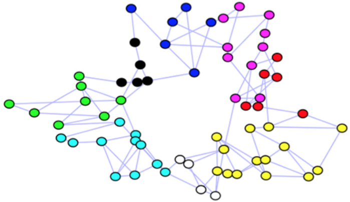
(e) Eigenvector, $u_3(n)$



(e) Clustering based on the eigenvector, $u_3(n)$

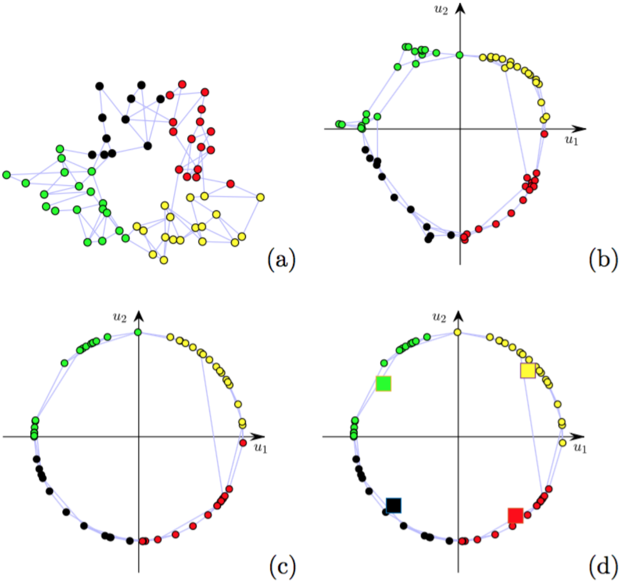


(g) Clustering based on the eigenvectors $u_1(n)$ and $u_2(n)$



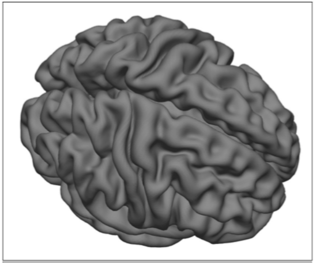
(h) Clustering based on the eigenvectors $u_1(n)$, $u_2(n)$ and $u_3(n)$

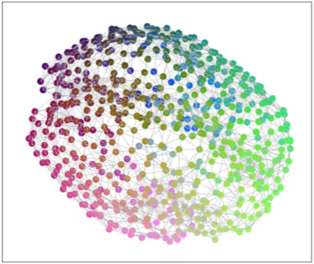
Spectral Vertex Clustering (Continue)



Spectral Vertex Clustering (Continue)

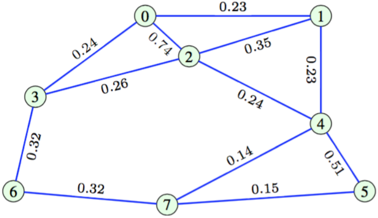
- Employ $\mathbf{q}_n = [\mathbf{u}_1(n), \mathbf{u}_2(n), \mathbf{u}_3(n)]$ as coordinates for the RGB scheme.





Random Walks on Graph (Diffusion Mapping)

1. A random walk on graph $\mathcal{G}$ is a sequence of vertices $v_0, v_1, ..., v_t, ...$, where each v_{t+1} is chosen to be a random neighbor v_t, $(v_t, v_{t+1}) \in \mathcal{B}$, with transition probability $P_{v_t, v_{t+1}} = \frac{\mathbf{W}_{v_t, v_{t+1}}}{\mathbf{D}_{v_t, v_t}}$
2. $\sum_i P_{ij} = 1$, $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$
3. The diffusion cloud for node i is $\mathbf{p}_i = [P_{i0}, P_{i1}, ..., P_{i(N-1)}]$
4. Let $p_i(t)$ be the probability that a walk is at node i at moment t
5. $p_j(t+1) = \sum_i \mathbf{P}_{ij} p_i(t)$
6. Let $\mathbf{p}(0) = [p_0(0)...p_{N-1}(0)]$ be the initial distribution
7. After t steps, $\mathbf{p}(t) = \mathbf{p}(0)\mathbf{P}^t$
8. Transition matrix of t step is $\mathbf{P}^t = (\mathbf{D}^{-1}\mathbf{W})^t$



$$\mathbf{P} = \begin{array}{c} \mathbf{p}_0 \\ \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \\ \mathbf{p}_4 \\ \mathbf{p}_5 \\ \mathbf{p}_6 \\ \mathbf{p}_7 \\ \\ \end{array} \begin{array}{c} \left[\begin{array}{cccccccc} 0 & 0.19 & 0.61 & 0.20 & 0 & 0 & 0 & 0 \\ 0.28 & 0 & 0.43 & 0 & 0.28 & 0 & 0 & 0 \\ 0.47 & 0.22 & 0 & 0.16 & 0.15 & 0 & 0 & 0 \\ 0.29 & 0 & 0.32 & 0 & 0 & 0 & 0.39 & 0 \\ 0 & 0.21 & 0.21 & 0 & 0 & 0.46 & 0 & 0.12 \\ 0 & 0 & 0 & 0 & 0.77 & 0 & 0 & 0.23 \\ 0 & 0 & 0 & 0.50 & 0 & 0 & 0 & 0.50 \\ 0 & 0 & 0 & 0 & 0.23 & 0.25 & 0.52 & 0 \end{array}\right] \\ \begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \end{array}$$

The Random Walk Matrix

- We assume that there is no zero degree vertex
- $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$ is not symmetric
- The normalized weight matrix $\mathbf{W}_N = \mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}$ is symmetric, and $\mathbf{W}_N = \mathbf{D}^{\frac{1}{2}}\left(\mathbf{D}^{-1}\mathbf{W}\right)\mathbf{D}^{-\frac{1}{2}} = \mathbf{D}^{\frac{1}{2}}\mathbf{P}\mathbf{D}^{-\frac{1}{2}}$
- $\mathbf{P} = \mathbf{D}^{-\frac{1}{2}}\mathbf{N}\mathbf{D}^{\frac{1}{2}} = \underbrace{\mathbf{D}^{-\frac{1}{2}}\mathbf{U}}_{\mathbf{\Phi}}\mathbf{\Lambda}\underbrace{\mathbf{U}^{-1}\mathbf{D}^{\frac{1}{2}}}_{\mathbf{\Psi}}$, since $\mathbf{U}$ gives orthonormal eigenvectors for $\mathbf{W}_N$
- By definition, $\mathbf{P}\mathbb{1} = \mathbb{1}$, which gives us ϕ_0
- $\psi_0 = \mathbf{D}^{\frac{1}{2}}\mathbf{u}_0 = \mathbf{D}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}}\phi_0 = \mathbf{D}\phi_0 = [\mathbf{D}_{0,0}, \ldots, \mathbf{D}_{N-1,N-1}]^\intercal$
- $\pi \stackrel{\text{def}}{=} [\mathbf{D}_{00}/\mathrm{Vol}(\mathcal{G}), \ldots, \mathbf{D}_{N-1,N-1}/\mathrm{Vol}(\mathcal{G})]^\intercal$ is the stationary distribution, where $\mathrm{Vol}(\mathcal{G}) = \sum_{i=0}^{N-1}\mathbf{D}_{i,i}$ is the volume of graph $\mathcal{G}$
- The stationary distribution π is defined for all graph without zero degree vertex, including disconnected graph and bipartite

The Stationary Distribution

1. Draw a vertex u from distribution π
2. Let m be a neighbour of u picked uniformly at random
3. Return m. m is also distributed as π since $\mathbf{P}\pi = \pi$

- 1 and 2 is picking a directed edge randomly: $\frac{\mathbf{D}_{i,i}}{\mathsf{Vol}(\mathcal{G})} \times \frac{1}{\mathbf{D}_{i,i}} = \frac{1}{2|\mathcal{B}|}$
- By repeating 1 $\sim$ 3, $u \to \underbrace{m \to ... \to u}_{t}$ gives a sequence of vertices
- If the initial vertex u is not picked from π, as $t \to \infty$, the distribution of u is π unless $\mathcal{G}$ is disconnected or a bipartite

Diffusion Mapping

1. "Generalized eigenvectors" of graph Laplacian are from $\mathbf{L}\mathbf{u}^{(G)} = \lambda^{(G)}\mathbf{D}\mathbf{u}^{(G)}$
2. Eigen-analysis on $\mathbf{P}$, its eigenvalue $\lambda^{(P)} = 1 - \lambda^{(G)}$

$$\begin{aligned} \mathbf{D}^{-1}\mathbf{W}\mathbf{u} &= \lambda^{(P)}\mathbf{u} \\ \mathbf{D}^{-1}(\mathbf{D} - \mathbf{L})\mathbf{u} &= \lambda^{(P)}\mathbf{u} \\ (\mathbf{I} - \mathbf{D}^{-1}\mathbf{L})\mathbf{u} &= \lambda^{(P)}\mathbf{u} \\ \mathbf{D}(1 - \lambda^{(P)})\mathbf{u} &= \mathbf{L}\mathbf{u} \end{aligned}$$

3. Eigen-analysis on $\mathbf{P}^t$ gives $\lambda^{(P)} = (1 - \lambda^{(G)})^t$, while the eigenvectors remain the same for $t = 1$.
4. Diffusion mapping for t step is

$$\mathbf{q}_n^{(t)} = [u_1^{(G)}(n), u_2^{(G)}(n), ..., u_{N-1}^{(G)}(n)](\mathbf{I} - \mathbf{\Lambda}^{(G)})^t$$

or with $\mathbf{P} = \mathbf{\Phi}\mathbf{\Lambda}^{(P)}\mathbf{\Psi}^\intercal$

$$\mathbf{q}_n^{(t)} = [\phi_1(n), \phi_2(n), ..., \phi_{N-1}(n)](\mathbf{\Lambda}^{(P)})^t$$

Diffusion Distances

- The Euclidean distance between $\mathbf{q}_n^{(t)}$ and $\mathbf{q}_m^{(t)}$ is $||\mathbf{q}_n - \mathbf{q}_m||_2$
- The diffusion cloud for node n after t steps $\mathbf{p}_n^{(t)}$ is the n^{th} row of $\mathbf{P}^t$
- Eigenvalue decomposition on $\mathbf{P}^t = \mathbf{\Phi}(\mathbf{\Lambda}^{(P)})^t\mathbf{\Psi}^\intercal$
- $\mathbf{p}_{nj}^{(t)} = \mathbf{P}^t_{nj} = \sum_{k=0}^{N-1}(\lambda_k^{(P)})^t\phi_k(n)\psi_k(j)$
- Calculate the weighted Euclidean distance between diffusion clouds of node n and m, where the weight is diag$(\frac{1}{\mathbf{D}_{0,0}}, ..., \frac{1}{\mathbf{D}_{N-1,N-1}})$. Its square is:

$$||(\mathbf{p}_n^{(t)} - \mathbf{p}_m^{(t)})\mathbf{D}^{-\frac{1}{2}}||_2^2 = \sum_{k=0}^{N-1}(\lambda_k^{(P)})^{2t}(\phi_k(n) - \phi_k(m))^2 = ||\mathbf{q}_n - \mathbf{q}_m||_2^2$$

- $\mathbf{D}^{-\frac{1}{2}}\mathbf{\Psi}$ is unitary
- The diffusion distance after t steps between n and m is $D_f^{(t)}(n,m)$

$$D_f^{(t)}(n,m)^2 \overset{\text{def}}{=} (\sum_{i=0}^{N-1}\mathbf{D}_{ii}) \times ||(\mathbf{p}_n^{(t)} - \mathbf{p}_m^{(t)})\mathbf{D}^{-\frac{1}{2}}||_2^2 = (\sum_{i=0}^{N-1}\mathbf{D}_{ii}) \times ||\mathbf{q}_n - \mathbf{q}_m||_2^2$$

Diffusion Process

- Define some quantity for node i at time t as $x_i(t)$
- The quantity for node i at time $t + \delta t$, is equals to $x_i(t)$ with the amount changes within neighbours during δt. And the amount of such flow is determined by elapsed time and a constant C.

$$x_i(t + \delta t) = x_i(t) + \sum_{j=0}^{N-1} \mathbf{W}_{ij}(x_j(t) - x_i(t)) \times C\delta t$$

$$\begin{aligned}
\frac{\mathrm{d}x_i}{\mathrm{d}t} &= C \sum_{j=0}^{N-1} \mathbf{W}_{ij}(x_j(t) - x_i(t)) \\
&= C(\sum_{j=0}^{N-1} \mathbf{W}_{ij} x_j(t) - \sum_{j=0}^{N-1} \mathbf{W}_{ij} x_i(t)) \\
&= C(\sum_{j=0}^{N-1} \mathbf{W}_{ij} x_j(t) - \mathbf{D}_{ii} x_i(t)) \\
&= -C \sum_{j=0}^{N-1} \mathbf{L}_{ij} x_j
\end{aligned}$$

Diffusion Process (Continue)

- In matrix-vector notation $\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} + C\mathbf{L}\mathbf{x} = \mathbf{0}$
- To solve this first-order matrix differential equation, write $\mathbf{x}$ as a linear combination of eigenvectors $\mathbf{u}_i$ of $\mathbf{L}$, with time-dependent $\mathbf{x} = \sum_{i=0}^{N-1} \alpha_i \mathbf{u}_i$

$$
\begin{aligned}
\mathbf{0} &= \frac{\mathrm{d}\left(\sum_{i=0}^{N-1} \alpha_i \mathbf{u}_i\right)}{\mathrm{d}t} + C\mathbf{L}\left(\sum_{i=0}^{N-1} \alpha_i \mathbf{u}_i\right) \\
&= \sum_{i=0}^{N-1} \left(\frac{\mathrm{d}\alpha_i}{\mathrm{d}t}\mathbf{u}_i + C\alpha_i \mathbf{L}\mathbf{u}_i\right) \\
&= \sum_{i=0}^{N-1} \left(\frac{\mathrm{d}\alpha_i}{\mathrm{d}t}\mathbf{u}_i + C\alpha_i \lambda_i \mathbf{u}_i\right) \\
\Rightarrow & \frac{\mathrm{d}\alpha_i}{\mathrm{d}t} + C\lambda_i \alpha_i = 0
\end{aligned}
$$

Diffusion Process (Continue)

Solution: $\mathbf{x}(t) = \sum_{i=0}^{N-1} \alpha_i(0) e^{-C\lambda_i t} \mathbf{u}_i$, since $\alpha_i(t) = \alpha_i(0) e^{-C\lambda_i t}$

1. $t \to \infty$, the term with $\lambda_0 = 0$, $\mathbf{u}_0 = \mathbb{1}$ will survive, $\lim_{t \to \infty} \mathbf{x}(t) = \alpha_0(0)\mathbb{1}$
2. Other terms will exponentially decay
3. Quantity per node becomes even as diffusion goes on.

Section 4

Conductance

The minimum value of $\phi(\mathcal{E})$ is called the Cheeger constant or conductance of a graph $\mathcal{G}$, where $\phi(\mathcal{E})$ is the Cheeger ratio of a subset $\mathcal{E}$

- $\phi(\mathcal{E}) \stackrel{def}{=} \frac{1}{\min\{\mathsf{Vol}(\mathcal{E}),\mathsf{Vol}(\mathcal{E}^c)\}}\mathsf{Cut}(\mathcal{E})$, $\mathsf{Cut}(\mathcal{E}) \stackrel{def}{=} \sum_{m\in\mathcal{E},n\in\mathcal{E}^c} \mathbf{W}_{mn}$
- $\phi(\mathcal{E})$ indicates the chance for a random walk to escape from or travel to $\mathcal{E}$

Let X be a sequence of vertices from a random walk that will reach the stationary distribution. We have $\mathbf{P}\{X(t) = i\} = \mathbf{D}_{ii}/\mathsf{Vol}(\mathcal{G})$. If $\mathsf{Vol}(\mathcal{E}) \leq \mathsf{Vol}(\mathcal{E}^c)$:

$$\begin{aligned}
\phi(\mathcal{E}) &= \frac{1}{\mathsf{Vol}(\mathcal{E})} \times \mathsf{Cut}(\mathcal{E}) \\
&= \frac{1}{\mathsf{Vol}(\mathcal{E})/\mathsf{Vol}(\mathcal{G})} \times \sum_{m\in\mathcal{E},n\in\mathcal{E}^c} \frac{\mathbf{W}_{mn}}{\mathbf{D}_{mm}} \times \frac{\mathbf{D}_{mm}}{\mathsf{Vol}(\mathcal{G})} \\
&= \mathbf{P}\{X(1) \in \mathcal{E} \cap X(0) \in \mathcal{E}^c\}/\mathbf{P}\{X(0) \in \mathcal{E}\} \\
&= \mathbf{P}\{X(1) \in \mathcal{E}^c | X(0) \in \mathcal{E}\}
\end{aligned}$$

If $\mathsf{Vol}(\mathcal{E}) > \mathsf{Vol}(\mathcal{E}^c)$, we have $\phi(\mathcal{E}) = \mathbf{P}\{X(1) \in \mathcal{E} | X(0) \in \mathcal{E}^c\}$

Volume Normalized Cut

- A more general form of the normalized cut may also involve vertex weights
- The volume normalized cut $\text{CutV}(\mathcal{E}, \mathcal{E}^c) = \left(\frac{1}{\text{Vol}(\mathcal{E})} + \frac{1}{\text{Vol}(\mathcal{E}^c)}\right) \text{Cut}(\mathcal{E})$
- $\text{CutV}(\mathcal{E}, \mathcal{E}^c)$ and $\phi(\mathcal{E})$ only differ by at most a quantity of 2

$$\text{CutV}(\mathcal{E}, \mathcal{E}^c) = \frac{\text{Vol}(\mathcal{G})}{\text{Vol}(\mathcal{E}) \times \text{Vol}(\mathcal{E}^c)} \text{Cut}(\mathcal{E}) = \frac{\text{Cut}(\mathcal{E})}{\text{Vol}(\mathcal{E}) \times (\text{Vol}(\mathcal{E}^c)/\text{Vol}(\mathcal{G}))}$$

- $\frac{1}{2} \leq \max\{\text{Vol}(\mathcal{E}), \text{Vol}(\mathcal{E}^c)\}/\text{Vol}(\mathcal{G}) \leq 1$
- $\phi(\mathcal{E}) \leq \text{CutV}(\mathcal{E}, \mathcal{E}^c) \leq 2\phi(\mathcal{E})$
- $\text{CutV}(\mathcal{E}, \mathcal{E}^c) = \mathbf{P}\{X(1) \in \mathcal{E} | X(0) \in \mathcal{E}^c\} + \mathbf{P}\{X(1) \in \mathcal{E}^c | X(0) \in \mathcal{E}\}$

Lower Bound via Spectral Relaxation

- If $\mathbf{f}$ is indicator function of $(\mathcal{E}, \mathcal{E}^c)$ with $\mathbf{f}^\intercal \mathbf{D} \mathbf{f} = 1$, $\frac{\mathbf{f}^\intercal \mathbf{L} \mathbf{f}}{\mathbf{f}^\intercal \mathbf{D} \mathbf{f}} = \mathsf{CutV}(\mathcal{E}, \mathcal{E}^c)$
- Let $\mathbf{g} = \mathbf{D}^{1/2}\mathbf{f}$, $\frac{\mathbf{f}^\intercal \mathbf{L} \mathbf{f}}{\mathbf{f}^\intercal \mathbf{D} \mathbf{f}} = \frac{\mathbf{g}^\intercal \mathbf{D}^{-1/2} \mathbf{L} \mathbf{D}^{-1/2} \mathbf{g}}{\mathbf{g}^\intercal \mathbf{D}^{-1/2} \mathbf{D} \mathbf{D}^{-1/2} \mathbf{g}} = \frac{\mathbf{g}^\intercal \mathbf{L}_N \mathbf{g}}{\mathbf{g}^\intercal \mathbf{g}}$

$$\min_{\mathbf{f} \perp \mathbb{1}, \mathbf{f}^\intercal \mathbf{D} \mathbf{f} = 1} \frac{\mathbf{f}^\intercal \mathbf{L} \mathbf{f}}{\mathbf{f}^\intercal \mathbf{D} \mathbf{f}} = \min_{\mathbf{g} \perp \mathbf{D}^{1/2} \mathbb{1}, \mathbf{g}^\intercal \mathbf{g} = 1} \frac{\mathbf{g}^\intercal \mathbf{L}_N \mathbf{g}}{\mathbf{g}^\intercal \mathbf{g}} = \lambda_1^{(N)}$$

$$\min \mathsf{CutV}(\mathcal{E}, \mathcal{E}^c) = \min_{\mathbf{f} \perp \mathbb{1}, \mathbf{f}^\intercal \mathbf{D} \mathbf{f} = 1, \mathbf{f} \text{ is indicator}} \frac{\mathbf{f}^\intercal \mathbf{L} \mathbf{f}}{\mathbf{f}^\intercal \mathbf{D} \mathbf{f}} \geq \min_{\mathbf{f} \perp \mathbb{1}, \mathbf{f}^\intercal \mathbf{D} \mathbf{f} = 1} \frac{\mathbf{f}^\intercal \mathbf{L} \mathbf{f}}{\mathbf{f}^\intercal \mathbf{D} \mathbf{f}} = \lambda_1^{(N)}$$

where $\lambda_1^{(N)}$ is the second smallest eigenvalue of $\mathbf{L}_N$

- $\min \mathsf{CutV}(\mathcal{E}, \mathcal{E}^c) \geq \lambda_1^{(N)}$

Lower Bound (Continue)

$\mathbf{u}_1^{(N)}$ achieves the smallest $\min\limits_{\mathbf{g}\perp\mathbf{D}^{1/2}\mathbb{1},\mathbf{g}^\intercal\mathbf{g}=1} \frac{\mathbf{g}^\intercal\mathbf{L}_N\mathbf{g}}{\mathbf{g}^\intercal\mathbf{g}}$

$\Rightarrow$ $\mathbf{u}_1^{(N)}$ gives the best way to put $\mathcal{G}$ on $\mathbb{R}$ that adjacent vertices in $\mathcal{G}$ are mapped to close values on $\mathbb{R}$ ($\mathbf{u}_1^{(N)}$ is a "smooth" signal on $\mathcal{G}$)

$\Rightarrow$ A one-dimensional representation of graph that optimally preserves local neighborhood information of $\mathbf{L}_N$

$\Rightarrow$ The chance that a random threshold separates two close vertices on $\mathbb{R}$ is small

---

$\mathbf{u}_2^{(N)}$ achieves the smallest $\min\limits_{\mathbf{g}\perp\mathbf{D}^{1/2}\mathbb{1},\mathbf{g}\perp\mathbf{u}_1^{(N)},\mathbf{g}^\intercal\mathbf{g}=1} \frac{\mathbf{g}^\intercal\mathbf{L}_N\mathbf{g}}{\mathbf{g}^\intercal\mathbf{g}}$

Find the best optimizer in remaining subspace ...

$\mathbf{u}_k^{(N)}$ achieves the smallest $\min\limits_{\mathbf{g}\perp\mathbf{D}^{1/2}\mathbb{1},\mathbf{g}\perp\mathbf{u}_1^{(N)},\ldots,\mathbf{g}\perp\mathbf{u}_{k-1}^{(N)},\mathbf{g}^\intercal\mathbf{g}=1} \frac{\mathbf{g}^\intercal\mathbf{L}_N\mathbf{g}}{\mathbf{g}^\intercal\mathbf{g}}$

$\mathbf{u}_0^{(N)}, \mathbf{u}_1^{(N)}$, ..., $\mathbf{u}_k^{(N)}$ is the $k-$dimensional orthonormal basis optimizing $R(\mathbf{L}_N)$

$\Rightarrow$ The spectral vector $\mathbf{q}_n^{(N)} = [\mathbf{u}_0^{(N)}(n), \mathbf{u}_1^{(N)}(n), ..., \mathbf{u}_k^{(N)}(n)]$, give optimal vertex positions in the reduced k- dimensional vertex space in the sense that they minimize an objective function which penalizes for the distance between neighboring vertices (in $\mathcal{G}$) in the spectral space.

$\Rightarrow$ We can use k-means to find meaningful clusters using $\mathbf{q}^{(N)}$

Preserving Upper Bound via Spectral Relaxation

For a $\mathbf{f}$, $\mathbf{f} \perp \mathbb{1}$, $\frac{\mathbf{f}^\intercal \mathbf{L} \mathbf{f}}{\mathbf{f}^\intercal \mathbf{D} \mathbf{f}} \leq \delta$, $\exists \tau \in \mathbb{R}$, $S = \{n \in \mathcal{B} : f(n) \leq \tau\}$, such that $\phi(S) \leq \sqrt{2\delta}$, which means $\mathsf{CutV}(S, S^c) \leq 2\phi(S) \leq 2\sqrt{2\delta} = O(\sqrt{\delta})$

- The Cheeger ratio of S: $\phi(S) = \frac{1}{\min\{\mathsf{Vol}(S), \mathsf{Vol}(S^c)\}} \mathsf{Cut}(S)$

Therefore, we can find a subset-wise constant indicator vector for (S, S^c) that corresponds to $\mathbf{u}_1^{(N)}$ with $\mathsf{CutV}(S, S^c) \leq O(\sqrt{\lambda_1^{(N)}})$

In addition, $\min \mathsf{CutV}(S, S^c) \geq \lambda_1^{(N)}$.

- If $\lambda_1^{(N)}$ is large, $\mathcal{G}$ does not have good clusters
- If $\lambda_1^{(N)}$ is small, we can use $\mathbf{u}_1^{(N)}$ to find a good clusters with two components

Upper Bound for Cheeger's Inequality (Proof)

Assumption: $\frac{\mathbf{f}^\intercal \mathbf{L} \mathbf{f}}{\mathbf{f}^\intercal \mathbf{D} \mathbf{f}} \leq \delta$, $\mathbf{f} \perp \mathbb{1}$
First Step: Find $\mathbf{z}$ by translating and scaling $\mathbf{f}$ while $\frac{\mathbf{z}^\intercal \mathbf{L} \mathbf{z}}{\mathbf{z}^\intercal \mathbf{D} \mathbf{z}} \leq \delta$ is also true

1 Translate $\mathbf{f}$ by a constant c along $\mathbf{u}_0 = \mathbb{1}$ to make the median element of the translated vector $\mathbf{z} = \mathbf{f} + c\mathbb{1}$ to be 0.

- Such translation preserves $\mathbf{f}^\intercal \mathbf{L} \mathbf{f}$
- Such translation increases $\mathbf{f}^\intercal \mathbf{D} \mathbf{f}$ since $(\mathbf{f} + c\mathbb{1})^\intercal \mathbf{D} (\mathbf{f} + c\mathbb{1}) = \mathbf{f}^\intercal \mathbf{D} \mathbf{f} + c^2 \mathbb{1}^\intercal \mathbf{D} \mathbb{1}$

2 Scale $\mathbf{z}$ and reorder it to be $z(0) \leq ... \leq z(N-1)$, $z(0)^2 + z(N-1)^2 = 1$

- Any scaling on $\mathbf{g}$ won't change $\frac{\mathbf{f}^\intercal \mathbf{L} \mathbf{f}}{\mathbf{f}^\intercal \mathbf{D} \mathbf{f}}$

Proof (Continue)

Second Step: Consider a random construction of S where the threshold $\tau \in (z(0), z(N-1))$ has a probability density $2|\tau|$

- Such probability density $2|\tau|$ is valid

$$\int_{z(0)}^{z(N-1)} 2|t|\,\mathrm{d}t = \int_{z(0)}^{0} 2|t|\,\mathrm{d}t + \int_{0}^{z(N-1)} 2|t|\,\mathrm{d}t = z(0)^2 + z(N-1)^2 = 1$$

- For $z(0) \le a < b \le z(N-1)$ Simplify and get

$$\mathbf{P}\{\tau \in (a,b)\} = \int_a^b 2|t|\,\mathrm{d}t = \begin{cases} a^2 + b^2, & \text{if } \mathsf{sign}(a) = -\mathsf{sign}(b) \;\; 1 \\ |b^2 - a^2|, & \text{if } \mathsf{sign}(a) = \mathsf{sign}(b) \;\; 2 \end{cases}$$

$$1 : a^2 + b^2 \le (a-b)^2 = |a-b|(|a| + |b|)$$

$$2 : |b^2 - a^2| = |a-b||a+b| = |a-b|(|a| + |b|)$$

- $\mathbf{P}\{\tau \in (a,b)\} \le |a-b|(|a| + |b|)$

Proof (Continue)

Third Step: Calculate the expectation of $\text{Cut}(S)$ with random variable τ

$$
\begin{aligned}
\mathbf{E}\{\text{Cut}(S)\} &= \mathbf{E}\{\frac{1}{2}\sum_{i,j}\mathbf{W}_{ij}\mathbf{1}_{(S,S^c)\text{ cuts edge }(i,j)}\} \\
&= \frac{1}{2}\sum_{i,j}\mathbf{W}_{ij}\mathbf{P}\{(S,S^c)\text{ cuts edge }(i,j)\} \\
&= \frac{1}{2}\sum_{i,j}\mathbf{W}_{ij}\mathbf{P}\{\tau \in (z(i), z(j))\} \\
&\leq \frac{1}{2}\sum_{i,j}\mathbf{W}_{ij}|z(i)-z(j)|(|z(i)|+|z(j)|) \\
&\leq \frac{1}{2}\sqrt{\mathbf{I}}\sqrt{\mathbf{II}}
\end{aligned}
$$

- The Cauchy–Schwarz inequality states that $|\mathbf{x}^\intercal \mathbf{y}| \leq ||\mathbf{x}||_2 \times ||\mathbf{y}||_2$

Proof (Continue)

$$
\begin{aligned}
\mathbf{I} &= \sum_{i,j} \mathbf{W}_{ij}(z(i) - z(j))^2 = 2\mathbf{z}^\intercal \mathbf{L}\mathbf{z} \leq 2\delta \mathbf{z}^\intercal \mathbf{D}\mathbf{z} \\
\mathbf{II} &= \sum_{i,j} \mathbf{W}_{ij}(|z(i)| + |z(j)|)^2 \leq \sum_{i,j} \mathbf{W}_{ij}(z(i)^2 + z(j)^2) \\
&= 4\sum_{i} \mathbf{D}_{ii} z(i)^2 = 4\mathbf{z}^\intercal \mathbf{D}\mathbf{z}
\end{aligned}
$$

$$
\mathbf{E}\{\mathsf{Cut}(S)\} \leq \frac{1}{2}\sqrt{2\delta \mathbf{z}^\intercal \mathbf{D}\mathbf{z}}\sqrt{4\mathbf{z}^\intercal \mathbf{D}\mathbf{z}} = \sqrt{2\delta}\mathbf{z}^\intercal \mathbf{D}\mathbf{z}
$$

Proof (Continue)

Fourth Step: Calculate expectation of $\mathsf{min}\{\mathsf{Vol}(S), \mathsf{Vol}(S^c)\}$

- The median of $\mathbf{z}$ is 0
  $\Rightarrow$ if $\tau \leq 0$, $\mathsf{Vol}(S) \leq \mathsf{Vol}(S^c)$; if $\tau \geq 0$, $\mathsf{Vol}(S) \geq \mathsf{Vol}(S^c)$
- $\sum_{i<m} \mathbf{P}\{z(i) < t < 0\} + \sum_{i \geq m} \mathbf{P}\{0 < t < z(i)\} = \sum_i z(i)^2$

$$
\begin{aligned}
\mathbf{E}\{\mathsf{min}\{\mathsf{Vol}(S), \mathsf{Vol}(S^c)\}\} &= \sum_i \mathbf{D}_{ii} \mathbf{P}\{z(i) \text{ is in the set with smaller volume}\} \\
&= \sum_i \mathbf{D}_{ii} z(i)^2 \\
&= \mathbf{z}^\mathsf{T} \mathbf{D} \mathbf{z}
\end{aligned}
$$

Proof (Continue)

Conclusion

$$\mathbf{E}\{\mathsf{Cut}(S)\} \leq \sqrt{2\delta}\mathbf{z}^\intercal\mathbf{D}\mathbf{z}$$
$$\mathbf{E}\{\mathsf{Cut}(S)\} - \mathbf{E}\{\mathsf{min}\{\mathsf{Vol}(S), \mathsf{Vol}(S^c)\}\} \times \sqrt{2\delta} \leq 0$$
$$\mathbf{E}\{\mathsf{Cut}(S) - \mathsf{min}\{\mathsf{Vol}(S), \mathsf{Vol}(S^c)\} \times \sqrt{2\delta}\} \leq 0$$

$$\exists\, S \text{ such that } \mathsf{Cut}(S) - \mathsf{min}\{\mathsf{Vol}(S), \mathsf{Vol}(S^c)\sqrt{2\delta} \leq 0$$
$$\Rightarrow \phi(S) = \frac{\mathsf{Cut}(S)}{\mathsf{min}\{\mathsf{Vol}(S), \mathsf{Vol}(S^c)} \leq \sqrt{2\delta}$$
$$\Rightarrow \mathsf{CutV}(S, S^c) \leq 2\phi(S) \leq 2\sqrt{2\delta}$$

Cheeger's Inequality

- The conductance or the Cheeger constant of graph $\mathcal{G}$: $\phi(\mathcal{G}) = \min_{S \in \mathcal{B}} \phi(S)$

$$\min_{S \in \mathcal{B}} \mathrm{CutV}(S, S^c) \geq \min_{\mathbf{f} \perp \mathbb{1}, \mathbf{f}^\intercal \mathbf{D} \mathbf{f} = 1} \frac{\mathbf{f}^\intercal \mathbf{L} \mathbf{f}}{\mathbf{f}^\intercal \mathbf{D} \mathbf{f}} = \lambda_1^{(N)}$$

$$\Rightarrow \forall\, \phi(S) \geq \frac{1}{2} \lambda_1^{(N)} \Rightarrow \phi(\mathcal{G}) \geq \frac{1}{2} \lambda_1^{(N)}$$

$$\frac{\mathbf{u}_1^{(N)\intercal} \mathbf{L} \mathbf{u}_1^{(N)}}{\mathbf{u}_1^{(N)\intercal} \mathbf{D} \mathbf{u}_1^{(N)}} = \lambda_1^{(N)}$$

$$\Rightarrow \exists\, \tau \text{ such that } \phi(S) \leq \sqrt{2\lambda_1^{(N)}} \text{ where } S = \{n \in \mathcal{B} : \mathbf{u}_1^{(N)}(n) \leq \tau\}$$

$$\Rightarrow \phi(\mathcal{G}) \leq \sqrt{2\lambda_1^{(N)}}$$

- $\frac{1}{2} \lambda_1^{(N)} \leq \phi(\mathcal{G}) \leq \sqrt{2\lambda_1^{(N)}}$

Thank You for Your Attention!