Spectral Clustering

Part 3: The Normalized Laplacian

Ng Yen Kaow

More constraint for balance

- Further constraints can be added to the eigenvalue system
  - The next problem, Graph Partitioning, will use this strategy
  - However, the resultant eigenvalue system will no longer be standard

Graph Partitioning Problem

- Given edge weight matrix $W = (w_{ij})$ and vertex mass matrix M with diagonal elements (m_i), a 2-partitioning of an undirected graph $G = (V, E)$ is a partition of V into two groups S and $\bar{S}$ such that $\mathrm{cut}(S, \bar{S}) = \sum_{i \in S, j \in \bar{S}} w_{ij}$ is minimized under the constraint that $\sum_{i \in S} m_i = \sum_{i \in \bar{S}} m_i$, or $1^{\mathrm{T}} M x = 0$
  - If let $m_i = 1$ for all i, then $\sum_{i \in S} m_i = \sum_{i \in \bar{S}} m_i$ is the same as $|S| = |\bar{S}|$, the balance requirement

Constrained optimization problem

- Minimize $x^{\mathrm{T}}Lx$ where $L = D' - W$ subject to $x^{\mathrm{T}}M \in \{1, -1\}$ and $\mathbf{1}^{\mathrm{T}}Mx = 0$
  - $x_i \in \{1, -1\}$ and $\mathbf{1}^{\mathrm{T}}Mx = 0$ together enforce balance in the solution

  - However, problem is NP-hard
    - Recall that even the minimum bisection problem, where all edges and vertices have the same weight, is NP-hard

Relaxed Rayleigh quotient version

- Minimize $x^{\mathrm{T}} L x$ where $L = D' - W$
  subject to $x^{\mathrm{T}} M x = \sum_i m_i$ and $\mathbf{1}^{\mathrm{T}} M x = 0$
  - $x_i \in \{1, -1\} \Rightarrow x^{\mathrm{T}} M x = \sum_i m_i$ but not the other way around
  - Balance no longer enforced but that's the least of our worry for now because instead of the standard eigensystem
- Optimization must now be achieved through solving the generalized eigensystem

$$Lx = \lambda M x$$

Relaxed Rayleigh quotient version

- Minimize $x^{\mathrm{T}}Lx$ where $L = D' - W$ subject to $x^{\mathrm{T}}Mx = \sum_i m_i$ and $\mathbf{1}^{\mathrm{T}}Mx = 0$
- Optimize through $Lx = \lambda Mx$
- Since $L\mathbf{1} = 0$, $\mu_k = \mathbf{1}$
  - However, eigenvectors in the solutions are not orthogonal but rather, M-orthogonal ($\mu_i M \mu_j = 0$ for $i \neq j$)
    - $\mathbf{1}^{\mathrm{T}}M\mu_{k-1} = 0$ fulfilled
- Convert to a standard eigenvalue system $M^{-1/2}LM^{-1/2}x = \lambda x$ to compute

Convert to $M^{-1/2}LM^{-1/2}x = \lambda x$

- Minimize $x^{\mathrm{T}}Lx$ where $L = D' - W$
  subject to $x^{\mathrm{T}}Mx = \sum_i m_i$ and $\mathbf{1}^{\mathrm{T}}Mx = 0$
- Let $y = M^{1/2}x$, that is, $x = M^{-1/2}y$

$$x^{\mathrm{T}}Lx \Rightarrow y^{\mathrm{T}}M^{-1/2}LM^{-1/2}y$$

$$x^{\mathrm{T}}Mx = \sum_i m_i \Rightarrow y^{\mathrm{T}}y = \sum_i m_i$$

$$\mathbf{1}^{\mathrm{T}}Mx = 0 \Rightarrow \mathbf{1}^{\mathrm{T}}M^{1/2}y = 0$$

Hence equivalently

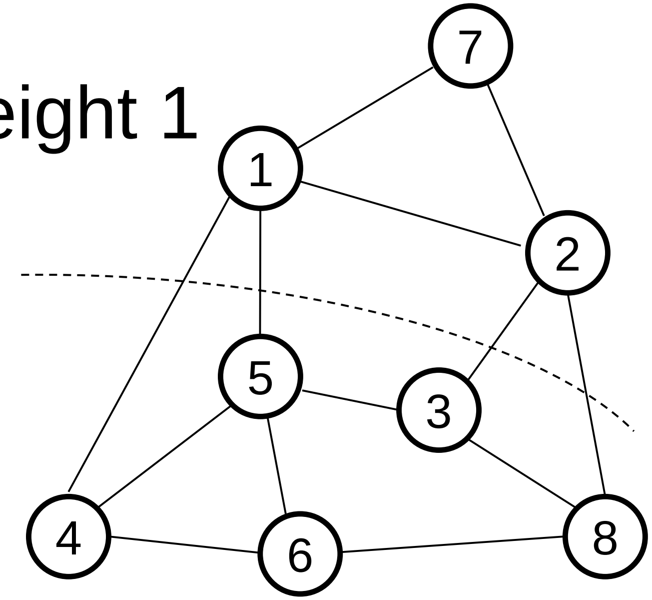
- Minimize $yM^{-1/2}LM^{-1/2}y$
  subject to $y^{\mathrm{T}}y = \sum_i m_i$ and $\mathbf{1}^{\mathrm{T}}M^{1/2}y = 0$

Convert to $M^{-1/2}LM^{-1/2}x = \lambda x$

- Minimize $yM^{-1/2}LM^{-1/2}y$
  subject to $y^{\mathrm{T}}y = 1$ and $\mathbf{1}^{\mathrm{T}}M^{1/2}y = 0$
- As $\mathbf{1}$ is a eigenvector for $Lx = \lambda Mx$ with eigenvalue 0, $M^{1/2}\mathbf{1}$ is a eigenvector for this system with eigenvalue 0 (smallest)
  - Since eigenvectors of this system are orthogonal, $\left(M^{1/2}\mathbf{1}\right)\mu_{k-1} = 0$
    $\Rightarrow \mathbf{1}^{\mathrm{T}}M^{1/2}y = 0$ fulfilled
  - In fact the eigenvalues for this system are the same as those for $Lx = \lambda Mx$, even though the eigenvectors are different (related by $y = M^{1/2}x$)

Eigendecomposition

- All edges and vertices with weight 1



λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
5.9390	5.1420	4.6660	4.0	3.0500	1.8100	1.3940	0.0

μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_6	μ_6
0.5677	-0.1583	-0.4862	0.3536	0.2315	-0.2855	0.1766	0.3536
-0.4281	0.6222	-0.2059	0.3536	0.0622	0.2469	0.2690	0.3536
0.3517	0.1203	0.2984	-0.3536	0.5170	0.5007	-0.0694	0.3536
-0.0855	0.0612	0.6267	0.3536	0.1159	-0.4899	-0.3044	0.3536
-0.5514	-0.3549	-0.3566	-0.3536	0.3216	-0.1795	-0.2392	0.3536
0.2351	0.3822	-0.2014	-0.3536	-0.5589	-0.1183	-0.4263	0.3536
-0.0354	-0.1476	0.2596	-0.3536	-0.2798	-0.2029	0.7349	0.3536
-0.0540	-0.5251	0.0654	0.3536	-0.4096	0.5286	-0.1411	0.3536

Generalize eigenvalue system

- Use of generalized eigenvalue system for spectral clustering first proposed in
  - Donath and Homan, "*Algorithms for partitioning of graphs and computer logic based on eigenvectors of connection matrices*", 1972, IBM Technical Disclosure Bulletin 15(3):938–944
- Also used in Normalized Cut
  - Which is currently almost synonymous with spectral clustering

Normalized Cut Problem

- Given weight matrix $W = (w_{ij})$ and weighted degree matrix $D' = (d_i)$, the normalized cut of an undirected graph $G = (V, E)$ is a partition of V into two groups S and $\bar{S}$ such that

$$\mathrm{ncut}(S, \bar{S}) = \mathrm{cut}(S, \bar{S}) \left(\frac{1}{\mathrm{vol}(S)} + \frac{1}{\mathrm{vol}(\bar{S})} \right)$$

is minimized, where $\mathrm{vol}(S) = \sum_{i \in S} d_i$, that is, sum of all the weights of the edges adjacent to vertices in S, and $\mathrm{cut}(S, \bar{S}) = \sum_{i \in S, j \in \bar{S}} w_{ij}$

Normalized Cut

- Represent a partition S, $\bar{S}$ of V with $x \in \mathbb{R}^n$, where

$$x_i = \begin{cases} \dfrac{1}{\mathrm{vol}(S)} & \text{if } i \in S \\ -\dfrac{1}{\mathrm{vol}(\bar{S})} & \text{if } i \in \bar{S} \end{cases}$$

1. $x^{\mathrm{T}} L x = \sum_{ij} w_{ij} \left(x_i - x_j\right)^2 = \left(\frac{1}{\mathrm{vol}(S)} + \frac{1}{\mathrm{vol}(\bar{S})}\right)^2 \sum_{ij} w_{ij}$

$$= \left(\frac{1}{\mathrm{vol}(S)} + \frac{1}{\mathrm{vol}(\bar{S})}\right)^2 \mathrm{cut}\left(S, \overline{S}\right)$$

2. $x^{\mathrm{T}} D' x = \sum_i d_i (x_i)^2 = \sum_{i \in S} \frac{d_i}{\mathrm{vol}(S)^2} + \sum_{i \in \bar{S}} \frac{d_i}{\mathrm{vol}(\bar{S})^2} = \frac{1}{\mathrm{vol}(S)} + \frac{1}{\mathrm{vol}(\bar{S})}$

$$1 + 2 \Rightarrow \frac{x^{\mathrm{T}} L x}{x^{\mathrm{T}} D' x} = \mathrm{cut}(S, \bar{S}) \left(\frac{1}{\mathrm{vol}(S)} + \frac{1}{\mathrm{vol}(\bar{S})}\right) = \mathrm{ncut}(S, \bar{S})$$

Constrained optimization problem

- Minimize $x^{\mathrm{T}}Lx$ where $L = D' - W$

  subject to $x_i \in \left\{\frac{1}{\mathrm{vol}(S)}, -\frac{1}{\mathrm{vol}(\bar{S})}\right\}$,

  $x^{\mathrm{T}}D'x = 1$, and

  $\mathbf{1}^{\mathrm{T}}D'x = 0$
- Problem is NP-hard
- Note:
  - $\mathbf{1}^{\mathrm{T}}D'x = \sum_{i\in S}\frac{d_i}{\mathrm{vol}(S)} - \sum_{i\in\bar{S}}\frac{d_i}{\mathrm{vol}(\bar{S})} = 1 - 1 = 0$
  - $\frac{1}{\mathrm{vol}(S)}, -\frac{1}{\mathrm{vol}(\bar{S})}$ are not the only possible choices
    - See https://arxiv.org/abs/1311.2492

Relaxed Rayleigh quotient version

- Minimize $x^{\mathrm{T}}Lx$

  subject to $x^{\mathrm{T}}D'x = 1$ and $\mathbf{1}^{\mathrm{T}}D'x = 0$

Through the same reasoning as in graph partitioning problem, equivalently solve

- Minimize $y(D')^{-1/2}L(D')^{-1/2}y$

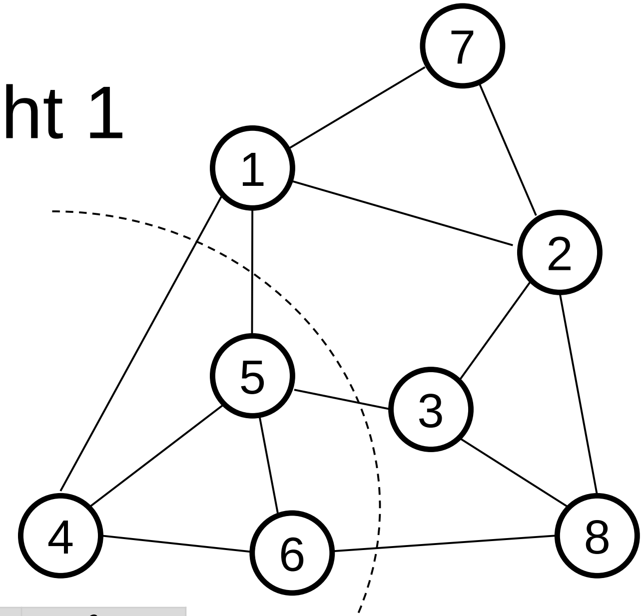
  subject to $y^{\mathrm{T}}y = 1$ and $\mathbf{1}^{\mathrm{T}}(D')^{1/2}y = 0$

  where $y = (D')^{1/2}x$

  - $(D')^{-1/2}L(D')^{-1/2}$ is now commonly known as the **normalized Laplacian**

Eigendecomposition

- Edges and vertices have weight 1



λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
1.6760	1.5100	1.42700	1.3100	0.9900	0.5880	0.4990	0.0

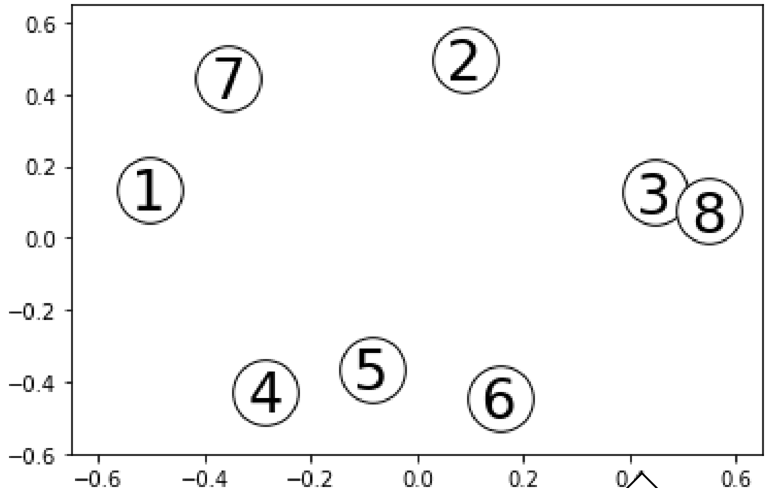
μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8
0.3485	0.0034	0.6240	-0.2451	-0.0704	-0.5023	0.1342	0.3922
-0.0304	0.6546	-0.3393	-0.2014	0.0768	0.0885	0.4973	0.3922
0.4129	-0.3896	-0.1906	-0.0484	-0.5545	0.4474	0.1265	0.3397
-0.2148	-0.2574	-0.4363	-0.5537	0.0989	-0.2859	-0.4286	0.3397
-0.4292	0.2801	0.1122	0.4236	-0.5021	-0.0836	-0.3638	0.3922
0.5058	0.1486	-0.0793	0.3598	0.4989	0.1541	-0.4454	0.3397
-0.1662	-0.4557	-0.2360	0.5096	0.2180	-0.3552	0.4457	0.2774
-0.4397	-0.2128	0.4406	-0.1475	0.3513	0.5487	0.0744	0.3397

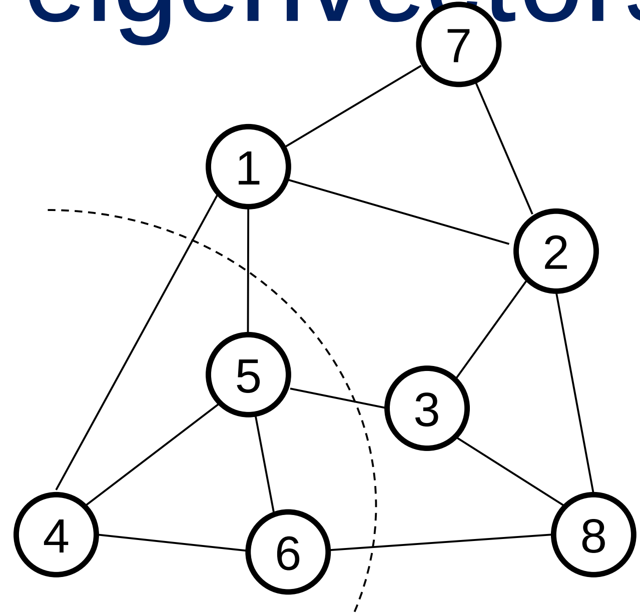
Shi and Malik (1997, 2000)

- Proposed the NP-hard ncut problem
- Found Laplacian for ncut
- Related ncut Laplacian to generalized eigenvalue system, resulting in the now ubiquitous **normalized Laplacian**
  - However, the first use of the generalized eigenvalue system for spectral clustering was in 1972
- Clustering with multiple eigenvectors (Shi and Malik 2000)

Clustering w/ multiple eigenvectors

- For normalized Laplacian



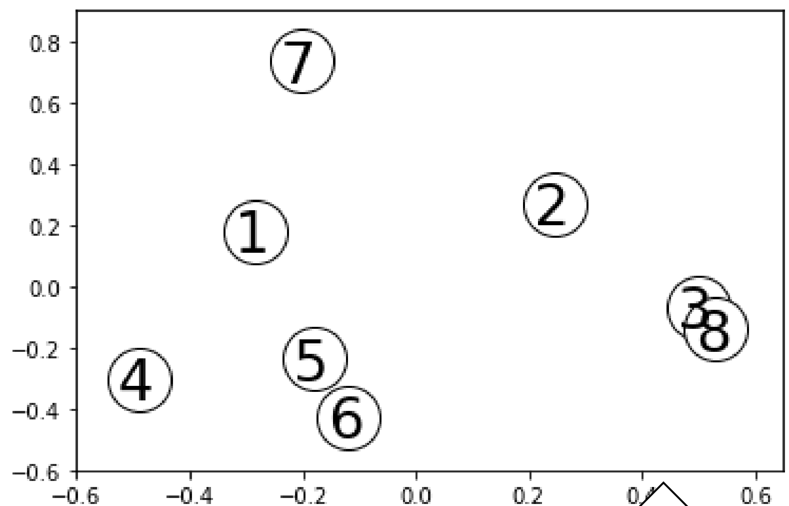


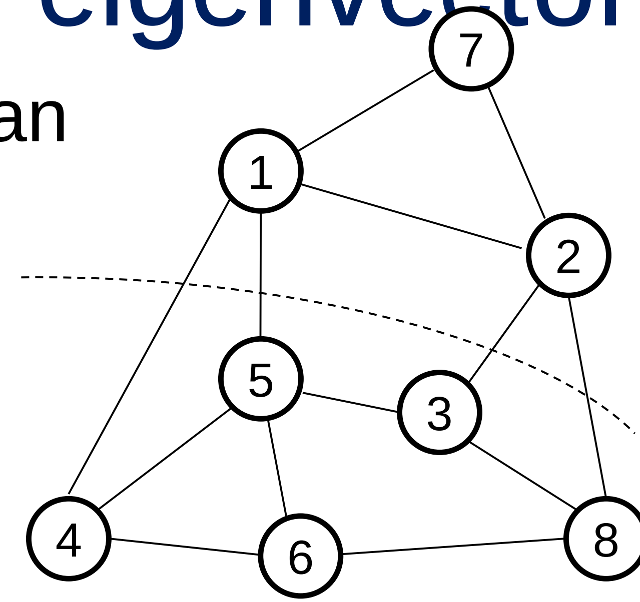
μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8
0.3485	0.0034	0.6240	-0.2451	-0.0704	-0.5023	0.1342	0.3922
-0.0304	0.6546	-0.3393	-0.2014	0.0768	0.0885	0.4973	0.3922
0.4129	-0.3896	-0.1906	-0.0484	-0.5545	0.4474	0.1265	0.3397
-0.2148	-0.2574	-0.4363	-0.5537	0.0989	-0.2859	-0.4286	0.3397
-0.4292	0.2801	0.1122	0.4236	-0.5021	-0.0836	-0.3638	0.3922
0.5058	0.1486	-0.0793	0.3598	0.4989	0.1541	-0.4454	0.3397
-0.1662	-0.4557	-0.2360	0.5096	0.2180	-0.3552	0.4457	0.2774
-0.4397	-0.2128	0.4406	-0.1475	0.3513	0.5487	0.0744	0.3397

Use the values from the top few eigenvectors for clustering (with, for example, *k*-means)

Clustering w/ multiple eigenvectors

- For graph partitioning Laplacian





Edges and vertices have weight 1

μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_6	μ_6
0.5677	-0.1583	-0.4862	0.3536	0.2315	-0.2855	0.1766	0.3536
-0.4281	0.6222	-0.2059	0.3536	0.0622	0.2469	0.2690	0.3536
0.3517	0.1203	0.2984	-0.3536	0.5170	0.5007	-0.0694	0.3536
-0.0855	0.0612	0.6267	0.3536	0.1159	-0.4899	-0.3044	0.3536
-0.5514	-0.3549	-0.3566	-0.3536	0.3216	-0.1795	-0.2392	0.3536
0.2351	0.3822	-0.2014	-0.3536	-0.5589	-0.1183	-0.4263	0.3536
-0.0354	-0.1476	0.2596	-0.3536	-0.2798	-0.2029	0.7349	0.3536
-0.0540	-0.5251	0.0654	0.3536	-0.4096	0.5286	-0.1411	0.3536

The resultant eigenvectors are less suitable for clustering

Theoretical justification

- Why does normalized Laplacian work
  - Probabilistic (random walk) justification by Maila and Shi (2000)
- Why does *k*-means on multiple eigenvectors work for the normalized Laplacian
  - Ng *et al.* (2001) show conditions for method to work

More clustering methods

- Based on the Fiedler vector
  - Sign cut or zero threshold cut
  - Median cut (ensures balance)
  - Sweep/criterion cut
    - Sort vertices by Fiedler vector values and cut at the lowest value of some cost function
  - Jump/gap cut
    - Sort vertices by Fiedler vector values and cut at the point of largest gap
- Based on multiple eigenvectors
  - Simultaneous *k*-way (Shi and Malik 2000)
  - *k*-means (Ng *et al.* 2001)

Equivalent Laplacian

- Ng *et al.* (2001) used $L' = D'^{-\frac{1}{2}}(W)D'^{-\frac{1}{2}}$ instead of the normalized Laplacian for analysis
  - $L' = I - L$ ($L =$ normalized Laplacian)

$$L = D'^{-1/2}(D' - W)D'^{-1/2}$$
$$= D'^{-1/2}(D')D'^{-\frac{1}{2}} - D'^{-\frac{1}{2}}(W)D'^{-\frac{1}{2}}$$
$$= I - D'^{-\frac{1}{2}}(W)D'^{-\frac{1}{2}} = I - L$$

- Results in the same eigenvectors but eigenvalues become $1 - \lambda_1, \dots, 1 - \lambda_k$