CS722/822: Machine Learning

Instructor: Jiangwen Sun
Computer Science Department

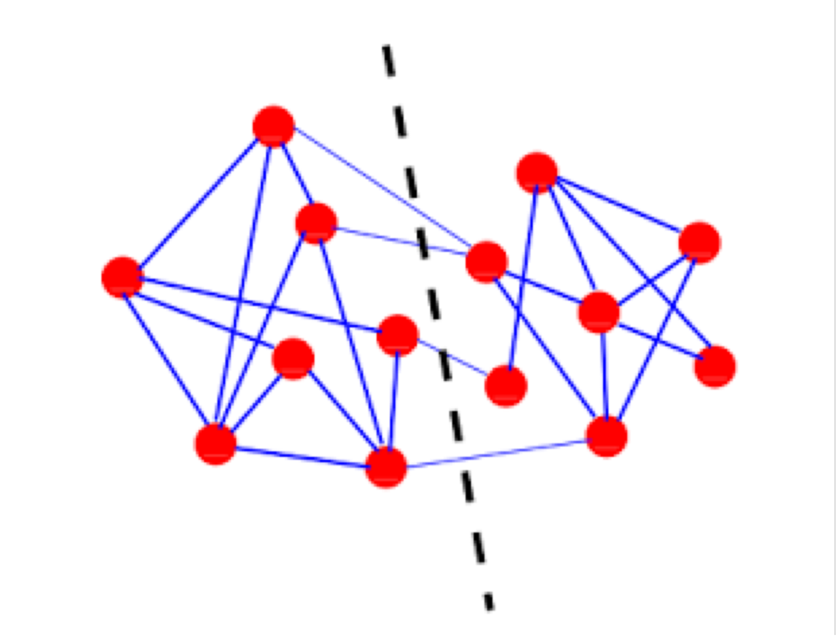
Spectral clustering

Outline

- Basics of graph cut
- Two-way cut
  - RatioCut
- k-way cut
  - RatioCut
  - Ncut
- Accuracy of spectral clustering

Graph Cut

- $G = (V, E)$ is a graph with n vertices and m edges
- Partition vertices into two groups, such that
  - Many within-group edges/vertices
  - Few cross-group edges

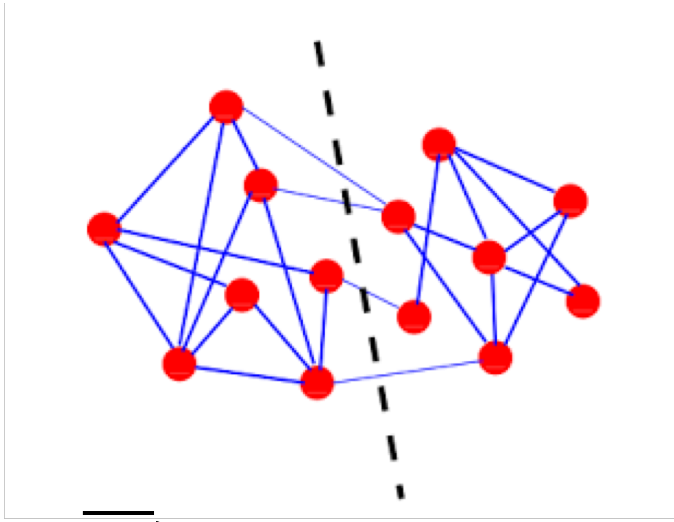


- Search all possible cuts
- Optimization problem

Formulating Graph Cut

- $V = A \cup \bar{A}$
- Weight of crossing-edges between A and $\bar{A}$:

$$W(A, \bar{A}) = \sum_{\substack{i \in A \\ j \in \bar{A}}} w_{ij}$$



- $\mathrm{cut}(A, \bar{A}) = W(A, \bar{A}) = \frac{1}{2} \sum_{i=1}^{2} W(A_i, \overline{A_i})$
- $\mathrm{cut}(A_1, A_2, \cdots, A_k) = \frac{1}{2} \sum_{i=1}^{k} W(A_i, \overline{A_i})$

RatioCut

- Minimize crossing edges
- Maximize within-group vertices

- $\mathrm{RatioCut}(A_1, \cdots, A_k) = \sum_{i=1}^{k} \frac{\mathrm{cut}(A_i, \overline{A_i})}{|A_i|}$

Insights into RatioCut

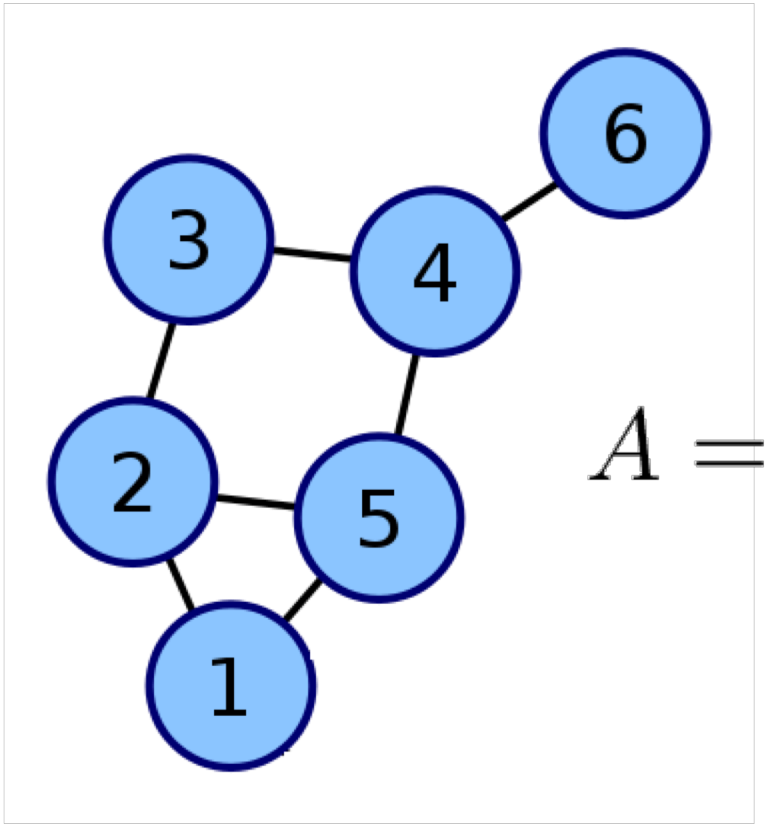
- Want Balanced Partition

- NP-Hard problem
  - For integer cluster labels $1, 2, \ldots, k$

- Spectral Clustering solves the *relaxed* version
  - Reformulate with real-valued vectors
  - Find optimal solution
  - Round to integer labels

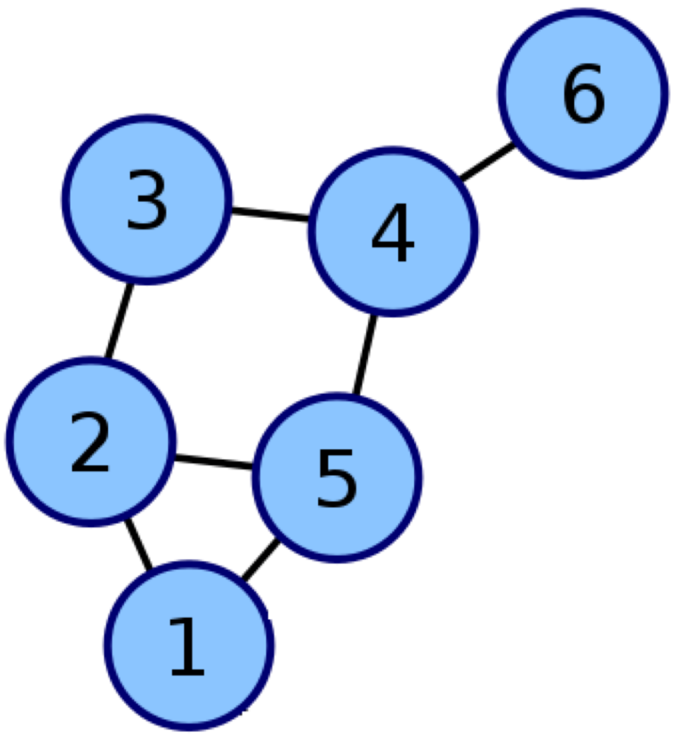
GRAPH LAPLACIAN

Adjacency Matrix



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

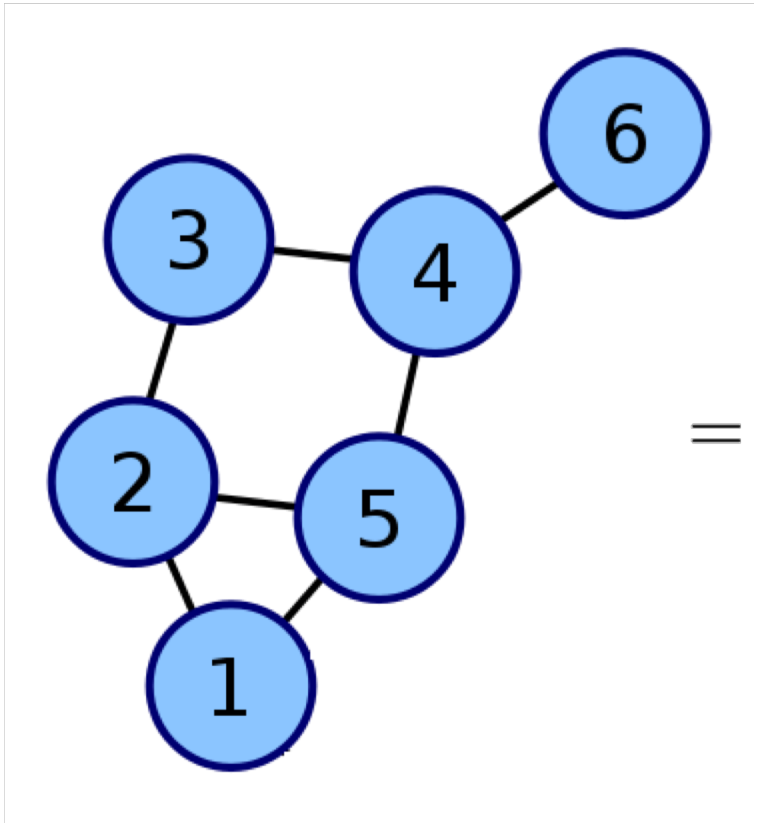
Degree Matrix



$$D = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$d_{ij} = \begin{cases} \sum_k a_{ik}, & i = j \\ 0, & i \neq j \end{cases}$$

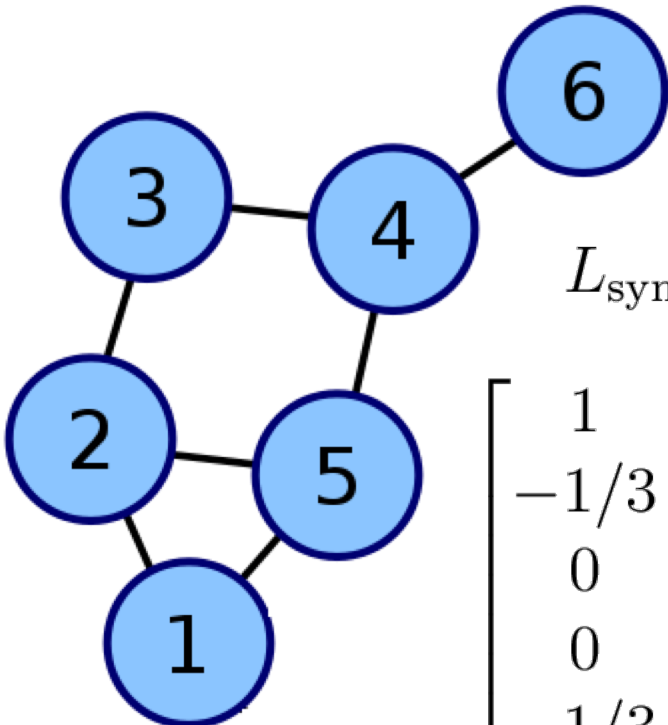
Laplacian Matrix



$$L = D - A$$

$$= \begin{bmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}$$

Normalized Laplacian Matrix



$$L_{\text{sym}} = D^{-1/2} L D^{-1/2} =$$

$$\begin{bmatrix} 1 & -1/2 & 0 & 0 & -1/2 & 0 \\ -1/3 & 1 & -1/3 & 0 & -1/3 & 0 \\ 0 & -1/2 & 1 & -1/2 & 0 & 0 \\ 0 & 0 & -1/3 & 1 & -1/3 & -1/3 \\ -1/3 & -1/3 & 0 & -1/3 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}$$

Edge Vector

$$b_e = b_{ij} = \begin{array}{c} \\ \left[\begin{matrix} \vdots \\ 1 \\ \vdots \\ -1 \\ \vdots \end{matrix}\right] \end{array} \begin{array}{l} \\ \leftarrow i \text{ th entry} \\ \\ \leftarrow j \text{ th entry} \\ \\ \end{array}$$

$$\begin{aligned} b_e^T x &= \begin{bmatrix} \cdots & 1 & \cdots & -1 & \cdots \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= (x_i - x_j) \end{aligned}$$

Edge Laplacians

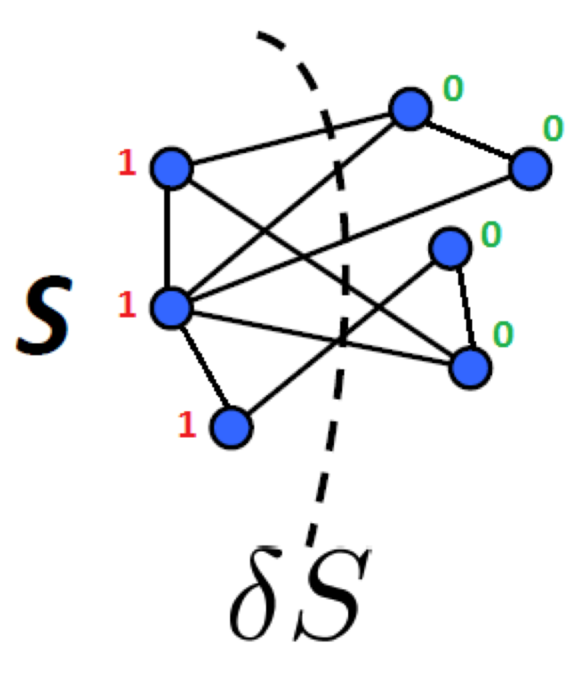
$$L_e = b_e b e^T = \begin{bmatrix} \ddots & \cdots & \cdots & \cdots & \vdots \\ \vdots & 1 & \cdots & -1 & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & -1 & \cdots & 1 & \vdots \\ \vdots & \cdots & \cdots & \cdots & \ddots \end{bmatrix}$$

$$\begin{aligned} L &= \textstyle\sum_e w_e L_e \\ &= \textstyle\sum_e w_e b_e b_e^T \end{aligned}$$

Laplacian Quadratic Form, $x^T Lx$

$$
\begin{aligned}
& x^T L x \\
&= x^T \left(\sum_{e=(i,j)} w_e b_e b_e^T \right) x \\
&= \sum_{e=(i,j)} (x^T b_e) w_e (b_e^T x) \\
&= \sum_{e=(i,j)} (b_e^T x)^T w_e (b_e^T x) \\
&= \sum_{e=(i,j)} (x_i - x_j) w_e (x_i - x_j) \\
&= \sum_{e=(i,j)} w_e (x_i - x_j)^2
\end{aligned}
$$

If $x \in \{0,1\}^n$ denotes a vertex partition,
$x^T Lx$ = cut-weight



How this relates to the RatioCut?

Recall Ratio Cut

- Minimize crossing edges
- Maximize in-part vertices

- $\mathrm{RatioCut}(A_1, \cdots, A_k)$: $\min \sum_{i=1}^{k} \frac{\mathrm{cut}(A_i, \overline{A_i})}{|A_i|}$

RATIOCUT FOR $k = 2$

RatioCut to Laplacian Quadratic Form (1)

$$
\begin{aligned}
&\mathrm{RatioCut}(\mathrm{A},\bar{\mathrm{A}}) \cdot |\mathrm{V}| \\
&= \left(\frac{\mathrm{cut}(\mathrm{A},\bar{\mathrm{A}})}{|\mathrm{A}|} + \frac{\mathrm{cut}(\bar{\mathrm{A}},\mathrm{A})}{|\bar{\mathrm{A}}|}\right) \cdot |\mathrm{V}| \\
&= \mathrm{cut}(\mathrm{A},\bar{\mathrm{A}}) \cdot \left(\frac{|\mathrm{V}|}{|\mathrm{A}|} + \frac{|\mathrm{V}|}{|\bar{\mathrm{A}}|}\right) \\
&= \mathrm{cut}(\mathrm{A},\bar{\mathrm{A}}) \cdot \left(\frac{|\mathrm{A}| + |\bar{\mathrm{A}}|}{|\mathrm{A}|} + \frac{|\mathrm{A}| + |\bar{\mathrm{A}}|}{|\bar{\mathrm{A}}|}\right) \\
&= \mathrm{cut}(\mathrm{A},\bar{\mathrm{A}}) \cdot \left(\frac{|\bar{\mathrm{A}}|}{|\mathrm{A}|} + \frac{|\mathrm{A}|}{|\bar{\mathrm{A}}|} + 2\right)
\end{aligned}
$$

RatioCut to Laplacian Quadratic Form (2)

$$
\begin{aligned}
&\mathrm{RatioCut}(\mathrm{A}, \bar{\mathrm{A}}) \cdot |\mathrm{V}| \\
&= \left(\sum_{\mathrm{i} \in \mathrm{A}, \mathrm{j} \in \bar{\mathrm{A}}} \mathrm{w_{ij}} \right) \cdot \left(\frac{|\bar{\mathrm{A}}|}{|\mathrm{A}|} + \frac{|\mathrm{A}|}{|\bar{\mathrm{A}}|} + 2 \right) \\
&= \left(\frac{1}{2} \sum_{\mathrm{i} \in \mathrm{A}, \mathrm{j} \in \bar{\mathrm{A}}} \mathrm{w_{ij}} + \frac{1}{2} \sum_{\mathrm{i} \in \mathrm{A}, \mathrm{j} \in \bar{\mathrm{A}}} \mathrm{w_{ij}} \right) \cdot \left(\sqrt{\frac{|\bar{\mathrm{A}}|}{|\mathrm{A}|}} + \sqrt{\frac{|\mathrm{A}|}{|\bar{\mathrm{A}}|}} \right)^2 \\
&= \frac{1}{2} \sum_{\mathrm{i} \in \mathrm{A}, \mathrm{j} \in \bar{\mathrm{A}}} \mathrm{w_{ij}} \left(\sqrt{\frac{|\bar{\mathrm{A}}|}{|\mathrm{A}|}} + \sqrt{\frac{|\mathrm{A}|}{|\bar{\mathrm{A}}|}} \right)^2 + \frac{1}{2} \sum_{\mathrm{i} \in \mathrm{A}, \mathrm{j} \in \bar{\mathrm{A}}} \mathrm{w_{ij}} \left(-\sqrt{\frac{|\bar{\mathrm{A}}|}{|\mathrm{A}|}} - \sqrt{\frac{|\mathrm{A}|}{|\bar{\mathrm{A}}|}} \right)^2
\end{aligned}
$$

RatioCut to Laplacian Quadratic Form (3)

$$\text{Let} \quad f_i = \begin{cases} \sqrt{\frac{|\bar{A}|}{|A|}} & \text{if } v_i \in A \\ -\sqrt{\frac{|A|}{|\bar{A}|}} & \text{if } v_i \in \bar{A} \end{cases}$$

Then

$$\begin{aligned}
&\text{RatioCut}(A, \bar{A}) \cdot |V| \\
&= \frac{1}{2} \sum_{i \in A, j \in \bar{A}} w_{ij}(f_i - f_j)^2 + \frac{1}{2} \sum_{i \in A, j \in \bar{A}} w_{ij}(f_j - f_i)^2 \\
&= \frac{1}{2} \sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2 \\
&= \frac{1}{2} \cdot \left(2 \cdot \sum_{i \ \leq j}^{n} w_{ij}(f_i - f_j)^2 \right) \\
&= f^T L f
\end{aligned}$$

What Did We Do?

- We showed that $\min \mathrm{RatioCut} \equiv \min f^T L f$

- Properties of f:
  - f is orthogonal to the constant, $f \perp 1$

$$f^T 1 = \sum f_i = |A| \cdot \sqrt{\frac{|\bar{A}|}{|A|}} + |\bar{A}| \cdot (-\sqrt{\frac{|A|}{|\bar{A}|}}) = \sqrt{|A||\bar{A}|} - \sqrt{|A||\bar{A}|} = 0$$

  - Length of f is $\sqrt{n}$

$$\|f\|^2 = f^T f = |A| \cdot \frac{|\bar{A}|}{|A|} + |\bar{A}| \cdot \frac{|A|}{|\bar{A}|} = |A| + |\bar{A}| = n$$

Reformulated Minimization Problem

- Discrete optimization

$$\min_{A} f' L f, \qquad \text{subject to } f \perp 1, \left||f|\right| = \sqrt{n}$$

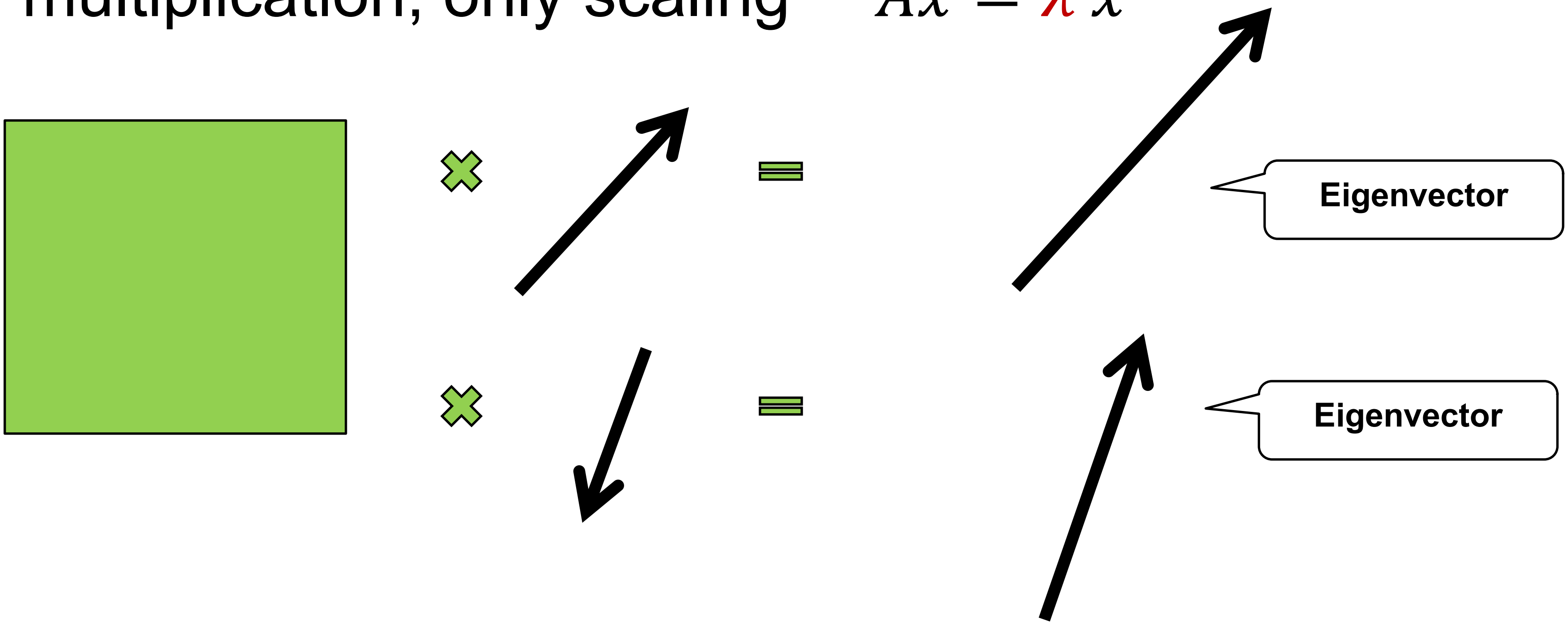
- Relax to arbitrary real vectors
  - because Integer programming is NP-hard

$$\min_{f \in \mathbb{R}^n} f' L f, \qquad \text{subject to } f \perp 1, \left||f|\right| = \sqrt{n}$$

Eigenvalues and Eigenvectors

$A \leftarrow n \times n$ matrix, $x \leftarrow n \times 1$ vector

Eigenvector: direction same/negative after multiplication, only scaling $Ax = \lambda x$



Eigenvalue: amount of distortion $\lambda = x^T A x / x^T x$

Rayleigh Quotient and Eigenvalues

$$q(A,x) = \frac{x^T A x}{x^T x}$$

- S_1= Set of all vectors
  - Vector v_1 attains minimum $\lambda_1 = q(A,x)\ \forall x \in\ S_1$
  - Set of first smallest eigenvectors
- S_2= Set of all vectors $\perp\ \{v_1\}$
  - Vector v_2 attains minimum $\lambda_2 = q(A,x)\ \forall x \in\ S_2$
  - Set of second smallest eigenvectors

…

- S_n= Set of all vectors $\perp \{v_1, v_2, \cdots, v_{n-1}\}$
  - v_n attains minimum $\lambda_n = q(A,x)\ \forall x \in\ S_n$
  - Set of n-th smallest eigenvectors

$$\lambda_1 < \lambda_2 < \cdots < \lambda_n$$

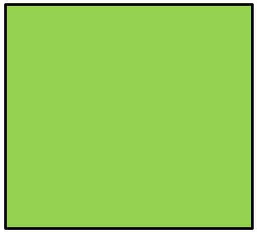
Rayleigh-Ritz Theorem (simplified)

- the k-th smallest eigenvector v_k is the n-dimensional vector x
- such that
  - x is perpendicular to all previous eigenvectors

    and
  - x minimizes the quantity $\frac{x^T A x}{x^T x}$

Smallest Eigenvalue of the Laplacian

- Rows/columns of L sum to 0



$$\times \begin{bmatrix} \mathbf{1} \\ \vdots \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} = 0 \times \begin{bmatrix} \mathbf{1} \\ \vdots \\ \mathbf{1} \end{bmatrix}$$

- $\lambda_1 = 0$ and $v_1 = 1$, the constant vector

- Any constant vector is the first eigenvector

Second Eigenvector of the Laplacian

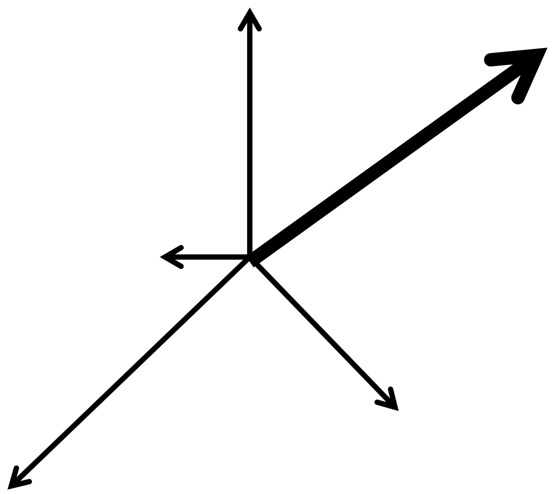
- Since eigenvectors are perpendicular to each other,
  - $v_2 \perp 1$
  - "Fiedler vector"… named after Miroslav Fiedler

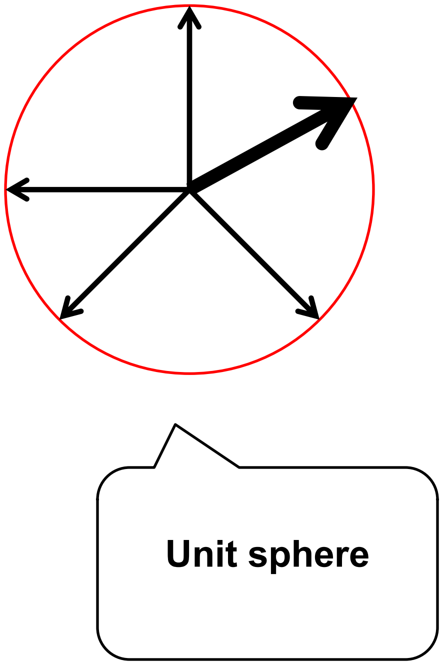
- $\lambda_2 = \text{argmin} \ (x^T Lx / \ x^T x)$ over all $x \in R^n, x \perp 1$

- Condition similar to the vector f chosen when minimizing RatioCut

Eigenvector and Scaling

- Consider an eigenvector





- Scale the vector to length 1
  - Still eigenvector
  - With the same eigenvalue
- Find unit-length eigenvectors on the unit-sphere

v_2 and Laplacian Quadratic Form

- $\lambda_2 = \text{argmin} \ \ (x^T Lx / \ x^T x)$ over all $x \in R^n, x \ \perp 1$

$$\Downarrow$$

- $\lambda_2 = \text{argmin} \ \ (x^T Lx)$ over all $x \in R^n, x \ \perp 1, x^T x = 1$

- $\min_{f \in \mathbb{R}^n} f^T Lf, \text{ subject to } f \perp 1, ||f|| = 1$

$$\Uparrow$$

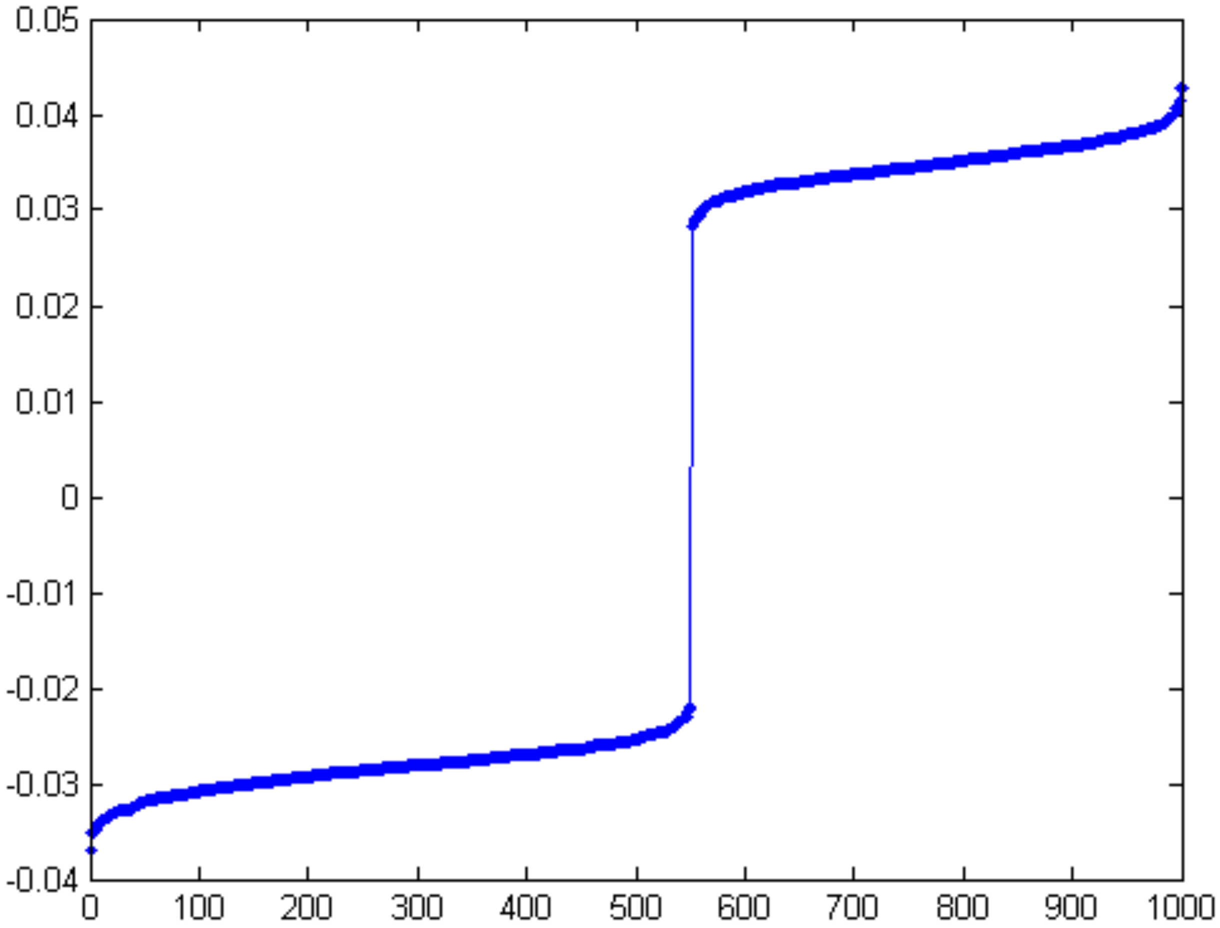
- $\min_{f \in \mathbb{R}^n} f^T Lf, \text{ subject to } f \perp 1, ||f|| = \sqrt{n}$

Big Picture

- $\text{Minimizing RatioCut} \equiv \text{Finding } v_2 \text{ of L}$
  - v_2 is the eigenvector corresponding to the second smallest eigenvalue, λ_2
- Clustering can be made based on v_2
  - Simply by rounding to integer labels, or
  - Running k-means

This is a mapping from n-dimension to 1-dimension

Plot of the Second Eigenvector



RATIOCUT FOR $k \geq 3$

Partition Indicator Vectors (1)

- Partitions $A_1, A_2, \dots, A_k$
- Partition A_j has its indicator vector h_j

$$h_j = \begin{bmatrix} 0 \\ \vdots \\ 1/\sqrt{|A_j|} \\ 0 \\ \vdots \\ 0 \\ 1/\sqrt{|A_j|} \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \\ \\ \leftarrow \text{ This vertex in } A_j \\ \\ \\ \\ \leftarrow \text{ This vertex in } A_j \\ \\ \\ \end{matrix}$$

Partition Indicator Vectors (2)

- Note that $$h_j^T h_j = |A_j| \cdot (1/\sqrt{|A_j|})^2 = 1$$
- Note that $$h_i^T h_j = 0 \text{ for all } i \neq j$$
- Thus vectors $h_1, h_2, \cdots, h_k$ are orthonormal

$$h_j = \begin{bmatrix} 0 \\ \vdots \\ 1/\sqrt{|A_j|} \\ 0 \\ \vdots \\ 0 \\ 1/\sqrt{|A_j|} \\ \vdots \\ 0 \end{bmatrix} \begin{array}{l} \\ \\ \leftarrow \text{ This vertex in } A_j \\ \\ \\ \\ \leftarrow \text{ This vertex in } A_j \\ \\ \\ \end{array}$$

RatioCut for k>2

- It can be shown that

$$h_i^T L h_i = \frac{\mathrm{cut}(A_i, \bar{A}_i)}{|A_i|}$$

- Then

$$\mathrm{RatioCut}(A_1, \cdots, A_k) = \sum_{i=1}^{k} \frac{\mathrm{cut}(A_i, \overline{A_i})}{|A_i|} = \sum_i h_i^T L h_i = \mathrm{Trace}(H^T L H)$$

RatioCut for k>2

- Now we have a trace minimization problem

$$\min_{A_1,\ldots,A_k} \operatorname{Tr}(H'LH) \text{ subject to } H'H = I$$

- Relax to take real values

$$\min_{H \in \mathbb{R}^{n \times k}} \operatorname{Tr}(H'LH) \text{ subject to } H'H = I.$$

- The minimum objective is the sum of the k smallest eigenvectors of L

- So each optimal h_i is the eigenvector associated with the i-th smallest eigenvalue?

RatioCut for k>2

- Now we know the $n \times k$ matrix H
  - by computing k smallest eigenvectors of L
- Treat each row of H as a k-dimensional mapping of the corresponding n-dimensional point.
- Use `k-means` to partition rows of H
- Assign cluster label to each vertex

NCUT

Normalized Cut

- Minimize # of crossing edges
- Maximize # of within-group edges

- $\mathrm{NCut}(A_1, \cdots, A_k) = \sum_{i=1}^{k} \frac{\mathrm{cut}(A_i, \overline{A_i})}{\mathrm{vol}(A_i)}$

where

- $\mathrm{vol}(v) = \deg(v)$
- $\mathrm{vol}(A) = \sum_{v \in A} \deg(v)$

Cluster Indicator Vectors ($k = 2$ case)

- RatioCut

$$f_i = \begin{cases} \sqrt{|\overline{A}|/|A|} & \text{if } v_i \in A \\ -\sqrt{|A|/|\overline{A}|} & \text{if } v_i \in \overline{A}. \end{cases}$$

  - $f \perp 1$
  - $f^T f = n$

- Ncut

$$f_i = \begin{cases} \sqrt{\frac{\mathrm{vol}(\overline{A})}{\mathrm{vol}\, A}} & \text{if } v_i \in A \\ -\sqrt{\frac{\mathrm{vol}(A)}{\mathrm{vol}(\overline{A})}} & \text{if } v_i \in \overline{A}. \end{cases}$$

  - $Df \perp 1$
  - $f^T D f = \mathrm{vol}(V)$

Minimization Problem

- minimize $f^T L f$

  over all $f \in R^n$

subject to

- $Df \perp 1$
- $f^T D f = \mathrm{vol(V)}$

- *Constraints different from* RatioCut

Work with $g = D^{1/2} f$

- Let $\mathrm{d_i} = \deg(v_i)$
- Let $g = D^{1/2} f$, that is $g_i = \sqrt{d_i} \cdot f_i$
- Then $f = D^{-\frac{1}{2}} g$
- And

$$f^T L f = \left(D^{-\frac{1}{2}} g\right)^T L \left(D^{-\frac{1}{2}} g\right) \quad = g^T \left(D^{-\frac{1}{2}} L D^{-\frac{1}{2}}\right) g$$

- And $f^T D f = g^T g = \mathrm{vol}(V)$
- $Df \perp 1 \equiv g \perp D^{1/2} 1$

Relaxed Minimization Problem with g

- minimize $\quad g^T \left(D^{-\frac{1}{2}} L D^{-\frac{1}{2}}\right) g$
  over all $g \in R^n$

subject to

  - $g \perp D^{1/2} 1$
  - $g^T g = \mathrm{vol(V)}$

- $D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$ is the *normalized* Laplacian, L_{sym}
- $D^{1/2} 1$ is the first eigenvector of L_{sym} with eigenvalue 0

Minimization problem with f

- minimize $\quad f^T L f$
  over all $f \in R^n$

subject to

- $Df \perp 1$
- $f^T D f = \mathrm{vol(V)}$

Ncut

- Like RatioCut, the solution to NCut is the second eigenvector of L_{sym}

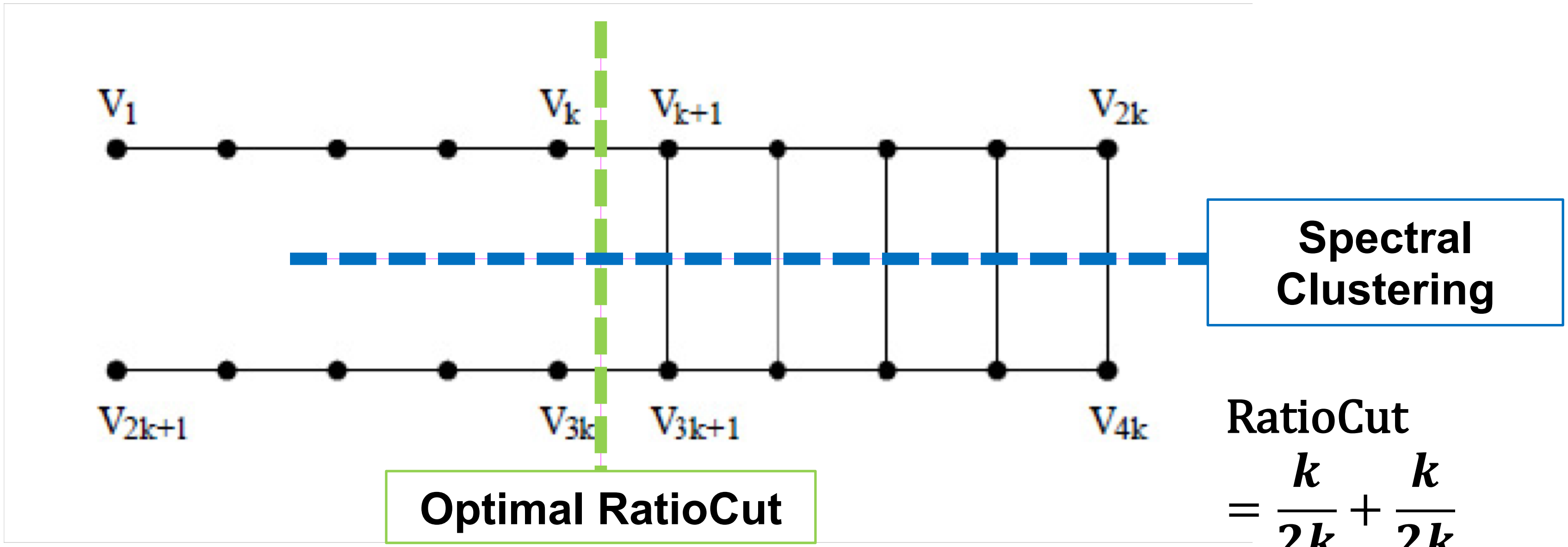
- Analysis for $k \geq 3$ is similar to that of RatioCut

ACCURACY OF SPECTRAL CLUSTERING

Big Picture

- Finding balanced cut is NP-hard
- Relax the integer cluster label vectors to real-valued ones
- Solve the relaxed balanced cut problem, which is equivalent to an eigen decomposition problem
- The relaxation not guaranteed to give optimal solution

Badness of Spectral Clustering



$$\text{RatioCut} = \frac{k}{2k} + \frac{k}{2k} = 1$$

$$\text{RatioCut} = \frac{2}{2k} + \frac{2}{2k} = \frac{2}{k}$$

Spectral Clustering result is $k/2$ times worse than optimal result