Machine Learning by ambedkar@IISc

- Spectral Clustering

Spectral Methods

What is....?

What are spectral methods?

- Underlying objects in a problem can be represented as matrices
- Eigenvalues and eigenvectors of these matrices become clue to a solution.

What are eigenvalues and vectors?

- $\lambda \in \mathbb{C}$ is said to be an eigenvalue of $n \times n$ matrix M if it satisfies $Mv = \lambda v$ for $v \neq 0$.
- v said to be eigenvector of M corresponding to λ.

Can eigenvalues and eigenvectors make a person rich?

- Yes!

- Google page rank algorithm

- Must read: (`K. Bryan and T. Leise, $25,000,000,000 Eigenvector: The Linear Algebra behind Google, SIAM review, 2006`)

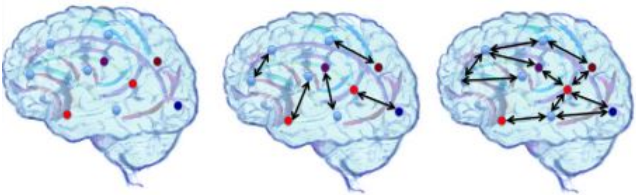
Human Brain

Credit: *Christiaan Vermeleun, www.td.org.*

Human Brain

- Possibly the most complex network known to man
- 100 billion neurons (nodes)
- 100 trillion connections (edges)
- *How can we go about making sense of all this?*

Understanding Human Brain

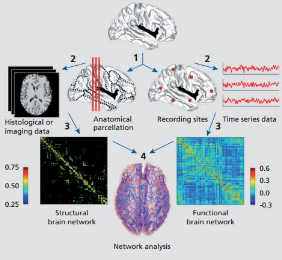


Credit: *Stam et. al, "The organization of physiological brain networks.", Clinical neurophysiology*

- One viewpoint: Study the brain from a network science perspective.
- Model the structural/functional connectivity of brain regions as "Brain Networks"[1].
- Lot of data to work with: fMRI, EEG, MEG etc.

[1] Park and Friston, Science, 2013

Brain Networks: Community Structure

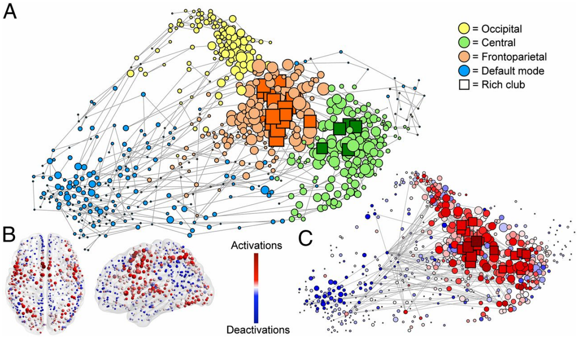


dit: *Sporns, 2013*

- A common property of Brain Networks is **segregation** of neurons based on anatomical or functional characteristics[a]
- In graph theory framework, this community structure can be studied with **cluster analysis**.

[a] (Sporns, 2013)

Clustering over Brain Networks

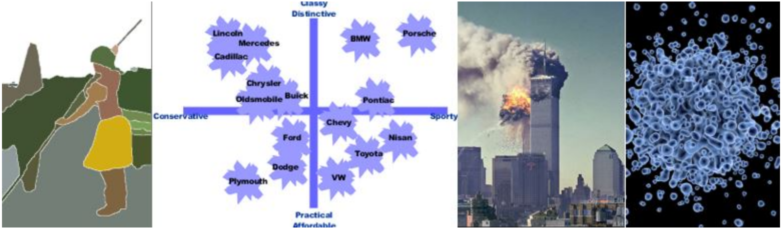


Credit[2]

- A: Functional coactivation network - Different 'Functional' Clusters
- B, C: Red Nodes represent the 'hub' nodes in the network

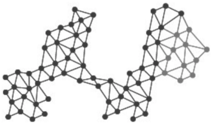
[2] Crossley et al. "Cognitive relevance of the community structure of the human brain functional coactivation network." PNAS (2013)

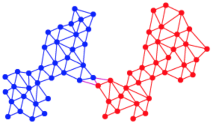
Clustering over Networks: Applications



- Image segmentation
- Market segmentation in consumer/business networks
- Detection of Terrorist Groups in Online Social Networks
- Epidemic spreading on networks

Graph Partitioning[3]

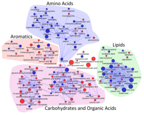




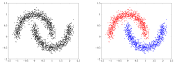
Objective:

- High connectivity within clusters
- Few edges across clusters (small cut)
- Balanced partitions

Applications:



Network partitioning



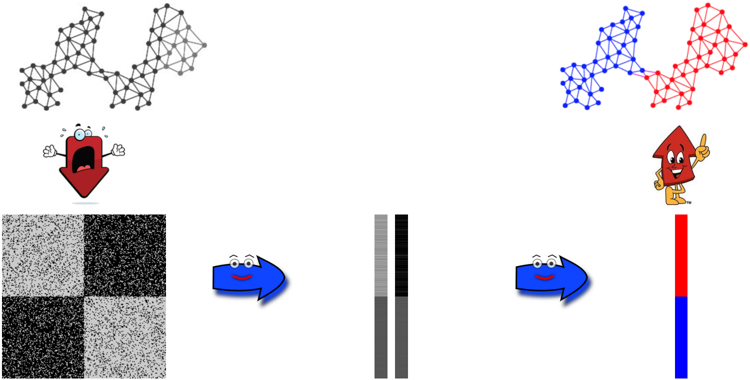
Data clustering

Image segmentation

[3] Drawings and pictures are borrowed from Debarghya

Spectral Graph partitioning[4]

Input Graph | Good balanced cut

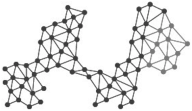


(Normalized) Adjacency matrix | Find k dominant eigenvectors | Run k-means on rows

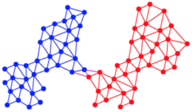
[4]Drawings and pictures are borrowed from Debarghya

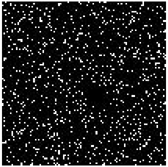
Spectral Graph partitioning [5]

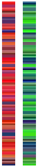
Input Graph



Good balanced cut







(Normalized)
Adjacency matrix

Find k dominant
eigenvectors

Run k-means
on rows

[5] Drawings and pictures borrowed from Debarghya

A quick LA recall

M a real valued $n \times n$ matrix.

- $\lambda \in \mathbb{C}$ is said to be eigenvalue of M if it satisfies $Mv = \lambda v$ for $v \neq 0$. v said to be eigenvector of M.
- Spectrum of M is the set of eigenvalues along with their multiplicities.

M a real valued $n \times n$ *symmetric* matrix

- If u, v are eigenvectors of distinct eigenvalues then u and v are orthogonal.
- Eigenvalues of M are real
- M is diagonalizable (there exists an invertible matrix P such that $P^{-1}MP$ is diagonal)
- There exists L such that $LL^T = L^T L = I$ such that LAL^T is diagonal.

A quick LA recall

M a real valued $n \times n$ matrix.

- $\lambda \in \mathbb{C}$ is said to be eigenvalue of M if it satisfies $Mv = \lambda v$ for $v \neq 0$. v said to be eigenvector of M.
- Spectrum of M is the set of eigenvalues along with their multiplicities.

M a real valued $n \times n$ *symmetric* matrix

- If u, v are eigenvectors of distinct eigenvalues then u and v are orthogonal.
- Eigenvalues of M are real
- M is diagonalizable (there exists an invertible matrix P such that $P^{-1}MP$ is diagonal)
- There exists L such that $LL^T = L^TL = I$ such that LAL^T is diagonal.

A quick LA recall

M a real valued $n \times n$ matrix.

- $\lambda \in \mathbb{C}$ is said to be eigenvalue of M if it satisfies $Mv = \lambda v$ for $v \neq 0$. v said to be eigenvector of M.
- Spectrum of M is the set of eigenvalues along with their multiplicities.

M a real valued $n \times n$ *symmetric* matrix

- If u, v are eigenvectors of distinct eigenvalues then u and v are orthogonal.
- Eigenvalues of M are real
- M is diagonalizable (there exists an invertible matrix P such that $P^{-1}MP$ is diagonal)
- There exists L such that $LL^T = L^TL = I$ such that LAL^T is diagonal.

A quick LA recall

M a real valued $n \times n$ matrix.

- $\lambda \in \mathbb{C}$ is said to be eigenvalue of M if it satisfies $Mv = \lambda v$ for $v \neq 0$. v said to be eigenvector of M.
- Spectrum of M is the set of eigenvalues along with their multiplicities.

M a real valued $n \times n$ *symmetric* matrix

- If u, v are eigenvectors of distinct eigenvalues then u and v are orthogonal.
- Eigenvalues of M are real
- M is diagonalizable (there exists an invertible matrix P such that $P^{-1}MP$ is diagonal)
- There exists L such that $LL^T = L^T L = I$ such that LAL^T is diagonal.

A quick LA recall

M a real valued $n \times n$ matrix.

- $\lambda \in \mathbb{C}$ is said to be eigenvalue of M if it satisfies $Mv = \lambda v$ for $v \neq 0$. v said to be eigenvector of M.
- Spectrum of M is the set of eigenvalues along with their multiplicities.

M a real valued $n \times n$ *symmetric* matrix

- If u, v are eigenvectors of distinct eigenvalues then u and v are orthogonal.
- Eigenvalues of M are real
- M is diagonalizable (there exists an invertible matrix P such that $P^{-1}MP$ is diagonal)
- There exists L such that $LL^T = L^TL = I$ such that LAL^T is diagonal.

A quick LA recall

M a real valued $n \times n$ matrix.

- $\lambda \in \mathbb{C}$ is said to be eigenvalue of M if it satisfies $Mv = \lambda v$ for $v \neq 0$. v said to be eigenvector of M.
- Spectrum of M is the set of eigenvalues along with their multiplicities.

M a real valued $n \times n$ *symmetric* matrix

- If u, v are eigenvectors of distinct eigenvalues then u and v are orthogonal.
- Eigenvalues of M are real
- M is diagonalizable (there exists an invertible matrix P such that $P^{-1}MP$ is diagonal)
- There exists L such that $LL^T = L^TL = I$ such that LAL^T is diagonal.

A quick LA recall

M a real valued $n \times n$ matrix.

- $\lambda \in \mathbb{C}$ is said to be eigenvalue of M if it satisfies $Mv = \lambda v$ for $v \neq 0$. v said to be eigenvector of M.
- Spectrum of M is the set of eigenvalues along with their multiplicities.

M a real valued $n \times n$ *symmetric* matrix

- If u, v are eigenvectors of distinct eigenvalues then u and v are orthogonal.
- Eigenvalues of M are real
- M is diagonalizable (there exists an invertible matrix P such that $P^{-1}MP$ is diagonal)
- There exists L such that $LL^T = L^TL = I$ such that LAL^T is diagonal.

A quick LA recall

M a real valued $n \times n$ matrix.

- $\lambda \in \mathbb{C}$ is said to be eigenvalue of M if it satisfies $Mv = \lambda v$ for $v \neq 0$. v said to be eigenvector of M.
- Spectrum of M is the set of eigenvalues along with their multiplicities.

M a real valued $n \times n$ *symmetric* matrix

- If u, v are eigenvectors of distinct eigenvalues then u and v are orthogonal.
- Eigenvalues of M are real
- M is diagonalizable (there exists an invertible matrix P such that $P^{-1}MP$ is diagonal)
- There exists L such that $LL^T = L^TL = I$ such that LAL^T is diagonal.

Some matrices related to graphs

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$.

- **Adjacency Matrix:** $A \in \mathbb{R}^{n \times n}$ such that

$$A_{ij} = \begin{cases} 0 & \text{if} \quad i = j, \\ 1 & \text{if} \quad (i,j) \in E, \\ 0 & \text{if} \quad (i,j) \notin E. \end{cases}$$

- **Degree Matrix:** $D \in \mathbb{R}^{n \times n}$ is diagonal matrix such that $D_{ii} = \deg(i)$
- **Incidence Matrix:** $B \in \mathbb{R}^{n \times e}$, where rows indexed by vertices and columns indexed by edges and $B_{ij} = 1$ if vertex i lies on edge j.
- **Laplacian Matrix:** $L \in \mathbb{R}^{n \times n}$ is defined as $L = D - A$
- **Normalized Laplacian:** $L \in \mathbb{R}^{n \times n}$ is defined as

Some matrices related to graphs

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$.

- **Adjacency Matrix:** $A \in \mathbb{R}^{n \times n}$ such that

$$A_{ij} = \begin{cases} 0 & \text{if} \quad i = j, \\ 1 & \text{if} \quad (i,j) \in E, \\ 0 & \text{if} \quad (i,j) \notin E. \end{cases}$$

- **Degree Matrix:** $D \in \mathbb{R}^{n \times n}$ is diagonal matrix such that $D_{ii} = \deg(i)$
- **Incidence Matrix:** $B \in \mathbb{R}^{n \times e}$, where rows indexed by vertices and columns indexed by edges and $B_{ij} = 1$ if vertex i lies on edge j.
- **Laplacian Matrix:** $L \in \mathbb{R}^{n \times n}$ is defined as $L = D - A$
- **Normalized Laplacian:** $L \in \mathbb{R}^{n \times n}$ is defined as

Some matrices related to graphs

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$.

- **Adjacency Matrix:** $A \in \mathbb{R}^{n \times n}$ such that

$$A_{ij} = \begin{cases} 0 & \text{if} \quad i = j, \\ 1 & \text{if} \quad (i,j) \in E, \\ 0 & \text{if} \quad (i,j) \notin E. \end{cases}$$

- **Degree Matrix:** $D \in \mathbb{R}^{n \times n}$ is diagonal matrix such that $D_{ii} = \deg(i)$
- **Incidence Matrix:** $B \in \mathbb{R}^{n \times e}$, where rows indexed by vertices and columns indexed by edges and $B_{ij} = 1$ if vertex i lies on edge j.
- **Laplacian Matrix:** $L \in \mathbb{R}^{n \times n}$ is defined as $L = D - A$
- **Normalized Laplacian:** $L \in \mathbb{R}^{n \times n}$ is defined as

Some matrices related to graphs

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$.

- **Adjacency Matrix:** $A \in \mathbb{R}^{n \times n}$ such that

$$A_{ij} = \begin{cases} 0 & \text{if} \quad i = j, \\ 1 & \text{if} \quad (i, j) \in E, \\ 0 & \text{if} \quad (i, j) \notin E. \end{cases}$$

- **Degree Matrix:** $D \in \mathbb{R}^{n \times n}$ is diagonal matrix such that $D_{ii} = \deg(i)$
- **Incidence Matrix:** $B \in \mathbb{R}^{n \times e}$, where rows indexed by vertices and columns indexed by edges and $B_{ij} = 1$ if vertex i lies on edge j.
- **Laplacian Matrix:** $L \in \mathbb{R}^{n \times n}$ is defined as $L = D - A$
- **Normalized Laplacian:** $L \in \mathbb{R}^{n \times n}$ is defined as

Some matrices related to graphs

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$.

- **Adjacency Matrix:** $A \in \mathbb{R}^{n \times n}$ such that

$$A_{ij} = \begin{cases} 0 & \text{if} \quad i = j, \\ 1 & \text{if} \quad (i, j) \in E, \\ 0 & \text{if} \quad (i, j) \notin E. \end{cases}$$

- **Degree Matrix:** $D \in \mathbb{R}^{n \times n}$ is diagonal matrix such that $D_{ii} = \deg(i)$
- **Incidence Matrix:** $B \in \mathbb{R}^{n \times e}$, where rows indexed by vertices and columns indexed by edges and $B_{ij} = 1$ if vertex i lies on edge j.
- **Laplacian Matrix:** $L \in \mathbb{R}^{n \times n}$ is defined as $L = D - A$
- **Normalized Laplacian:** $L \in \mathbb{R}^{n \times n}$ is defined as

Some matrices related to graphs

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$.

- **Adjacency Matrix:** $A \in \mathbb{R}^{n \times n}$ such that

$$A_{ij} = \begin{cases} 0 & \text{if} \quad i = j, \\ 1 & \text{if} \quad (i,j) \in E, \\ 0 & \text{if} \quad (i,j) \notin E. \end{cases}$$

- **Degree Matrix:** $D \in \mathbb{R}^{n \times n}$ is diagonal matrix such that $D_{ii} = \deg(i)$
- **Incidence Matrix:** $B \in \mathbb{R}^{n \times e}$, where rows indexed by vertices and columns indexed by edges and $B_{ij} = 1$ if vertex i lies on edge j.
- **Laplacian Matrix:** $L \in \mathbb{R}^{n \times n}$ is defined as $L = D - A$
- **Normalized Laplacian:** $L \in \mathbb{R}^{n \times n}$ is defined as

Graph Laplacian

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. **Laplacian:** $L \in \mathbb{R}^{n \times n}$ such that

$$L_{ij} = \begin{cases} d_i & \text{if} \quad i = j, \\ -1 & \text{if} \quad (i,j) \in E, \\ 0 & \text{if} \quad (i,j) \notin E. \end{cases}$$

Theorem

Let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be eigenvalues of L. Then

1. L is symmetric and positive semidefinite
2. $\lambda_1 = 0$
3. $\lambda_2 > 0$ iff G is connected
4. $\lambda_k = 0$ and $\lambda_{k+1} > 0$ iff G has exactly k-disjoint

Graph Laplacian

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. **Laplacian:** $L \in \mathbb{R}^{n\times n}$ such that

$$L_{ij} = \begin{cases} d_i & \text{if} \quad i = j, \\ -1 & \text{if} \quad (i,j) \in E, \\ 0 & \text{if} \quad (i,j) \notin E. \end{cases}$$

Theorem

Let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be eigenvalues of L. Then

1. L is symmetric and positive semidefinite
2. $\lambda_1 = 0$
3. $\lambda_2 > 0$ iff G is connected
4. $\lambda_k = 0$ and $\lambda_{k+1} > 0$ iff G has exactly k-disjoint

Graph Laplacian

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. **Laplacian:** $L \in \mathbb{R}^{n \times n}$ such that

$$L_{ij} = \begin{cases} d_i & \text{if} \quad i = j, \\ -1 & \text{if} \quad (i,j) \in E, \\ 0 & \text{if} \quad (i,j) \notin E. \end{cases}$$

Theorem

Let $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ be eigenvalues of L. Then

1. L is symmetric and positive semidefinite
2. $\lambda_1 = 0$
3. $\lambda_2 > 0$ iff G is connected
4. $\lambda_k = 0$ and $\lambda_{k+1} > 0$ iff G has exactly k-disjoint

Cuts

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. Let $V_1 \subset V$.

Boundary: The boundary of V_1 is defined as

$$\delta V_1 = \{(i,j) \in E : i \in V_1 \text{ and } j \notin V_1\}$$

- **Cut:**

$$\text{Cut}(V_1) = |\delta V_1|$$

- **Expansion Cut**

$$\text{ExpansionCut}(V_1, V - V_1) = \frac{|\delta V_1|}{\min\{|V_1|, |V - V_1|\}}$$

- **Ratio Cut:**

$$\text{RatioCut}(V_1, V - V_1) = \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V - V_1|}$$

Cuts

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. Let $V_1 \subset V$.

Boundary: The boundary of V_1 is defined as

$$\delta V_1 = \{(i, j) \in E : \ i \in V_1 \ \text{and} \ j \notin V_1\}$$

- **Cut:**

$$\text{Cut}(V_1) = |\delta V_1|$$

- **Expansion Cut**

$$\text{ExpansionCut}(V_1, V - V_1) = \frac{|\delta V_1|}{\min\{|V_1|, |V - V_1|\}}$$

- **Ratio Cut:**

$$\text{RatioCut}(V_1, V - V_1) = \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V - V_1|}$$

Cuts

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. Let $V_1 \subset V$.

Boundary: The boundary of V_1 is defined as

$$\delta V_1 = \{(i, j) \in E : \ i \in V_1 \ \text{and} \ j \notin V_1\}$$

- **Cut:**

$$\text{Cut}(V_1) = |\delta V_1|$$

- **Expansion Cut**

$$\text{ExpansionCut}(V_1, V - V_1) = \frac{|\delta V_1|}{\min\{|V_1|, |V - V_1|\}}$$

- **Ratio Cut:**

$$\text{RatioCut}(V_1, V - V_1) = \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V - V_1|}$$

Cuts

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. Let $V_1 \subset V$.

Boundary: The boundary of V_1 is defined as

$$\delta V_1 = \{(i, j) \in E : \; i \in V_1 \; \text{and} \; j \notin V_1\}$$

- **Cut:**

$$\text{Cut}(V_1) = |\delta V_1|$$

- **Expansion Cut**

$$\text{ExpansionCut}(V_1, V - V_1) = \frac{|\delta V_1|}{\min\{|V_1|, |V - V_1|\}}$$

- **Ratio Cut:**

$$\text{RatioCut}(V_1, V - V_1) = \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V - V_1|}$$

Cuts

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. Let $V_1 \subset V$.

Boundary: The boundary of V_1 is defined as

$$\delta V_1 = \{(i,j) \in E : \ i \in V_1 \text{ and } j \notin V_1\}$$

- **Cut:**

$$\text{Cut}(V_1) = |\delta V_1|$$

- **Expansion Cut**

$$\text{ExpansionCut}(V_1, V - V_1) = \frac{|\delta V_1|}{\min\{|V_1|, |V - V_1|\}}$$

- **Ratio Cut:**

$$\text{RatioCut}(V_1, V - V_1) = \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V - V_1|}$$

Metrics for partitioning

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. Let $V_1 \subset V$.

Boundary: The boundary of V_1 is defined as

$$\delta V_1 = \{(i, j) \in E : i \in V_1 \text{ and } j \notin V_1\}$$

- **Edge Expansion:**

$$\phi_G = \min_{|V_1| \leq \frac{|V|}{2}} \frac{|\delta V_1|}{|V_1|}$$

- **Ratio Cut:**

$$\eta_G = \min_{|V_1| \leq \frac{|V|}{2}} \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V - V_1|}$$

Metrics for partitioning

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. Let $V_1 \subset V$.

Boundary: The boundary of V_1 is defined as

$$\delta V_1 = \{(i, j) \in E : \ i \in V_1 \ \text{and} \ j \notin V_1\}$$

- **Edge Expansion:**

$$\phi_G = \min_{|V_1| \le \frac{|V|}{2}} \frac{|\delta V_1|}{|V_1|}$$

- **Ratio Cut:**

$$\eta_G = \min_{|V_1| \le \frac{|V|}{2}} \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V - V_1|}$$

Metrics for partitioning

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. Let $V_1 \subset V$.

Boundary: The boundary of V_1 is defined as

$$\delta V_1 = \{(i,j) \in E : \; i \in V_1 \text{ and } j \notin V_1\}$$

- **Edge Expansion:**

$$\phi_G = \min_{|V_1| \leq \frac{|V|}{2}} \frac{|\delta V_1|}{|V_1|}$$

- **Ratio Cut:**

$$\eta_G = \min_{|V_1| \leq \frac{|V|}{2}} \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V - V_1|}$$

Metrics for partitioning

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. Let $V_1 \subset V$.
Boundary: The boundary of V_1 is defined as

$$\delta V_1 = \{(i,j) \in E : \ i \in V_1 \ \text{and} \ j \notin V_1\}$$

- **Edge Expansion:**

$$\phi_G = \min_{|V_1| \leq \frac{|V|}{2}} \frac{|\delta V_1|}{|V_1|}$$

- **Ratio Cut:**

$$\eta_G = \min_{|V_1| \leq \frac{|V|}{2}} \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V - V_1|}$$

A simple calculation of $x^T L x$

$$
\begin{aligned}
x^T L x &= x^T D x - x^T A x \\
&= \sum_{i=1}^{n} d_i x_i^2 - \sum_{i,j=1}^{n} A_{ij} x_i x_j \\
&= \sum_{i=1}^{n} d_i x_i^2 - \sum_{(i,j)\in E} x_i x_j + x_j x_i \\
&= \sum_{(i,j)\in E} (x_i^2 + x_j^2) - \sum_{(i,j)\in E} x_i x_j + x_j x_i \\
&= \sum_{(i,j)\in E} (x_i - x_j)^2
\end{aligned}
$$

Rayleigh Principle or Courant-Fisher Theorem

Theorem

Let M be a symmetric matrix and let $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_n$ be eigenvalues of M. Then

$$\theta_k = \max_{n-k+1 \dim T} \min_{x \in T, x \neq 0} \frac{x^T M x}{x^T x}$$

Theorem

Let L be the Laplacian of a graph $G = (V, E)$. Then

$$\lambda_2 = \min_{x \perp 1} \frac{x^T M x}{x^T x}$$

Rayleigh Principle or Courant-Fisher Theorem

Theorem

Let M be a symmetric matrix and let $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_n$ be eigenvalues of M. Then

$$\theta_k = \max_{n-k+1 \dim T} \quad \min_{x \in T, x \neq 0} \frac{x^T M x}{x^T x}$$

Theorem

Let L be the Laplacian of a graph $G = (V, E)$. Then

$$\lambda_2 = \min_{x \perp 1} \frac{x^T M x}{x^T x}$$

Cheeger's Inequality

Definition (Cheeger's Constant)

Let $G = (V, E)$ be a graph and consider a graph bisection problem. Then

$$\phi_G = \min_{|V_1| \leq \frac{n}{2}} \frac{|\delta V_1|}{|V_1|}$$

Theorem (Cheeger's Inequality)

Let $d_{\max}$ denote the maximum degree of G and λ_2 be the second smallest eigenvalue of the Laplacian L of G. Then

$$\frac{\lambda_2}{2} \leq \phi_G \leq \sqrt{2\lambda_2 d_{\max}}$$

Note: Look at proofs of Mohar and Spielman

Cheeger's Inequality

Definition (Cheeger's Constant)

Let $G = (V, E)$ be a graph and consider a graph bisection problem. Then

$$\phi_G = \min_{|V_1| \leq \frac{n}{2}} \frac{|\delta V_1|}{|V_1|}$$

Theorem (Cheeger's Inequality)

Let $d_{\max}$ denote the maximum degree of G and λ_2 be the second smallest eigenvalue of the Laplacian L of G. Then

$$\frac{\lambda_2}{2} \leq \phi_G \leq \sqrt{2\lambda_2 d_{\max}}$$

Note: Look at proofs of Mohar and Spielman

Cheeger's Inequality

Definition (Cheeger's Constant)

Let $G = (V, E)$ be a graph and consider a graph bisection problem. Then

$$\phi_G = \min_{|V_1| \leq \frac{n}{2}} \frac{|\delta V_1|}{|V_1|}$$

Theorem (Cheeger's Inequality)

Let $d_{\max}$ denote the maximum degree of G and λ_2 be the second smallest eigenvalue of the Laplacian L of G. Then

$$\frac{\lambda_2}{2} \leq \phi_G \leq \sqrt{2\lambda_2 d_{\max}}$$

Note: Look at proofs of Mohar and Spielman

Cheeger's Inequality (Contd...)

Definition (Cheeger's Constant)

Let $G = (V, E)$ be a graph and consider a graph bisection problem. Then

$$\phi_G = \min_{|V_1| \leq \frac{n}{2}} \frac{|\delta V_1|}{|V_1|}$$

Theorem (Cheeger's Inequality)

Let $d_{\max}$ denote the maximum degree of G and λ_2 be the second smallest eigenvalue of the Laplacian L of G. Then

$$2\phi_G \leq \lambda_2 \leq \frac{{\phi_G}^2}{2}$$

Note: Look at proofs of Mohar and Spielman

Cheeger's Inequality (Contd...)

Definition (Cheeger's Constant)

Let $G = (V, E)$ be a graph and consider a graph bisection problem. Then

$$\phi_G = \min_{|V_1| \leq \frac{n}{2}} \frac{|\delta V_1|}{|V_1|}$$

Theorem (Cheeger's Inequality)

Let $d_{\max}$ denote the maximum degree of G and λ_2 be the second smallest eigenvalue of the Laplacian L of G. Then

$$2\phi_G \leq \lambda_2 \leq \frac{{\phi_G}^2}{2}$$

Note: Look at proofs of Mohar and Spielman

Cheeger's Inequality (Contd...)

Definition (Cheeger's Constant)

Let $G = (V, E)$ be a graph and consider a graph bisection problem. Then

$$\phi_G = \min_{|V_1| \leq \frac{n}{2}} \frac{|\delta V_1|}{|V_1|}$$

Theorem (Cheeger's Inequality)

Let $d_{\max}$ denote the maximum degree of G and λ_2 be the second smallest eigenvalue of the Laplacian L of G. Then

$$2\phi_G \leq \lambda_2 \leq \frac{{\phi_G}^2}{2}$$

Note: Look at proofs of Mohar and Spielman

Recall **Ratio Cut:**

$$\mathrm{RCut}(V_1, V_1^c) = \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V_1^c|}$$

A simple calculation shall give us this:

Define $y \in \mathbb{R}^n$ as

$$y_i = \begin{cases} \sqrt{\frac{|V_1^c|}{|V_1||V|}} & \text{if} \quad i \in V_1, \\ -\sqrt{\frac{|V_1|}{|V_1||V|}} & \text{if} \quad i \notin V_1. \end{cases} \tag{1}$$

Then

$$y^T L y = \mathrm{Rcut}(V_1, V_1^c)$$

Let say $\mathcal{Y}^*$ as subset of $\mathbb{R}^n$ denote various y defined as in (*) for various subsets of V_1 of V.

Graph Bisection

Recall **Ratio Cut:**

$$\mathrm{RCut}(V_1, V_1^c) = \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V_1^c|}$$

A simple calculation shall give us this:

Define $y \in \mathbb{R}^n$ as

$$y_i = \begin{cases} \sqrt{\frac{|V_1^c|}{|V_1||V|}} & \text{if} \quad i \in V_1, \\ -\sqrt{\frac{|V_1|}{|V_1||V|}} & \text{if} \quad i \notin V_1. \end{cases} \tag{1}$$

Then

$$y^T L y = \mathrm{Rcut}(V_1, V_1^c)$$

Let say $\mathcal{Y}^*$ as subset of $\mathbb{R}^n$ denote various y defined as in (*) for various subsets of V_1 of V.

Recall **Ratio Cut:**

$$\mathrm{RCut}(V_1, V_1^c) = \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V_1^c|}$$

A simple calculation shall give us this:

Define $y \in \mathbb{R}^n$ as

$$y_i = \begin{cases} \sqrt{\frac{|V_1^c|}{|V_1||V|}} & \text{if} \quad i \in V_1, \\ \\ -\sqrt{\frac{|V_1|}{|V_1||V|}} & \text{if} \quad i \notin V_1. \end{cases} \tag{1}$$

Then

$$y^T L y = \mathrm{Rcut}(V_1, V_1^c)$$

Let say $\mathcal{Y}^*$ as subset of $\mathbb{R}^n$ denote various y defined as in (*) for various subsets of V_1 of V.

Graph Bisection

Recall **Ratio Cut:**

$$\mathrm{RCut}(V_1, V_1^c) = \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V_1^c|}$$

A simple calculation shall give us this:

Define $y \in \mathbb{R}^n$ as

$$y_i = \begin{cases} \sqrt{\frac{|V_1^c|}{|V_1||V|}} & \text{if} \quad i \in V_1, \\ -\sqrt{\frac{|V_1|}{|V_1||V|}} & \text{if} \quad i \notin V_1. \end{cases} \tag{1}$$

Then

$$y^T L y = \mathrm{Rcut}(V_1, V_1^c)$$

Let say $\mathcal{Y}^*$ as subset of $\mathbb{R}^n$ denote various y defined as in (*) for various subsets of V_1 of V.

Graph Bisection (contd..)

Objective:

$$\min_{y \in \mathcal{Y}^*} y^T L y$$

Trivial Relaxation:

$$\min_{y \in \mathbb{R}^n} y^T L y$$

Not very useful as $1^T L 1 = 0$

Nice Relaxation:

Since $y^T 1 = \sum_{i \in V} y_i = 0$, y is orthogonal to 1. Also since $y^T y = \sum_{i \in V} y_i^2 = 1$, y is a unit norm vector. Hence the relaxed problem can be

$$\min_{y \perp 1} \frac{y^T L y}{y^T y}$$

Graph Bisection (contd..)

Objective:

$$\min_{y \in \mathcal{Y}^*} y^T L y$$

Trivial Relaxation:

$$\min_{y \in \mathbb{R}^n} y^T L y$$

Not very useful as $1^T L 1 = 0$

Nice Relaxation:

Since $y^T 1 = \sum_{i \in V} y_i = 0$, y is orthogonal to 1. Also since $y^T y = \sum_{i \in V} y_i^2 = 1$, y is a unit norm vector. Hence the relaxed problem can be

$$\min_{y \perp 1} \frac{y^T L y}{y^T y}$$

Graph Bisection (contd..)

Objective:

$$\min_{y \in \mathcal{Y}^*} y^T L y$$

Trivial Relaxation:

$$\min_{y \in \mathbb{R}^n} y^T L y$$

Not very useful as $1^T L 1 = 0$

Nice Relaxation:

Since $y^T 1 = \sum_{i \in V} y_i = 0$, y is orthogonal to 1. Also since $y^T y = \sum_{i \in V} y_i^2 = 1$, y is a unit norm vector. Hence the relaxed problem can be

$$\min_{y \perp 1} \frac{y^T L y}{y^T y}$$

Objective:

$$\min_{y \in \mathcal{Y}^*} y^T L y$$

Trivial Relaxation:

$$\min_{y \in \mathbb{R}^n} y^T L y$$

Not very useful as $1^T L 1 = 0$

Nice Relaxation:

Since $y^T 1 = \sum_{i \in V} y_i = 0$, y is orthogonal to 1. Also since $y^T y = \sum_{i \in V} y_i^2 = 1$, y is a unit norm vector. Hence the relaxed problem can be

$$\min_{y \perp 1} \frac{y^T L y}{y^T y}$$

Graph k-way partitioning

Ratio Cut:

$$\text{Rcut}(V_1, \ldots, V_k) = \sum_{\ell=1}^{k} \frac{|\delta V_\ell|}{|V_\ell|}$$

Lets define Y: Define $y \in \mathbb{R}^{n \times k}$ such that

$$Y_{i\ell} = \begin{cases} \frac{1}{\sqrt{|V_\ell|}} & \text{if} \quad i \in V_\ell, \\ 0 & \text{otherwise.} \end{cases} \qquad (**)$$

Claim: $Y^T Y = I$

Claim: $\text{Rcut}(V_1, \ldots, V_k) = \text{Trace}(Y^T L Y)$

Graph k-way partitioning

Ratio Cut:

$$\text{Rcut}(V_1, \dots, V_k) = \sum_{\ell=1}^{k} \frac{|\delta V_\ell|}{|V_\ell|}$$

Lets define Y: Define $y \in \mathbb{R}^{n \times k}$ such that

$$Y_{i\ell} = \begin{cases} \frac{1}{\sqrt{|V_\ell|}} & \text{if} \quad i \in V_\ell, \\ 0 & \text{otherwise.} \end{cases} \qquad (**)$$

Claim: $Y^T Y = I$

Claim: $\text{Rcut}(V_1, \dots, V_k) = \text{Trace}(Y^T L Y)$

Graph k-way partitioning

Ratio Cut:

$$\text{Rcut}(V_1, \ldots, V_k) = \sum_{\ell=1}^{k} \frac{|\delta V_\ell|}{|V_\ell|}$$

Lets define Y: Define $y \in \mathbb{R}^{n \times k}$ such that

$$Y_{i\ell} = \begin{cases} \frac{1}{\sqrt{|V_\ell|}} & \text{if} \quad i \in V_\ell, \\ \\ 0 & \text{otherwise.} \end{cases} \qquad (**)$$

Claim: $Y^T Y = I$

Claim: $\text{Rcut}(V_1, \ldots, V_k) = \text{Trace}(Y^T L Y)$

Graph k-way partitioning

Ratio Cut:

$$\text{Rcut}(V_1, \ldots, V_k) = \sum_{\ell=1}^{k} \frac{|\delta V_\ell|}{|V_\ell|}$$

Lets define Y: Define $y \in \mathbb{R}^{n \times k}$ such that

$$Y_{i\ell} = \begin{cases} \frac{1}{\sqrt{|V_\ell|}} & \text{if} \quad i \in V_\ell, \\ 0 & \text{otherwise.} \end{cases} \qquad (**)$$

Claim: $Y^T Y = I$

Claim: $\text{Rcut}(V_1, \ldots, V_k) = \text{Trace}(Y^T L Y)$

Graph k-way partitioning

- Objective

$$\min_{Y \in \mathcal{Y}^{**}} \text{Trace}(Y^T L Y)$$

- Relaxation

$$\min_{\substack{Y \in \mathbb{R}^n \\ Y^T Y = I}} \text{Trace}(Y^T L Y)$$

- Optimal Value

$$Y^{\text{opt}} = [v_1 \dots v_k]$$

matrix of k leading orthonormal eigenvectors of L

Graph k-way partitioning

- Objective

$$\min_{Y \in \mathcal{Y}^{**}} \operatorname{Trace}(Y^T L Y)$$

- Relaxation

$$\min_{\substack{Y \in \mathbb{R}^n \\ Y^T Y = I}} \operatorname{Trace}(Y^T L Y)$$

- Optimal Value

$$Y^{\text{opt}} = [v_1 \ldots v_k]$$

matrix of k leading orthonormal eigenvectors of L

Graph k-way partitioning

- Objective

$$\min_{Y \in \mathcal{Y}^{**}} \text{Trace}(Y^T L Y)$$

- Relaxation

$$\min_{\substack{Y \in \mathbb{R}^n \\ Y^T Y = I}} \text{Trace}(Y^T L Y)$$

- Optimal Value

$$Y^{\text{opt}} = [v_1 \dots v_k]$$

matrix of k leading orthonormal eigenvectors of L

With Normaized Cuts

Normalized Cut:

$$\text{Ncut}(V_1, \ldots, V_k) = \sum_{\ell=1}^{k} \frac{|\delta V_\ell|}{\text{Vol}(V_\ell)}$$

where $\text{Vol}(V_\ell) = \sum_{i \in V_\ell} \deg(i)$

Lets define Y again: Define $y \in \mathbb{R}^{n \times k}$ such that

$$Y_{i\ell} = \begin{cases} \frac{1}{\sqrt{\text{Vol}(V_\ell)}} & \text{if} \quad i \in V_\ell, \\ 0 & \text{otherwise.} \end{cases} \qquad (***)$$

Claim: $Y^T D Y = I$

Claim: $\text{Ncut}(V_1, \ldots, V_k) = \text{Trace}(Y^T L Y)$

With Normazied Cuts

Normalized Cut:

$$\text{Ncut}(V_1, \ldots, V_k) = \sum_{\ell=1}^{k} \frac{|\delta V_\ell|}{\text{Vol}(V_\ell)}$$

where $\text{Vol}(V_\ell) = \sum_{i \in V_\ell} \deg(i)$

Lets define Y again: Define $y \in \mathbb{R}^{n \times k}$ such that

$$Y_{i\ell} = \begin{cases} \frac{1}{\sqrt{\text{Vol}(V_\ell)}} & \text{if} \quad i \in V_\ell, \\ 0 & \text{otherwise.} \end{cases} \qquad (***)$$

Claim: $Y^T DY = I$

Claim: $\text{Ncut}(V_1, \ldots, V_k) = \text{Trace}(Y^T LY)$

Normalized Cut:

$$\text{Ncut}(V_1, \ldots, V_k) = \sum_{\ell=1}^{k} \frac{|\delta V_\ell|}{\text{Vol}(V_\ell)}$$

where $\text{Vol}(V_\ell) = \sum_{i \in V_\ell} \deg(i)$

Lets define Y again: Define $y \in \mathbb{R}^{n \times k}$ such that

$$Y_{i\ell} = \begin{cases} \frac{1}{\sqrt{\text{Vol}(V_\ell)}} & \text{if} \quad i \in V_\ell, \\ 0 & \text{otherwise.} \end{cases} \qquad (***)$$

Claim: $Y^T D Y = I$

Claim: $\text{Ncut}(V_1, \ldots, V_k) = \text{Trace}(Y^T L Y)$

Normalized Cut:

$$\text{Ncut}(V_1, \ldots, V_k) = \sum_{\ell=1}^{k} \frac{|\delta V_\ell|}{\text{Vol}(V_\ell)}$$

where $\text{Vol}(V_\ell) = \sum_{i \in V_\ell} \deg(i)$

Lets define Y again: Define $y \in \mathbb{R}^{n \times k}$ such that

$$Y_{i\ell} = \begin{cases} \frac{1}{\sqrt{\text{Vol}(V_\ell)}} & \text{if} \quad i \in V_\ell, \\ \\ 0 & \text{otherwise.} \end{cases} \qquad (***)$$

Claim: $Y^T D Y = I$

Claim: $\text{Ncut}(V_1, \ldots, V_k) = \text{Trace}(Y^T L Y)$

With normalized cuts

- Objective

$$\min_{Y \in \mathcal{Y}^{***}} \operatorname{Trace}(Y^T L Y)$$

- Relaxation

$$\min_{\substack{Y \in \mathbb{R}^n \\ Y^T D Y = I}} \operatorname{Trace}(Y^T L Y)$$

- By substituting $\widetilde{Y} = D^{\frac{1}{2}} Y$ the objective translates to

$$\min_{\substack{\widetilde{Y} \in \mathbb{R}^n \\ \widetilde{Y}^T \widetilde{Y} = I}} \operatorname{Trace}(\widetilde{Y}^T D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \widetilde{Y})$$

With normalized cuts

- Objective

$$\min_{Y \in \mathcal{Y}^{***}} \text{Trace}(Y^T L Y)$$

- Relaxation

$$\min_{\substack{Y \in \mathbb{R}^n \\ Y^T D Y = I}} \text{Trace}(Y^T L Y)$$

- By substituting $\widetilde{Y} = D^{\frac{1}{2}} Y$ the objective translates to

$$\min_{\substack{\widetilde{Y} \in \mathbb{R}^n \\ \widetilde{Y}^T \widetilde{Y} = I}} \text{Trace}(\widetilde{Y}^T D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \widetilde{Y})$$

With normalized cuts

- Objective

$$\min_{Y \in \mathcal{Y}^{***}} \text{Trace}(Y^T L Y)$$

- Relaxation

$$\min_{\substack{Y \in \mathbb{R}^n \\ Y^T D Y = I}} \text{Trace}(Y^T L Y)$$

- By substituting $\widetilde{Y} = D^{\frac{1}{2}} Y$ the objective translates to

$$\min_{\substack{\widetilde{Y} \in \mathbb{R}^n \\ \widetilde{Y}^T \widetilde{Y} = I}} \text{Trace}(\widetilde{Y}^T D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \widetilde{Y})$$

Algorithm

1. Compute graph Laplacian or normalized graph Laplacian
2. Compute k-leading eigenvectors $Y \in \mathbb{R}^{n \times k}$ of L
3. Normalize rows of Y and say it is $\bar{Y}$
4. Run k-means on rows of $\bar{Y}$
5. according to this partition V

K-means Step

$$S^* = \underset{\substack{S \in \mathbb{R}^{n \times k} \\ S \text{has at most } k \text{ distinct rows}}}{\arg\max} ||\bar{Y} - S||_F^2$$

Algorithm

1. Compute graph Laplacian or normalized graph Laplacian
2. Compute k-leading eigenvectors $Y \in \mathbb{R}^{n\times k}$ of L
3. Normalize rows of Y and say it is $\bar{Y}$
4. Run k-means on rows of $\bar{Y}$
5. according to this partition V

K-means Step

$$S^* = \underset{\substack{S\in\mathbb{R}^{n\times k} \\ S \text{has at most } k \text{ distinct rows}}}{\arg\max} ||\bar{Y} - S||_F^2$$

Spectral Clustering Algorithm

Algorithm

1. Compute graph Laplacian or normalized graph Laplacian
2. Compute k-leading eigenvectors $Y \in \mathbb{R}^{n \times k}$ of L
3. Normalize rows of Y and say it is $\bar{Y}$
4. Run k-means on rows of $\bar{Y}$
5. according to this partition V

K-means Step

$$S^* = \underset{\substack{S \in \mathbb{R}^{n \times k} \\ S \text{has at most } k \text{ distinct rows}}}{\arg\max} ||\bar{Y} - S||_F^2$$

Algorithm

1. Compute graph Laplacian or normalized graph Laplacian
2. Compute k-leading eigenvectors $Y \in \mathbb{R}^{n \times k}$ of L
3. Normalize rows of Y and say it is $\bar{Y}$
4. Run k-means on rows of $\bar{Y}$
5. according to this partition V

K-means Step

$$S^* = \underset{\substack{S \in \mathbb{R}^{n \times k} \\ S \text{has at most } k \text{ distinct rows}}}{\arg\max} ||\bar{Y} - S||_F^2$$

Spectral Clustering Algorithm

Algorithm

1. Compute graph Laplacian or normalized graph Laplacian
2. Compute k-leading eigenvectors $Y \in \mathbb{R}^{n \times k}$ of L
3. Normalize rows of Y and say it is $\bar{Y}$
4. Run k-means on rows of $\bar{Y}$
5. according to this partition V

K-means Step

$$S^* = \arg\max_{\substack{S \in \mathbb{R}^{n \times k} \\ S \text{has at most } k \text{ distinct rows}}} ||\bar{Y} - S||_F^2$$

Spectral Clustering Algorithm

Algorithm

1. Compute graph Laplacian or normalized graph Laplacian
2. Compute k-leading eigenvectors $Y \in \mathbb{R}^{n \times k}$ of L
3. Normalize rows of Y and say it is $\bar{Y}$
4. Run k-means on rows of $\bar{Y}$
5. according to this partition V

K-means Step

$$S^* = \underset{\substack{S \in \mathbb{R}^{n \times k} \\ S \text{has at most } k \text{ distinct rows}}}{\arg\max} ||\bar{Y} - S||_F^2$$

Algorithm 1 Spectral Clustering Algorithm

Input: Similarity matrix $\mathbf{A} \in \mathbb{R}^{+m \times m}$ and number of clusters k

Output: Cluster assignment vector $\mathbf{c} \in \{1, \ldots k\}^m$

Compute a diagonal matrix $\mathbf{D}$ such that $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$

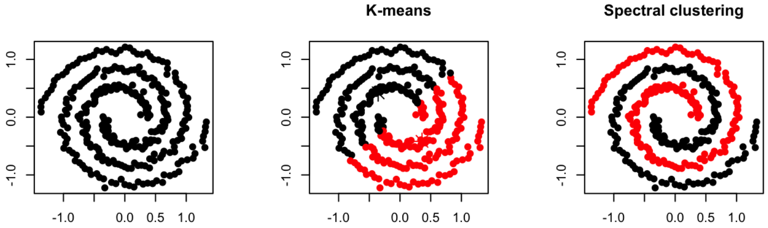
Compute $\mathbf{L} = \mathbf{D} - \mathbf{A}$

Find $\mathbf{U} \in \mathbb{R}^{m \times k}$ containing top k eigenvectors of $\mathbf{L}$ as columns

Compute $\tilde{\mathbf{U}} \in \mathbb{R}^{m \times k}$ such that $\tilde{\mathbf{U}}_i = \frac{\mathbf{U}_i}{||\mathbf{U}_i||}$, where $\mathbf{U}_i$ is the i^{th} row of $\mathbf{U}$

Obtain $\mathbf{c}$ by clustering the rows of $\tilde{\mathbf{U}}$ using k-Means

Clustering - Spectral Clustering (contd. . .)



Spectral clustering can detect non-convex clusters where k-Means fails[6]

[6] Image Source: `http://scalefreegan.github.io`

Clustering - Other Issues

- How to select the number of clusters?
  - Elbo method, Bayesian model selection, information theoretic methods etc.
- Which algorithm to use?
  - Different algorithms offer different perspectives
  - Since clustering is exploratory in nature, must try different algorithms
- How to evaluate the quality of clustering?
  - Ground truth available: Accuracy, Normalized Mutual Information (NMI) score etc.
  - Ground truth unavailable: Modularity, Log Likelihood, Silhouette coefficient etc.