COMPSCI 514: Algorithms for Data Science

Cameron Musco

University of Massachusetts Amherst. Fall 2024.

Lecture 18

Logistics

- Problem Set 3 solutions have been posted.
- Problem Set 4 will be released soon.

Summary

Last Class: SVD and Applications of Low-Rank Approximation

$X = U\Sigma V^T$

- SVD and connections to eigendecomposition and optimal low-rank approximation.
- Matrix completion
- Entity Embeddings. words → vectors

Summary

Last Class: SVD and Applications of Low-Rank Approximation

- SVD and connections to eigendecomposition and optimal low-rank approximation.
- Matrix completion
- Entity Embeddings.

This Class: Linear Algebraic Techniques for Graph Analysis

- Start on graph clustering for community detection and non-linear clustering.
- Spectral clustering: finding good cuts via Laplacian eigenvectors.

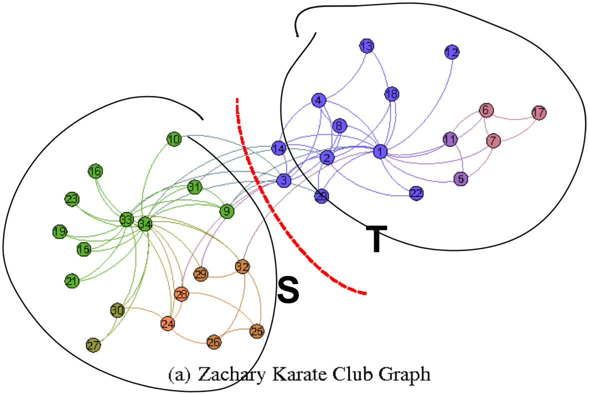
Spectral Clustering

A very common task is to partition or cluster vertices in a graph based on similarity/connectivity.

Spectral Clustering

A very common task is to partition or cluster vertices in a graph based on similarity/connectivity.

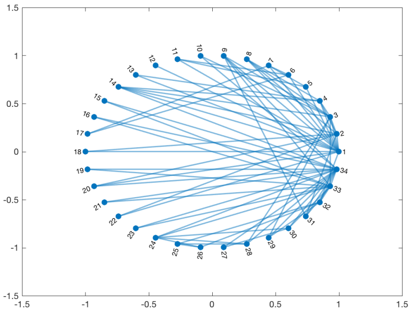
Community detection in naturally occurring networks.



(a) Zachary Karate Club Graph

A very common task is to partition or cluster vertices in a graph based on similarity/connectivity.

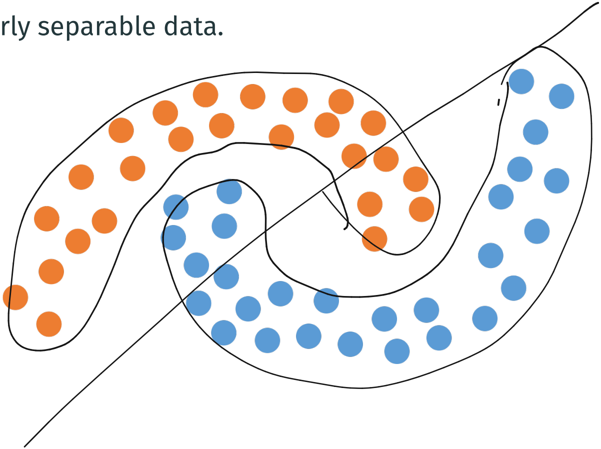
Community detection in naturally occurring networks.



Spectral Clustering

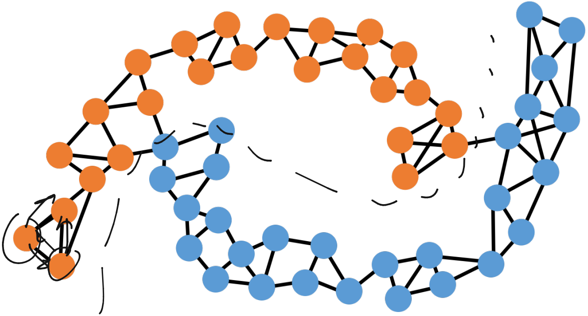
A very common task is to partition or cluster vertices in a graph based on similarity/connectivity.

Non-linearly separable data.



A very common task is to partition or cluster vertices in a graph based on similarity/connectivity.

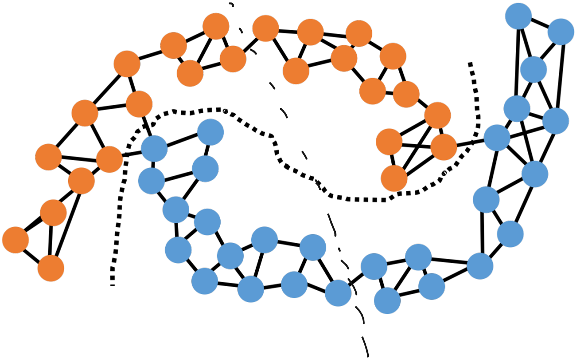
Non-linearly separable data.



Spectral Clustering

A very common task is to partition or cluster vertices in a graph based on similarity/connectivity.

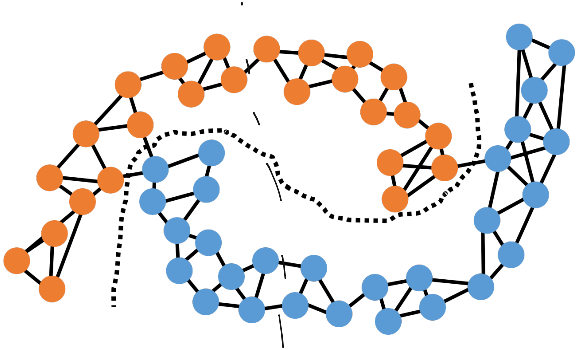
Non-linearly separable data.



Spectral Clustering

A very common task is to partition or cluster vertices in a graph based on similarity/connectivity.

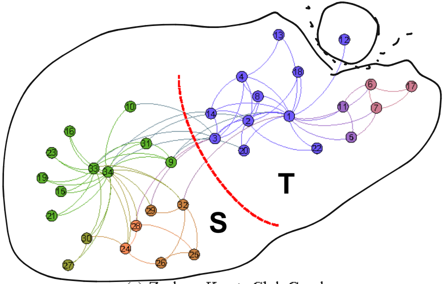
Non-linearly separable data.



Next Few Classes: Find this cut using eigendecomposition. First – motivate why this type of approach makes sense.

Cut Minimization

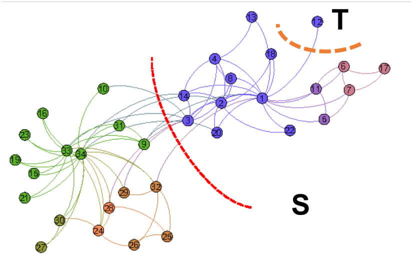
Simple Idea: Partition clusters along minimum cut in graph.



(a) Zachary Karate Club Graph

Cut Minimization

Simple Idea: Partition clusters along minimum cut in graph.

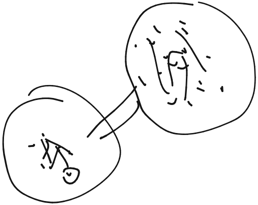


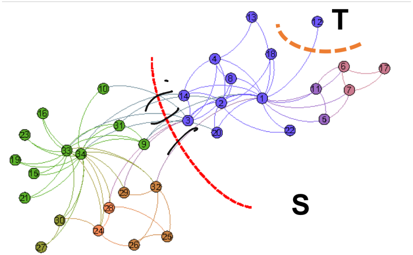
(a) Zachary Karate Club Graph

Small cuts are often not informative.

Cut Minimization

Simple Idea: Partition clusters along minimum cut in graph.





(a) Zachary Karate Club Graph

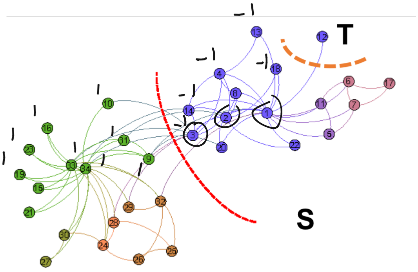
Small cuts are often not informative.

Solution: Encourage cuts that separate large sections of the graph.

Cut Minimization

Simple Idea: Partition clusters along minimum cut in graph.

$$v = \begin{bmatrix} -1 \\ -1 \\ -1 \\ \vdots \\ 1 \\ -1 \\ 1 \end{bmatrix}$$



(a) Zachary Karate Club Graph

Small cuts are often not informative.

Solution: Encourage cuts that separate large sections of the graph.

- Let $\vec{v} \in \mathbb{R}^n$ be a cut indicator: $\vec{v}(i) = 1$ if $i \in S$. $\vec{v}(i) = -1$ if $i \in T$. Want $\vec{v}$ to have roughly equal numbers of 1s and -1s. I.e., $\sum_{i=1}^{n} v(i) =$ $\vec{v}^T\vec{1} \approx 0$.

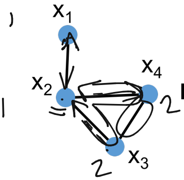
The Laplacian View

$v^T L v = 2^2 \cdot n\text{-}1 \approx 4 \cdot n$

$v^T L v = 4$

For a graph with adjacency matrix **A** and degree matrix **D**, **L** = **D** − **A** is the graph Laplacian.

$n \times n$



$$\underset{\mathbf{D}}{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}} - \underset{\mathbf{A}}{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}} = \underset{\mathbf{L}}{\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix}}$$

$\in \mathbb{R}^n$

For any vector $\vec{v}$, its 'smoothness' over the graph is given by:

$v \in \mathbb{R}^4$

$$\sum_{(i,j)\in E} (\vec{v}(i) - \vec{v}(j))^2 = \vec{v}^T \mathbf{L} \vec{v}.$$

$v = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}$

$v^T L v = 0^2 + (-1)^2 + (-1)^2 + 0^2 = 2$

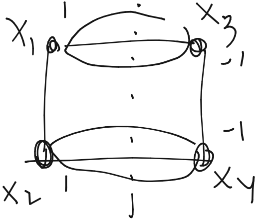
$v^T L v = 1^2 (n-1) = n-1$

For a cut indicator vector $\vec{v} \in \{-1, 1\}^n$ with $\vec{v}(i) = -1$ for $i \in S$ and $\vec{v}(i) = 1$ for $i \in T$:

1. $\vec{v}^T \mathbf{L} \vec{V} = \sum_{(i,j)\in E} (\vec{v}(i) - \vec{v}(j))^2 = 4 \cdot cut(S, T)$.

$v^T L v = 2^2 + 2^2$

$= 8$



$V = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$

The Laplacian View

For a cut indicator vector $\vec{v} \in \{-1, 1\}^n$ with $\vec{v}(i) = -1$ for $i \in S$ and $\vec{v}(i) = 1$ for $i \in T$:

1. $\underline{\vec{v}^T \mathbf{L} \vec{V}} = \sum_{(i,j)\in E} (\vec{v}(i) - \vec{v}(j))^2 = 4 \cdot cut(S, T)$.
2. $\underline{\vec{v}^T \vec{1}} = |T| - |S|$.

want both small

The Laplacian View

For a cut indicator vector $\vec{v} \in \{-1, 1\}^n$ with $\vec{v}(i) = -1$ for $i \in S$ and $\vec{v}(i) = 1$ for $i \in T$:

1. $\vec{v}^T \mathbf{L} \vec{V} = \sum_{(i,j) \in E} (\vec{v}(i) - \vec{v}(j))^2 = 4 \cdot cut(S, T)$.
2. $\vec{v}^T \vec{1} = |T| - |S|$.

Want to minimize both $\vec{v}^T \mathbf{L} \vec{v}$ (cut size) and $\vec{v}^T \vec{1}$ (imbalance).

The Laplacian View

For a cut indicator vector $\vec{v} \in \{-1, 1\}^n$ with $\vec{v}(i) = -1$ for $i \in S$ and $\vec{v}(i) = 1$ for $i \in T$:

1. $\vec{v}^T \mathbf{L} \vec{V} = \sum_{(i,j)\in E} (\vec{v}(i) - \vec{v}(j))^2 = 4 \cdot cut(S, T)$.
2. $\vec{v}^T \vec{1} = |T| - |S|$.

Want to minimize both $\vec{v}^T \mathbf{L} \vec{v}$ (cut size) and $\vec{v}^T \vec{1}$ (imbalance).

Next Step: See how this dual minimization problem is naturally solved (sort of) by eigendecomposition.

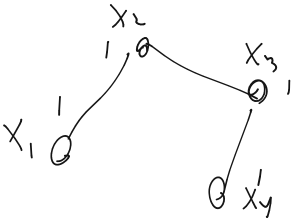
The smallest eigenvector of the Laplacian is:

$$\vec{v}_n = \frac{1}{\sqrt{n}} \cdot \vec{1} = \underset{v \in \mathbb{R}^n \text{ with } \|\vec{v}\|=1}{\arg\min} \vec{v}^T \mathbf{L} \vec{v}$$

$\lambda_1 = \max v^T L v$

$\lambda_2 = \max_{v \perp v_1} v^T L v$

with eigenvalue $\lambda_n(\mathbf{L}) = \vec{v}_n^T \mathbf{L} \vec{v}_n = 0$. Why?



$v = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ $\quad v^T L v = 0$

n: number of nodes in graph, $\mathbf{A} \in \mathbb{R}^{n \times n}$: adjacency matrix, $\mathbf{D} \in \mathbb{R}^{n \times n}$: diagonal degree matrix, $\mathbf{L} \in \mathbb{R}^{n \times n}$: Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$.

Second Smallest Laplacian Eigenvector

By Courant-Fischer, the second smallest eigenvector is given by:

$$\vec{v}_{n-1} = \operatorname*{arg\,min}_{v \in \mathbb{R}^n \text{ with } \|\vec{v}\|=1,\ \vec{v}_n^T \vec{v}=0} \vec{v}^T \mathbf{L} \vec{v}.$$

n: number of nodes in graph, $\mathbf{A} \in \mathbb{R}^{n \times n}$: adjacency matrix, $\mathbf{D} \in \mathbb{R}^{n \times n}$: diagonal degree matrix, $\mathbf{L} \in \mathbb{R}^{n \times n}$: Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$. S, T: vertex sets on different sides of cut.

Second Smallest Laplacian Eigenvector

By Courant-Fischer, the second smallest eigenvector is given by:

$$\vec{v}_{n-1} = \underset{v \in \mathbb{R}^n \text{ with } \|\vec{v}\|=1, \ \vec{v}_n^T \vec{v}=0}{\arg\min} \vec{v}^T \mathbf{L} \vec{v}.$$

— cut size

perfect balance

If $\vec{v}_{n-1}$ were in $\left\{-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right\}^n$ it would have:

- $\vec{v}_{n-1}^T L \vec{v}_{n-1} = \frac{4}{\sqrt{n}} \cdot cut(S,T)$ as small as possible given that $\vec{v}_{n-1}^T \vec{v}_n = \frac{1}{\sqrt{n}} \vec{v}_{n-1}^T \vec{1} = \frac{|T|-|S|}{n} = 0$.

$V_{n-1}^T V_n = 0$

$V_n = \begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix} \frac{1}{\sqrt{6}}$

$V_{n-1}^T 1 = 0$ $\quad [1]$

$\sum_{i=1}^{n} V_{n-1}(i) = 0$

$V_{n-1} = \begin{bmatrix}1\\1\\-1\\-1\end{bmatrix} \cdot \frac{1}{\sqrt{6}}$

> n: number of nodes in graph, $\mathbf{A} \in \mathbb{R}^{n \times n}$: adjacency matrix, $\mathbf{D} \in \mathbb{R}^{n \times n}$: diagonal degree matrix, $\mathbf{L} \in \mathbb{R}^{n \times n}$: Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$. S, T: vertex sets on different sides of cut.

Second Smallest Laplacian Eigenvector

By Courant-Fischer, the second smallest eigenvector is given by:

$$\vec{v}_{n-1} = \underset{v \in \mathbb{R}^n \text{ with } \|\vec{v}\|=1,\ \vec{v}_n^T \vec{v}=0}{\arg\min} \vec{v}^T \mathbf{L} \vec{v}.$$

If $\vec{v}_{n-1}$ were in $\left\{-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right\}^n$ it would have:

- $\vec{v}_{n-1}^T L \vec{v}_{n-1} = \frac{4}{\sqrt{n}} \cdot cut(S,T)$ as small as possible given that $\vec{v}_{n-1}^T \vec{v}_n = \frac{1}{\sqrt{n}} \vec{v}_{n-1}^T \vec{1} = \frac{|T|-|S|}{n} = 0$.
- I.e., $\vec{v}_{n-1}$ would indicate the smallest perfectly balanced cut.

n: number of nodes in graph, $\mathbf{A} \in \mathbb{R}^{n \times n}$: adjacency matrix, $\mathbf{D} \in \mathbb{R}^{n \times n}$: diagonal degree matrix, $\mathbf{L} \in \mathbb{R}^{n \times n}$: Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$. S, T: vertex sets on different sides of cut.

Second Smallest Laplacian Eigenvector

By Courant-Fischer, the second smallest eigenvector is given by:

$$\vec{v}_{n-1} = \underset{v \in \mathbb{R}^n \text{ with } \|\vec{v}\|=1,\ \vec{v}_n^T \vec{v}=0}{\arg\min} \vec{v}^T \mathbf{L} \vec{v}.$$

If $\vec{v}_{n-1}$ were in $\left\{-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right\}^n$ it would have:

- $\vec{v}_{n-1}^T L \vec{v}_{n-1} = \frac{4}{\sqrt{n}} \cdot cut(S,T)$ as small as possible given that $\vec{v}_{n-1}^T \vec{v}_n = \frac{1}{\sqrt{n}} \vec{v}_{n-1}^T \vec{1} = \frac{|T|-|S|}{n} = 0$.
- I.e., $\vec{v}_{n-1}$ would indicate the smallest perfectly balanced cut.
- The eigenvector $\vec{v}_{n-1} \in \mathbb{R}^n$ is not generally binary, but still satisfies a 'relaxed' version of this property.

n: number of nodes in graph, $\mathbf{A} \in \mathbb{R}^{n \times n}$: adjacency matrix, $\mathbf{D} \in \mathbb{R}^{n \times n}$: diagonal degree matrix, $\mathbf{L} \in \mathbb{R}^{n \times n}$: Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$. S, T: vertex sets on different sides of cut.

Cutting With the Second Laplacian Eigenvector

Find a good partition of the graph by computing

$$\vec{v}_{n-1} = \mathop{\arg\min}_{v \in \mathbb{R}^d \text{ with } \|\vec{v}\|=1, \ \vec{v}^T\vec{1}=0} \vec{v}^T \mathbf{L} \vec{v}.$$

cut size

balance

Set S to be all nodes with $\vec{v}_{n-1}(i) < 0$, T to be all with $\vec{v}_2(i) \geq 0$.

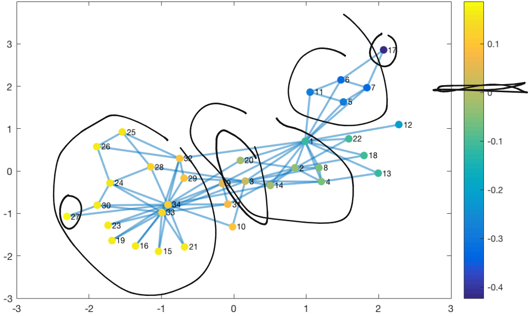
$$\begin{bmatrix} -.58 \\ 1 \\ 0.2 \\ -.03 \\ \vdots \end{bmatrix}$$

Cutting With the Second Laplacian Eigenvector

Find a good partition of the graph by computing

$$\vec{v}_{n-1} = \underset{v\in\mathbb{R}^d \text{ with } \|\vec{v}\|=1,\ \vec{v}^T\vec{1}=0}{\arg\min} \vec{v}^T \mathbf{L}\vec{v}.$$

Set S to be all nodes with $\vec{v}_{n-1}(i) < 0$, T to be all with $\vec{v}_2(i) \geq 0$.



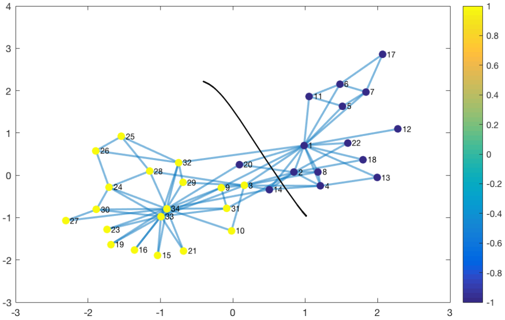
Cutting With the Second Laplacian Eigenvector

Find a good partition of the graph by computing

$$\boxed{\vec{v}_{n-1} = \underset{v \in \mathbb{R}^d \text{ with } \|\vec{v}\|=1,\ \vec{v}^T\vec{1}=0}{\arg\min} \vec{v}^T \mathbf{L} \vec{v}.}$$

generalize to other thresholds

Set S to be all nodes with $\vec{v}_{n-1}(i) < 0$, T to be all with $\vec{v}_{2}(i) \geq 0$.



Spectral Partitioning in Practice

The Shi-Malik normalized cuts algorithm is one of the most commonly used variants of this approach, using the normalized Laplacian $\bar{\mathbf{L}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$.

$v^T L v$

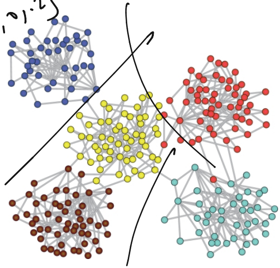
n: number of nodes in graph, $\mathbf{A} \in \mathbb{R}^{n \times n}$: adjacency matrix, $\mathbf{D} \in \mathbb{R}^{n \times n}$: diagonal degree matrix, $\mathbf{L} \in \mathbb{R}^{n \times n}$: Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$.

Spectral Partitioning in Practice

The Shi-Malik normalized cuts algorithm is one of the most commonly used variants of this approach, using the normalized Laplacian $\bar{\mathbf{L}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$.

Important Consideration: What to do when we want to split the graph into more than two parts?

[.5, .03, 1, 0, .2]



n: number of nodes in graph, $\mathbf{A} \in \mathbb{R}^{n \times n}$: adjacency matrix, $\mathbf{D} \in \mathbb{R}^{n \times n}$: diagonal degree matrix, $\mathbf{L} \in \mathbb{R}^{n \times n}$: Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$.

Spectral Partitioning in Practice

The Shi-Malik normalized cuts algorithm is one of the most commonly used variants of this approach, using the normalized Laplacian $\bar{\mathbf{L}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$.

Important Consideration: What to do when we want to split the graph into more than two parts?

Spectral Clustering:

n: number of nodes in graph, $\mathbf{A} \in \mathbb{R}^{n \times n}$: adjacency matrix, $\mathbf{D} \in \mathbb{R}^{n \times n}$: diagonal degree matrix, $\mathbf{L} \in \mathbb{R}^{n \times n}$: Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$.

Spectral Partitioning in Practice

The Shi-Malik normalized cuts algorithm is one of the most commonly used variants of this approach, using the normalized Laplacian $\overline{\mathbf{L}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$.

Important Consideration: What to do when we want to split the graph into more than two parts?

Spectral Clustering:

- Compute smallest k nonzero eigenvectors $\vec{v}_{n-1}, \ldots, \vec{v}_{n-k}$ of $\overline{\mathbf{L}}$.

n: number of nodes in graph, $\mathbf{A} \in \mathbb{R}^{n \times n}$: adjacency matrix, $\mathbf{D} \in \mathbb{R}^{n \times n}$: diagonal degree matrix, $\mathbf{L} \in \mathbb{R}^{n \times n}$: Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$.

Spectral Partitioning in Practice

The Shi-Malik normalized cuts algorithm is one of the most commonly used variants of this approach, using the normalized Laplacian $\bar{\mathbf{L}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$.

Important Consideration: What to do when we want to split the graph into more than two parts?

Spectral Clustering:

$k = 2$

$$\begin{array}{cc} v_{n-1} & v_{n-2} \\ \end{array}$$
$$\begin{bmatrix} -1 & .2 \\ -.7 & .03 \\ 0 & \vdots \\ .3 & \\ 1 & \end{bmatrix}$$

- Compute smallest k nonzero eigenvectors $\vec{v}_{n-1}, \ldots, \vec{v}_{n-k}$ of $\bar{\mathbf{L}}$.
- Represent each node by its corresponding row in $\mathbf{V} \in \mathbb{R}^{n \times k}$ whose columns are $\vec{v}_{n-1}, \ldots \vec{v}_{n-k}$.

n: number of nodes in graph, $\mathbf{A} \in \mathbb{R}^{n \times n}$: adjacency matrix, $\mathbf{D} \in \mathbb{R}^{n \times n}$: diagonal degree matrix, $\mathbf{L} \in \mathbb{R}^{n \times n}$: Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$.

Spectral Partitioning in Practice

The Shi-Malik normalized cuts algorithm is one of the most commonly used variants of this approach, using the normalized Laplacian $\bar{\mathbf{L}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$.

Important Consideration: What to do when we want to split the graph into more than two parts?

Spectral Clustering:

$V^T L V$

- Compute smallest k nonzero eigenvectors $\vec{v}_{n-1}, \ldots, \vec{v}_{n-k}$ of $\bar{\mathbf{L}}$.
- Represent each node by its corresponding row in $\mathbf{V} \in \mathbb{R}^{n \times k}$ whose columns are $\vec{v}_{n-1}, \ldots \vec{v}_{n-k}$.
- Cluster these rows using k-means clustering (or really any clustering method).

n: number of nodes in graph, $\mathbf{A} \in \mathbb{R}^{n \times n}$: adjacency matrix, $\mathbf{D} \in \mathbb{R}^{n \times n}$: diagonal degree matrix, $\mathbf{L} \in \mathbb{R}^{n \times n}$: Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$.

Laplacian Embedding

The smallest eigenvectors of $\mathbf{L} = \mathbf{D} - \mathbf{A}$ give the orthogonal 'functions' that are smoothest over the graph. I.e., minimize

$$\vec{v}^T \mathbf{L} \vec{v} = \sum_{(i,j) \in E} [\vec{v}(i) - \vec{v}(j)]^2.$$

Laplacian Embedding

The smallest eigenvectors of $\mathsf{L} = \mathsf{D} - \mathsf{A}$ give the orthogonal 'functions' that are smoothest over the graph. I.e., minimize

$$\vec{v}^T \mathsf{L} \vec{v} = \sum_{(i,j) \in E} [\vec{v}(i) - \vec{v}(j)]^2.$$

Embedding points with coordinates given by $[\vec{v}_{n-1}(j), \vec{v}_{n-2}(j), \ldots, \vec{v}_{n-k}(j)]$ ensures that coordinates connected by edges have minimum total squared Euclidean distance.

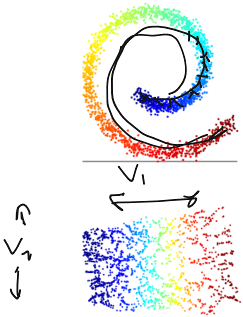
[2,3]

[-1,2]

The smallest eigenvectors of $\mathbf{L} = \mathbf{D} - \mathbf{A}$ give the orthogonal 'functions' that are smoothest over the graph. I.e., minimize

$$\vec{v}^T \mathbf{L} \vec{v} = \sum_{(i,j) \in E} [\vec{v}(i) - \vec{v}(j)]^2.$$

Embedding points with coordinates given by $[\vec{v}_{n-1}(j), \vec{v}_{n-2}(j), \ldots, \vec{v}_{n-k}(j)]$ ensures that coordinates connected by edges have minimum total squared Euclidean distance.



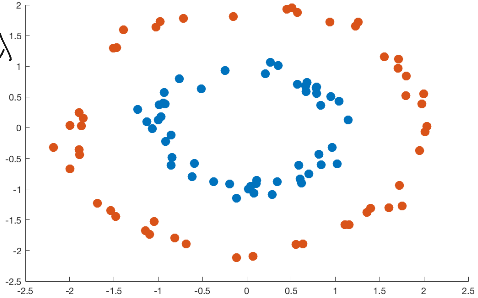
- Spectral Clustering
- Laplacian Eigenmaps
- Locally linear embedding
- Isomap
- Node2Vec, DeepWalk, etc. (variants on Laplacian)

agglomerative clustering

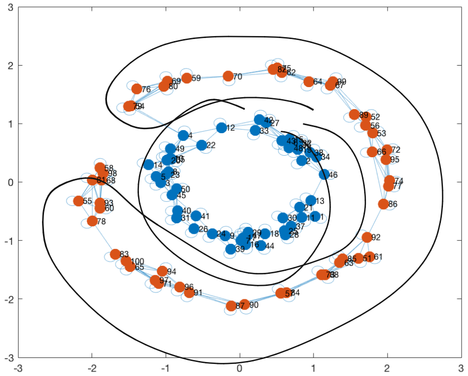
kernel method

pick the kernel

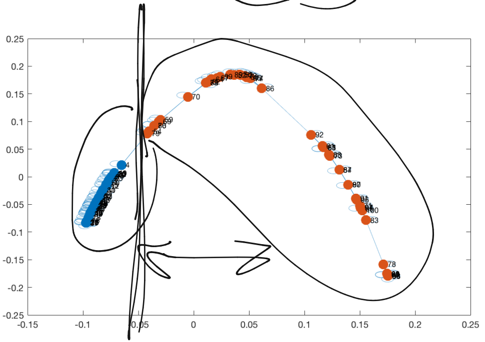
Original Data: (not linearly separable)



k-Nearest Neighbors Graph:



Embedding with eigenvectors $\underbrace{\vec{v}_{n-1}}, \underbrace{\vec{v}_{n-2}}$: (linearly separable)



V_{n-1}

Generative Models

So Far: Have argued that spectral clustering partitions a graph effectively, along a small cut that separates the graph into large pieces. But it is difficult to give any formal guarantee on the 'quality' of the partitioning in general graphs.

Generative Models

So Far: Have argued that spectral clustering partitions a graph effectively, along a small cut that separates the graph into large pieces. But it is difficult to give any formal guarantee on the 'quality' of the partitioning in general graphs.

λ_{n-1}

(Fiedler value

small = well-partitioned

big = not well-partitioned

Common Approach: Give a natural generative model for random inputs and analyze how the algorithm performs on inputs drawn from this model.

- Very common in algorithm design for data analysis/machine learning (can be used to justify least squares regression, k-means clustering, PCA, etc.)
- We'll do this next time, introducing the Stochastic Block Model.