

One Week Online Workshop on Statistics and Machine Learning in Practice
Brahmananda Keshab Chandra College

SPECTRAL CLUSTERING

APARAJITA KHAN
SENIOR RESEARCH FELLOW
INDIAN STATISTICAL INSTITUTE
JULY 29, 2020

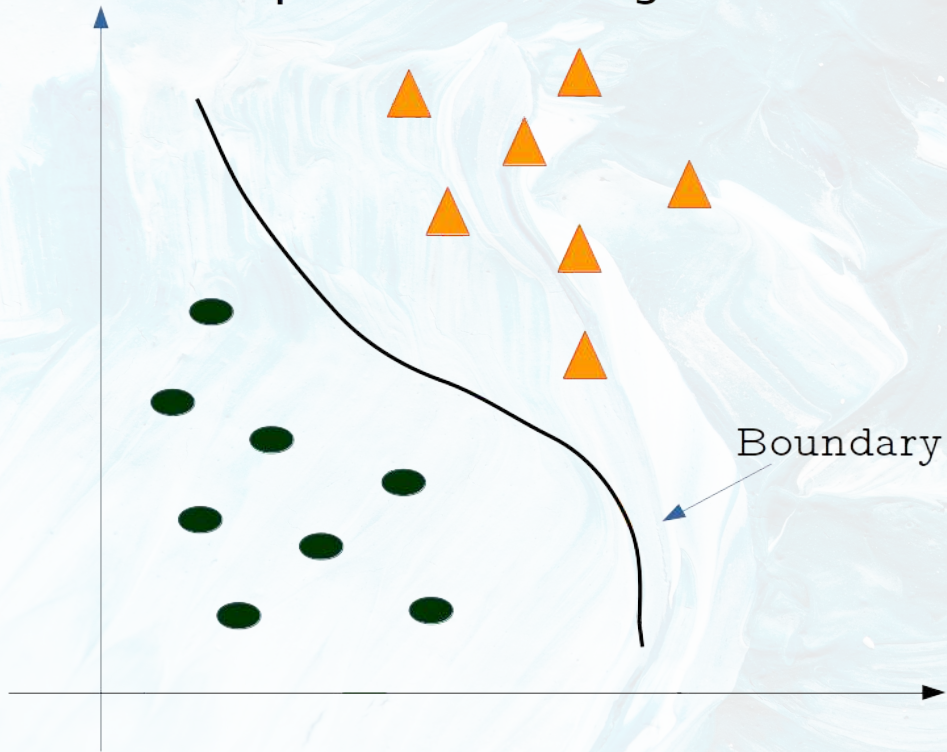
CLUSTERING

The study of natural groupings in data

CLUSTERING

The study of natural groupings in data

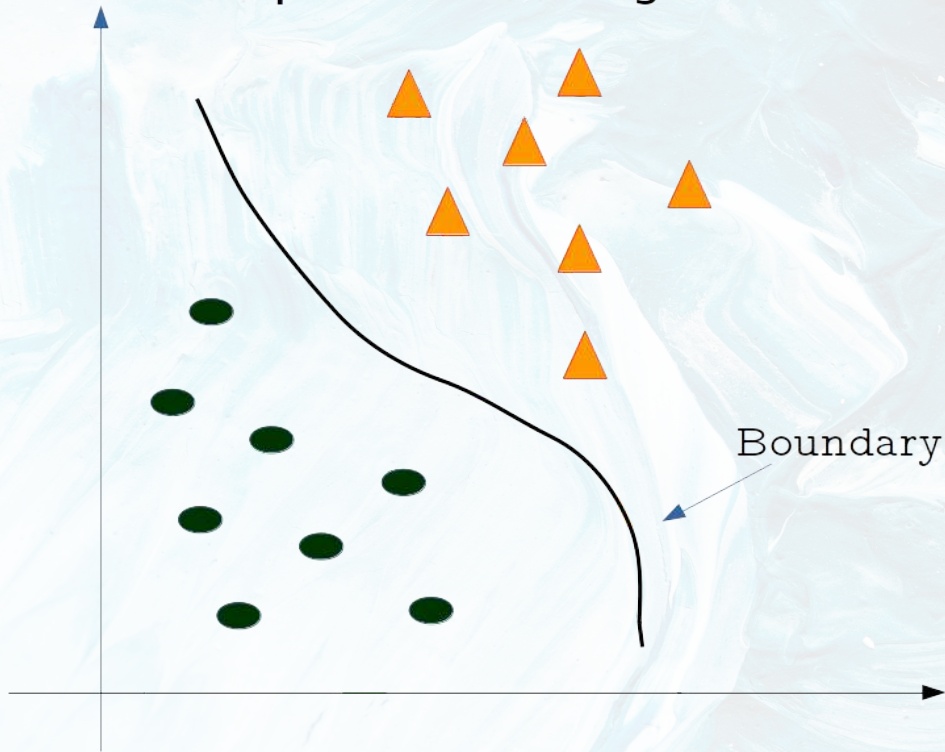
Supervised Learning



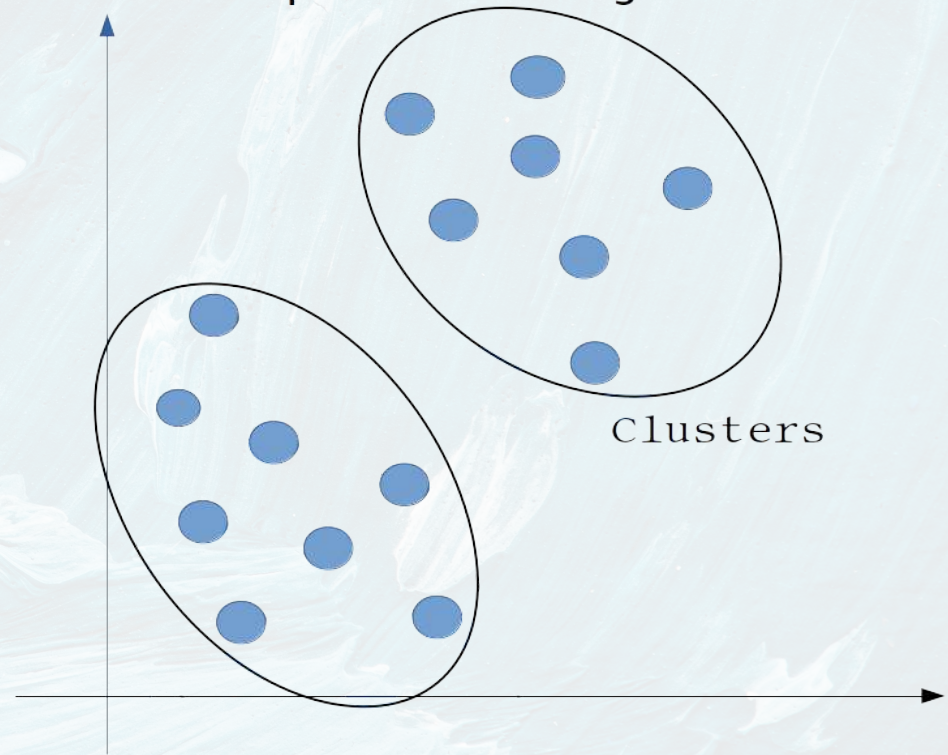
CLUSTERING

The study of natural groupings in data

Supervised Learning



Unsupervised Learning



CLUSTERING: Applications

Grouping similar news articles

https://www.indiatoday.in/news.html 60%

NEWS

INDIA

Rajasthan crisis: Governer under pressure from BJP, says Congress

- Over 3,300 coronavirus patients 'untraceable' in Bengaluru amid spike in cases
- Covid bed occupancy coming down, few people need hospitalisation now: Kejriwal
- Coronavirus in India: Bodies of Covid patients burnt in open in Patna
- Watch | India battles Covid-19
- Kargil Vijay Diwas 2020: From Rajnath Singh to Rahul Gandhi, leaders pay tribute to brave martyrs
- Pressman House: The building at heart of Maharashtra political controversy
- Kolkata: Ambulance driver demands Rs 9,200 from coronavirus patients for 6-km journey to hospital

SPORTS

IPL 2020's knock-on effect on UAE economy will be huge: Kumar Sangakkara

- It was about changing the momentum of innings: Stuart Broad on his 33-ball fifty
- Have never thought of bringing the ODI style of play in Test cricket: Stuart Broad
- England vs West Indies: Broad, Stokes put hosts on top in Test match
- Racism a topical and burning issue for Indians and Sri Lankans: Kumar Sangakkara
- Ganguly has an astute cricket brain, he will be a fair ICC chief: Sangakkara
- If MS Dhoni thinks he can still win matches for India, he should play: Gautam Gambhir
- What's done is done: Irfan Pathan lashes out at Steve Bucknor over 2008 howlers

MOVIES

AR Rahman: A gang in Bollywood is spreading false rumours about me

- Rajkumar Rao gets angry after watching Dil Bechara
- Arjun, Zoya, Karthi: The new trio: Vote for them
- The Akhtars on nepotism and celebrity culture
- Sonu Sood on e-Mind Rocks 2020: I wasn't

TRENDING NEWS

Dr PK Mahanandi: The untouchable boy who became art advisor for Swedish government

- Lund University has had enough of Indians on Facebook
- Carryminati's YouTube account hacked, hacker asks for bitcoin donations
- Adorable video of elephant playing with its human friend goes viral

Covaxin enters human trials at AIIMS: All you need to know

Rajasthan crisis explained in 10 point: From Pilot, Gehlot to ED and Malinga

SUGGESTED STORIES

WATCH RIGHT NOW

Watch: Hyderabad's Osmania Hospital once again flooded after heavy rain

Assam floods: Nearly 27 people affected, 91 killed in deluge

Don't buy it from black market: Delhi CM Kejriwal on Operation Plasma Bazaar | EXCLUSIVE

WATCH: India's coronavirus tally exceeds 12 lakh

Why not Manesar hotel: Chidambaram asks ED after raids on Ashok Gehlot's brother

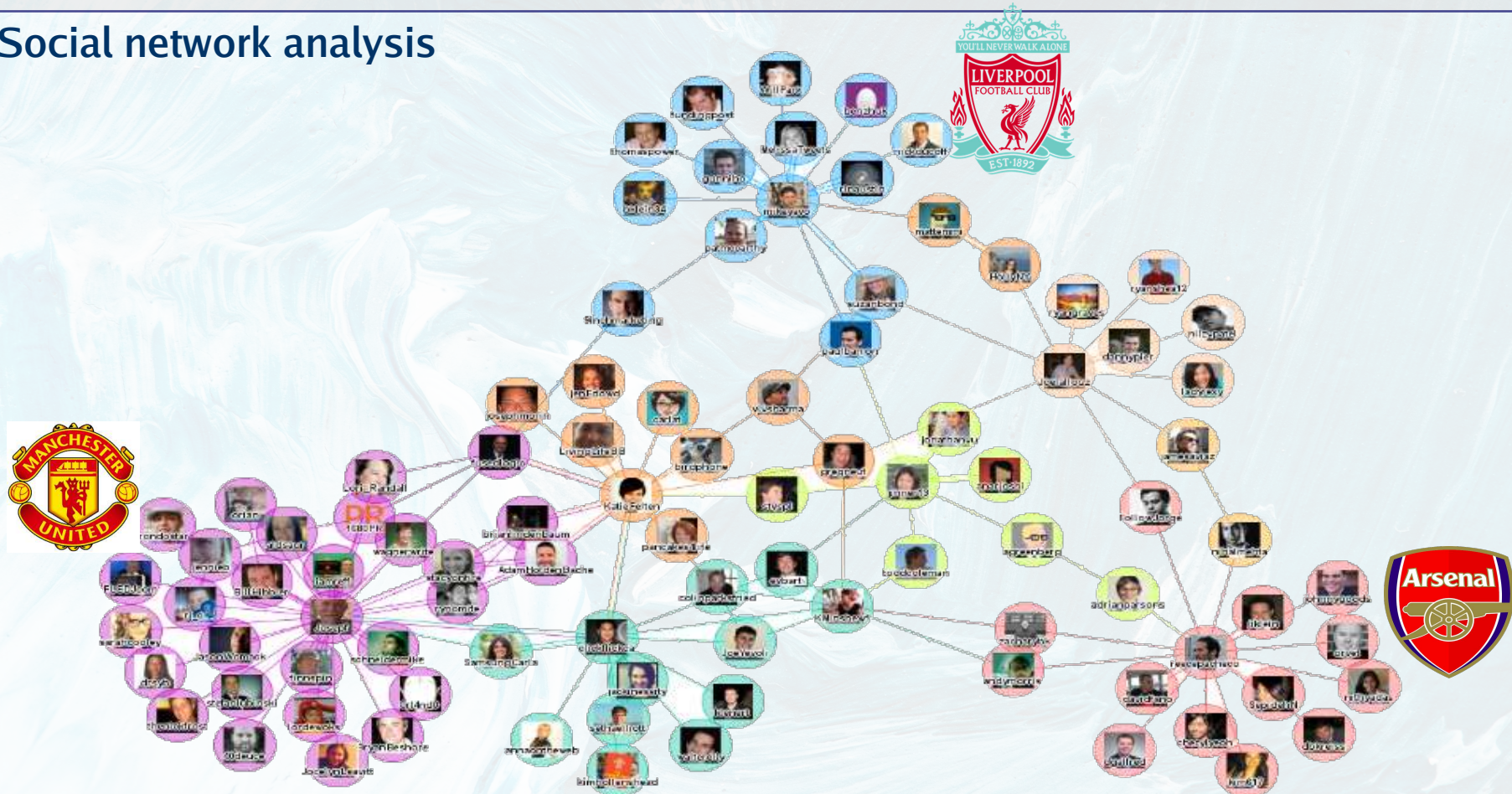
India/Corona cluster

sports cluster

Entertainment cluster

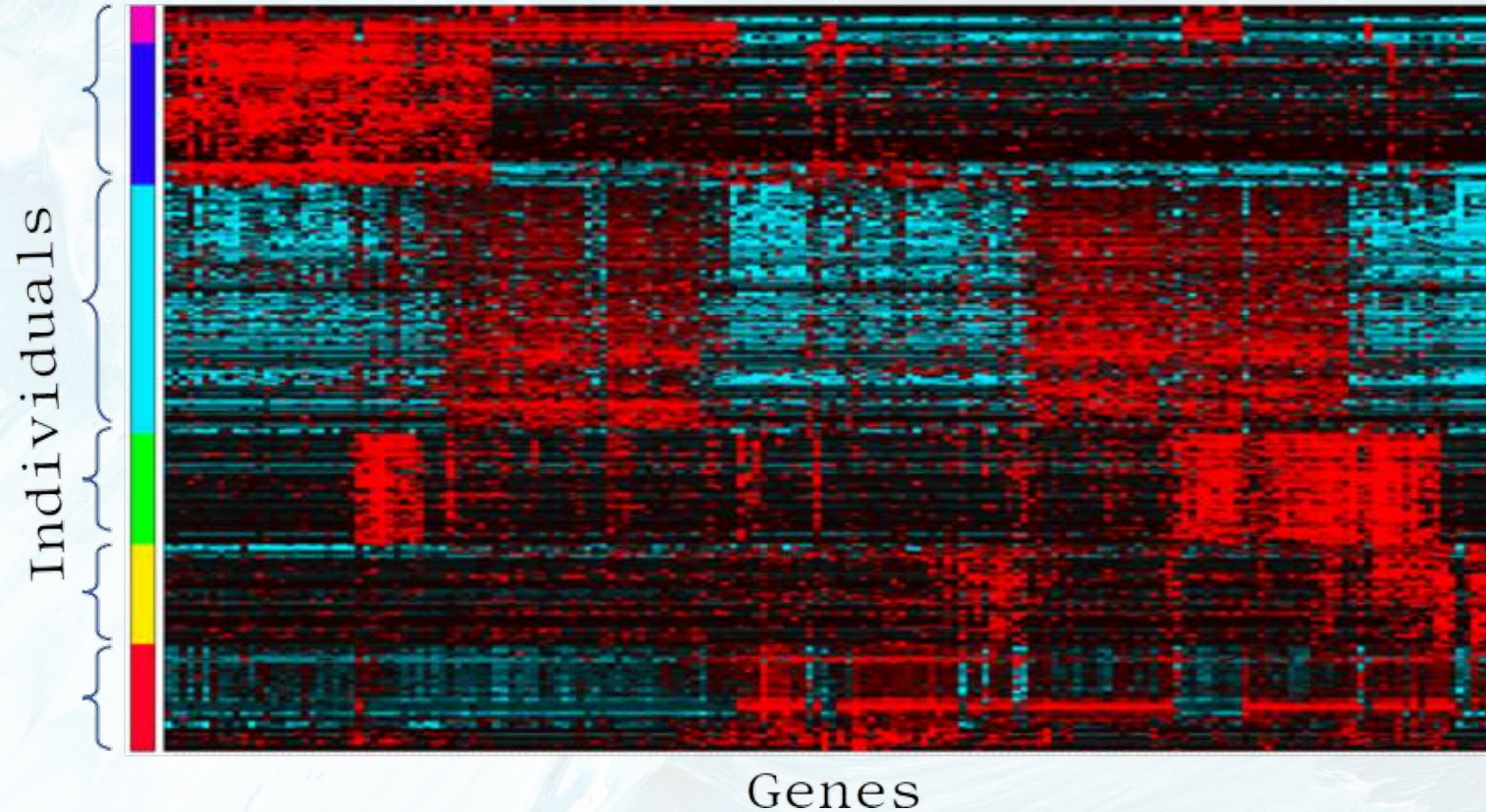
CLUSTERING: Applications

Social network analysis



CLUSTERING: Applications

Bioinformatics: Identifying disease subtypes, patient groups



CLUSTERING: Applications

Image segmentation



FEATURE-SPACE VS GRAPH CLUSTERING

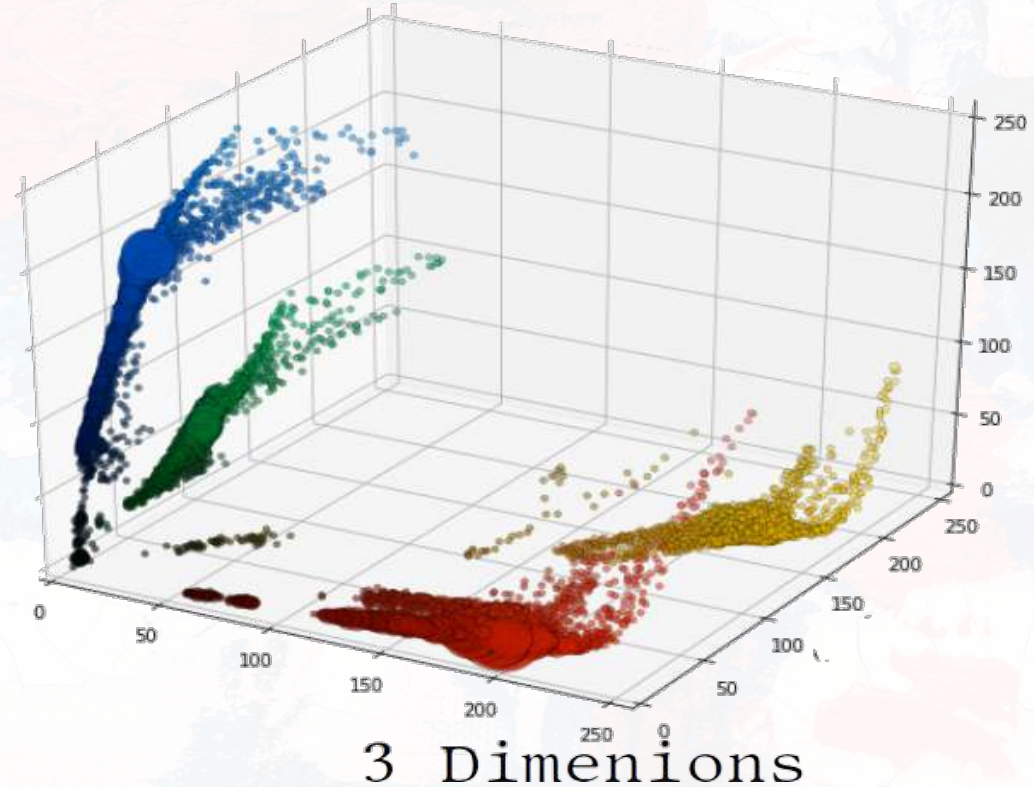
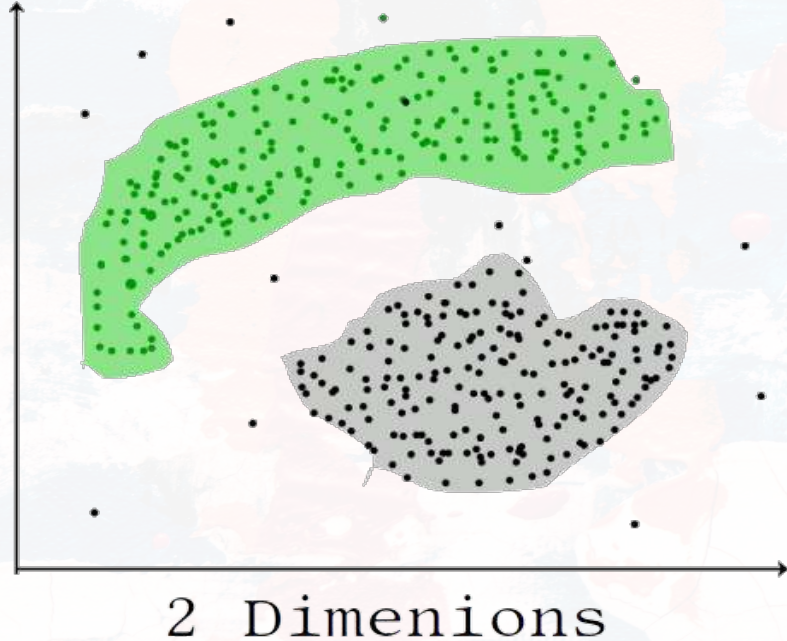
Feature-space: k -Means Clustering

$\mathbf{X} \in \mathbb{R}^{n \times d}$ n samples, d features

FEATURE-SPACE VS GRAPH CLUSTERING

Feature-space: k -Means Clustering

$\mathbf{X} \in \mathbb{R}^{n \times d}$ n samples, d features



FEATURE-SPACE VS GRAPH CLUSTERING

Feature-space: *k*-Means Clustering

- But, what about high dimensions?

Disease subtyping: ~1,000 samples ~20,000 genes

Object recognition: 1024x1024 images -> ~1M pixels

FEATURE-SPACE VS GRAPH CLUSTERING

Feature-space: *k*-Means Clustering

- But, what about high dimensions?

Disease subtyping: ~1,000 samples ~20,000 genes

Object recognition: 1024x1024 images -> ~1M pixels

- In such high dimensions:

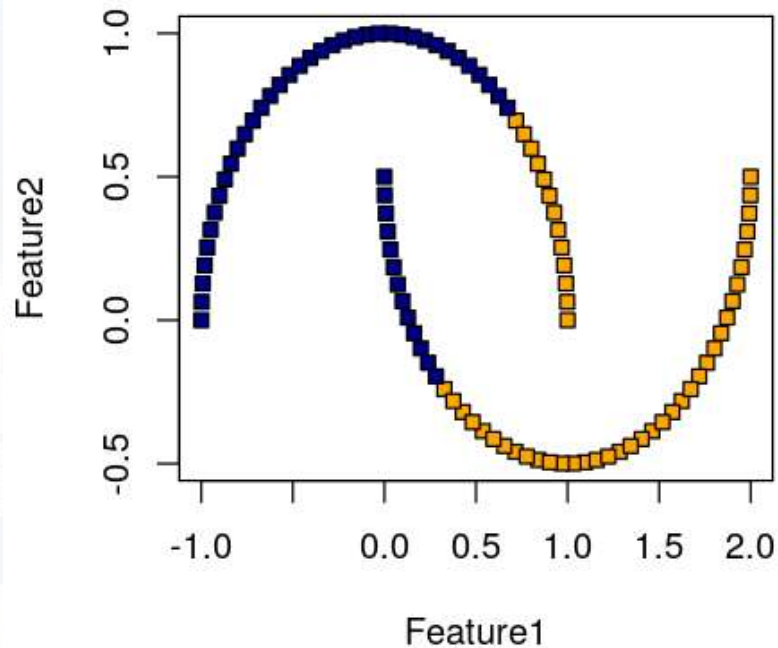
Data becomes geometrically sparse

Distance between nearby points **roughly** same as distance between far away points

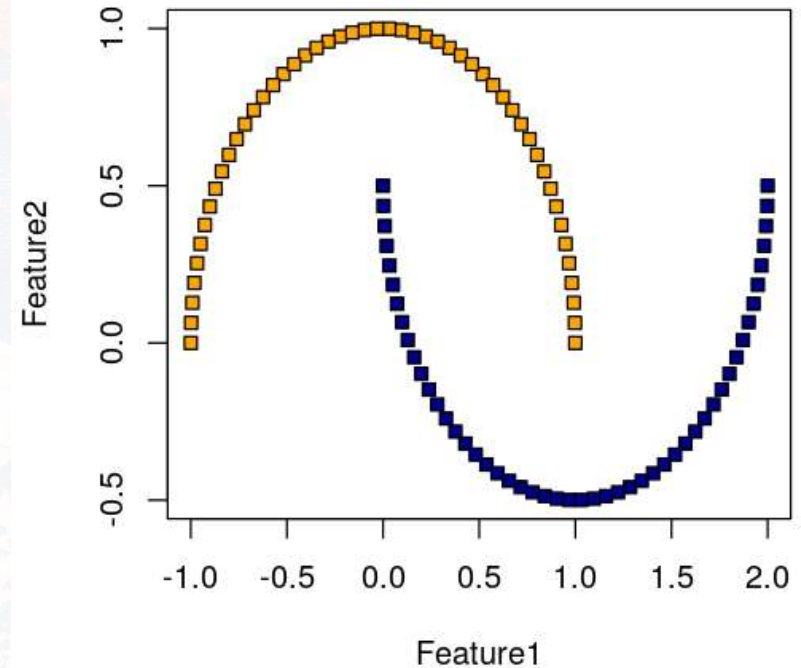
FEATURE-SPACE VS GRAPH CLUSTERING

Handling non-linearity

Cluster assignment for k-Means



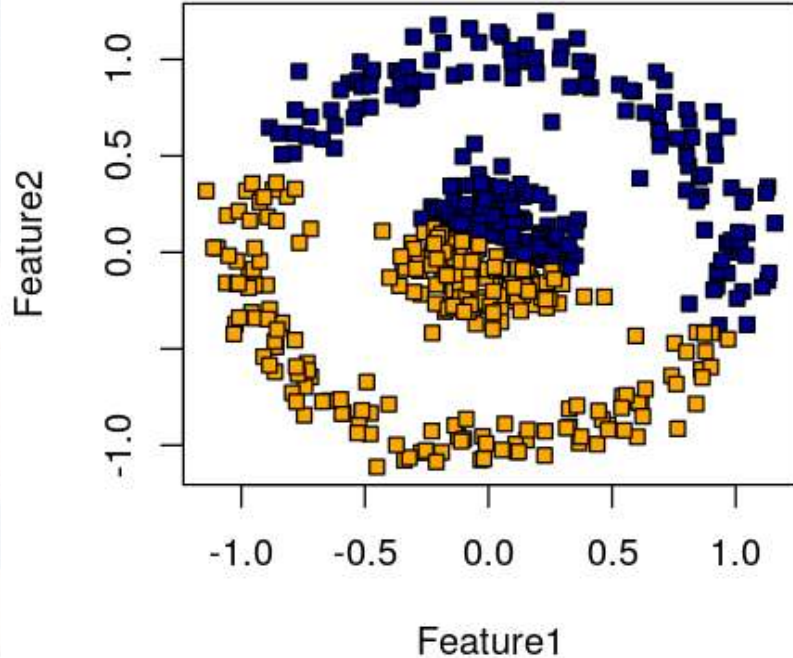
Cluster assignment for spectral clustering



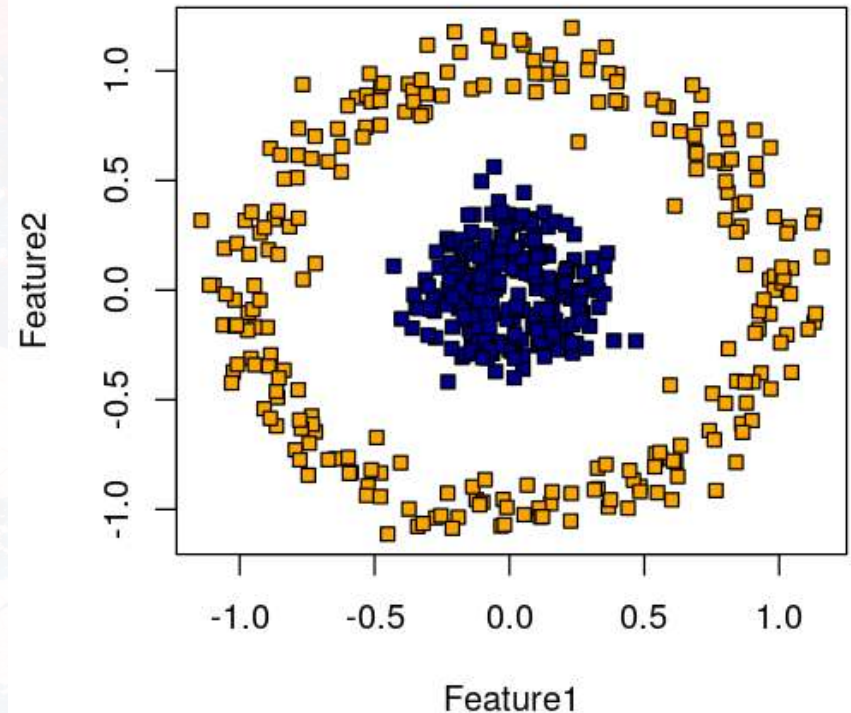
FEATURE-SPACE VS GRAPH CLUSTERING

Handling non-linearity

Cluster assignment for k-Means

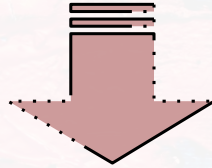


Cluster assignment for spectral clustering



FEATURE-SPACE VS GRAPH CLUSTERING

FEATURE-SPACE BASED REPRESENTATION



GRAPH BASED REPRESENTATION

FEATURE-SPACE VS GRAPH CLUSTERING

Feature-space

$$\mathbf{X} \in \mathbb{R}^{n \times d}$$

Samples: n

Features/Dimension: d

Graph

FEATURE-SPACE VS EIGENSPACE CLUSTERING

Feature-space

$$\mathbf{X} \in \mathbb{R}^{n \times d}$$

Samples: n
Features/Dimension: d

$$\kappa(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)$$



Gaussian similarity kernel

Graph

$$\mathbf{W} \in \mathbb{R}^{n \times n}$$

Pairwise
similarity
matrix

FEATURE-SPACE VS GRAPH CLUSTERING

Feature-space

$$\mathbf{X} \in \mathbb{R}^{n \times d}$$

Samples: n
Features/Dimension: d

$$\kappa(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)$$

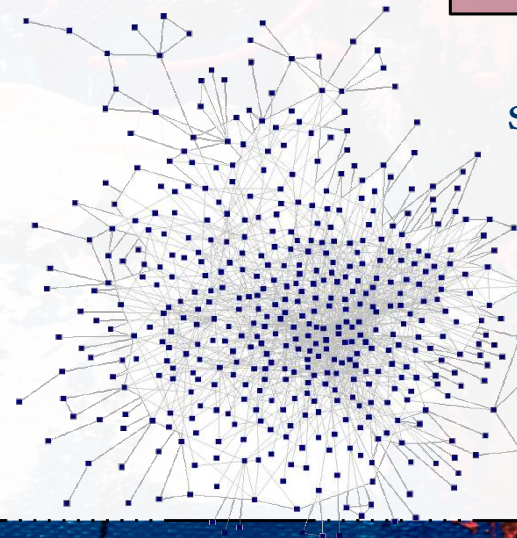


Gaussian similarity kernel

Graph

$$\mathbf{W} \in \mathbb{R}^{n \times n}$$

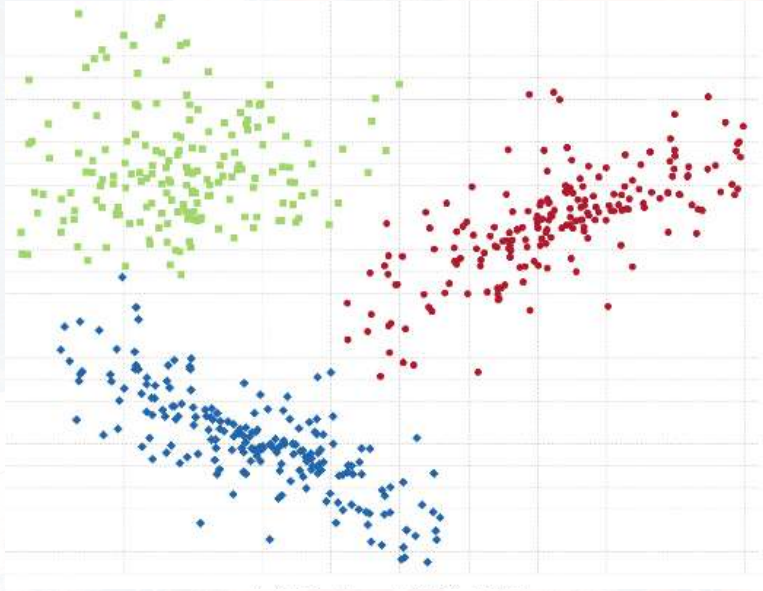
Pairwise
similarity
matrix



G \nearrow weighted
adjacency

FEATURE-SPACE VS GRAPH CLUSTERING

Feature-space

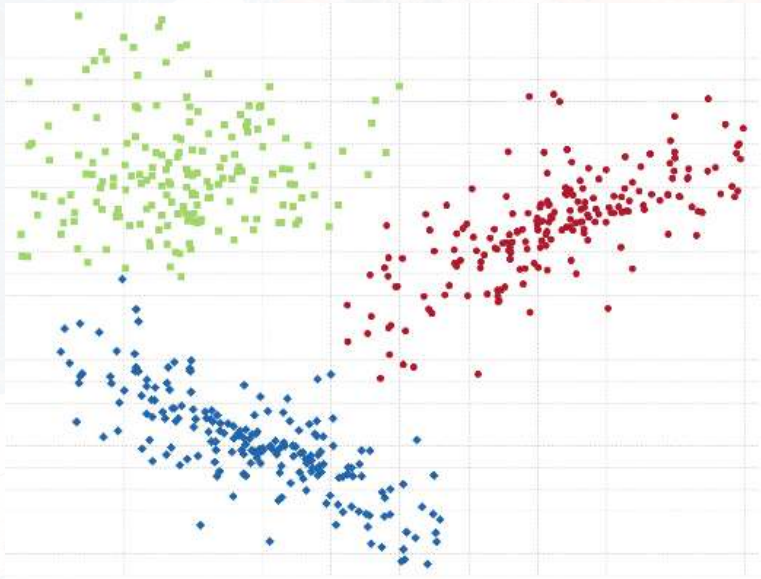


Graph

Clusters in dataset

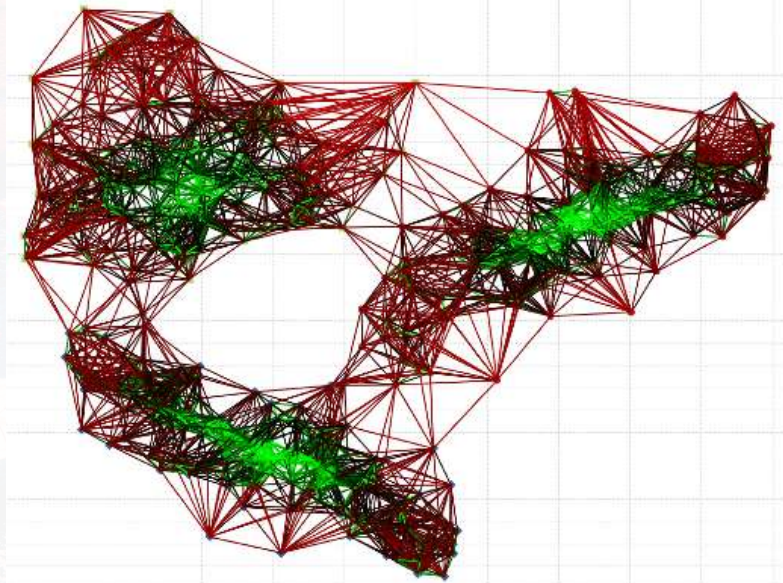
FEATURE-SPACE VS GRAPH CLUSTERING

Feature-space



Clusters in dataset

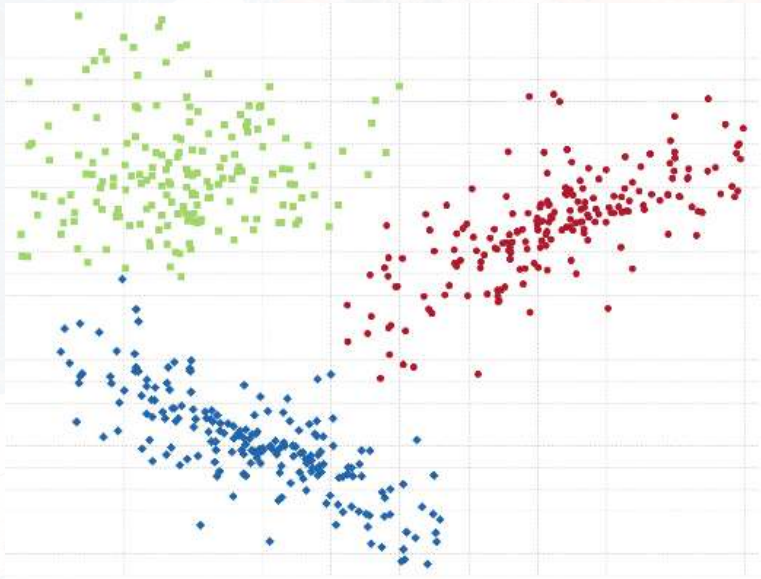
Graph



Strongly connected components in graph

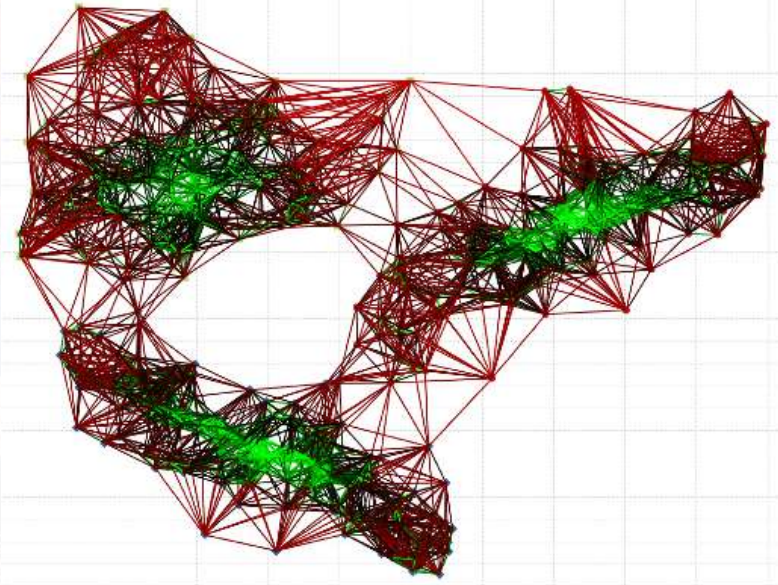
FEATURE-SPACE VS GRAPH CLUSTERING

Feature-space



$$\text{minimize } \sum_{j=1}^k \sum_{i \in \mathcal{C}_j} \left\| x_i - \frac{1}{|\mathcal{C}_j|} \sum_{t \in \mathcal{C}_j} x_t \right\|_2^2$$

Graph



Graph-cut problem

THE GRAPH-CUT

Find a k -way partition of graph such that:

Edges between different components have very low weight

(points in different clusters are dissimilar)

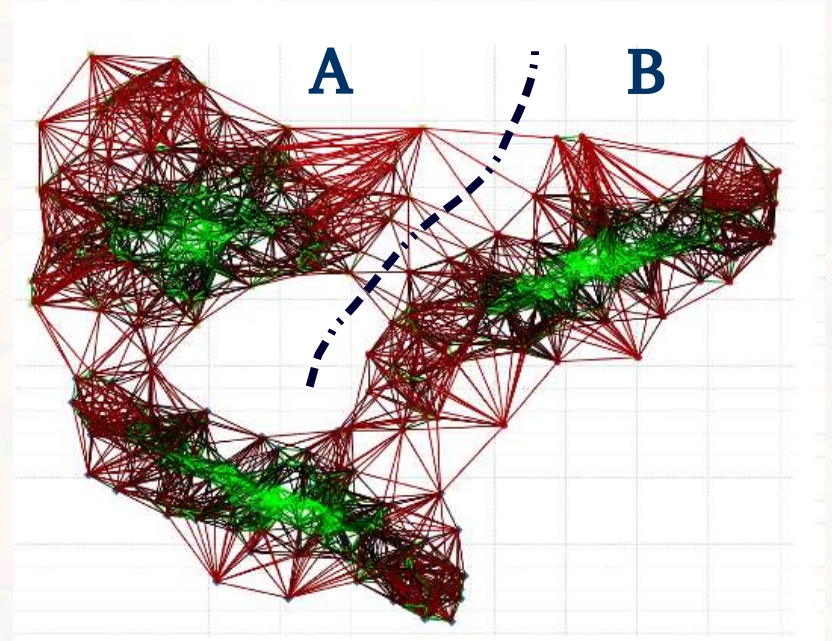
Edges within a component have high weight

(points within same cluster are similar)

THE GRAPH-CUT

$$\mathbf{W} = [w_{ij}]_{n \times n}$$

$$\text{cut } W(A, B) := \sum_{i \in A, j \in B} w_{ij}$$



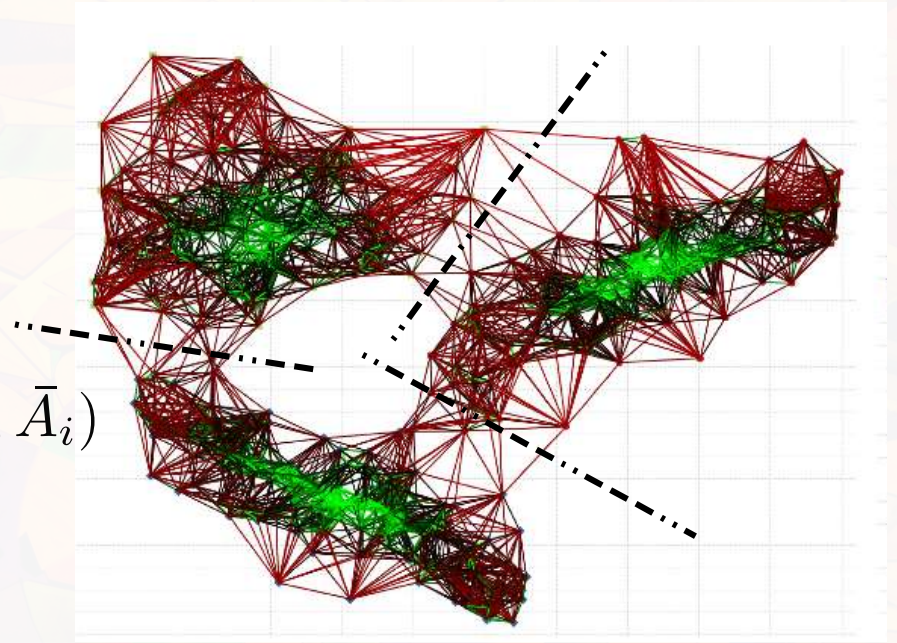
THE GRAPH-CUT

$$\mathbf{W} = [w_{ij}]_{n \times n}$$

$$\text{cut } W(A, B) := \sum_{i \in A, j \in B} w_{ij}$$

k-way cut:

$$\text{minimize cut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k W(A_i, \bar{A}_i)$$



THE GRAPH-CUT

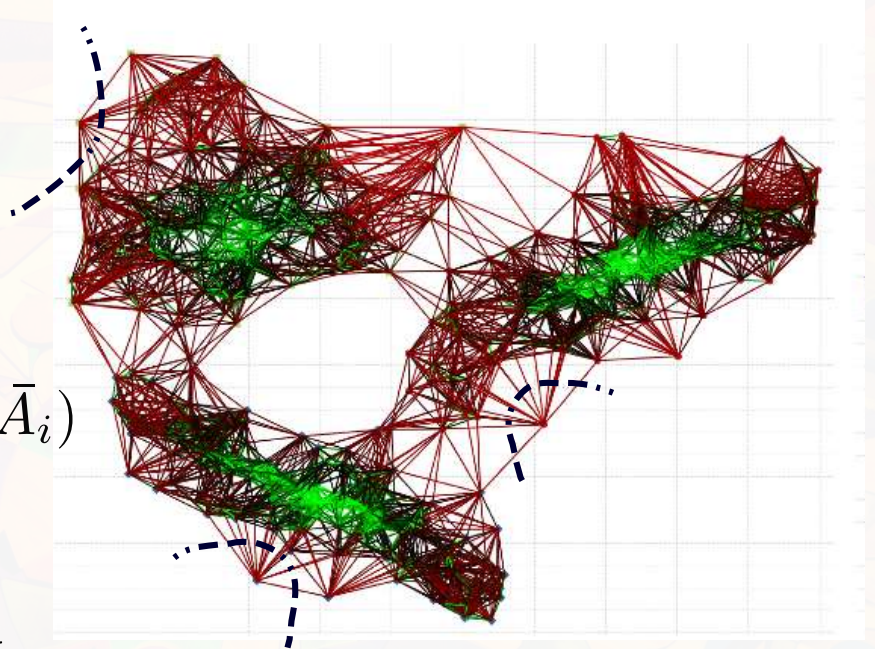
$$\mathbf{W} = [w_{ij}]_{n \times n}$$

$$\text{cut } W(A, B) := \sum_{i \in A, j \in B} w_{ij}$$

k-way cut:

$$\text{minimize cut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k W(A_i, \bar{A}_i)$$

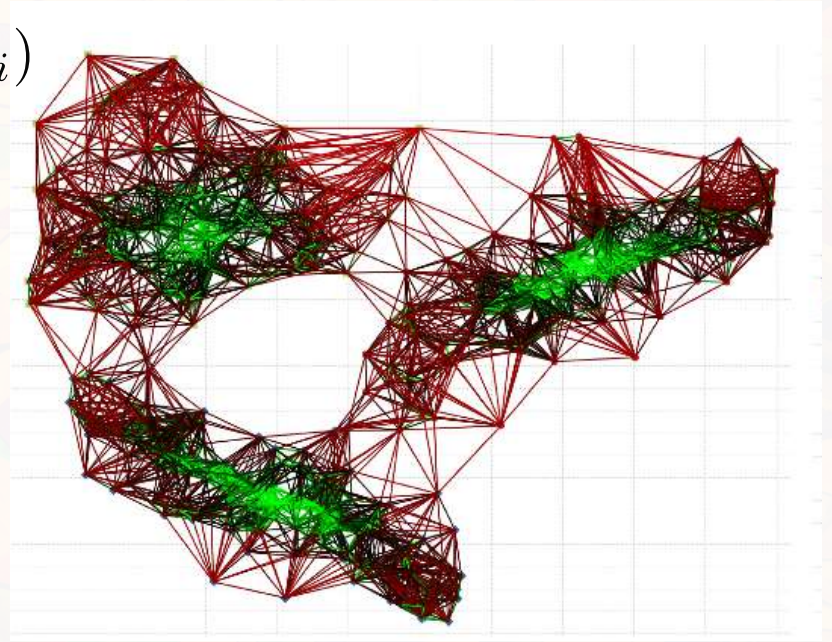
- In many cases, mincut simply separates individual vertices from rest of the graph



THE GRAPH-CUT

minimize $\text{cut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k W(A_i, \bar{A}_i)$

Explicitly request subsets
to be “reasonably large”



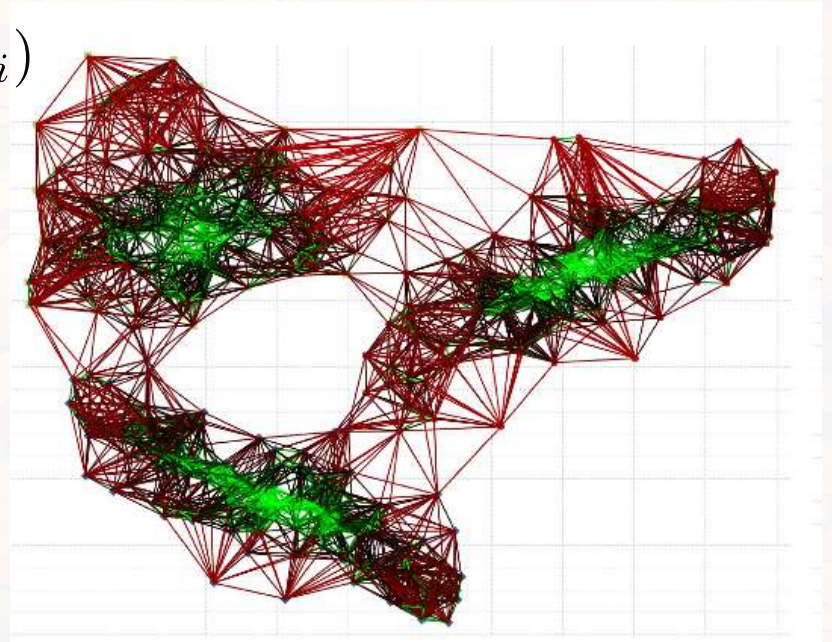
THE GRAPH-CUT

minimize $\text{cut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k W(A_i, \bar{A}_i)$

Explicitly request subsets
to be “reasonably large”

degree of vertex v_i : $d_i = \sum_{j=1}^n w_{ij}$

size of subset: $\text{vol}(A_i) := \sum_{i \in A} d_i$



THE GRAPH-CUT

minimize $\text{cut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k W(A_i, \bar{A}_i)$

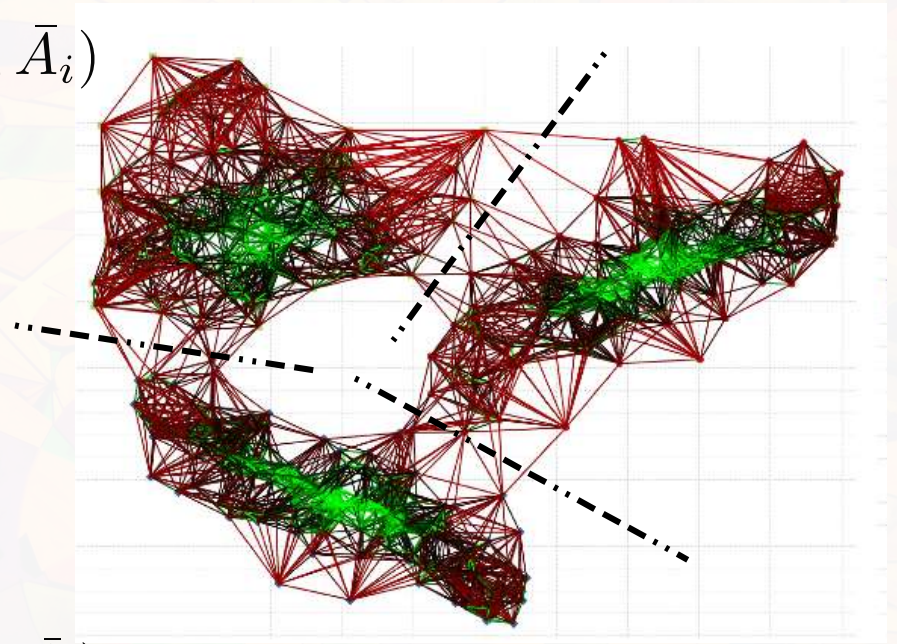
Explicitly request subsets
to be “reasonably large”

degree of vertex v_i : $d_i = \sum_{j=1}^n w_{ij}$

size of subset: $\text{vol}(A_i) := \sum_{i \in A} d_i$

Normalized cut

minimize $\text{Ncut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{\text{vol}(A_i)}$



SOLVING THE GRAPH-CUT

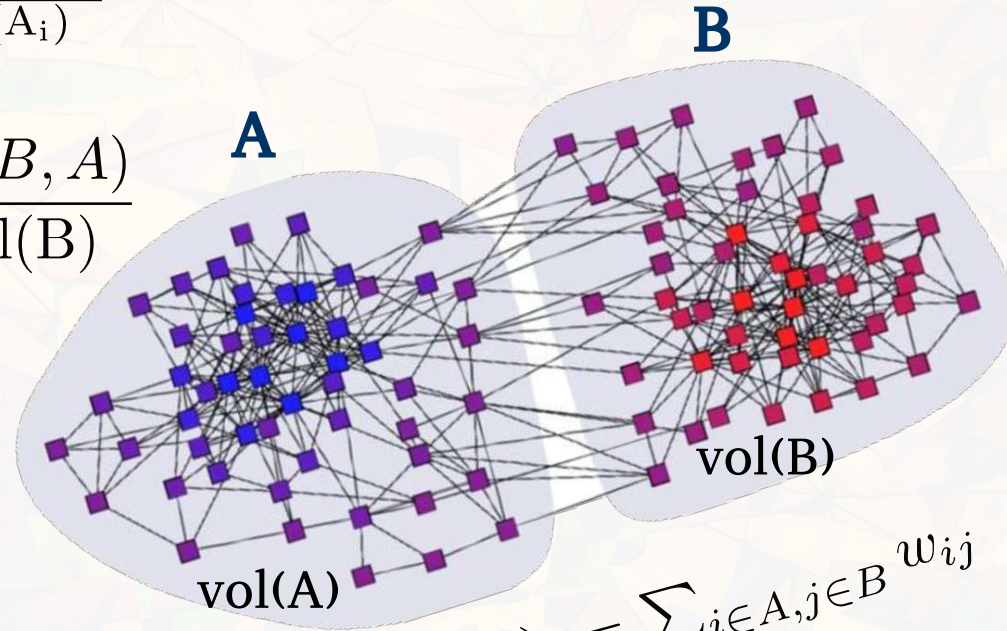
$$\text{minimize } \text{Ncut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{\text{vol}(A_i)}$$

SOLVING THE GRAPH-CUT

$$\text{minimize Ncut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{\text{vol}(A_i)}$$

Solving 2-way cut

$$\text{minimize Ncut}(A, B) := \frac{W(A, B)}{\text{vol}(A)} + \frac{W(B, A)}{\text{vol}(B)}$$



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

SOLVING THE GRAPH-CUT

$$\text{minimize } \text{Ncut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{\text{vol}(A_i)}$$

Solving 2-way cut

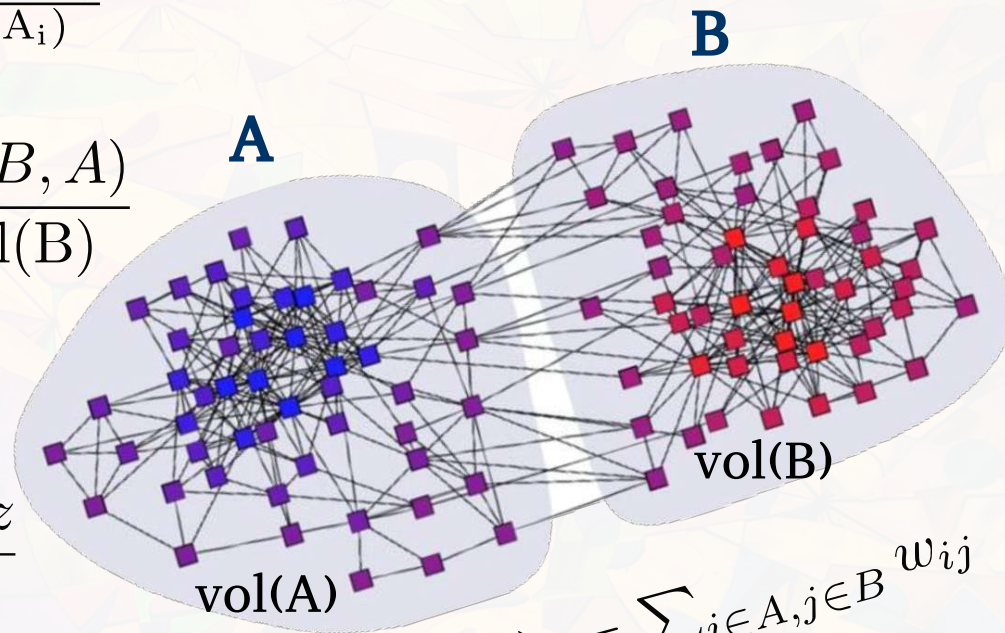
$$\text{minimize } \text{Ncut}(A, B) := \frac{W(A, B)}{\text{vol}(A)} + \frac{W(B, A)}{\text{vol}(B)}$$



Reduces to

$$\text{minimize}_{z \in \mathbb{R}^n} \frac{z^T \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W}) \mathbf{D}^{-\frac{1}{2}} z}{z^T z}$$

$$\text{such that } z^T \left(\mathbf{D}^{\frac{1}{2}} \mathbf{1} \right) = 0$$



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

SOLVING THE GRAPH-CUT

$$\begin{aligned} & \underset{z \in \mathbb{R}^n}{\text{minimize}} \quad \frac{z^T \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W}) \mathbf{D}^{-\frac{1}{2}} z}{z^T z} \\ & \text{such that } z^T \left(\mathbf{D}^{\frac{1}{2}} \mathbf{1} \right) = 0 \end{aligned}$$

SOLVING THE GRAPH-CUT

$$\begin{aligned} & \underset{z \in \mathbb{R}^n}{\text{minimize}} \quad \frac{z^T \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W}) \mathbf{D}^{-\frac{1}{2}} z}{z^T z} \\ & \text{such that } z^T \left(\mathbf{D}^{\frac{1}{2}} \mathbf{1} \right) = 0 \end{aligned}$$

\mathbf{W} = Pairwise-similarity matrix of size $n \times n$

$$\mathbf{D} = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix} \text{ Degree matrix}$$

SOLVING THE GRAPH-CUT

$$\begin{aligned} &\underset{z \in \mathbb{R}^n}{\text{minimize}} \quad \frac{z^T \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W}) \mathbf{D}^{-\frac{1}{2}} z}{z^T z} \\ &\text{such that } z^T \left(\mathbf{D}^{\frac{1}{2}} \mathbf{1} \right) = 0 \end{aligned}$$

The normalized graph Laplacian

$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W}) \mathbf{D}^{-\frac{1}{2}}$$

SOLVING THE GRAPH-CUT

$$\begin{aligned} & \underset{z \in \mathbb{R}^n}{\text{minimize}} \quad \frac{z^T \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W}) \mathbf{D}^{-\frac{1}{2}} z}{z^T z} \\ & \text{such that } z^T \left(\mathbf{D}^{\frac{1}{2}} \mathbf{1} \right) = 0 \end{aligned}$$

The normalized graph Laplacian

$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W}) \mathbf{D}^{-\frac{1}{2}}$$

- Symmetric positive semi-definite

$$v^T \mathbf{L} v \geq 0, \forall v \in \mathbb{R}^n, v \neq \mathbf{0}$$

all the eigenvalues are ≥ 0

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

SOLVING THE GRAPH-CUT

$$\begin{aligned} & \underset{z \in \mathbb{R}^n}{\text{minimize}} \quad \frac{z^T \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W}) \mathbf{D}^{-\frac{1}{2}} z}{z^T z} \\ & \text{such that } z^T \left(\mathbf{D}^{\frac{1}{2}} \mathbf{1} \right) = 0 \end{aligned}$$

The normalized graph Laplacian

$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W}) \mathbf{D}^{-\frac{1}{2}}$$

- Symmetric positive semi-definite

$$v^T \mathbf{L} v \geq 0, \forall v \in \mathbb{R}^n, v \neq \mathbf{0}$$

all the eigenvalues are ≥ 0

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

Rayleigh quotient:

Let A be a real symmetric matrix
Constraint: v is orthogonal to $j-1$
smallest eigenvectors v_1, \dots, v_{j-1} , the

$$\text{quotient } \frac{v^T A v}{v^T v}$$

is minimized by next smallest
eigenvector v_j and eigenvalue λ_j

SOLVING THE GRAPH-CUT

$$\underset{z \in \mathbb{R}^n}{\text{minimize}} \frac{z^T \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W}) \mathbf{D}^{-\frac{1}{2}} z}{z^T z}$$

such that $z^T \left(\mathbf{D}^{\frac{1}{2}} \mathbf{1} \right) = 0$

Minimized by:
second smallest eigenvector
of \mathbf{L} and its eigenvalue

The normalized graph Laplacian

$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W}) \mathbf{D}^{-\frac{1}{2}}$$

- Symmetric positive semi-definite

$$v^T \mathbf{L} v \geq 0, \forall v \in \mathbb{R}^n, v \neq \mathbf{0}$$

all the eigenvalues are ≥ 0

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

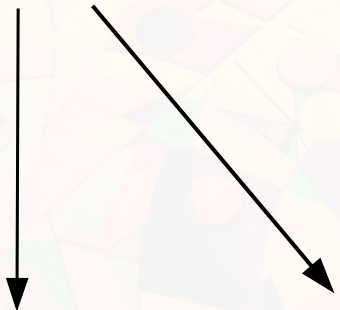
SOLVING THE GRAPH-CUT

$$\underset{z \in \mathbb{R}^n}{\text{minimize}} \quad \frac{z^T \mathbf{L} z}{z^T z} \quad \text{such that} \quad z^T \left(\mathbf{D}^{\frac{1}{2}} \mathbf{1} \right) = 0$$

SOLVING THE GRAPH-CUT

$$\underset{z \in \mathbb{R}^n}{\text{minimize}} \frac{z^T \mathbf{L} z}{z^T z} \text{ such that } z^T (\mathbf{D}^{\frac{1}{2}} \mathbf{1}) = 0$$

Eigen decomposition $\mathbf{L} = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^T$


$$\begin{bmatrix} \sqrt{d_1} & | & & & | \\ \vdots & u_2 & u_3 & \dots & u_n \\ \sqrt{d_n} & | & | & & | \end{bmatrix} \quad \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \lambda_n \end{bmatrix}$$

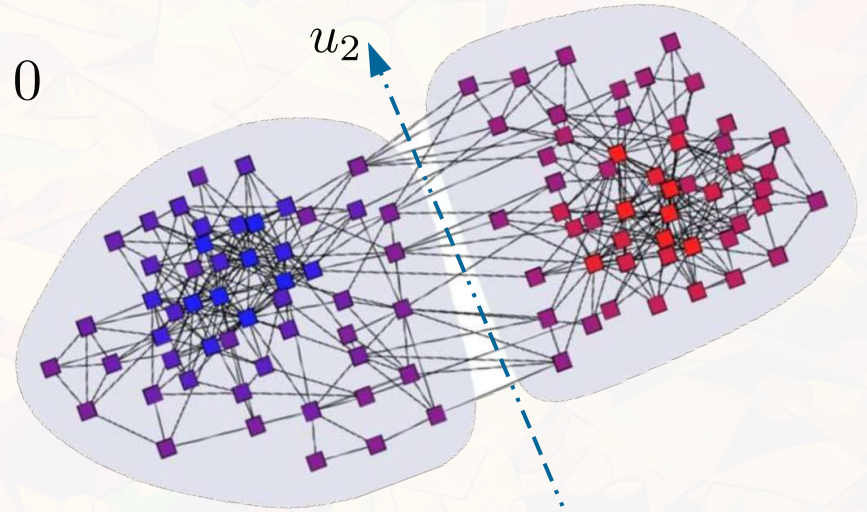
(arranged in ascending order)

SOLVING THE GRAPH-CUT

$$\underset{z \in \mathbb{R}^n}{\text{minimize}} \frac{z^T \mathbf{L} z}{z^T z} \text{ such that } z^T (\mathbf{D}^{\frac{1}{2}} \mathbf{1}) = 0$$

Eigen decomposition $\mathbf{L} = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^T$

$$\begin{bmatrix} \sqrt{d_1} & | & & & | \\ \vdots & & & & \\ \sqrt{d_n} & | & u_3 & \dots & | \\ & | & & & | \end{bmatrix} \begin{matrix} | \\ u_2 \\ | \end{matrix} \begin{matrix} | \\ u_3 \\ | \end{matrix} \dots \begin{matrix} | \\ u_n \\ | \end{matrix} \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \lambda_n \end{bmatrix}$$



SOLVING THE GRAPH-CUT

Solving k-way cut

$$\text{minimize Ncut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{\text{vol}(A_i)}$$

SOLVING THE GRAPH-CUT

Solving k-way cut

$$\text{minimize } \text{Ncut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{\text{vol}(A_i)}$$



$$\text{minimize } \text{trace}(U^T \mathbf{L} U)$$

$$\text{such that } U \in \mathbb{R}^{n \times k}, U^T U = \mathbf{I}_k$$

SOLVING THE GRAPH-CUT

Solving k-way cut

$$\text{minimize } \text{Ncut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{\text{vol}(A_i)}$$



$$\text{minimize } \text{trace}(U^T \mathbf{L} U)$$

$$\text{such that } U \in \mathbb{R}^{n \times k}, U^T U = \mathbf{I}_k$$

$$= \sum_{i=1}^k \lambda_i(\mathbf{L}) \quad [\text{Ky-Fan theorem}]$$

$$U^* = \text{Eigenvectors corresponding to } k \text{ smallest eigenvalues of } \mathbf{L}$$

SOLVING THE GRAPH-CUT

Solving k-way cut

$$\text{minimize } \text{Ncut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{\text{vol}(A_i)}$$



$$\text{minimize } \text{trace}(U^T \mathbf{L} U)$$

$$\text{such that } U \in \mathbb{R}^{n \times k}, U^T U = \mathbf{I}_k$$

$$\text{Eigen decomposition } \mathbf{L} = U \Sigma U^T$$

$$\begin{bmatrix} | & | & & | & | & & | \\ u_1 & u_2 & \dots & u_k & u_{k+1} & \dots & u_n \\ | & | & & | & | & & | \end{bmatrix}$$

SOLVING THE GRAPH-CUT

Solving k-way cut

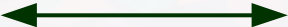
$$\text{minimize } \text{Ncut}(A_1, A_2, \dots, A_k) := \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{\text{vol}(A_i)}$$

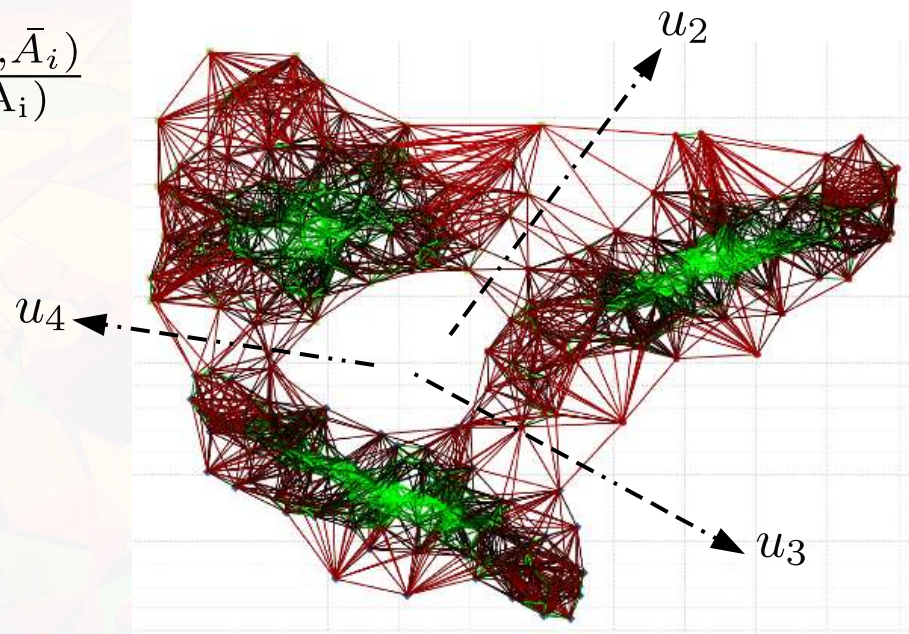


$$\text{minimize } \text{trace}(U^T \mathbf{L} U)$$

$$\text{such that } U \in \mathbb{R}^{n \times k}, U^T U = \mathbf{I}_k$$

$$\text{Eigen decomposition } \mathbf{L} = U \Sigma U^T$$

$$\begin{bmatrix} | & | & & | & | & & | \\ u_1 & u_2 & \dots & u_k & u_{k+1} & \dots & u_n \\ | & | & & | & | & & | \end{bmatrix}$$




SPECTRAL CLUSTERING

Why the word “Spectral”?

SPECTRAL CLUSTERING

Why the word “Spectral”?

- Derived from “spectrum”

SPECTRAL CLUSTERING

Why the word “Spectral”?

- Derived from “spectrum”
- Eigenvalues of a matrix are called its spectrum

SPECTRAL CLUSTERING

Why the word “Spectral”?

- Derived from “spectrum”
- Eigenvalues of a matrix are called its spectrum
- Here, clustering solutions are obtained from eigenvalues and eigenvectors of some matrix L

SPECTRAL CLUSTERING

Why the word “Spectral”?

- Derived from “spectrum”
- Eigenvalues of a matrix are called its spectrum
- Here, clustering solutions are obtained from eigenvalues and eigenvectors of some matrix L
- Clustering using *spectrum* of L

“Spectral Clustering”!

SPECTRAL CLUSTERING ALGORITHM

Normalized spectral clustering by Ng, Jordan, and Weiss (2002)

Input Similarity matrix \mathbf{W} , number of clusters k .

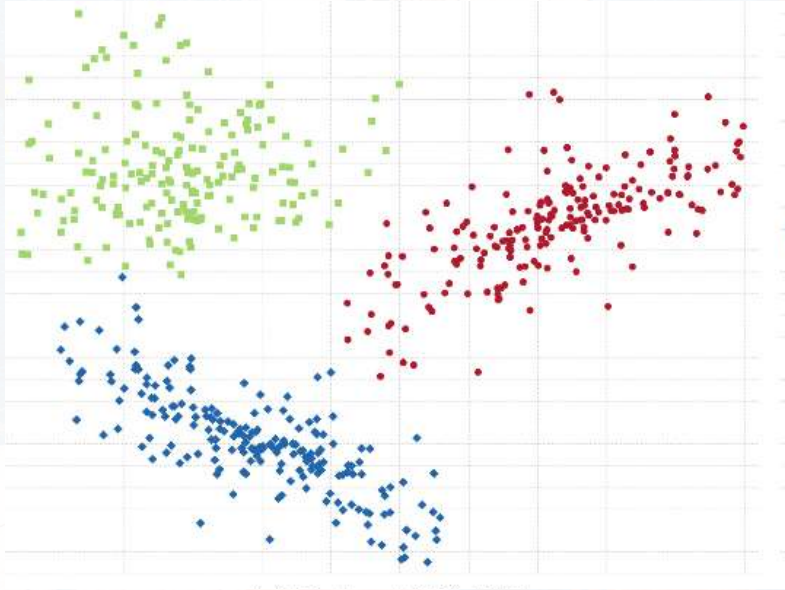
Output Clusters A_1, \dots, A_k .

1. Construct degree matrix \mathbf{D} and normalized Laplacian $\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-\frac{1}{2}}$
2. Find eigenvectors $U = [u_1 \dots u_k]$ corresponding to k smallest eigenvalues of matrix \mathbf{L} .
3. Perform clustering on the rows of U using k -means algorithm.

Return clusters A_1, \dots, A_k from k -means clustering.

SPECTRAL CLUSTERING

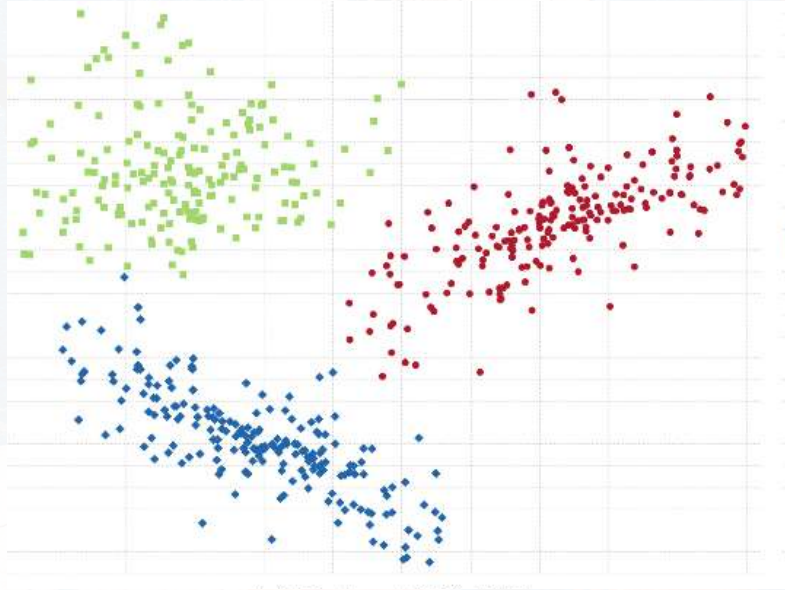
Feature-space



$$\mathbf{X} \in \mathbb{R}^{n \times d}$$

SPECTRAL CLUSTERING

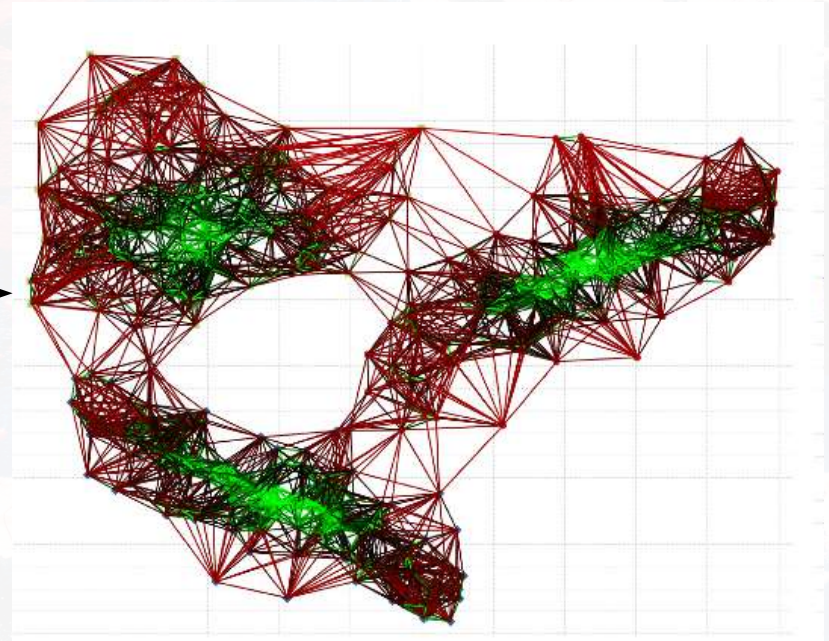
Feature-space



$$\mathbf{X} \in \mathbb{R}^{n \times d}$$

similarity
measure

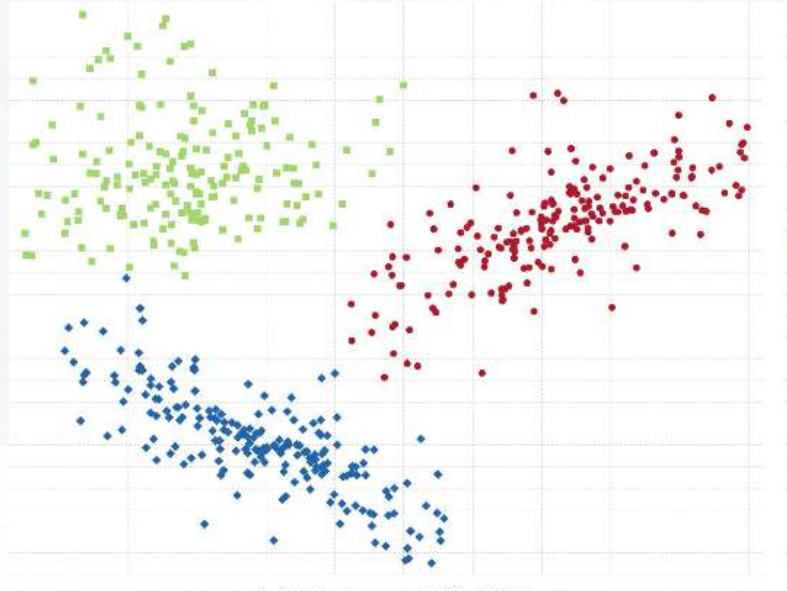
Graph



$$\mathbf{W} \text{ and Laplacian } \mathbf{L} \in \mathbb{R}^{n \times n}$$

SPECTRAL CLUSTERING

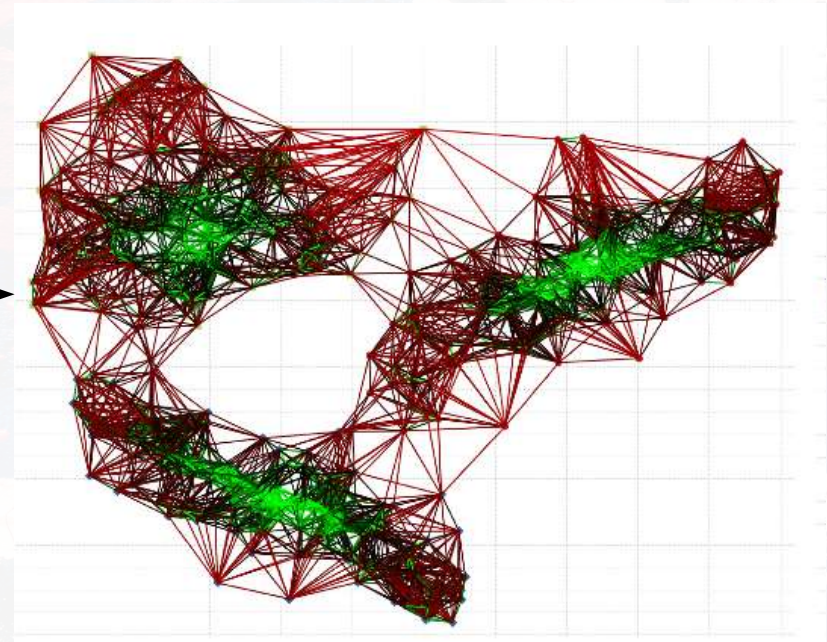
Feature-space



$$\mathbf{X} \in \mathbb{R}^{n \times d}$$

similarity
measure

Graph



$$\mathbf{W} \text{ and Laplacian } \mathbf{L} \in \mathbb{R}^{n \times n}$$

k – means clustering on low-rank U ← Eigenvectors $U \in \mathbb{R}^{n \times k}$

SPECTRAL CLUSTERING

We will study the
performance of spectral
clustering in lab session.

The background of the slide features a marbled pattern in shades of light blue, teal, and white. A large, semi-transparent grey rectangular box is centered horizontally and vertically, serving as a backdrop for the text. The text "THANK You" is written in a dark blue, serif font, with "THANK" in all caps and "You" in title case. The entire composition is framed by a thin, dark blue border.

THANK You