Spectral Clustering

Hayan Lee

Research Fellow @ Simons Institute for the Theory of Computing, UC Berkeley
Simons Postdoctral Fellow @ JGI, Lawrence Berkeley National Laboratory

Outline

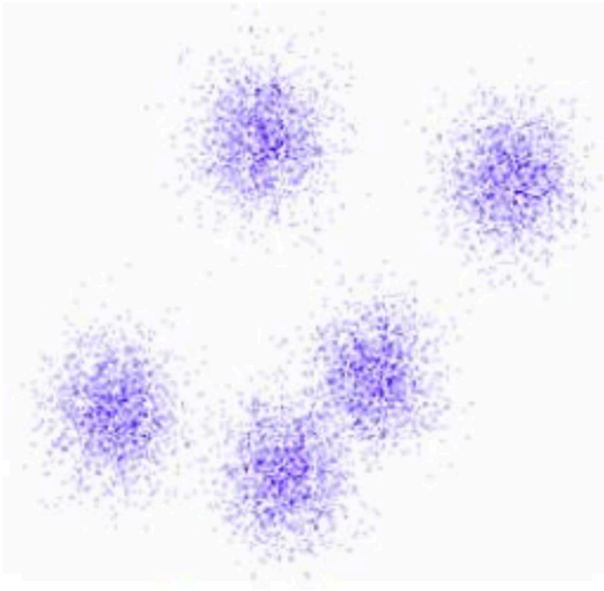
- Background
  - K-means clustering vs. spectral clustering
  - Adjacent matrix, degree matrix, graph Laplcian
  - Graph partitioning problem : min-cut, normalized min-cut
- Spectral Clustering
  - Algorithm
  - How to interpret
  - How to understand intuitively
  - How to apply to metagenome assembly

Background

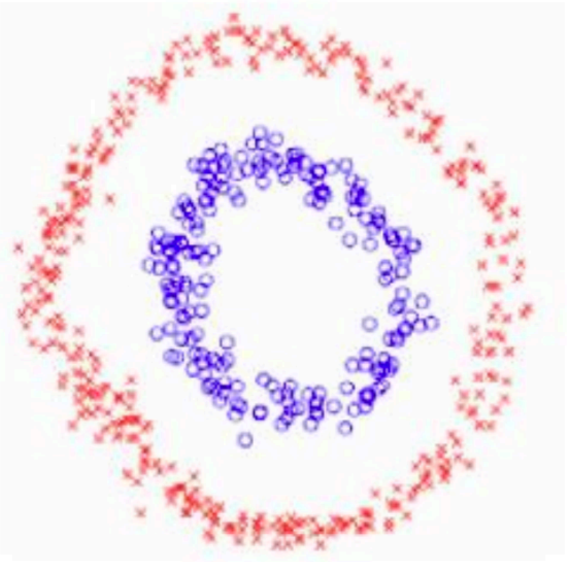
Clustering

K-means	Spectral Clustering
− Requires to assume K in advance − Poor clustering performance over non-convex data − Rely on randomized approach − Greedy Algorithm • Only finds local minima • Needs multiple start from various initial states − EM approach • Needs iterative process	+ No assumptions need to be made • Convex or non-convex + Do not need to run multiple times + No greedy algorithm + Deterministic + No EM approach • No iteration is required • Very good interpretation • Well backed by mathematics e.g. graph Laplacian, SVD etc.

Data Clustering Criteria



By compactness
e.g. K-means



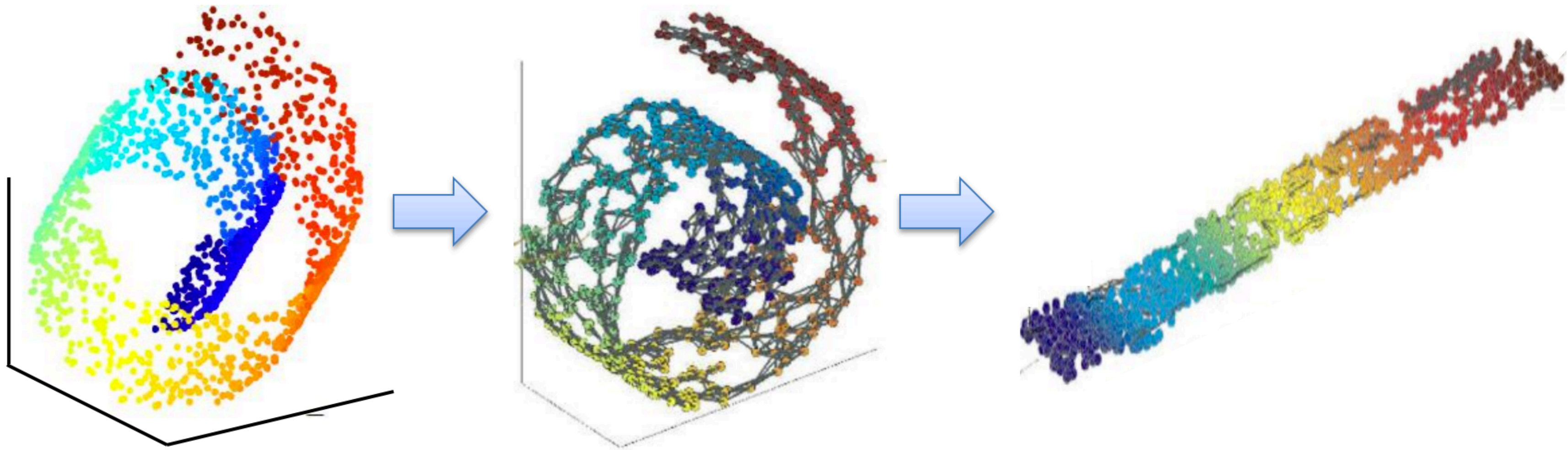
By connectivity (similarity)
e.g. Spectral clustering

Data Clustering Criteria

By compactness
e.g. K-means

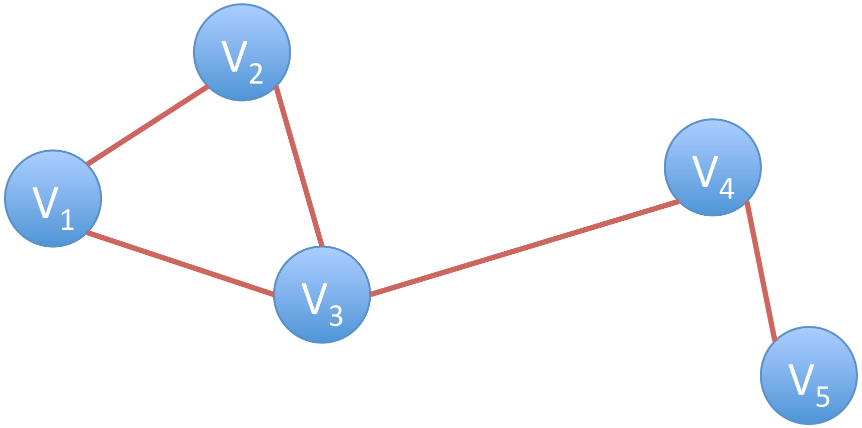
By connectivity (similarity)
e.g. Spectral clustering

How does it work?



Original data points in high dimension → Capture similarity (local information) → Project data points to low dimension space, keeping local similarity

Adjacent Matrix

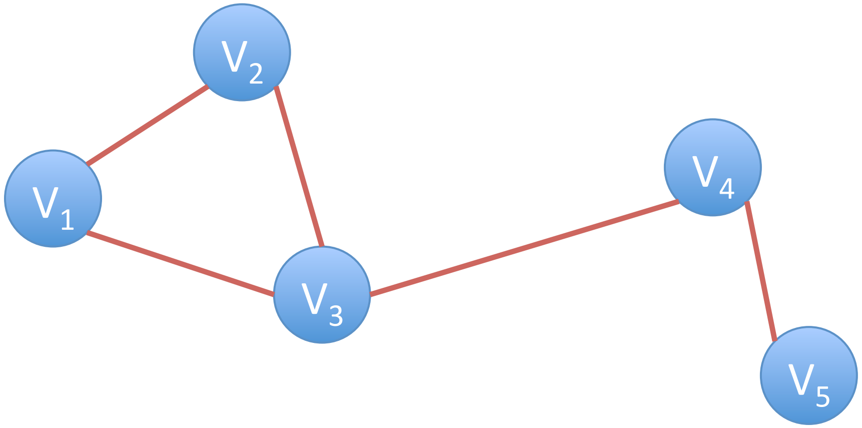


How to capture similarity

1. Just make something up if you're a domain expert
2. ε–neighbors
3. Use kernel e.g. Gaussian kernel similarity function

$$w = \begin{pmatrix} 5 & 4 & 4 & 0 & 0 \\ 4 & 5 & 4 & 0 & 0 \\ 4 & 4 & 5 & 1 & 0 \\ 0 & 0 & 1 & 5 & 4 \\ 0 & 0 & 0 & 4 & 5 \end{pmatrix}$$

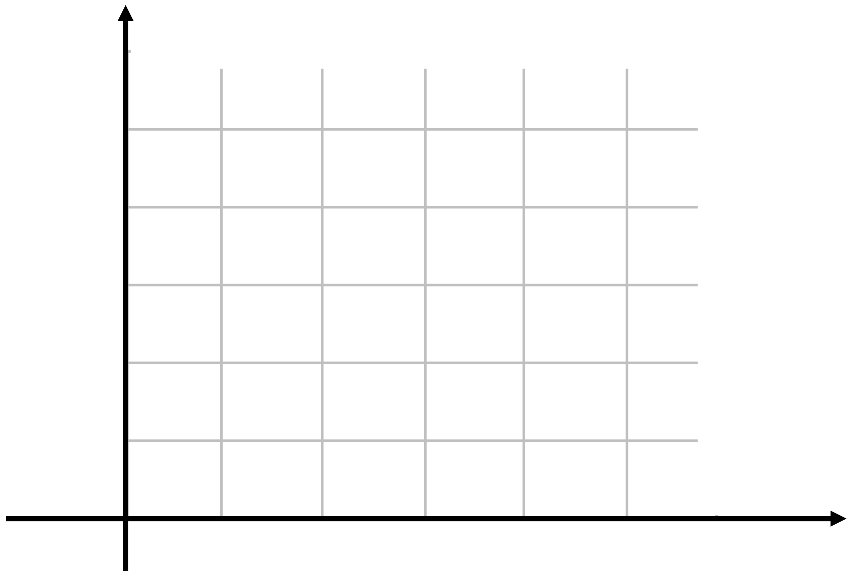
Degree Matrix



$$D = \begin{pmatrix} d_1 & 0 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 & 0 \\ 0 & 0 & d_3 & 0 & 0 \\ 0 & 0 & 0 & d_4 & 0 \\ 0 & 0 & 0 & 0 & d_5 \end{pmatrix}$$

Partitioning a graph into two clusters

x	y
0	0
0	1
1	0
4	2
5	2
5	3



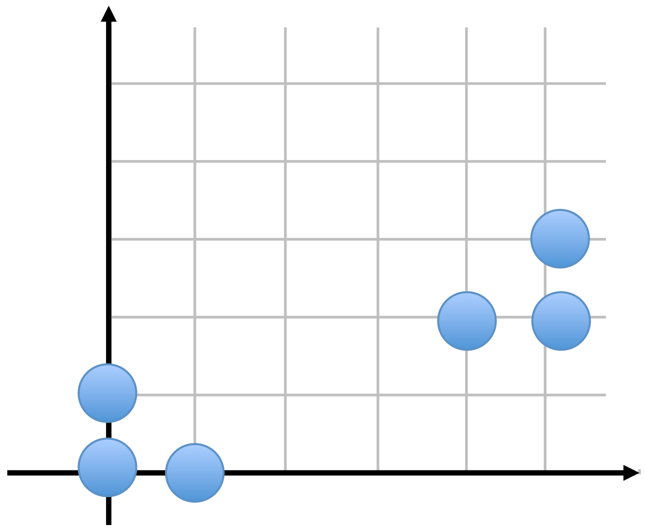
Gaussian kernel

$$w_{ij} = e^{-\frac{\|v_i - v_j\|^2}{2\sigma^2}}$$

$$w = \begin{pmatrix} 1 & 0.6 & 0.6 & 0 & 0 & 0 \\ 0.6 & 1 & 0.37 & 0.0002 & 0 & 0 \\ 0.6 & 0.37 & 1 & 0.0015 & 0 & 0 \\ 0 & 0.0002 & 0.0015 & 1 & 0.6 & 0.37 \\ 0 & 0 & 0 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0 & 0.37 & 0.6 & 1 \end{pmatrix}$$

Partitioning a graph into two clusters

x	y
0	0
0	1
1	0
4	2
5	2
5	3



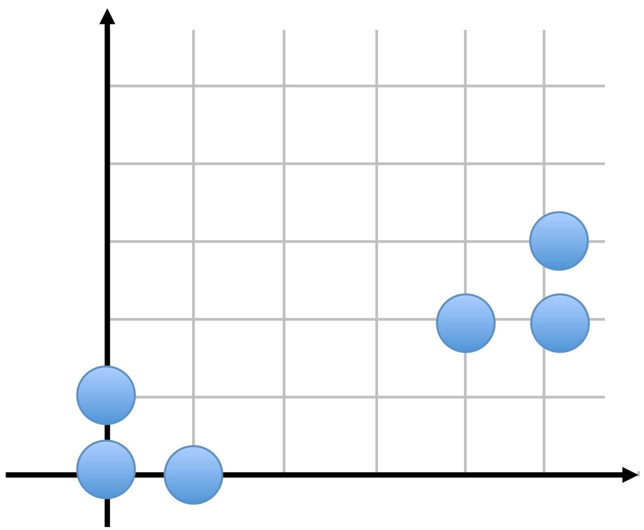
$$w = \begin{pmatrix} 1 & 0.6 & 0.6 & 0 & 0 & 0 \\ 0.6 & 1 & 0.37 & 0.0002 & 0 & 0 \\ 0.6 & 0.37 & 1 & 0.0015 & 0 & 0 \\ 0 & 0.0002 & 0.0015 & 1 & 0.6 & 0.37 \\ 0 & 0 & 0 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0 & 0.37 & 0.6 & 1 \end{pmatrix}$$

$$R = \sum w_{ij}(f_i - f_j)^2$$

$$f = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad f = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

Partitioning a graph into two clusters

x	y
0	0
0	1
1	0
4	2
5	2
5	3



$$w = \begin{pmatrix} 1 & 0.6 & 0.6 & 0 & 0 & 0 \\ 0.6 & 1 & 0.37 & 0.0002 & 0 & 0 \\ 0.6 & 0.37 & 1 & 0.0015 & 0 & 0 \\ 0 & 0.0002 & 0.0015 & 1 & 0.6 & 0.37 \\ 0 & 0 & 0 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0 & 0.37 & 0.6 & 1 \end{pmatrix}$$

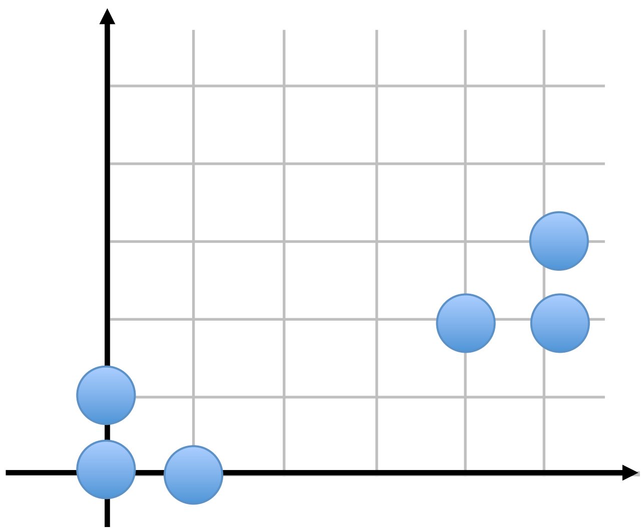
$$f = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$R = \sum w_{ij}(f_i - f_j)^2$$

1. Ignore any weights if data points are in the same cluster
2. Sum weights between clusters. If it’s small, it means clusters are good.

Min-Cut Problem

x	y
0	0
0	1
1	0
4	2
5	2
5	3



$$w = \begin{pmatrix} 1 & 0.6 & 0.6 & 0 & 0 & 0 \\ 0.6 & 1 & 0.37 & 0.0002 & 0 & 0 \\ 0.6 & 0.37 & 1 & 0.0015 & 0 & 0 \\ 0 & 0.0002 & 0.0015 & 1 & 0.6 & 0.37 \\ 0 & 0 & 0 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0 & 0.37 & 0.6 & 1 \end{pmatrix}$$

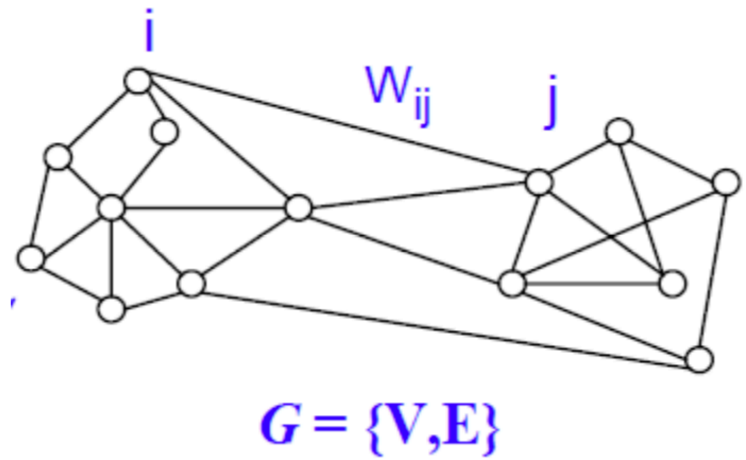
$$R = \sum w_{ij}(f_i - f_j)^2$$

$$\underset{f}{\operatorname{argmin}} \sum w_{ij}(f_i - f_j)^2$$

We have a polynomial time solution O(VE)

1. Ignore any weights if data points are in the same cluster.
2. Sum weights between clusters. If it's small, it means clusters are good.
3. This equation measures the relationship between two clusters.
4. We want to find f such that this expression is as small as possible.

Normalized Min-Cut Problem



$$R = \sum w_{ij}(f_i - f_j)^2 \left(\frac{1}{vol(A)} + \frac{1}{vol(B)} \right)$$

$$vol(A) = \sum_{i \in A} d_i$$

$$vol(B) = \sum_{i \in B} d_i$$

This is NP-hard!
Spectral clustering is a relaxation of these.

Spectral Clustering

Spectral Clustering Algorithm

- Input
  - Data points

Hopland Soil Metagenomic Data

- 1 TB
- 1 sample
- Multi-sample and abundance based approach cannot be used

Spectral Clustering Algorithm

- Input : Data points
- Build similarity graph W and degree matrix D
- Compute graph Laplacian L
- Computer the first K eigenvectors v_1, v_2, … v_K of the matrix
- Build the matrix $V \in R^{n \times k}$ with eigenvectors as columns
- Cluster the points Z_i with the k-means algorithms in R^k

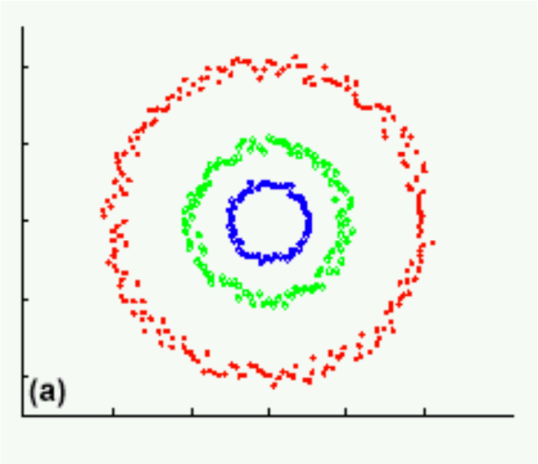
Dimensionality Reduction

$$n \times n \to n \times k$$

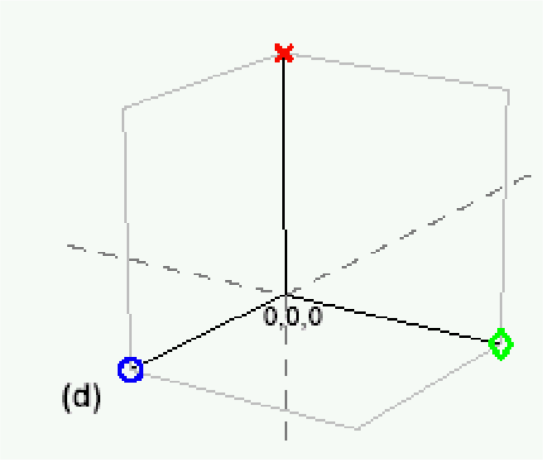
$$n \to k$$

How to interpret (1)

Original data

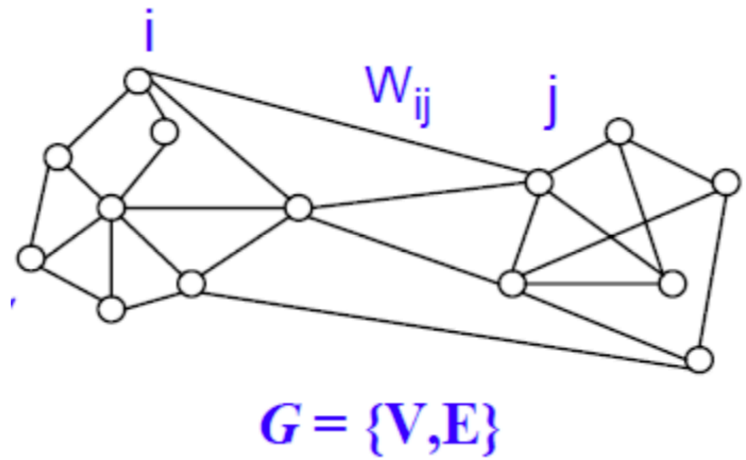


Projected data



Data are projected into a lower-dimensional space where they can be easily separable with simple clustering algorithm such as K-means clustering.

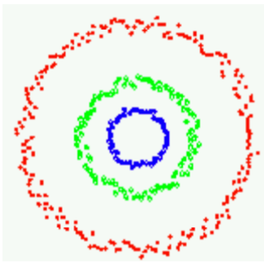
How to understand intuitively (1)



If the graph is totally connected, the first Laplacian eigenvector is constant, meaning all 1s

$$f_1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{pmatrix}$$

How to understand intuitively (2)

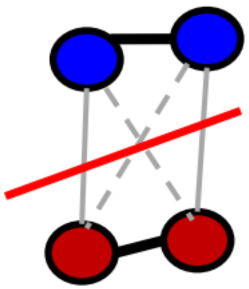


OR

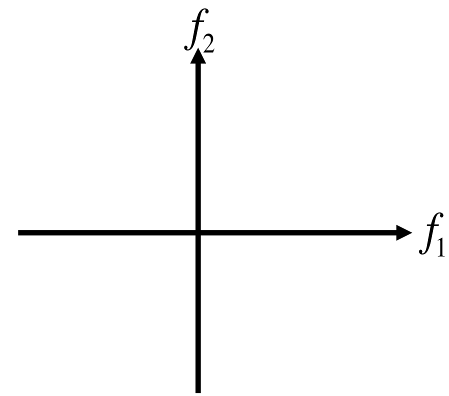
$$L = \begin{bmatrix} L_1 & \ddots & 0 \\ \ddots & L_2 & \ddots \\ 0 & \ddots & L_3 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

If the graph is disconnected and have k connected components, the graph Laplacian consists of diagonal blocks and the first K Laplacian eigenvectors are 1s of corresponding blocks.

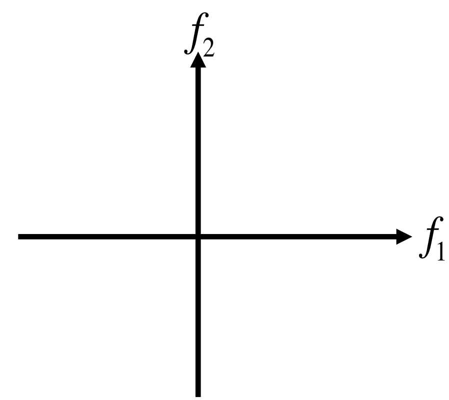
How to understand intuitively (3)



$$w = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \quad f_1 = \begin{pmatrix} 0.71 \\ 0.71 \\ 0 \\ 0 \end{pmatrix} \quad f_2 = \begin{pmatrix} 0 \\ 0 \\ 0.71 \\ 0.71 \end{pmatrix}$$

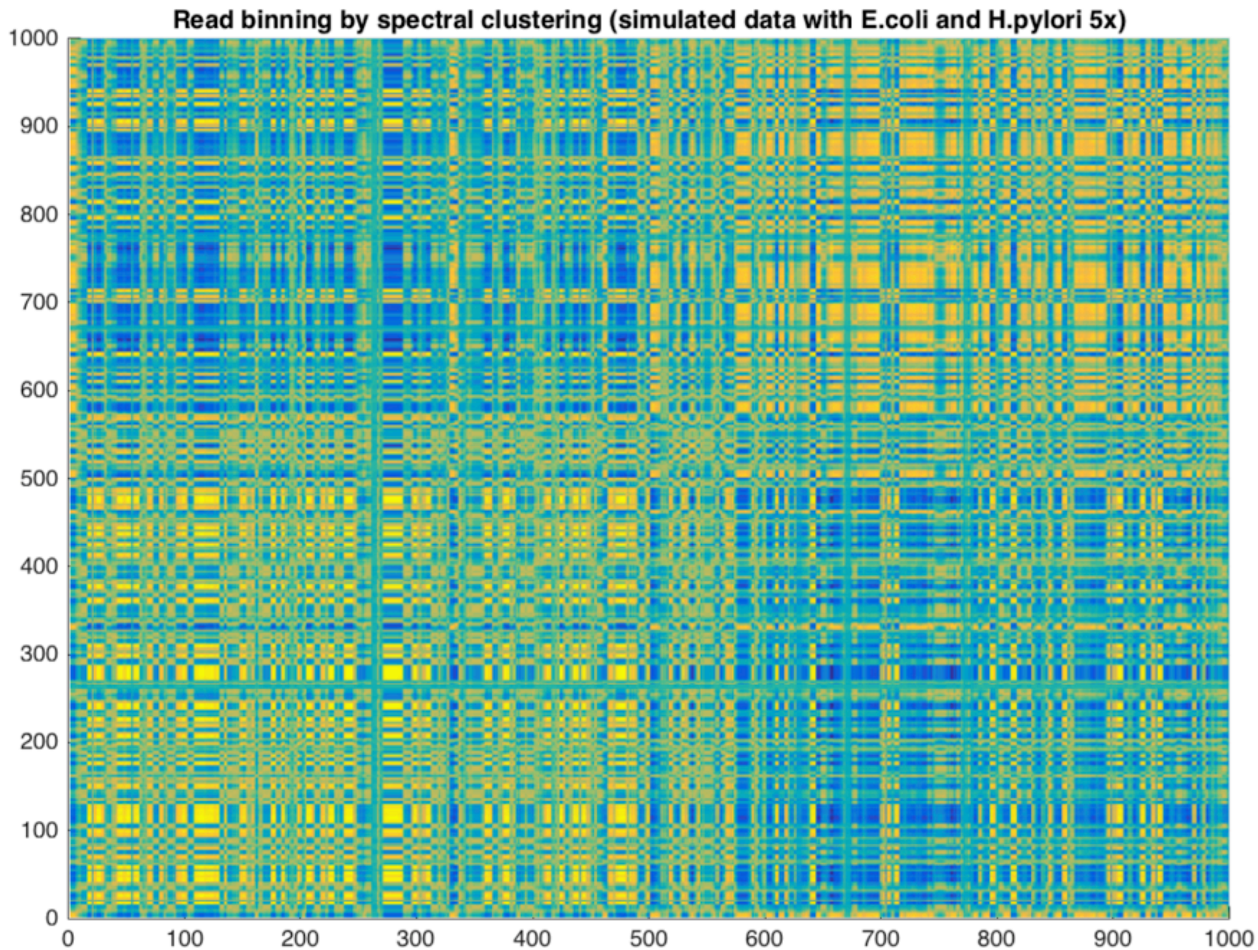


$$w = \begin{pmatrix} 1 & 1 & 0.2 & 0 \\ 1 & 1 & 0 & 0.1 \\ 0.2 & 0 & 1 & 1 \\ 0 & 0.1 & 1 & 1 \end{pmatrix} \quad f_1 = \begin{pmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{pmatrix} \quad f_2 = \begin{pmatrix} 0.47 \\ 0.52 \\ -0.47 \\ -0.52 \end{pmatrix}$$



Applications : Metagenome assembly

Spectral clustering on simulated data



Q&A