IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

# Spectral clustering

Students : Pape Samba DIALLO
Van-Khoa NGUYEN

31/10/2017

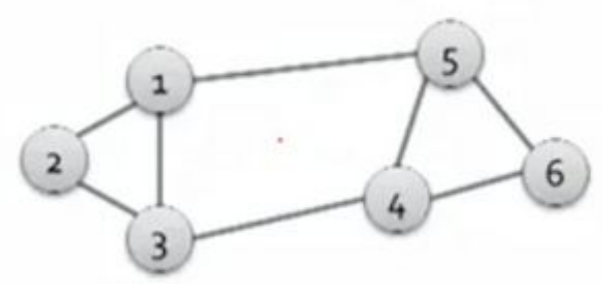
# 1. What is spectral clustering ? (1)



**Adjacency matrix (A) :**

Given a graph with *n* vertices and *m* nodes, the adjacency matrix is a square n*n matrix with the property:

A[i][j] = 1 if there is an edge between node i and node j, 0 otherwise




**K-neareast neighbors**

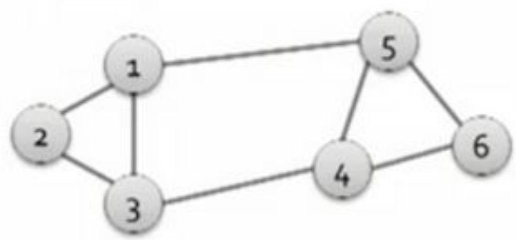
| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 0 | 1 | 0 |
| 2 | 1 | 0 | 1 | 0 | 0 | 0 |
| 3 | 1 | 1 | 0 | 1 | 0 | 0 |
| 4 | 0 | 0 | 1 | 0 | 1 | 1 |
| 5 | 1 | 0 | 0 | 1 | 0 | 1 |
| 6 | 0 | 0 | 0 | 1 | 1 | 0 |

## Degree matrix (D)

The degree matrix is a n*n diagonal matrix with the property

d[i][i] = the number of adjacent edges in node i or the **degree of node i**

d[i][j] = 0




| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 3 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 2 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 3 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 3 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 3 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 2 |

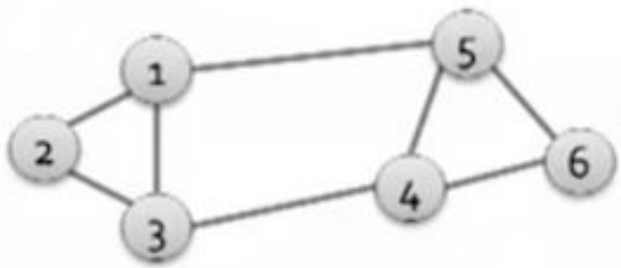
# 1. What is spectral clustering ? (3)



## Laplacian matrix (L)

The laplacian matrix is a n*n matrix defined as: **L = D -A**

Its **eigen values are positive real numbers** and the **eigen vectors are real and orthogonal** (the dot product of the 2 vectors is 0)




| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 3 | -1 | -1 | 0 | -1 | 0 |
| 2 | -1 | 2 | -1 | 0 | 0 | 0 |
| 3 | -1 | -1 | 3 | -1 | 0 | 0 |
| 4 | 0 | 0 | -1 | 3 | -1 | -1 |
| 5 | -1 | 0 | 0 | -1 | 3 | -1 |
| 6 | 0 | 0 | 0 | -1 | -1 | 2 |

## Conductance

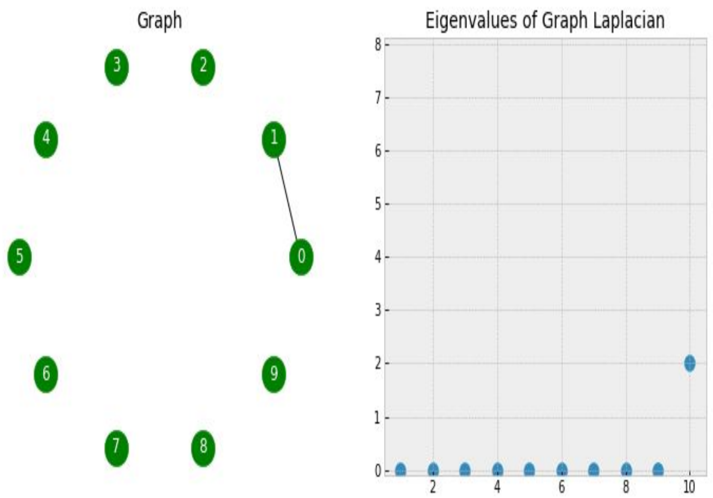
Calculating the eigen values and eigen vectors of A with x ( n dimensional vector with the values of the nodes): **A * x = lambda * x**

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad y_i = \sum_{j=1}^{n} A_{ij} x_j = \sum_{(i,j) \in E} x_j$$

We are ready to summarize the spectral clustering steps:

- ➢ Compute the Laplacian matrix L of the input graph G
- ➢ Compute the eigen values (lambda) and eigen vectors (x) such that L* x = lambda * x
- ➢ Select n eigenvectors corresponding to the smallest eigenvalues and redefine the input space as a n dimensional subspace
- ➢ Find clusters in this subspace using various clustering algorithms, such as **k-means**

# 1. Eigenvalues of the Laplacian graph



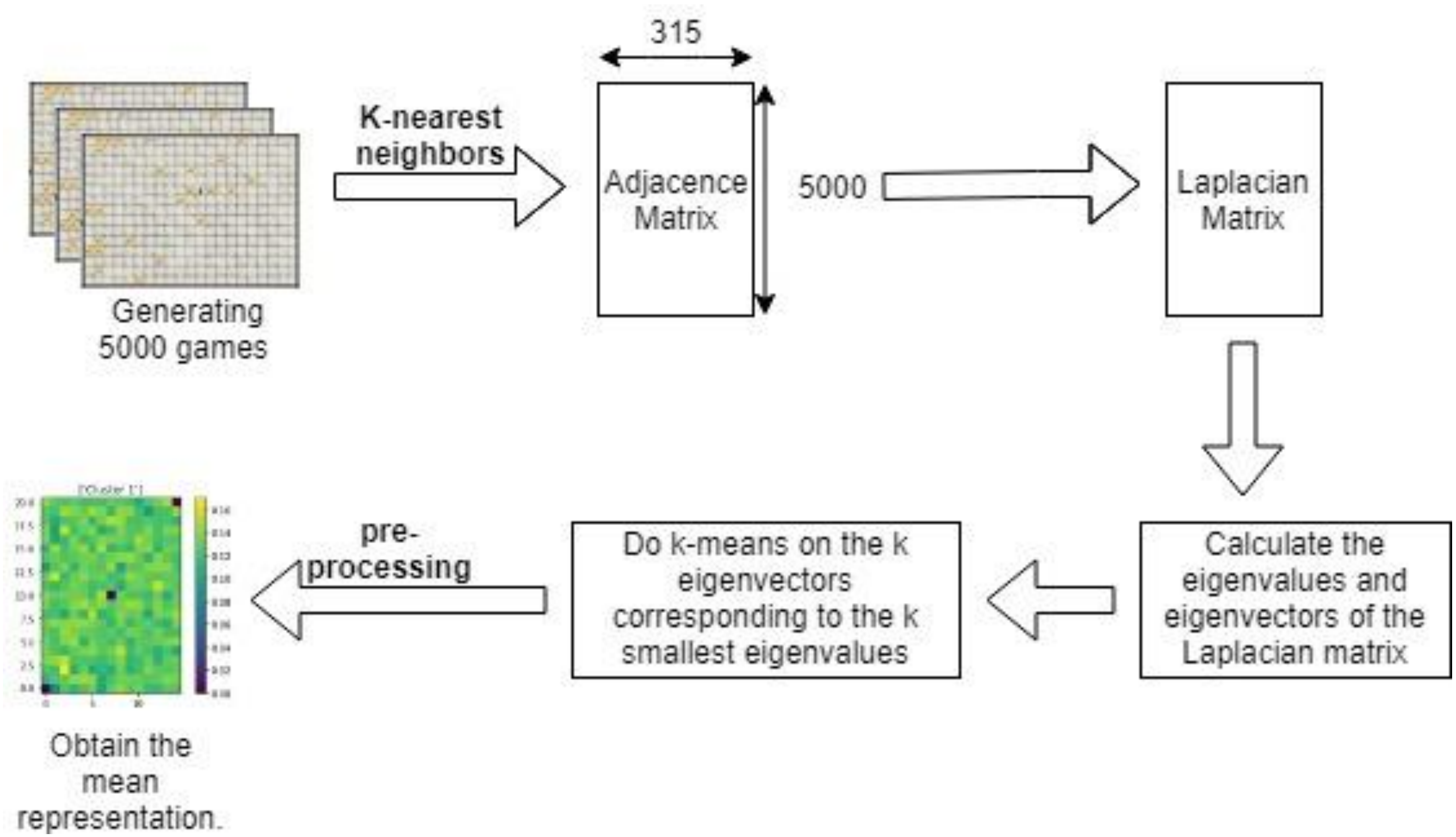

- Ten eigenvalues corresponding to ten points in the graph.

- Number of 0 eigenvalues corresponding to the number of connected components in the graph.

- The more eigenvalues are nears to 0, the more we have the number the cluster in graph.

# 2. Applying the spectral clustering on the Pyrat DataSet

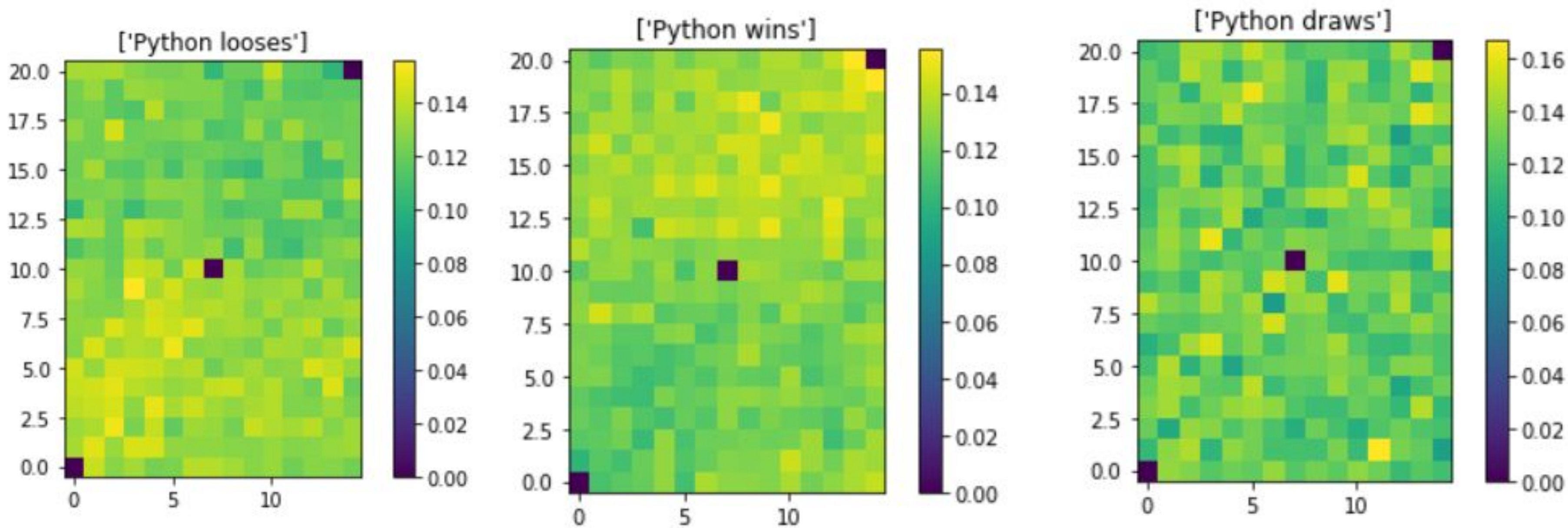
## Diagram of the processus



# 2. Applying the spectral clustering on the Pyrat DataSet



## The dataset ground truth explanation

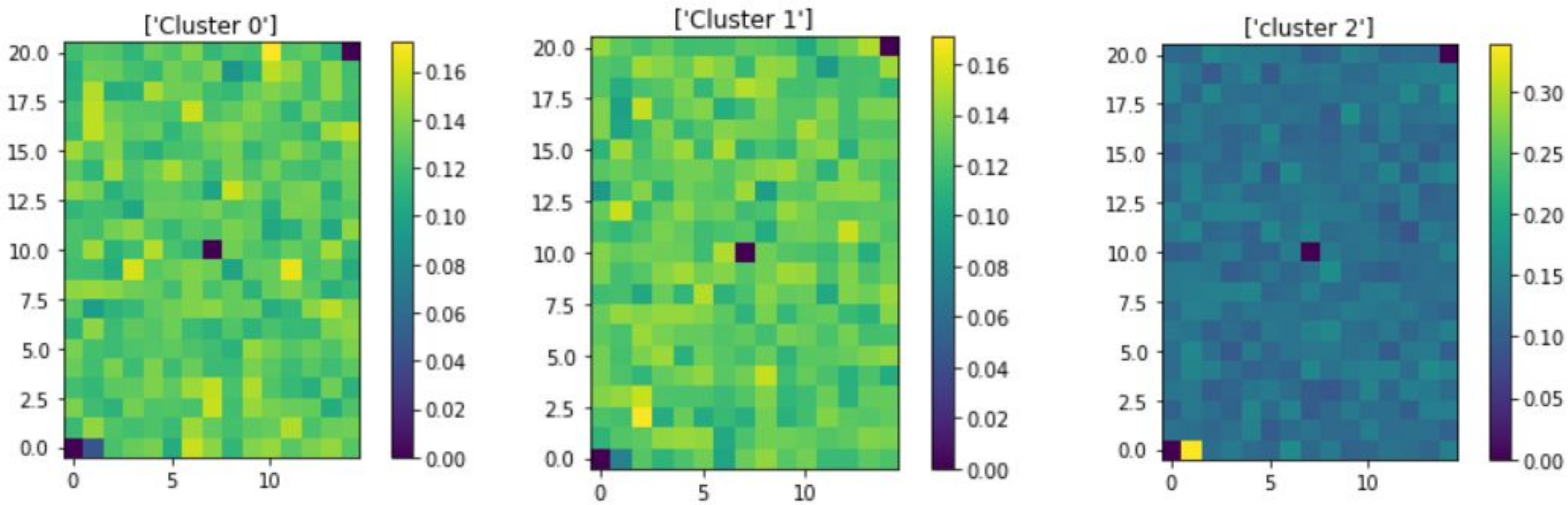



**The cheese density distribution in three cases**

# 2. Applying the spectral clustering on the Pyrat DataSet



## The results of implementing the spectral clustering (n_cluster = 3)




**Three different clusters obtained by spectral clustering and the mean representation of them.**

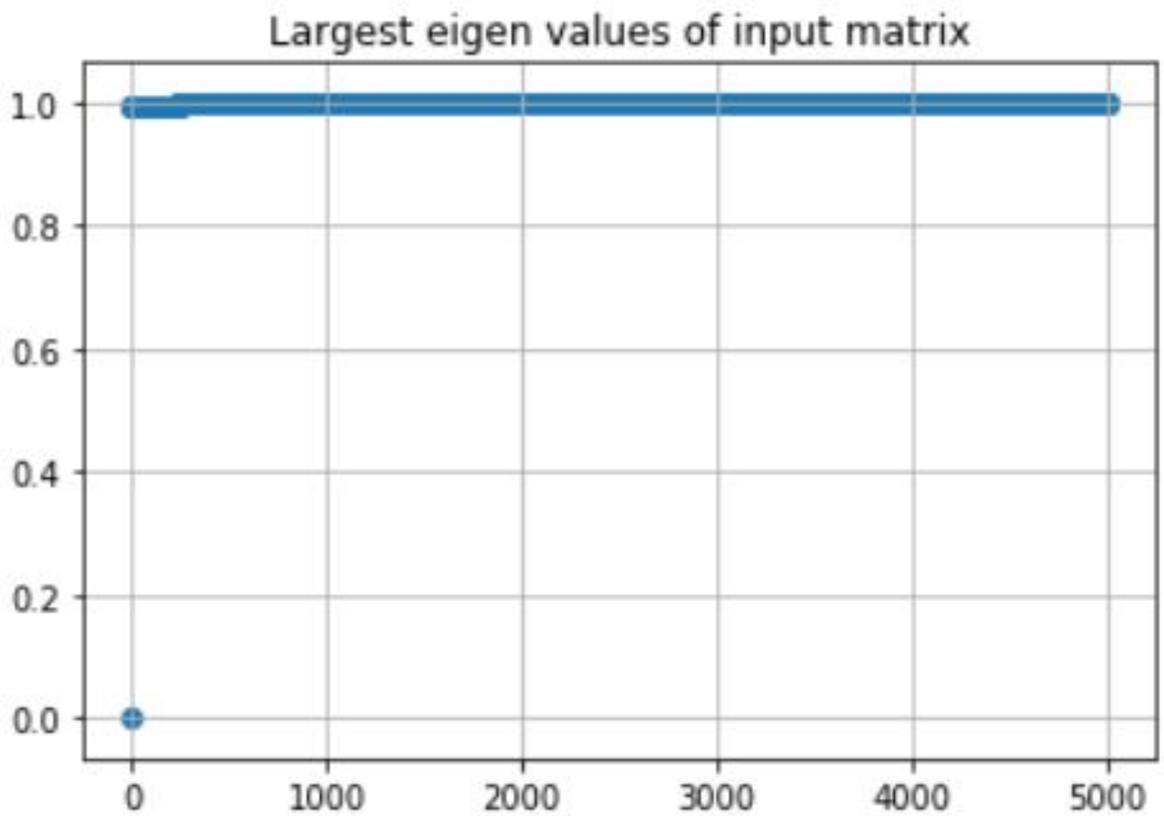
## The self-tuning spectral clustering and the optimal number of clusters

- The idea behind the self tuning spectral clustering is **determine the optimal number of clusters** and also **the similarity metric σi** used in the computation of the affinity matrix (Self-Tuning Spectral Clustering- Lihi Zelnik-Manor and Pietro Perona)

- In the paper "a tutorial on spectral clustering — ulrike von luxburg" Eigengap heuristic suggests the number of clusters k is usually given by the value of k that **maximizes the eigengap** (difference between consecutive eigenvalues). **The larger this eigengap** is, the closer the eigenvectors of the ideal case and hence **the better spectral clustering works**.

# 2. Applying the spectral clustering on the Pyrat DataSet

## The self-tuning spectral clustering and the optimal number of clusters




Optimal number of cluster 1 (Pyrat Dataset)

# References




- Spectral graph clustering and optimal number of clusters estimation at https://towardsdatascience.com/spectral-graph-clustering-and-optimal-number-of-clusters-estimation-32704189afbe

- Spectral Clustering at https://towardsdatascience.com/spectral-clustering-aba2640c0d5b