Data Mining - ID2222

Homework 4

Graph Spectra

André Silva - Jérémy Navarro

November 30, 2020

1 Overview

We chose to use *Matlab* for the development of the project.

Information about the structure of the project, as well as instructions on how to run it follow:

```
$ unzip homework4.zip
$ cd homework4/
$ matlab main.m
```

The project is composed by:

- `main.m`: The executable script which performs experimentation on the dataset.
- `example1.dat`: First dataset with k=4 clusters.
- `example2.dat`: Second dataset with k=2 clusters.

2 Implementation

2.1 Read the file

First we read the file.

2.2 Convert edge list to adjacency matrix

Then, we convert the graph representation from an edge list into an adjacency matrix, A. The diagonal in A is empty.

2.3 Construct $L = D^{-1/2}AD^{-1/2}$

Matrix D is a diagonal matrix. For every element d_{ii} the value is the sum of the row i of matrix A.

2.4 Eigenvector of L

We form matrix X, composed, in each column, of the eigenvectors of matrix L.

2.5 Normalization

We create matrix Y by normalizing X.

2.6 k-clustering

We apply K-means clustering to matrix Y. We consider each row as a point.

2.7 Final clustering

Finally, we assign each point to a cluster depending on the previous step.

3 Results

3.1 Clustering

After running the algorithm on the datasets, we can extract clusters from them. In Figure 1 we can see that all 4 clusters are separated. In Figure 2 even if the 2 clusters are connected, we can sort every nodes in the correct cluster.

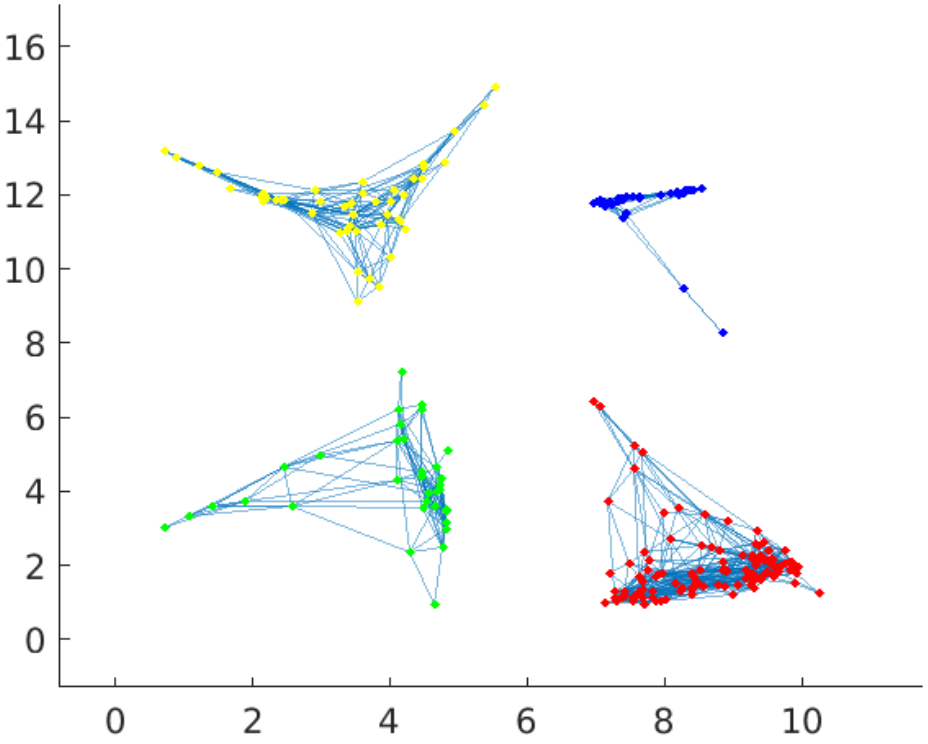

Figure 1: Dataset 1 with cluster highlighted ($k = 4$)

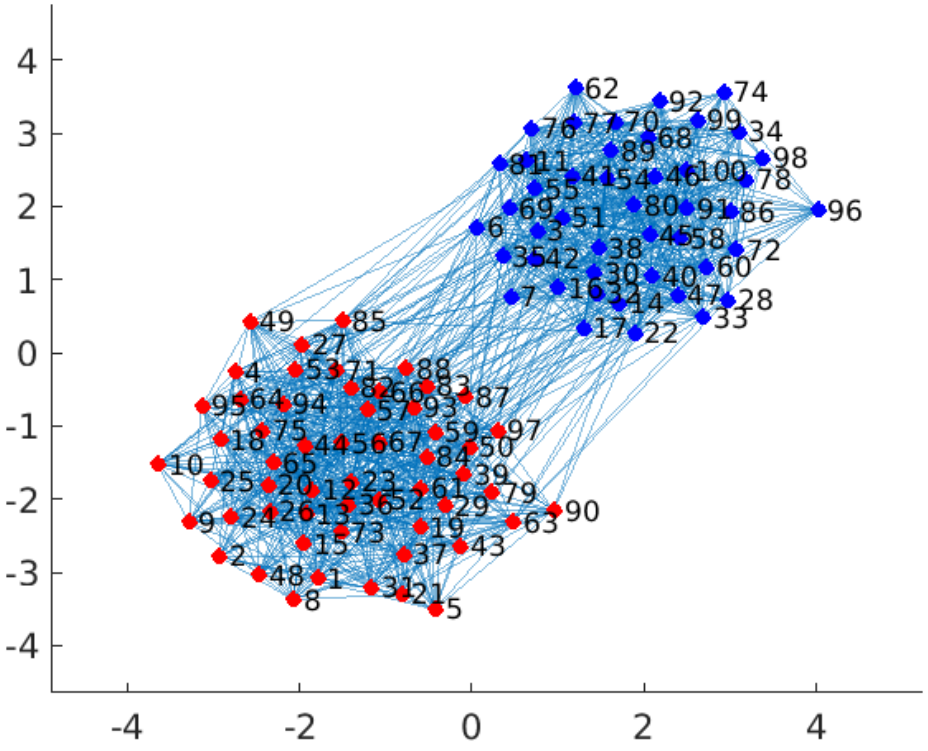

Figure 2: Dataset 2 with cluster highlighted ($k = 2$)

3.2 Sparsity pattern

We can see in, Figure 3, that when there is no connection between clusters the sparsity matrix is composed by non-overlapping blocs. However, in Figure 4 clusters are connected and highly overlap, making it hard to see two well defined blocs.

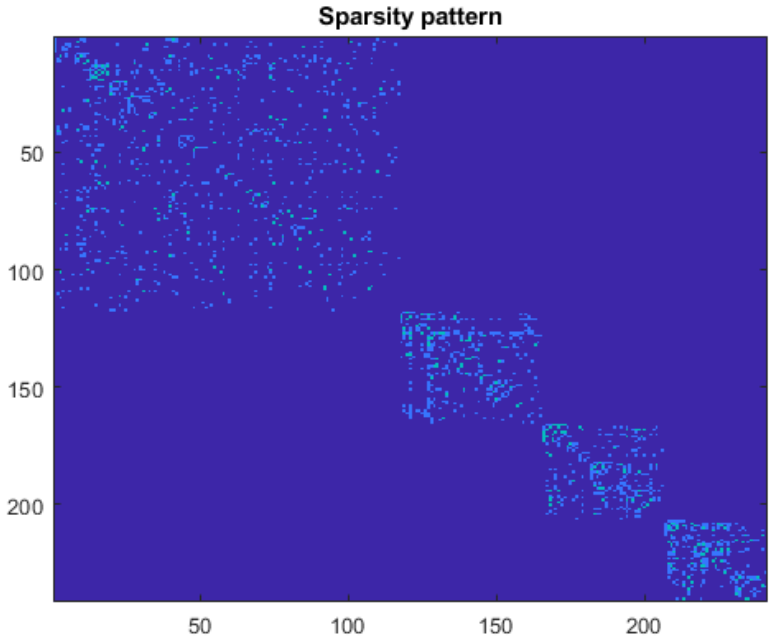

Figure 3: Sparsity pattern for dataset 1. Cluster are not connected.

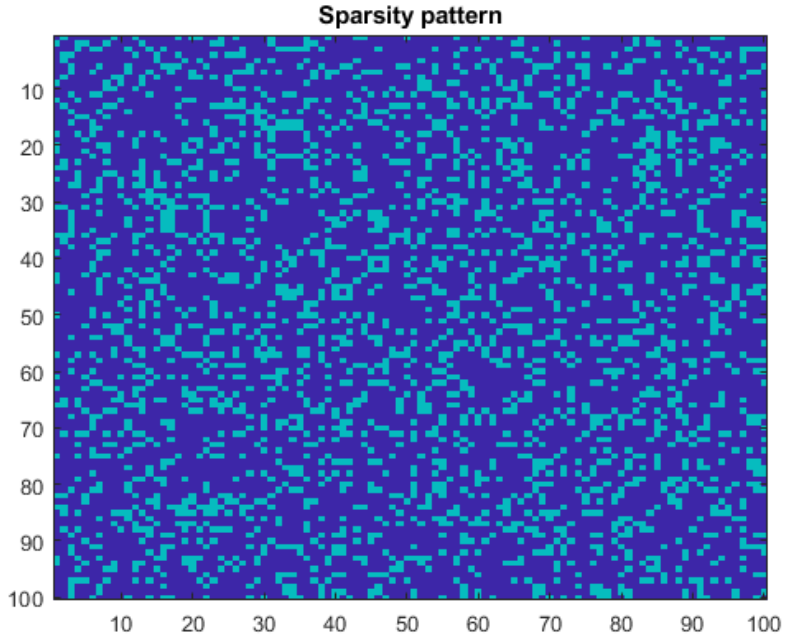

Figure 4: Sparsity pattern for dataset 2. Clusters are strongly connected.

3.3 Fiedler vector

We visualize the fiedler vector for each of the datasets.

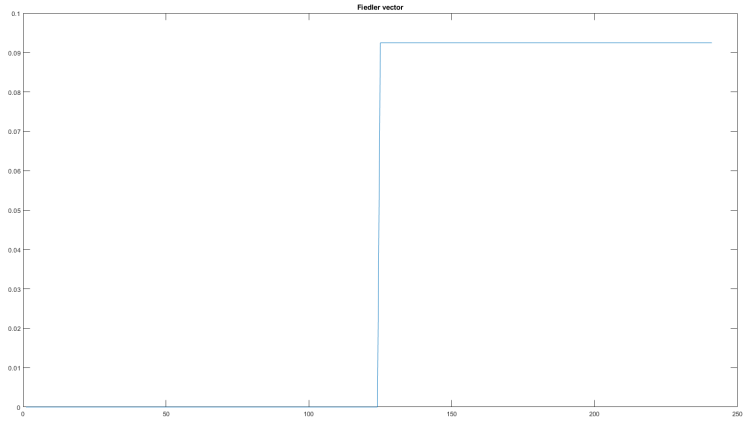

Figure 5: Fiedler vector for dataset 1

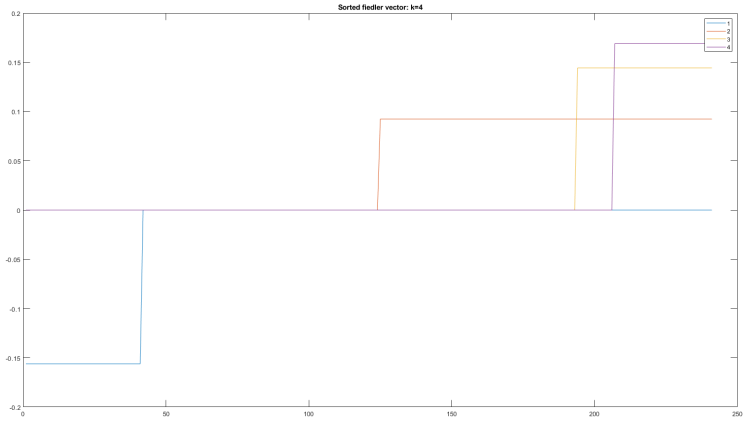

Figure 6: k-eigenvectors for dataset 1

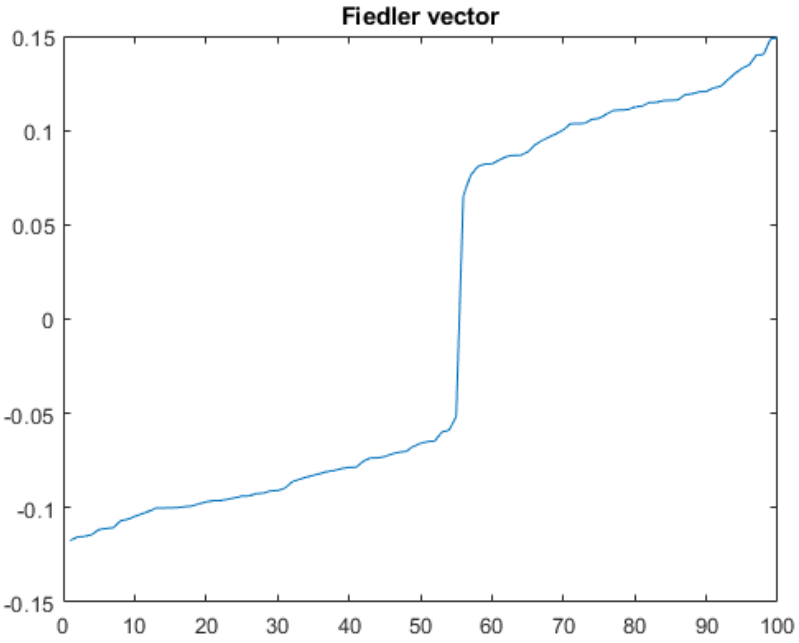

Figure 7: Fiedler vector for dataset 2

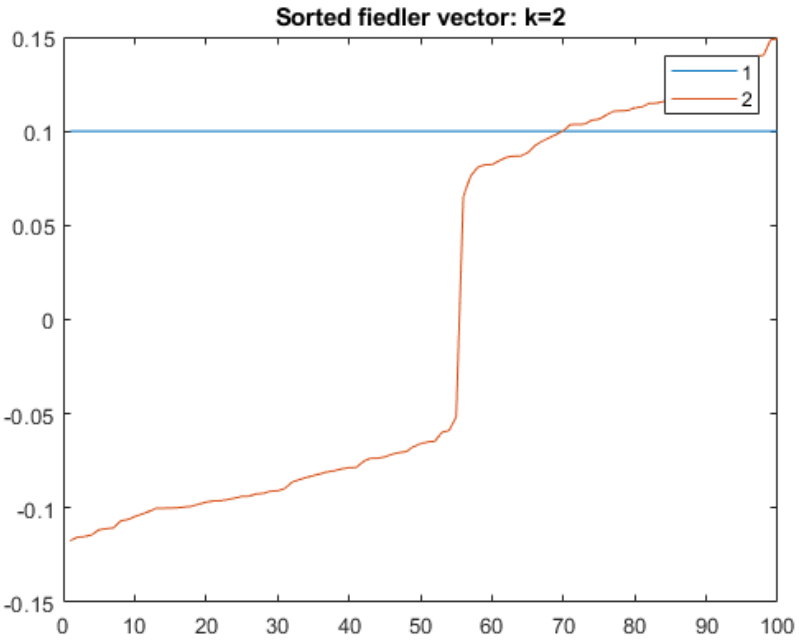

Figure 8: k-eigenvectors for dataset 2