ID2222 Homework 4

Nikita Roldugins, Vianey Darsel

December 2, 2022

1 Introduction

This homework is to implement the algorithm purposed in this article [1] from Andrew Ng, Michael Jordan, and Yair Weiss.

The implementation was made on Matlab and the code is in the linked file. It contains a function HM4_GraphSpectra that takes the number of the example as a parameter and plots the results of clustering. The data is in a folder "data" which is in the folder of the Matlab file.

2 Implementation

To do the clustering based on graph spectra. We have followed the following steps:

- convert the list of edges to an adjacency matrix A (which is also the affinity matrix here) of n nodes.
- compute the sum of the affinity of each node (and verify that the sum is non-null, Which means that D is invertible).
- compute $D^{\frac{1}{2}}$ and then laplacian matrix $L = D^{\frac{1}{2}}AD^{\frac{1}{2}}$
- get the eigenvalues of L, sort them and find the highest gap between two consecutive eigenvalues. That gives the number k of clusters.
- keep the k eigenvectors with value in $\mathbb{R}^n$ (X).
- normalize these vectors (Y).
- use kmeans internal function to cluster the n points in $\mathbb{R}^k$ $(clust)$.
- reorder the nodes in order to sort by the clusters.

3 Results

A Example 1

Our algorithm provides us with good results in both examples.

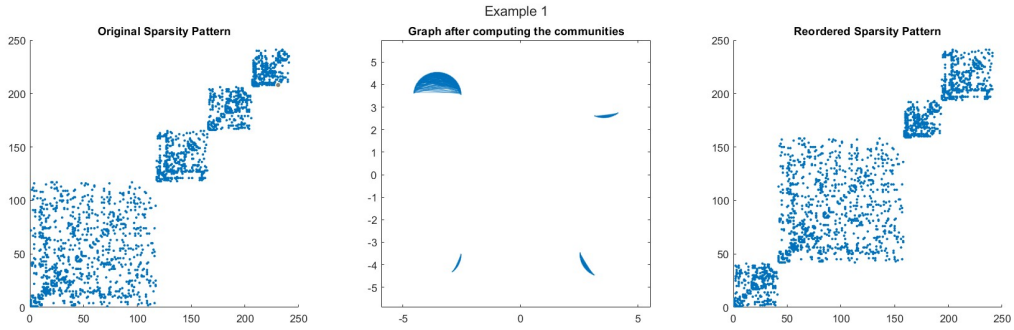

Figure 1: Global plots on Graph of example 1

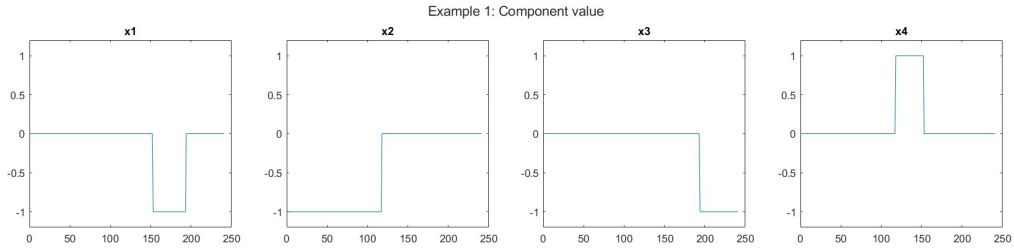

Figure 2: Values in eigenvectors space of the nodes (Example 1)

In the first example, the data was already sorted. We can see that the graph is partitioned into 4 subgraphs (Figure 1). This can be seen in both the plot of the graph and the sparsity pattern. Moreover, we see clearly the clusters when we look at the coordinates of the nodes in the eigenvectors bases (Figure 2).

B Example 1

In the second example, the cluster are not pre-computed (Figure 3). But the algorithm is able to find and divide the graph in two clusters. Regarding the final Sparsity pattern, it seems that the

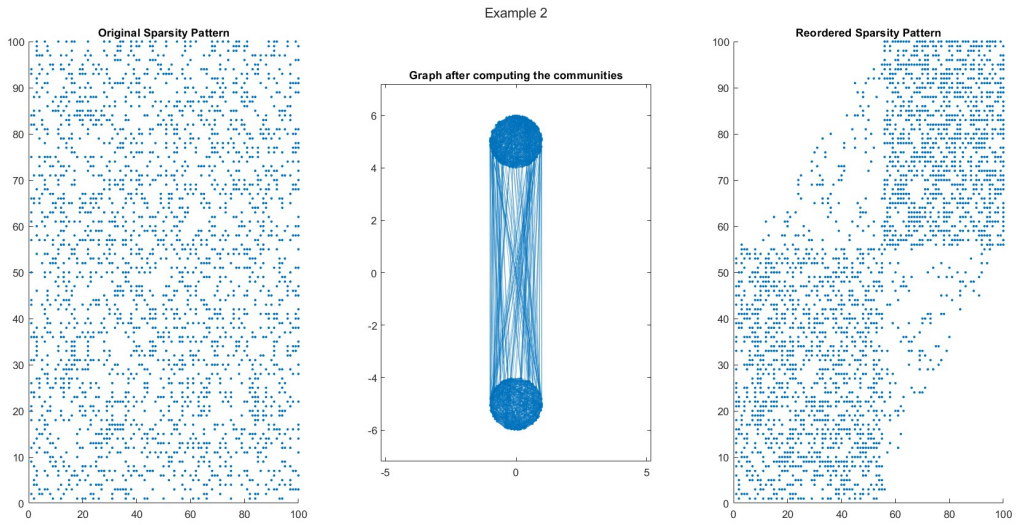

Figure 3: Global plots on Graph of example 1

algorithm has well divided the graph. The Figure 4 able us to understand how the kmeans algorithm

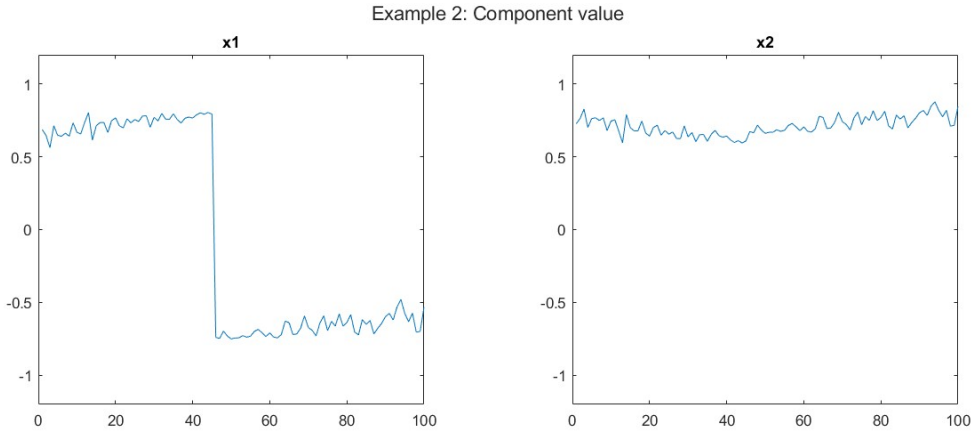

Figure 4: Values in eigenvectors space of the nodes (Example 1)

has created the clusters, based on the coordinate x_1 of a point/node. If it is positive, it belongs to the first cluster. Otherwise to the second.

References

[1] NG, Andrew, JORDAN, Michael, et WEISS, Yair. On spectral clustering: Analysis and an algorithm. Advances in neural information processing systems, 2001, vol. 14.