Data Mining Homework 4

Graph Spectra

Group 22 - Blanca Bastardés Climent, Alice Ciallella

Short description

The implementation of this assignment has been done in MATLAB. All the code is provided in the script **hw4.m**. The approach followed in this assignment consists in clustering multiple eigenvectors to classify spectral data.

Instructions to run the code

We provide the following folders and files:

- **hw4.m:** A matlab script containing code for clustering a spectral graph.
- **data/*:** Two files corresponding to two different graphs. The graph example1.dat is a real graph and example2.dat is a synthetic one.

Results

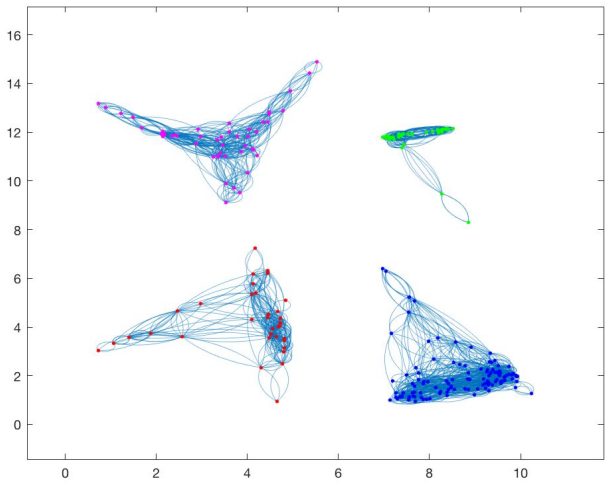
Two different approaches to compute the affinity matrix were used. The first one was using the adjacency matrix, in which there will be 1's where the edge exists and 0's where there is no link between nodes. The second approach was to use the distance between the nodes using the shortest path. No differences in the results can be observed between the two methods; we believe this is due to the fact that the clusters are well separated and the clustering task is relatively easy. With more complex graphs we expect the distances method to perform better than the adjacency matrix.

The number of clusters *k* has been chosen randomly in the first run and then adjusted after the plot of the results. Below we can observe the clusters for the two graphs.

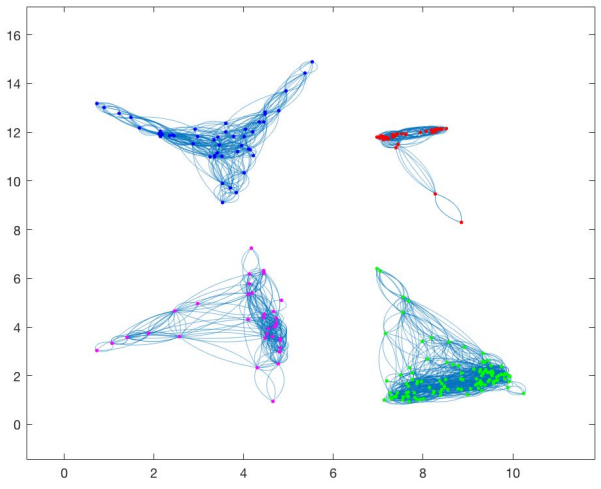
In Graph 1 we can distinctly observe four clusters, each represents an independent network making the graph disconnected. In Graph 2, we can notice two main clusters, which are linked through several edges meaning that the graph is connected.

(Plots in the following page)

- **Graph 1**

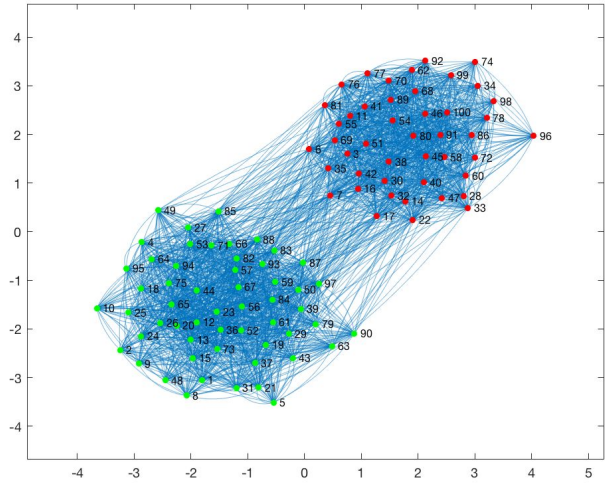


Clustering with adjacency matrix

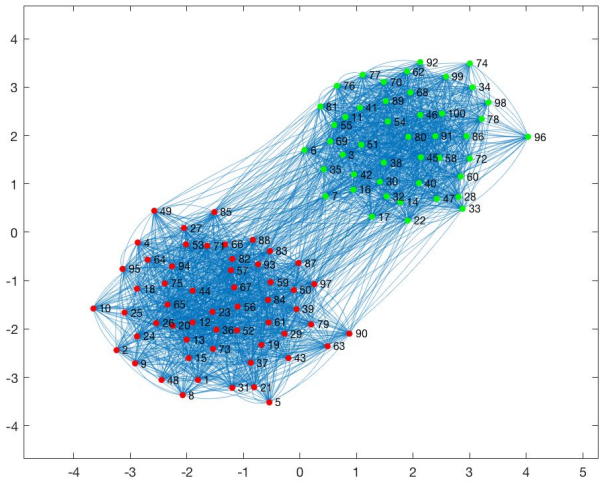


Clustering with affinity matrix considering distances

- **Graph 2**



Clustering with adjacency matrix



Clustering with affinity matrix considering distances