Homework 4: Graph Spectra

Group 11:

Boyu Xue (boyux@kth.se), Kristupas Bajarunas(kvba@kth.com)

Implement of the Algorithm:

1. Import the data of comma-separated edge list in Matlab.
2. Separate the data into two part: starting nodes and ending node, which correspond to the start and end of one edge.
3. Add the edge to the graph.
4. Create the adjacency matrix (affinity matrix) A which is a symmetric matrix.
5. Create the degree matrix D, which is a diagonal matrix and has the degree of the node on the diagonal.
6. Create the normalized laplacian matrix L by using equation $\mathrm{L} = \mathrm{D}^{-\frac{1}{2}}\mathrm{A}\mathrm{D}^{-\frac{1}{2}}$.
7. Calculate the eigenvectors and eigenvalues of L.
8. Find the k largest eigenvectors by calculating the optimal value of k. This is done by looking at the biggest eigengap. For dataset "example1", k=4. For dataset "example2", k=2.
9. Create the matrix $X = [X_1 X_2 \ldots X_k] \in R^{n\times k}$ by stacking the eigenvectors in columns.
10. Form the matrix Y from X by renormalizing each of X's rows to have unit length. Using the equation $Y_{ij} = X_{ij}/(\Sigma_j X_{ij}^2)^{\frac{1}{2}}$.
11. Treating each row of Y as a point in R^k , cluster them into k clusters via K-means.
12. Display graphs with clusters. Dataset "example1" has four independent clusters. Dataset "example2" has 2 clusters with a few edges connected with each other.

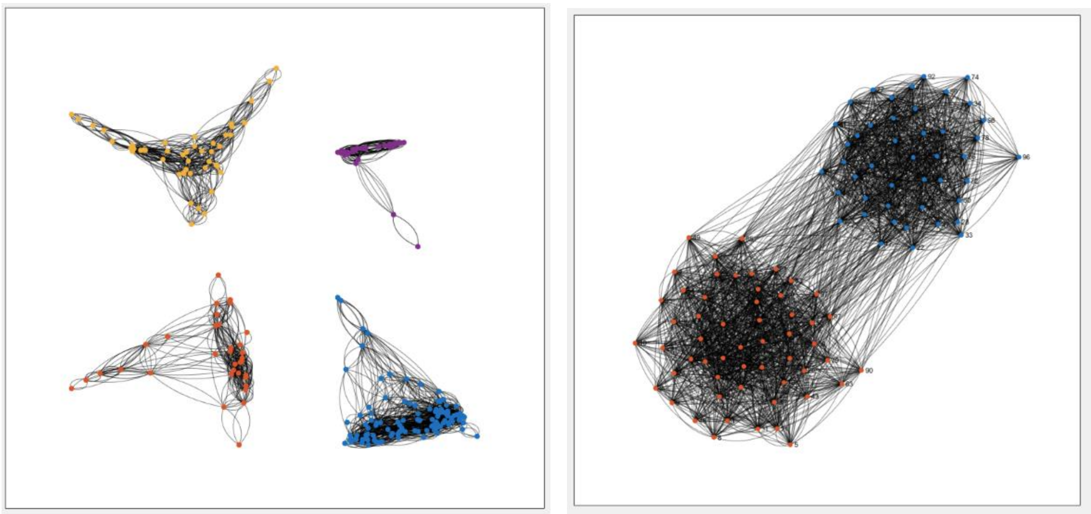

Fig.1. (a) Clusters of Dataset "example1". (b) Clusters of Dataset "example2".

13. Construct Fiedler Vector and plot the sorted Fiedler vector. This is the eigenvector corresponding to the second largest eigenvalue of the normal Laplacian matrix L=D-A.

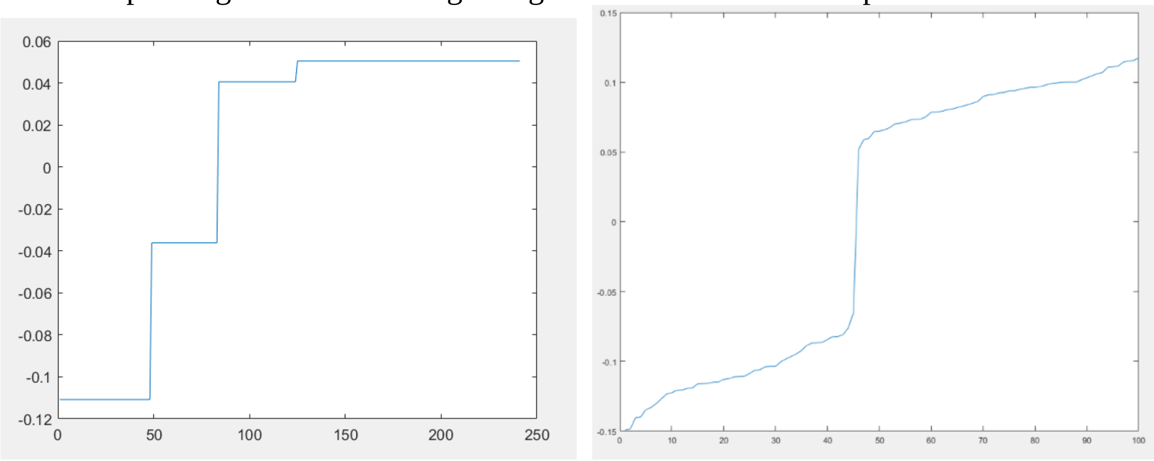

Fig.2. (a) Fiedler Vector of "example1". (b) Fiedler Vector of "example2".

14. Using the Fiedler Vector method to find communities and display the sparsity pattern. Dataset "example1" has four sparsity pattern. we cannot see independent communities which have no or very few connections with other communities.

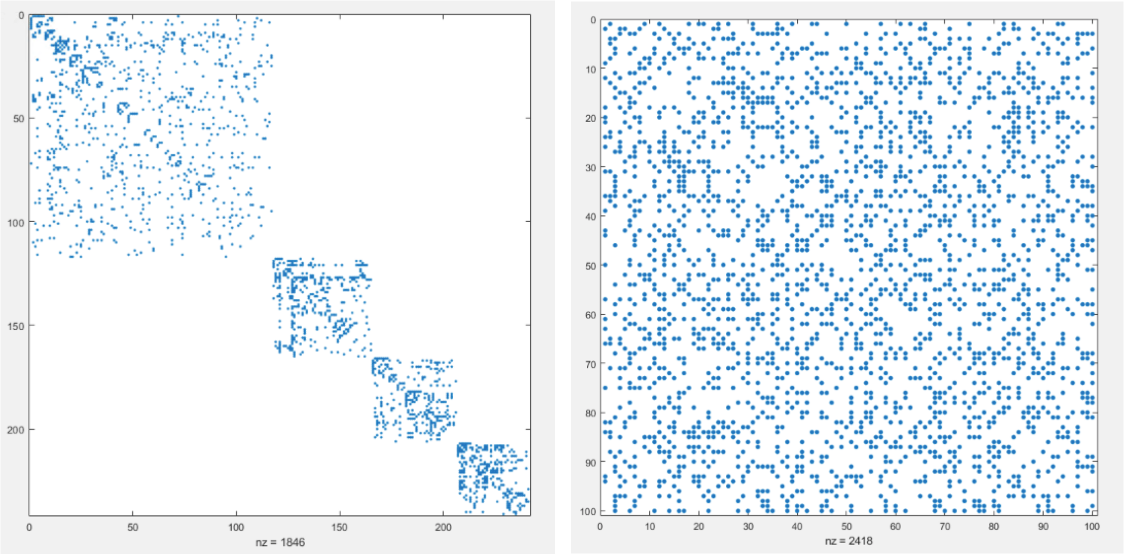

Fig.3. (a) Sparsity pattern of "example1". (b) Sparsity pattern of "example2".

Conclusion:

Dataset "example1" has four independent cluster, which means there is no connection between different clusters. And we can see the plot of the Fiedler vector of "example1" has four stages which correspond to four values of Field vector. And from the sparsity pattern of "example1", we can also see there is no connection between different communities.

Dataset "example2" has two clusters with some connection between them. We can see there is no stage in the plot of Fiedler vector of "example1", instead of the stages, the Fiedler vector increasing gradually with the increase of the counting point (100 points in total). And from the sparsity pattern of "example2", we cannot see independent communities which have no or very few connections with other communities. So maybe there is only one cluster in "example2".