Homework 4: Graph Spectra

Homework Group 54: Xu Wang(xuwa@kth.se)

- *Purpose*

In this homework, the spectral graph clustering algorithm described in nips01-spectral.pdf is implemented and we are going to use the algorithm to cluster two sample graphs, one real and one synthesized.

- Algorithm

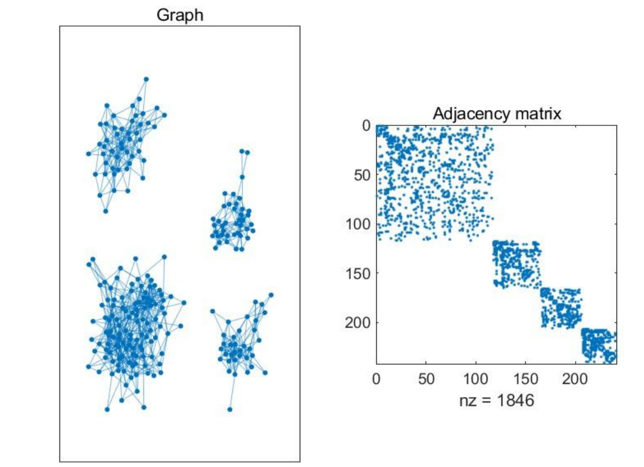
Given a set of points $S = \{s_1, \ldots, s_n\}$ in $\mathbb{R}^l$ that we want to cluster into k subsets:

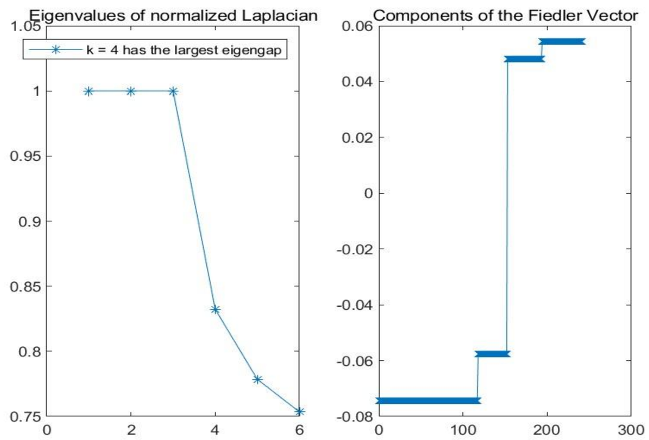
1. Form the affinity matrix $A \in \mathbb{R}^{n \times n}$ defined by $A_{ij} = \exp(-||s_i - s_j||^2/2\sigma^2)$ if $i \neq j$, and $A_{ii} = 0$.
2. Define D to be the diagonal matrix whose (i,i)-element is the sum of A's i-th row, and construct the matrix $L = D^{-1/2}AD^{-1/2}$.[1]
3. Find $x_1, x_2, \ldots, x_k$, the k largest eigenvectors of L (chosen to be orthogonal to each other in the case of repeated eigenvalues), and form the matrix $X = [x_1 x_2 \ldots x_k] \in \mathbb{R}^{n \times k}$ by stacking the eigenvectors in columns.
4. Form the matrix Y from X by renormalizing each of X's rows to have unit length (i.e. $Y_{ij} = X_{ij}/(\sum_j X_{ij}^2)^{1/2}$).
5. Treating each row of Y as a point in $\mathbb{R}^k$, cluster them into k clusters via K-means or any other algorithm (that attempts to minimize distortion).
6. Finally, assign the original point s_i to cluster j if and only if row i of the matrix Y was assigned to cluster j.

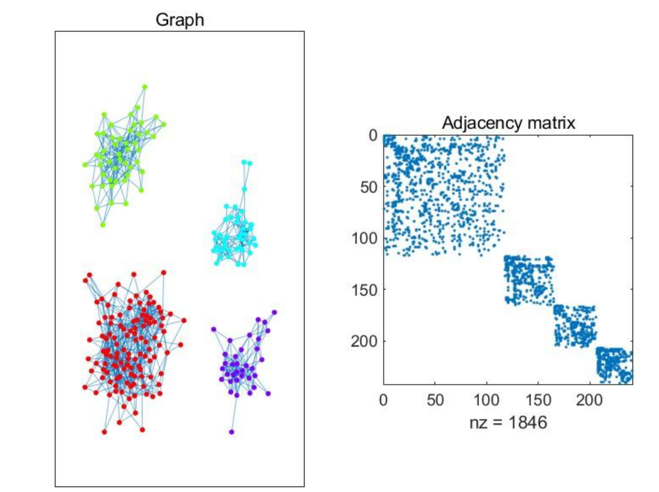
In this homework, we simply use the adjacency matrix to represent the graph since there are no weights on edges. The number of clusters k is determined by maximizing the eigengap of the eigenvalues of the normalized Laplacian matrix described in step 2.

Fiedler Vector: The eigenvector corresponding to the second smallest eigenvalue of the Laplacian matrix, L, is called Fiedler Vector. If the graph has two modules, it bisects the graph into only two communities based on the sign of the corresponding vector entry.

- *Example 1(Real)*







- *Example 2(synthetic)*