Graph Mining CSF426
Lab session 8 (evaluative)
Time: 5 pm – 7 pm
Date: 17-10-2024
Instructor IC - Vinti Agarwal

Instructions: All questions need to be answered. **You are required to write programs in jupyter notebook and submit .ipynb on canvas**. Please rename your solutions in format <ID-NAME-LABNO>. For theoretical questions, you can type answers in the jupyter notebook itself. There is no need to create a separate text file. **[Total Marks =10]**

Objective: Compare the iterative spectral clustering approach with k-means based spectral clustering approach.

Dataset: The dataset provided to you is known as "Karate club graph". The dataset is in the form of a graph consisting of 34 nodes divided into 2 classes ("Mr. Hi" & "Officer") and 156 weighted edges.

Tasks:

***Iterative Spectral clustering*:**

Perform spectral clustering to create four different clusters by following the below mentioned steps:

a. **Constructing the Laplacian Matrix:** Compute the Laplacian matrix L from the given adjacency matrix A of the graph.
b. **Spectral Decomposition of the Laplacian Matrix:** Calculate the eigenvalues and eigenvectors of the Laplacian matrix L.
c. **Spectral Clustering Based on the Fiedler Eigenvector:** Partition the data into two clusters (C1 and C2) using spectral clustering based on the Fiedler eigenvector (vector corresponding to second smallest eigenvalue of Laplacian matrix).
d. **Iterative Spectral Clustering:** Perform iterative spectral clustering by recursively partitioning the data into smaller clusters based on the Fiedler eigenvectors of subgraphs. That is, further divide both C1 and C2 into two clusters each (say C11, C12, C21, C22) by creating a subgraph of the nodes in cluster.
e. Evaluating Cluster Purity: Compute the purity of the obtained clusters to assess the quality of the clustering results. Also, plot the clusters.

Spectral clustering using k-means clustering algorithm:

a. **Constructing the Laplacian Matrix:** Compute the Laplacian matrix L from the given adjacency matrix A of the graph.
b. **Spectral Decomposition of the Laplacian Matrix**: Calculate the eigenvalues and eigenvectors of the Laplacian matrix L.
c. **Constructing a Feature Matrix from Eigenvectors:** Create a feature matrix X by concatenating the first m eigenvectors of the Laplacian matrix. The value of m can be determined by examining the sorted eigenvalues and selecting an appropriate cutoff point.
d. **K-Means Clustering:** Divide the data into four clusters using the k-means clustering algorithm on the feature matrix X. In this case, the feature matrix X represents the node embeddings, and the algorithm will group nodes with similar feature values into the same cluster.
e. **Evaluating Cluster Purity:** Compute the purity of the obtained clusters to assess the quality of the clustering results. Also, plot the clusters.