Homework 4: Graph Spectra

Data Mining – 4

Lakshmi Srinidh Pachabotla

Group - 56

December 02, 2024

1. Introduction

Clustering is a foundational task in unsupervised learning used to group data points based on their similarities. Spectral clustering is an advanced technique that utilizes graph theory and eigenvalue decomposition to identify communities in data. Unlike traditional clustering methods, spectral clustering is particularly effective for non-convex clusters or irregularly shaped data.

In this assignment, the spectral clustering algorithm is implemented and analysed as described in the paper "On Spectral Clustering: Analysis and an Algorithm" by Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. The datasets used are:

- ❖ A real-world graph (example1.dat), representing a social network of medical professionals.
- ❖ A synthetic graph (example2.dat), is generated to exhibit complex clustering properties.

The goal was to identify clusters in datasets using the graph Laplacian, analyse the results through eigenvalue gaps, fielder vectors, and visualizations, and validate the effectiveness of special clustering.

2. Solution and Methodology

- Graph Representation: Datasets were passed as edgeless took construct undirected graphs. Adjacency matrix (A) encodes connections between nodes. The degree matrix (D) is the diagonal matrix capturing the degree of each node. Graph Laplacian is constructed as L = D - A, capturing the structure of the graph.

- Eigenvalue and Eigenvector Computation: The eigenvalues and eigenvectors of the Laplacian were computed as the smallest eigenvalues provided insights into the graph structure and the second smallest eigenvector, that is Fielder vector, was used to determine the clusters.

- Dimensionality Reduction and Clustering: The eigenvectors corresponding to the smallest eigenvalues (excluding the first trivial eigenvector) were used to embed the graph in a lower dimensional space. K means clustering was applied to this representation to partition the notes.

- Visualization and Analysis: Clustered graphs are the notes that are color-coded based on their clusters. The Eigenvalue Gap is the plot of sorted Eugene values guided by the choice of the number of clusters. The fielder vector is the scatter plot of the fielder vector validated by the cluster separations. Sparsity pattern is the plots of the Laplacian matrix demonstrated intra-cluster connectivity and inter-cluster spare city.

3. **Graph Analysis**:

For the dataset example1.dat, the value gap plot revealed a significant separation in the eigenvalues, confirming the presence of three natural clusters in the graph. The clustered graph visualization further validated this observation by showing three distinct groups of notes, each with dense internal connections and minimal overlap with other clusters. The Fielder vector plot supported this by presenting three clear bands of values corresponding to the three clusters. Additionally, the sparsity pattern of the Laplacian matrix demonstrated a block-like structure where intra-cluster connections were densely represented while inter-cluster. Connections were sparse, reinforcing the validity of the clustering process.

In the case of example2.dat, value gap plots suggested 4 clusters that aligned with the clustering results. The clustered graph visualization highlighted 4 well-separated groups of notes showcasing the algorithm’s ability to handle complex and dense graph structures. The Fielder vector plot provided further evidence of this, displaying 4 distinct groupings of node values that mapped directly to the four clusters. The sparsity pattern of the Laplacian matrix revealed a dense structure with clearly defined regions corresponding to each cluster, indicating strong intra-cluster connections and limited inter-cluster interactions. Together this analysis demonstrated the effectiveness of spectral clustering in identifying communities within the graph.

4. Conclusion

This assignment demonstrated the effectiveness of spectral clustering in identifying communities within complex graph structures. The eigenvalue gap analysis and field vector plots provided strong validation for the chosen numbers of clusters (K = 3, for example1.dat, and K = 4, for example 2.dat).

Spectral clustering successfully grouped notes into distinct clusters, even in dense or irregularly structured graphs. Visualizations such as clustered graphs. Begin value gaps and sparsity patterns reinforced the accuracy of the results. This analysis highlights the practical applicability of spectral clustering in real-world and synthetic datasets.

In summary:

- Spectral clustering is a powerful tool for graph-based clustering tasks.
- Its ability to leverage the graph Laplacian and eigenvalue decomposition makes it particularly effective for non-convex and irregularly shaped data.