Report 4: Graph Spectra

Group 150: Lorenzo Deflorian, Riccardo Fragale

November 26, 2025

1   Introduction

The goal of this homework is to implement the spectral graph clustering algorithm described in this paper `https://ai.stanford.edu/~ang/papers/nips01-spectral.pdf` and use it to analyze two given graphs.

2   Implementation

The implementation is straightforward and can be found in the `miner/core/spectra/cluster_machi` module.

   Given the graph represented by it's adjacency matrix, the steps to follow are:

1. Remove loops from the graph by subtracting the diagonal of the graph from itself

2. Build the degree matrix D

3. Build the Laplacian matrix $L = D - A$

4. Compute the eigenvalues and eigenvectors of the Laplacian matrix L

5. We will take the first k eigenvectors corresponding to the k larges eigenvalues to build the feature matrix X

6. We will then form the normalized feature matrix Y to have unit length vectors for each row.

7. We will then cluster the data using the k-means algorithm.

   The degree matrix D is a diagonal matrix where the element (i, i) is the degree of the i-th node, hence it just tells us how many edges are connected to the i-th node.

   The Laplacian matrix L is a symmetric matrix that is used to compute the eigenvalues and eigenvectors of the graph. It is defined as $L = D - A$, where A is the adjacency matrix of the graph.

The idea behind the Laplacian matrix is to measure how far each node is from the other nodes. (... expand on this)

Once we have the feature matrix X, we can form the normalized feature matrix Y to have unit length vectors for each row.

The k-means algorithm is used to cluster the data into k clusters.

3   To cluster or not to cluster

One of the first questions that we can ask is: what is the optimal number of clusters to use?

To answer this question we can check the eigenvalues of the Laplacian matrix and compare the gaps between them. The idea is that the larger the gap, the more distinct the clusters are.

The implementation of this is straightforward and can be found in the `miner/core/spectra/gap_finder.py` module.

Let's take a look at the eigenvalues of the Laplacian matrix for the example 1 graph.

Table 1 shows the gaps between consecutive eigenvalues of the Laplacian matrix. The largest gap is between λ_4 and λ_5 with a value of 0.1942, suggesting that $k = 4$ might be an appropriate number of clusters.

Table 1: Gaps between consecutive eigenvalues of the Laplacian matrix

Eigenvalue Pair	Gap Value
$\lambda_1 - \lambda_2$	0.0000
$\lambda_2 - \lambda_3$	0.0000
$\lambda_3 - \lambda_4$	0.0000
$\lambda_4 - \lambda_5$	0.1942
$\lambda_5 - \lambda_6$	0.0012
$\lambda_6 - \lambda_7$	0.0483
$\lambda_7 - \lambda_8$	0.0093
$\lambda_8 - \lambda_9$	0.0043
$\lambda_9 - \lambda_{10}$	0.0044
$\lambda_{10} - \lambda_{11}$	0.0590
$\lambda_{11} - \lambda_{12}$	0.0030
$\lambda_{12} - \lambda_{13}$	0.0173
$\lambda_{13} - \lambda_{14}$	0.0038
$\lambda_{14} - \lambda_{15}$	0.0275

The largest gap is between λ_4 and λ_5 with a value of 0.1942, suggesting that $k = 4$ might be an appropriate number of clusters.

Let's plot the clusters for $k = 4$.

Figure 1: Clusters for $k = 4$

Let's now take a look at the goodness of the clusters. We will compute the inter-cluster edges count, the intra-cluster edges count, the expansion ratio and the conductance for different values of k and see if we actually did find the optimal number of clusters.

Table 2 shows the cluster quality metrics for different values of k. The best results are achieved when $k = 4$, which matches our earlier analysis based on eigenvalue gaps. For $k = 4$, we have zero inter-cluster edges and all edges are intra-cluster, resulting in perfect expansion ratio and conductance values of 0.00.

Table 2: Cluster quality metrics for different values of k

k	Inter-edges	Intra-edges	expansion ratio	conductance
2	0	923	0.00	0.00
3	0	923	0.00	0.00
4	0	923	0.00	0.00
5	12	911	0.71	0.41
6	35	888	0.71	0.32
7	126	797	0.71	0.50
8	163	760	0.93	0.48
9	191	732	1.61	0.62

Let's now take a look at the example 2 graph.

Table 3 shows the gaps between consecutive eigenvalues of the Laplacian matrix for example 2. The largest gap is between λ_2 and λ_3 with a value of 0.5451, suggesting that $k = 2$ might be an appropriate number of clusters.

Table 3: Gaps between consecutive eigenvalues of the Laplacian matrix for example 2

Eigenvalue Pair	Gap Value
$\lambda_1 - \lambda_2$	0.1656
$\lambda_2 - \lambda_3$	0.5451
$\lambda_3 - \lambda_4$	0.0187
$\lambda_4 - \lambda_5$	0.0173
$\lambda_5 - \lambda_6$	0.0042
$\lambda_6 - \lambda_7$	0.0098
$\lambda_7 - \lambda_8$	0.0028
$\lambda_8 - \lambda_9$	0.0094
$\lambda_9 - \lambda_{10}$	0.0136
$\lambda_{10} - \lambda_{11}$	0.0034
$\lambda_{11} - \lambda_{12}$	0.0106
$\lambda_{12} - \lambda_{13}$	0.0058
$\lambda_{13} - \lambda_{14}$	0.0078
$\lambda_{14} - \lambda_{15}$	0.0033

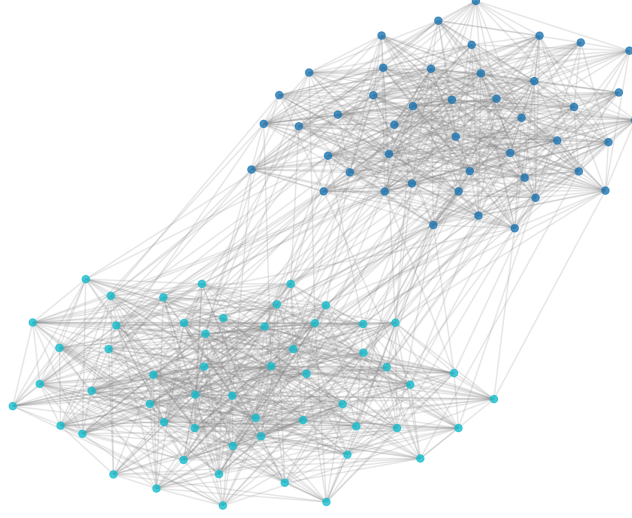If we plot the clusters for $k = 2$ we get the following result:

Figure 2: Clusters for $k = 2$

Let's now take a look at the goodness of the clusters for example 2. Table 4 shows the cluster quality metrics for different values of k. The best results are achieved when $k = 2$, which matches our earlier analysis based on eigenvalue gaps. For $k = 2$, we have the lowest number of inter-cluster edges (113) and the highest number of intra-cluster edges (1096), with the best expansion ratio and conductance values.

Table 4: Cluster quality metrics for different values of k (example 2)

k	Inter-edges	Intra-edges	expansion ratio	conductance
2	113	1096	0.23	0.19
3	355	854	1.80	0.64
4	569	640	2.09	0.68
5	644	565	3.90	0.80
6	704	505	5.08	0.84
7	749	460	18.40	0.95
8	825	384	7.53	0.88
9	844	365	10.31	0.91

4 Results

5 Conclusion