

Report 4: Graph Spectra

Group 150: Lorenzo Deflorian, Riccardo Fragale

November 24, 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_machine` module.

Given the graph represented by its 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 largest 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 $k = 4$.

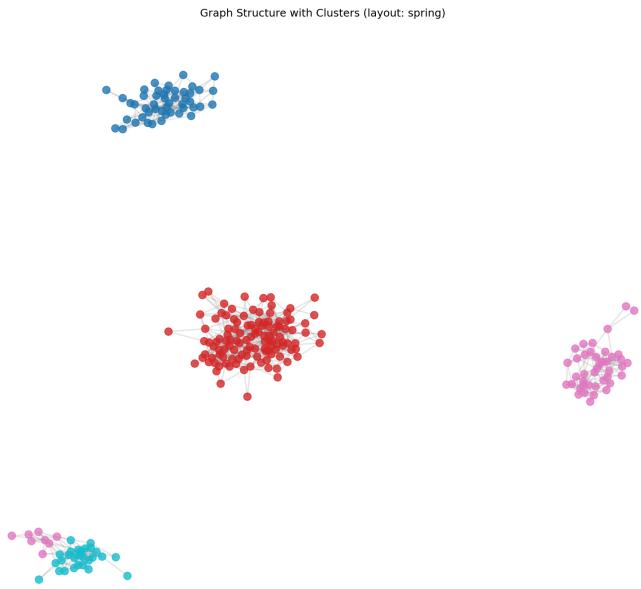


Figure 1: Clusters for $k = 4$

As we can see the clusters are quite good, the nodes are well separated and the four clusters are very distinct.

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

Table 2 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 2: 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:

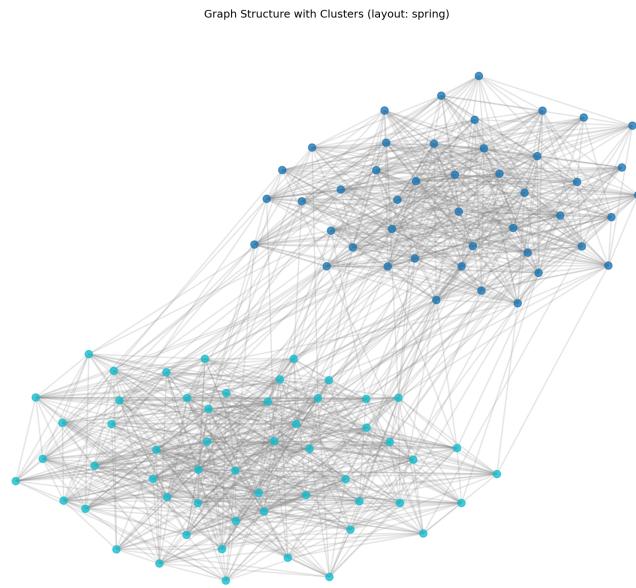


Figure 2: Clusters for $k = 2$

As we can see the clusters are not very good. The clusters are not very distinct and the clusters are not very well separated.

4 Results

5 Conclusion