Unnormalized Spectral Clustering on a Campus Graph

Zeynep Sena Yeşilyurt - 150230015
Computer Engineering Department
Istanbul Technical University
Istanbul, Turkey
yesilyurt23@itu.edu.tr

***Abstract*—In this project, spectral clustering is applied to the pedestrian road network of Istanbul Technical University (ITU) Maslak campus. The graph of the campus is generated using OpenStreetMap data with the help of the OSMnx library. The unnormalized graph Laplacian is computed, and its smallest eigenvectors are found using the inverse power method with orthogonalization. These eigenvectors are used as input for the KMeans algorithm to cluster the nodes. For comparison, coordinate-based KMeans clustering is also performed using only the physical positions of the nodes. The results are evaluated based on modularity scores and visual representations. According to the findings, spectral clustering gives better modularity and forms more meaningful clusters by using the graph's structure.**

***Index Terms*—spectral clustering, Laplacian, eigenvectors, modularity, KMeans, graph theory**

I. Introduction

Clustering is a common technique in machine learning that is used to group similar data points without using labeled information. One important type of clustering is spectral clustering, which is especially helpful when the data is represented as a graph. Instead of using distance, spectral clustering focuses on how nodes are connected in the graph. In this project, spectral clustering is applied to the pedestrian road network of Istanbul Technical University (ITU) Maslak campus. The map data is collected from OpenStreetMap using the OSMnx library, and a graph of the campus is created. The main aim is to divide the campus into meaningful regions by analyzing the connections between intersections (nodes) and roads (edges). To perform spectral clustering, the graph Laplacian matrix is built, which provides information about the structure of the graph. Then, the inverse power method with orthogonalization is used to find the smallest eigenvectors. These eigenvectors are given to the KMeans algorithm to form the final clusters. The results of spectral clustering are also compared with a simpler method: coordinate-based KMeans, which only uses the x and y positions of each node. To measure the quality of both methods, modularity scores are calculated and the clusters are shown on the map.

II. Methodology

In this section, the step-by-step process used to apply unnormalized spectral clustering to the ITU campus graph is explained. The method includes five main steps: building the graph from OpenStreetMap data, creating the necessary matrices such as the Laplacian, finding eigenvectors using the inverse power method, applying clustering in the spectral space, and finally checking the quality of the clusters using modularity. Each step is described in the following subsections.

A. Graph Construction and Matrix Preparation

In this project, the pedestrian road network of Istanbul Technical University (ITU) Maslak campus was used. The graph data was collected using the `OSMnx` Python library, which downloads road and intersection information from OpenStreetMap. In the graph, nodes represent intersections and edges show walkable paths between them.

Fig. 1. Pedestrian network of ITU campus visualized using OSMnx.

After the data was downloaded, the graph was converted into an undirected version so that spectral clustering could be applied. The adjacency matrix A was created from the graph, and the degree matrix D was calculated by summing the rows of A. Using these two matrices, the unnormalized graph Laplacian was computed as $L = D - A$.

The Laplacian matrix gives information about how nodes are connected in the network and plays an important role in spectral clustering. It helps analyze the relationships between nodes based on their connections instead of only their physical locations.

B. Spectral Clustering via Inverse Power Method

Spectral clustering is a technique that uses the eigenvalues and eigenvectors of a graph Laplacian matrix to find clusters in a network. In this project, unnormalized spectral clustering was applied to the campus road network using the smallest eigenvectors of the Laplacian matrix.

To find these eigenvectors, the *inverse power method* with orthogonalization was implemented. This numerical method makes it possible to calculate the smallest k eigenvectors without using any built-in eigensolver functions. The inverse power method works by solving a linear system of the form $(L-\mu I)^{-1}v$, where μ is a small shift value near zero and v is a random starting vector. To make sure that the eigenvectors are orthogonal, Gram-Schmidt orthogonalization was applied during each step.

The inverse power method was chosen instead of the standard power method because spectral clustering needs the smallest eigenvalues and their corresponding eigenvectors. The regular power method usually finds the largest eigenvalue, which is not useful in this case. The inverse power method helps find the smallest eigenvalues, which carry important structural information for detecting clusters in the graph.

After the k smallest eigenvectors were found, they were combined into a matrix and row-normalized. Then, the `KMeans` clustering algorithm was used to group the nodes based on these spectral features. The final clustering shows the structure of the graph and highlights connected regions in the campus network.

C. Coordinate-based KMeans Clustering

As a second method, the basic `KMeans` clustering algorithm was applied directly to the coordinates of the nodes. In this approach, only the x and y positions of each node on the map were used, and the connections between nodes in the graph were not considered.

This method is simpler and faster than spectral clustering because no graph structure or matrix calculation is needed. Instead, the nodes are grouped based on their physical positions.

Although the topological structure of the campus is not used in this method, meaningful regions can still be formed if the nodes are well distributed. Later, this method was compared with spectral clustering using modularity scores and visual results.

III. Experimental Results

In this section, the modularity scores and visual results of the clustering experiments are presented. Different values of k (the number of clusters) were tested for both spectral clustering and coordinate-based KMeans. The modularity score was used to measure how well the clusters match the structure of the graph. A higher score means better separation of communities.

Table I shows the modularity values for both methods across different values of k.

TABLE I
Modularity Scores for Different Values of k

Number of Clusters (k)	Spectral Clustering	Coordinate KMeans
2	0.4726	0.4604
3	0.5725	0.5805
4	0.6868	0.6482
5	0.7482	0.6562
6	0.7658	0.7412

Visual clustering results are shown in Figures 2–6. Each figure shows the graph colored by cluster labels. The spectral clustering result is shown on the left, and the coordinate-based KMeans result is shown on the right.

When the number of clusters increases, the separation between regions becomes clearer. For $k=5$ and $k=6$, spectral clustering provides better topological separation, while coordinate-based KMeans mostly reflects physical distance between nodes.

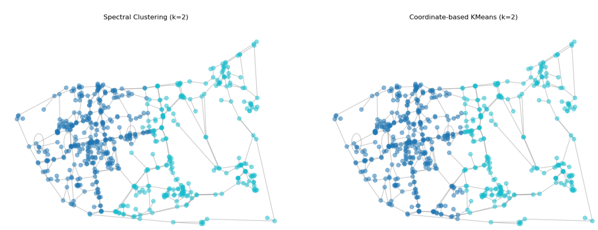

Fig. 2. Clustering results for $k=2$: Spectral (left), Coordinate-based KMeans (right).

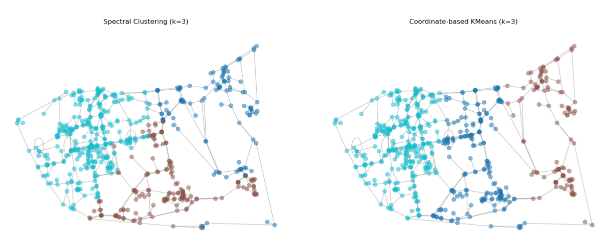

Fig. 3. Clustering results for $k=3$: Spectral (left), Coordinate-based KMeans (right).

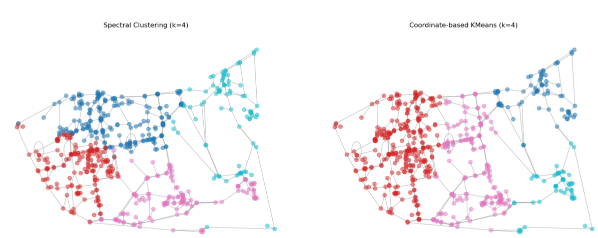

Fig. 4. Clustering results for $k=4$: Spectral (left), Coordinate-based KMeans (right).

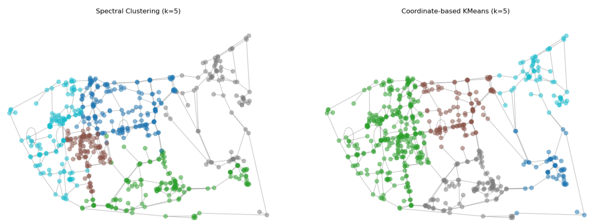

Fig. 5. Clustering results for $k = 5$: Spectral (left), Coordinate-based KMeans (right).

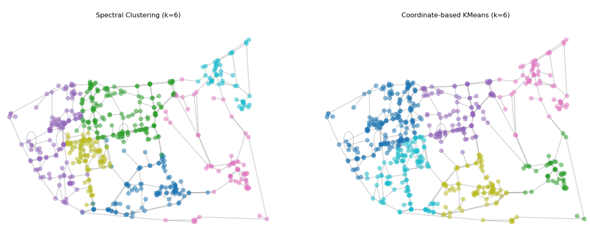

Fig. 6. Clustering results for $k = 6$: Spectral (left), Coordinate-based KMeans (right).

IV. Discussion

Figure 7 shows the modularity scores of both spectral clustering and coordinate-based KMeans for different values of k. As seen in the plot, the modularity of spectral clustering increases more steadily and reaches higher values, especially when $k \geq 4$. This means that spectral clustering is more successful in finding meaningful communities in the campus graph.

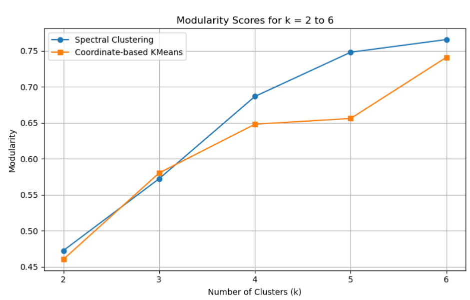

Fig. 7. Line plot of modularity scores for spectral and coordinate-based clustering methods.

In contrast, coordinate-based KMeans gives slightly lower scores for most values of k because it only uses the positions of the nodes and ignores the connections between them. The nodes are grouped by physical closeness, which does not always show the actual structure of the network.

Even though spectral clustering gives better modularity results, it also requires more computation time because of the eigenvector calculations in the inverse power method. This difference between accuracy and speed becomes more important when working with larger graphs or real-time systems.

In this project, the unnormalized Laplacian was used. In future work, using the normalized Laplacian or trying other types of spectral embeddings may lead to different results and better performance.

Overall, the results show that using the graph structure (as in spectral clustering) helps to form more meaningful groups, especially in complex spatial networks like a university campus.

V. Conclusion

In this project, unnormalized spectral clustering was applied to the pedestrian road network of Istanbul Technical University (ITU) Maslak campus. The smallest eigenvectors of the graph Laplacian were computed using the inverse power method, and clustering was performed using the KMeans algorithm. A simpler method, coordinate-based KMeans, was also applied to compare the results.

Based on the modularity scores, spectral clustering gave better results than coordinate-based clustering for most values of k. The visual outputs also supported this, as spectral clustering grouped the nodes more meaningfully by considering their connections in the graph.

These results suggest that using the structure of a graph is helpful when detecting communities in spatial networks. Although spectral clustering is more complex, it can provide more accurate and useful results when node connections are important.

In future studies, other versions of the Laplacian matrix or different spectral techniques can be tested to improve clustering quality or reduce computation time.

References

[1] Istanbul Technical University, “BLG202E: Numerical Methods in Computer Engineering – Term Project (Unnormalized Spectral Clustering),” 2025.

[2] U. Von Luxburg, “A tutorial on spectral clustering,” arXiv (Cornell University), Jan. 2007, doi: 10.48550/arxiv.0711.0189.

[3] G. Boeing, “OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks,” Computers Environment and Urban Systems, vol. 65, pp. 126–139, Jul. 2017, doi: 10.1016/j.compenvurbsys.2017.05.004.