SPECTRAL CLUSTERING

What is it?

Spectral clustering is a machine learning technique that groups data points into clusters based on their spectral properties, which capture the underlying structure of the data. It utilizes the eigenvalues and eigenvectors of a similarity or affinity matrix to partition the data into clusters, making it effective for discovering non-linear and complex patterns in the data.

Algorithm for Spectral Clustering

Form	Form a distance matrix
Transform	Transform the distance matrix into an affinity matrix A.
Compute	Compute the degree matrix D and the Laplacian matrix L = D-A.
Find	Find the eigenvalues and eigenvectors of L.
Use	Use the 'k' largest eigenvalues from the previous step to project the matrix into a lower dimensional space of dimension 'k'.
Cluster	Cluster the data points.

Why use Spectral Clustering?

Flexibility in handling Complex data. This means that it can handle data of different shapes and structures.

Spectral Clustering is based on connectivity of data. This type of clustering is useful as it helps find meaningful relationships from the dataset.

Some cons of Spectral Clustering

Higher complexity of computation. As we need to compute eigenvectors and eigenvalues of Laplacian Matrix, the process might become computationally expensive.

Choosing optimal parameters. As we need to choose parameters like 'k', it might get hard to choose the most optimal value of 'k'.

Uses for Spectral Clustering

Image Segmentation: Can be used to group pixels or regions with similar characteristics. This is useful in important research fields like computer vision.

Social Network Analysis: Can be used to identify communities or groups of nodes with dense connections and understand relationships within these networks.

Recommendation systems: Spectral clustering can be employed in recommendation systems to group users or items based on their preferences or similarities

EIGENVECTOR CENTRALITY

What is it?

In graph theory, eigenvector centrality is the measure of the influence of a node in a network. If a graph node is connected to a node with a high eigenvector centrality score, the graph node in question gets a higher centrality score as well.

How to compute Eigenvector Centrality

Mathematically, eigenvector centrality is computed using the eigenvector corresponding to the largest eigenvalue of the network. The adjacency matrix represents the connections or relationships between nodes in a network. The eigenvector centrality is given by the corresponding entry in the eigenvector.

Considerations to make

Centrality score for nodes with no incoming relationships will converge to 0.

Due to missing degree normalization, higher degree nodes have a very strong influence on their neighbours' score.

Uses for Eigenvector Centrality

In academic research, eigenvector centrality is used in citation networks to identify influential papers or authors. Nodes with high eigenvector centrality represent papers that are frequently cited by other influential papers, indicating their impact within the research community.

Eigenvector centrality has been extensively applied to study economic outcomes, including cooperation in social networks. In economic public goods problems, a person’s eigenvector centrality can be interpreted as how much that person’s preferences influence an efficient social outcome.