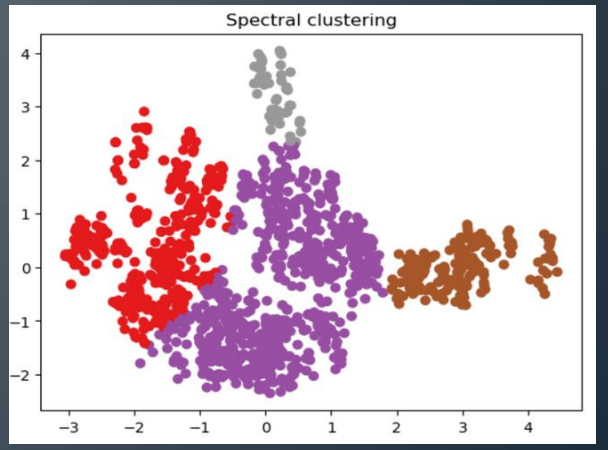
Spectral Clustering

Spectral Clustering

- In spectral clustering, set of nodes in a graph are identified based on edges that connect them.
- First, it uses information obtained from the spectrum of the similarities matrix to perform dimensionality reduction on the data.
- Next, it builds the Graph. Finally, it clusters the data.
- It has gained popularity across fields, including
  - ✓image segmentation
  - ✓social network analysis
  - ✓community detection



How does Spectral Clustering work?

- Make a graph to show the information. Euclidean distance, cosine similarity, and mutual information are a few examples of similarity metrics that can be used for this.
- Determine the graph's Laplacian matrix. The graph's connectedness is represented by the square Laplacian matrix. This is how it is defined: **L = D – W**
- where W is the graph's adjacency matrix, a square matrix that contains the similarity between each pair of nodes, and D is the graph's degree matrix, a diagonal matrix that contains the number of edges associated to each node.
- Find the Laplacian matrix's eigenvalues and eigenvectors. The inherent structure of the data is captured by the Laplacian matrix's eigenvalues and eigenvectors.
- Project the data points onto the Laplacian matrix's leading eigenvectors. The data is converted into a lower-dimensional space where the clusters are more distinctly divided by this dimensionality reduction stage.

Advantages of Spectral Clustering:

- **Scalability**: Spectral clustering can handle large datasets and high-dimensional data, as it reduces the dimensionality of the data before clustering.
- **Flexibility**: Spectral clustering can be applied to non-linearly separable data, as it does not rely on traditional distance-based clustering methods.
- **Robustness**: Spectral clustering can be more robust to noise and outliers in the data, as it considers the global structure of the data, rather than just local distances between data points.

Disadvantages of Spectral Clustering:

- **Complexity**: Spectral clustering can be computationally expensive, especially for large datasets, as it requires the calculation of eigenvectors and eigenvalues.
- **Model selection**: Choosing the right number of clusters and the right similarity matrix can be challenging and may require expert knowledge or trial and error.