Team Members: *Shruti Sureshan*(**M21CS015**), *Dhriti Prasanna Paul*(**M21CS056**)
Graph Clustering Algorithm

A graph cluster is a way of grouping data by forms of graphs. The assessment is based on two different forms. A graph is clustered in one of two ways: by clustering its nodes by their edges, or by their edges' distances. A vertex cluster is formed by grouping associated edges together based on their edge weights. A second method involves considering the graphs as points to be clustered and grouped on the basis of their similarity. Graph clustering falls under the category of graph mining algorithms. It is necessary to partition a graph in a way that dense clusters are closer to each other than others, and this can be achieved by using the concepts of Intracluster density (closer clusters will have a higher density) and Intercluster sparsity. The criteria to be followed are:

- In general, fewer inter-edges and more intra-edges indicate higher quality.
- There can never be any separation between cliques.
- There must be a connection between the clusters.

It is impossible to achieve all three properties of the perfect clustering scheme: Richness, Invariance of Scale, and Consistency, as described in Kleinberg's Impossibility Theorem. The Cut-based measures and counting measures mainly deal with conductance, expanded cuts, intercluster and intracluster conductance; performance and density are also computed. A graph clustering can be implemented according to various approaches: bottom-up, top-down, and location-based approaches. Starting with the singletons and merging clusters is the bottom-up approach. Starting with one cluster and splitting clusters is the top-down approach. During local optimization, we start with random clustering and migrate nodes.

We are always striving to minimize the distance squared, and the K means clustering method is the most popular. It consists of two variations: K-center (minimizing the maximum distance) and K-median (minimizing the average distance). The Greedy Agglomeration (merge) technique starts from singletons and does not merge unconnected clusters. Another technique is Clustering with a Minimum-Cut Tree, which creates groups with bigger intra-cluster cuts and smaller inter-cluster cuts. In-Network Percolation Clustering, the edges in the graph are removed iteratively until they reach a certain threshold before stopping. Here, the components initiate clusters.

A widely popular approach is the Markov Clustering algorithm along with random walks. The Markov chain is used to calculate random walks through a large graph simulating a long walk-through. Spectral Graph Theory is concerned with the correspondence between eigenvalues of the adjacency matrix and combinatorial properties of graphs. The concept of sparsity and sparsest cuts is defined, followed by an analysis of the eigenvalues of adjacency matrices in the context of optimization problems. The spectral partitioning technique is a popular heuristic for graph partitioning, as it identifies a partition by utilizing the approximate eigenvectors in the adjacency matrix.

Thank you.