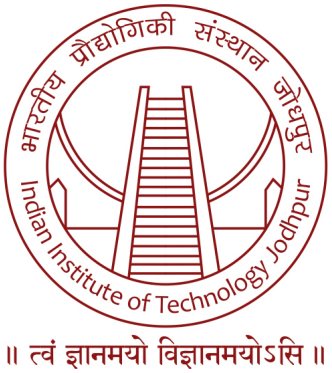
Intermediate Project Report

Report

on

Graph Clustering



by

Kishore Kumar Kalathur Chenchu (M21CS058)
Prabhala Sandhya Gayatri (M21CS060)
Tejaswee A (M21CS064)

Course Instructor

Dr. Anand Mishra
Assistant Professor
Department of Computer Science & Engineering
Indian Institute of Technology Jodhpur

1 Graph Clustering

Definition

Graph clustering is the process of partitioning a set of nodes present in a graph into disjoint groups or clusters so that the elements of each cluster are highly similar, but are very dissimilar to the elements of other clusters.

Objective

There exist many universally accepted graph clustering algorithms with the best strategies but usage of each algorithm depends on the use case. Given the constraints, optimizing these graph clustering algorithms is an NP-hard problem. Therefore, among the available best strategic graph clustering algorithms, our task is to explore a few of them that have a wider range of applications.

Graph Clustering Measures

The quality measures of a clustered graph is identified as follows:

1. Combination of less inter-cluster edges and more intra-cluster edges gives better and higher quality
2. Inseparable Cliques
3. Connected Clusters
4. Disjoint Cliques approaching maximum quality

Impossibility Theorem for Clustering

Given set S. Let $f : d \rightarrow \Gamma$ be a function on a distance function d on set S, returning a clustering Γ. No function f can simultaneously fulfill Scale Invariance, Richness and Consistency.[1]

There are cut based measures such as Conductance of a cut, Expansion of a cut, Inter-cluster conductance, Intra-cluster conductance and Counting based measures such as Density, Modularity.

Graph Clustering Approaches

The Graph Clustering Approaches are majorly classified as:

1. **Bottom-up:** It starts with singleton set of individual data points and merge them to form clusters.
2. **Top-down:** In this approach starting with one-cluster, we split to form smaller clusters in iterative manner.
3. **Local Optimization:** Starting with random clustering, we migrate towards the nodes.

There are various clustering algorithms that use the above mentioned approaches to form clusters. Top Down approach is used by K-Means algorithm to form clusters. Greedy Agglomeration, Markov Clustering Algorithm and Random Walks are known to use Bottom Up approach, as they start with singleton sets and merge to form clusters. Local Moving and Multilevel algorithm is a locally greedy algorithm. Therefore it is known to use Local Optimization Approach of clustering.

Other than the above mentioned algorithms there exist algorithms that are popular. Some of these are Clustering with Minimum-Cut Tree, Integer Linear Programs, Clique-Percolation and Spectral Partitioning algorithm.

References

[1] A. Marino, "Graph clustering algorithms." [Online]. Available: https://pages.di.unipi.it/marino/cluster18.pdf