Research Proposal for CSE304 Project

Tentative Title: Min-cut based Graph Clustering
Student Name: Deokhyeon Kim
Student ID: 20201032
Affiliation: Computer Science and Engineering, Ulsan National Institute of Science and Technology

1 Background and Motivation

Clustering is a fundamental task in data mining, used for discovering intrinsic structures in data. Since clustering problem does not even have formal definition, there are numerous ways to define clusters and various approaches to solve those problems. In this paper, I would like to utilize *min-cut* to divide a given graph into a designated number of clusters. Further, I would like to propose a new problem related to the min-cut.

2 Problem Statement

This proposal aims to address the following two problems.

- Problem 1: Construct k clusters efficiently by eliminating min-cut edges from given graph.
- Problem 2: Let, w[i][j] denotes the cost when Node i and Node j are in the different clusters. The problem is to find 3 clusters that minimize the sum of w[i][j] given that some nodes must belong to certain cluster.

3 Proposed Method

For problem 1, I would like to utilize the max-flow min-cut theorem to find the min-cut of each components. There are efficient data structures called the Gomory-Hu tree to find all-pairwise min-cut. We can find k clusters recursively by eliminating min-cut edges until desired number of clusters are found.

For problem 2, It can be shown that the problem is NP-hard. So I would like to propose approximation algorithms for this problem.

4 Experimental Plan

We plan to conduct experiments on benchmark datasets such as `MNIST`, `CIFAR-10`, and `Synthetic Gaussian Mixtures`. Metrics such as Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and runtime will be used to evaluate performance. Baseline comparisons will include vanilla K-means and DBSCAN on full data.

5 Expected Contributions

- A min-cut based approach to address graph clustering problem.
- Propose practical NP-hard problem related to min-cut.

6 Related Work

- Gomory and Hu, "Multi-terminal network flows", Journal of the Society for Industrial and Applied Mathematics, 1961.