Developing efficient heuristic approaches to cluster editing, inspired by other clustering problems

Angelos Zoumis

Supervisor: Dr Emir Demirović

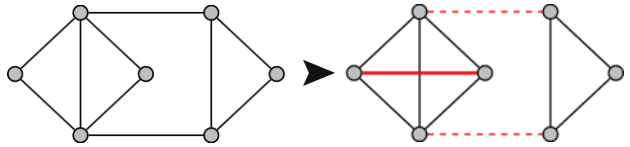
1. Problem Overview

Problem explanation

Clustering attempts to find the minimum number of **edge edits** (additions and removals), in order to transform a graph into a **cluster graph** (graph consisting of disconnected cliques).

Cluster editing visual example

Example of cluster editing transformation. This example requires 3 edits to transform into a cluster graph (“PACE 2021”, 2021)



2. Motivation

- The Cluster editing problem is **NP-Hard**. Therefore, solving on larger graphs is not possible in a reasonable time.
- The cluster editing problem has **many applications**, in sectors like **biology**.
- Using **existing algorithms** is an easy way to develop a heuristic approach for this problem

3. Algorithms

Base algorithms used

- **K-means algorithm**
- **Agglomerative hierarchical clustering**
- **Divisive hierarchical clustering**

Demands of the distance metric

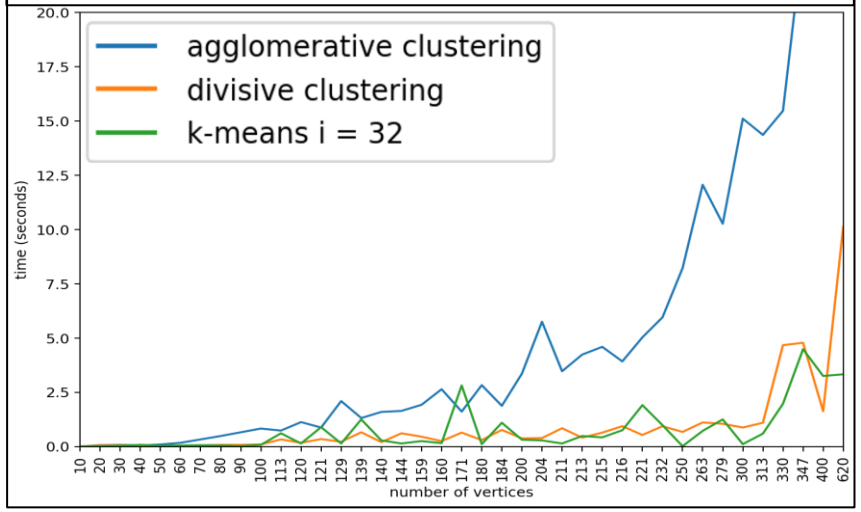
The algorithm for measuring distance should return a distance based on the **likelihood that two vertices are in the same cluster** in the optimal solution

Distance metric

- The final distance metric compares the **closed neighborhood** of the vertices we want to find the distance of.
- The closed neighborhood of a vertex consists of all vertices directly connected through an edge, plus itself

4. Results

Number of vertices of graph vs seconds taken to find the best answer



Edit difference from exact answer	Mean	Worst case
K-means	3.865%	68.75%
Agglomerative	20.03%	492.1%
Divisive	23.32%	109.1%

5. Conclusions

- Developing efficient heuristic approaches to cluster editing, inspired by other clustering problems is **feasible**
- In terms of time complexity and accuracy, the **k-means based algorithm performed the best**