Bachelor final project:
LP-based upper bounds for resource-constrained clustering

1 Introduction

In this project, a method is investigated for clustering an object set $\mathcal{A} = \{a_1, a_2, \ldots, a_N\} \subseteq \mathbf{R}^d$ for some $d \in \mathbb{N}$ into a set of clusters $\mathcal{C} = \{C_1, \ldots, C_K\}$, where $K \leq N$ and the elements of $\mathcal{C}$ are pairwise disjoint subsets of $\mathcal{A}$. In addition, we are given a subset $\{b_1, b_2, \ldots, b_K\} \subseteq \mathcal{A}$ such that $b_i \in C_i$ for each $i \in [K]$.

Blom [1] investigated a clustering method, which revolves around connecting objects by a set of disjoint trees T_i rooted at b_i for $i \in [K]$. Each cluster C_i is then intuitively given by the vertex set $V(T_i)$ of T_i.

2 Motivation

This clustering method is motivated by an application commonly found in logistics, namely the multi-depot vehicle routing problem (MDVRP, [2]). Within a clearly specified bounded region, a set of customers having a certain demand are served by multiple depots located strategically inside the region. Most natural algorithms to find solutions for MDVRP are so-called "cluster-first route-second heuristics". Each depot is assigned a subset of the customers, and conventional algorithms for single-depot vehicle routing problems, such as in [3], are used to solve these corresponding subproblems. In this context, our assumption means we can identify a depot for each cluster and create a tree connecting the depot to its subset of customers, where each location (customer or depot) is given by its latitude and longitude coordinates.

We assume that each depot has limited resources (e.g. number of vehicles) and therefore each tree should have bounded length (i.e., the sum of lengths of branches in the tree). Therefore, a resource parameter θ_d for each depot d is introduced. The goal is to find a set of disjoint trees T_d for each depot d such that as many customers as possible are covered by some tree and the resource parameters are satisfied.

3 Goal

Blom proved that this problem is strongly NP-hard, using a polynomial reduction from minimum weight k-cardinality tree problem [4]. Therefore, no polynomial exact algorithms exist for this problem unless P = NP. Blom also describes an IP formulation allowing for efficient Lagrangian queries to estimate the maximum objective value of this problem. However, practical results obtained from the Lagrange relaxation do not guarantee a good approximation of the maximum value. Therefore, interesting research ideas are as follows:

- Implement the linear relaxation of the integer programming formulation given in Blom. Although time complexity is higher for linear relaxation, it yields tighter upper bounds on the maximum value.
- Create a new heuristic for the resource-constrained clustering method.

This project is meant for ambitious students who have an interest in linear optimisation and combinatorial algorithms and who would like to do a final project that is both challenging in a theoretical and practical context. If you would like to know more about this project, contact Danny Blom either at d.a.m.p.blom@tue.nl or visit me in my office (MF4.096).

References

[1] D. Blom, "Upper and Lower Bounds for Rooted Forest Packings," *Master thesis (available upon request)*, 2019.

[2] G. B. Dantzig and J. H. Ramser, "The Truck Dispatching Problem," *Management Science*, vol. 6, no. 1, pp. 80–91, 1959.

[3] G. Clarke and J. Wright, "Scheduling of Vehicles from a Central Depot to a Number of Delivery Points," *Operations Research*, vol. 12, no. 4, pp. 568–581, 1964.

[4] M. Fischetti, H. W. Hamacher, K. Jørnsten, and F. Maffioli, "Weighted k-cardinality trees: Complexity and polyhedral structure," *Networks*, vol. 24, no. 1, pp. 11–21, 1994.