THE OPTIMAL CLUSTERING TREE PROBLEM.

By Dr. Michal Stern, joint work with Prof. Ephraim Korach.

Abstract

We consider the following problem: Given a complete graph $G=(V,E)$ with a cost on every edge and a given collection $\boldsymbol{S}$ of subsets of V. A spanning tree T is called a clustering tree (relative to $\boldsymbol{S}$) if each subset of the vertices in $\boldsymbol{S}$ induces a subtree in T. Our problem is to find a minimum cost clustering tree T. We call this problem the *OP* problem. One motivation for this problem is to construct a minimum cost communication tree network for a collection of non-disjoint groups of customers such that the network will provide "group fault tolerance" and "group privacy". We model this problem as a matroid and present a polynomial algorithm for it. For the case where the cardinality of the subsets in the collection does not exceed three, we provide a greedy algorithm, a linear algorithm and also a polyhedron description of the convex hull of all the feasible solutions. We consider also the *clustering-TSP-path* problem. In this case the problem is to construct a minimum cost TSP path instead of a tree. The problem in general is NP-hard and we solve some restricted cases. For the *stars-OP* problem, where each subset induces a complete star (instead of a subtree) of the constructed solution tree, we have a *structure theorem* and based on it a polynomial algorithm. We consider another version of the problem called the *Median Optimization Problem* (*MOP*): We have to find in G a tree T where each subset of the vertices in the collection induces a subtree (or a path or a star) in T such that the sum of all costs of the subtrees (or the paths or the stars) is minimum. For the subtrees and the stars cases we prove that a tree is optimal for *OP* if and only if it is optimal for *MOP*.