Feasibility of the Clustered Spanning Tree Problem Extended Abstract

Michal Stern — **Academic College of Tel Aviv-Yaffo and CRI University of Haifa**
Nili Guttmann-Beck — **Academic College of Tel Aviv-Yaffo**
Zeev Sorek — **Academic College of Tel Aviv-Yaffo**

Abstract

Let $H =< G, \mathcal{S} >$ be a hypergraph, where $G = (V, E)$ is a complete undirected graph and $\mathcal{S}$ is a set of clusters $S_i \subseteq V$, which may intersect. The Clustered Spanning Tree problem is to find whether there exists a spanning tree of G such that each cluster induces a subtree. In this paper we provide different techniques for this decision problem. The first technique can handle every instance hypergraph and is based on finding a maximum spanning tree in a special defined weighted version of G. The weights of the edges in the graph are defined according to the number of clusters containing each edge. All other techniques consider the intersection graph of H to decide whether a feasible solution exists. When applicable, these techniques achieve better complexity, as the intersection graph is significantly smaller compared with the original graph.

1 Introduction

Let $G = (V, E)$ be a complete undirected graph with vertex set V such that $|V| = n$ and complete edge set E. Let $H =< G, \mathcal{S} >$ be a hypergraph, where $\mathcal{S}$ is a set of clusters $S_1, \ldots, S_m$, $S_i \subseteq V$ for $i \in \{1, \ldots, m\}$, where the clusters may intersect. The Clustered Spanning Tree problem, denoted by CST, is to find an whether there exists a spanning tree of G, such that each cluster induces a subtree.

The main result in this paper is a unique technique for the essential and interesting question of whether a feasible solution exists for a given instance of the CST problem. This technique requires $O(n^2m)$ time complexity and can handle every instance hypergraph. In the first stage of the algorithm a new weighted graph is constructed, where the weight of each edge in the graph is equal to the number of clusters containing this edge. Next, we find a maximum spanning tree for this graph. A feasible solution for the CST problem exists if and only if the weight of this tree is $\sum_{i=1}^{m} |S_i| - m$. When equality holds, the maximum spanning tree offers a feasible solution.

This paper also introduces feasibility decisions using information derived from the hypergraph and the corresponding intersection graph. First, we prove that when H has a solution tree T^H, then induced subtrees of T^H are solution trees for the corresponding induced subproblems. This also proves that when an induced hypergraph does not have a solution tree, neither does H. For the special case where every vertex in V is contained in at most 2 clusters from $\mathcal{S}$, the CST problem has a feasible solution if and only if the corresponding intersection graph is a tree. As a result of the above claims, when the intersection graph of H contains a chordless cycle with at least 4 nodes, no feasible solution exists. The latter result agrees with the results summarized in [10].

When the intersection graph contains a cut-edge, we prove that deciding whether a feasible solution exists can be based on the decision made for each part of the intersection graph independently. The feasible solution tree for the given hypergraph is constructed using the feasible solution subtrees created from the corresponding subproblems and thus may significantly reduce the required complexity.

Throughout this paper we assume that the intersection graph of H is connected. Otherwise, a feasible solution tree for H can be constructed by adding edges between the feasible solutions of each connected component, if they exist.

In [4], [5], [11] and summarized in [10] it is shown that a hypergraph has a feasible solution tree if and only if it satisfies the Helly property and its intersection graph is chordal. In [6] it is proven that a graph is the intersection graph of subtrees of a tree if and only if it is chordal. In [12] a polynomial algorithm, which constructs a tree where each cluster spans a path, is presented. The most restricted problem where both the tree and subtrees are required to be paths, is in fact the Consecutive Ones Problem, which Booth and Leuker [3] solved in linear time using PQ-trees.

In the optimization *CST* problem the edges of E have weights and the objective is to find a feasible solution tree with minimum weight. This problem was solved by Korach and Stern in [8] where an optimum solution is found in $O(n^4m^2)$ time complexity, when a feasible solution exists. In addition, an abstraction of the problem using matroids is presented. For the restricted case where each cluster contains at most three vertices, there is a linear time algorithm and a polyhedral description of all feasible solutions. A special case of the optimization *CST* problem, where the optimum spanning tree solution is required to span a complete star on each cluster, is presented in [9]. A structure theorem which describes all feasible solutions and a polynomial algorithm for finding an optimum solution are presented, when the intersection graph is connected. Another related optimization problem is the clustering *TSP*-path, where the optimum solution tree is required to be a *TSP*-path. A lot of research has been investigated on the clustered *TSP*-path where the clusters are disjoint, for example bounded-approximation algorithms are presented in [1], [2] and [7].

2 Feasibility Based on a Weighted graph

This section introduces Algorithm Existence of Solution (ES) which is the main result of the paper. The algorithm, presented in Figure 1, either finds a solution tree for a hypergraph $H =< G, \mathcal{S} >$ or states that there is no feasible solution. The algorithm creates a weighted graph denoted by $G_{ES} = (V_{ES}, E_{ES})$, where $V_{ES} = V$ and E_{ES} contains an edge (v, u) if there exists a cluster S_i such that $\{v, u\} \subseteq S_i$. The weight $w(v, u)$ of an edge (v, u) is equal to the number of clusters containing both v and u. In this graph a maximum spanning tree T_{ES} is found. We prove that the hypergraph has a solution tree if and only if $w(T_{ES}) = \sum_{i=1}^{m} |S_i| - m$. When a solution exists, T_{ES} is a feasible solution tree.

Theorem 1. *Given a hypergraph $H =< G, \mathcal{S} >$, for every cluster $S_i \in \mathcal{S}$ and every T a spanning tree of G_{ES}, $w_i(T) \leq |S_i| - 1$ with $w_i(v, u) = 1$ if $\{v, u\} \subseteq S_i$ and $w_i(v, u) = 0$ otherwise. An equality holds if $T|S_i$ is a spanning tree of S_i. Moreover, $w(T) = \sum_{i=1}^{m} w_i(T) \leq \sum_{i=1}^{m} |S_i| - m$.*

Theorem 2. *Given a hypergraph $H =< G, \mathcal{S} >$ and T_{ES} a maximum spanning tree of G_{ES}, $w(T_{ES}) = \sum_{i=1}^{m} |S_i| - m$ if and only if T_{ES} is a feasible solution tree.*

Theorem 3. *The complexity of Algorithm ES (Figure 1) is $O(n^2m)$.*

3 Induced Hypergraphs

Consider a hypergraph $H =< G, \mathcal{S} >$ which is an instance for the *CST* problem. New instances of the problem are created by considering induced hypergraphs defined by collections of clusters

ES (Existence of Solution)
input *A hypergraph* $H =< G, \mathcal{S} >$, *where* $G = (V, E)$, $\mathcal{S} = \{S_1, \ldots, S_m\}$, $S_i \subseteq V$.
returns *A feasible solution tree or a a statement "No feasible solution".*
begin
Construct the following weighted graph $G_{ES} = (V_{ES}, E_{ES})$ *where* $V_{ES} \equiv V$,
and E_{ES} *contains the edge* (v, u) *(for* $v \neq u$*) if there exists a cluster* S_i *such that* $\{v, u\} \subseteq S_i$.
The weight of an edge (v, u) *is set:* $w(v, u) = |\{S_i : i \in \{1, \ldots, m\}, \{v, u\} \subseteq S_i\}|$.
Find T_{ES} *- a maximum spanning tree for* G_{ES}.
if $w(T) = \sum_{i=1}^{m} |S_i| - m$
  then return T_{ES}
  else return *"No feasible solution".*
end if
end *ES*

Figure 1: Algorithm ES

from $\mathcal{S}$. In this section we prove that when H has a solution tree T^H, then induced subtrees of T^H are solution trees for the corresponding induced subproblems. This also proves that when an induced hypergraph does not have a feasible solution tree, neither does H. Furthermore, when the intersection graph of H contains a cut-edge, the feasibility problem can be divided into smaller subproblems, where a feasible solution to the given instance exists if each subproblem has a feasible solution. The complexity decreases according to the sizes of the subproblems.

Theorem 4. *Let* $H =< G, \mathcal{S} >$ *be a hypergraph with a connected intersection graph and a solution tree* T^H. *If for* $\mathcal{S}' \subseteq \mathcal{S}$ *the intersection graph of* $\mathcal{S}'$ *is connected, then the induced tree of* T^H *on* $\bigcup_{S_i \in \mathcal{S}'} S_i$ *is a solution tree for the induced hypergraph* $H|\mathcal{S}'$.

Corollary 5. *Let* $H =< G, \mathcal{S} >$ *be a hypergraph with a connected intersection graph. If for* $\mathcal{S}' \subseteq \mathcal{S}$ *the intersection graph of* $\mathcal{S}'$ *is connected, and* $H|\mathcal{S}'$ *has no feasible solution tree, then* H *has no feasible solution tree.*

Theorem 6. *Let* $H =< G, \mathcal{S} >$ *be a hypergraph with a connected intersection graph. If the intersection graph contains a cut-edge (bridge) which divides the intersection graph into two connected components: the two intersection graphs of* $\mathcal{S}'$ *and* $\mathcal{S} \backslash \mathcal{S}'$, *then* H *has a feasible solution tree if and only if each one of* $H|\mathcal{S}'$ *and* $H|(\mathcal{S} \backslash \mathcal{S}')$ *has a feasible solution tree.*

4 Bounded number of containing clusters

The feasibility of the CST problem can be determined using information derived from different structures of the intersection graph. Since the intersection graph is significantly smaller compared with the weighted graph considered by Algorithm ES, methods for feasibility check based on the intersection graph can achieve better complexity.

It is proved here that when every vertex in V is contained in at most 2 clusters from $\mathcal{S}$, the CST problem has a feasible solution tree if and only if the intersection graph is a tree. As a result we get that when the intersection graph for an instance of the CST problem contains a chordless cycle with at least 4 nodes, then the instance has no feasible solution tree.

Definition 7. *For every $v \in V$ we define* $\boldsymbol{nc(v)} = |\{S_i : i \in \{1, \ldots, m\}, v \in S_i\}|$ *(the number of clusters which contain the vertex v).*

Observation 8. *In an intersection graph, a vertex $u \in V$ which belongs to k clusters ($nc(u) = k$) creates a k-sized clique. Hence, if the intersection graph of a hypergraph is a tree or a chordless cycle with at least 4 nodes, then $nc(v) \leq 2$ $\forall v \in V$.*

Theorem 9. *Let $H =< G, \mathcal{S} >$ be a hypergraph with $nc(v) \leq 2$ $\forall v \in V$ and a connected intersection graph. H has a feasible solution tree if and only if its intersection graph is a tree.*

Theorem 10. *Let $H =< G, \mathcal{S} >$ be a given hypergraph. The complexity of verifying that $nc(v) \leq 2$ $\forall v \in V$ and that the intersection graph of H is a tree, is $O(m^2 n)$.*

Corollary 11. *Let $H =< G, \mathcal{S} >$ be a hypergraph with a connected intersection graph of $\mathcal{S}$. If the intersection graph contains a chordless cycle with at least 4 nodes, then H has no feasible solution tree.*

The last Corollary is also proven using a completely different approach in [4], [5], [11] and summarized in [10].

References

[1] S. Anily, J. Bramel, and A. Hertz. A $\frac{5}{3}$-approximation algorithm for the clustered traveling salesman tour and path problems. *Oper. Res. Lett.*, 24(1-2):29–35, 1999.

[2] E. M. Arkin, R. Hassin, and L. Klein. Restricted delivery problems on a network. *Networks*, 29(4):205–216, 1997.

[3] K. S. Booth and G. S. Lueker. Testing for the consecutive ones property, interval graphs, and graph planarity using *PQ*-tree algorithms. *J. Comput. System Sci.*, 13(3):335–379, 1976. Working Papers presented at the ACM-SIGACT Symposium on the Theory of Computing (Albuquerque, N. M., 1975).

[4] P. Duchet. Propriété de Helly et problémes de représentation. *Colloqu. Internat. CNRS, Problemes Conbinatoires et Theorie du Graphs, Orsay, France* , 260: 117–118, 1976.

[5] C. Flament. Hypergraphes arborés. *Discrete Math.*, 21(3):223–227, 1978.

[6] F. Gavril. The intersection graphs of subtrees in trees are exactly the chordal graphs. *J. Combinatorial Theory Ser. B*, 16:47–56, 1974.

[7] N. Guttmann-Beck, R. Hassin, S. Khuller, and B. Raghavachari. Approximation algorithms with bounded performance guarantees for the clustered traveling salesman problem. *Algorithmica*, 28(4):422–437, 2000.

[8] E. Korach and M. Stern. The clustering matroid and the optimal clustering tree. *Math. Program.*, 98(1-3, Ser. B):385–414, 2003. Integer programming (Pittsburgh, PA, 2002).

[9] E. Korach and M. Stern. The complete optimal stars-clustering-tree problem. *Discrete Appl. Math.*, 156(4):444–450, 2008.

[10] T. A. McKee and F. R. McMorris. *Topics in intersection graph theory.* SIAM Monographs on Discrete Mathematics and Applications. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999.

[11] P. J. Slater. A characterization of soft hypergraphs. *Canad. Math. Bull.*, 21(3):335–337, 1978.

[12] R. Swaminathan and D. K. Wagner. On the consecutive-retrieval problem. *SIAM J. Comput.*, 23(2):398–414, 1994.