Extended formulations for the cardinality-constrained subtree of a tree problem

Agostinho Agra* Luis Gouveia⋆ Cristina Requejo◇

*Department of Mathematics and CEOC, University of Aveiro
Campus Universitário de Santiago, 3810-193 Aveiro

⋆Department of Statistics and Operations Research and Centro IO, University of Lisbon
Bloco C/6, Piso 4 - Campo Grande, 1749-016 Lisboa

◇Department of Mathematics and CEOC, University of Aveiro
Campus Universitário de Santiago, 3810-193 Aveiro

Abstract

We consider the problem of finding the most profitable subtree of a given tree with at most K nodes which is known as the Cardinality Subtree of a Tree Problem (CSTP). We present a new extended formulation with $\mathcal{O}(nK)$ two-indexed variables and $\mathcal{O}(nK)$ constraints. Each variable contains information on node inclusion as well as the order by which the node is included in the subtree solution. We show that the corresponding linear programming relaxation has integral extreme points.

Keywords: *subtree of a tree problem, cardinality-constraint, cardinality-constrained shortest path problem, extended formulations, ordering variables*

1 Introduction

We consider an undirected tree $T = (V, E)$ rooted at node 0 with node set $V = \{0, 1, \ldots, n\}$ and edge set E, a profit f_i associated with each node $i \in V$ and a positive integer K. We define the profit of a subtree as the sum of the profits of the nodes included in it. The Cardinality Constrained Subtree of a Tree Problem (CSTP) is to find, in T, a subtree $T' = (V', E')$ rooted at node 0 with maximum profit and such that $|V'| \leq K$. When $K = n+1$ the cardinality constraint becomes redundant and we obtain the (unconstrained) subtree of a tree problem which is studied in [6]. We assume w.l.o.g.: (i) $K < n+1$ and (ii) the unique path from node 0 to each node $j \in V \setminus \{0\}$ has length at most $K-1$.

Let p_j denote the predecessor of node j in T and, for all $j \in V$, consider the binary variables x_j indicating whether node j is in the solution. The CSTP can be formulated as the following integer linear programming problem.

$$\text{(CST)} \quad \max \quad \sum_{j \in V} f_j x_j$$

$$s.t. \quad x_0 = 1, \tag{1}$$

$$x_{p_j} \geq x_j, \ j = 1, \ldots, n \tag{2}$$

$$\sum_{j \in V} x_j \leq K \tag{3}$$

$$x_j \in \{0, 1\}, \ j = 1, \ldots, n \tag{4}$$

Constraint (1) ensures the root node is selected. This condition is assumed w.l.o.g. to simplify the formulations presented in Section 2. Constraints (2) establish that a node is in the solution only if its predecessor is in the solution. Constraints (3) state that the number of nodes in the solution is at most K. And constraints (4) are the variables integrality constraints.

This problem is a particular case of the constrained subtree of a tree problem (see, e.g., [5, 8, 9]), where (3) is replaced by the knapsack constraint $\sum_{j \in V} w_j x_j \leq b$ where $w_j \in \mathbb{Z}_+, j \in V$, $b \in \mathbb{Z}_+$. For that reason the CSTP can be solved in polynomial time by dynamic programming (see, e.g., [1, 5, 8]).

The CSTP was previously studied in [6] and [1]. Aghezzaf et al. [1] have given a description of co(CSTP) (where co(P) gives the convex hull of the integer solutions of problem P) for K = 3 and 4. However, by showing how to reduce the knapsack problem to a special case of the CSTP, they argue that providing a description of co(CSTP) would be at least as complicated as providing a complete linear description of the knapsack polytope with a constraint having a right-hand side equal to $K - 1$. This implies that, with exception to rather small values of K, such a description would be quite difficult to obtain explicitly. Alternatively, we can try to provide, implicitly, a description of the co(CSTP) by giving an extended formulation for the problem (a formulation that uses additional sets of variables) and such that the projection of the feasible set of the corresponding linear programming relaxation into the space of the x_j variables equals co(CSTP). In this paper we provide one such extended formulation. The underlying idea for deriving the formulation given in this paper is based on the following two observations: (i) our strong conviction that "good" variables to be introduced in such an extended formulation should have information on an "ordering" of the nodes in the subtree solution and (ii) the knowledge that this additional information should be given as an extra index of the original 0/1 variables. The second observation is based on knowledge taken from other works, where similar re-indexations have been used (see, e.g., ([2, 3]), to construct models with a strong linear programming relaxation. With the new set of variables, we will be able to obtain a compact and extended formulation for the CSTP which has $\mathcal{O}(nK)$ variables and $\mathcal{O}(nK)$ constraints and such that the projection of the linear programming feasible set of the formulation is equal to co(CSTP) for all values of K. To show this result we give an alternate characterization of the problem which permits us to provide a different extended formulation (although less compact than the previous one) for the problem and whose linear programming relaxation is easily proven to be tight (it is a shortest path formulation in a special graph). Since this shortest path formulation is defined in a space strictly containing the variable set of the ordering formulation, we, then, use projection to show that the set of feasible solutions of the linear programming relaxation of the "shortest path" formulation can be projected into the set of feasible solutions of the linear programming relaxation of the proposed and more compact "ordering" formulation.

This paper is organized as follows. In Section 2 we relate the CSTP with the cardinality constrained shortest path problem and as a consequence we obtain a shortest-path reformulation for the CSTP whose linear programming relaxation is easily proven to be tight. In Section 3 we study formulations using the new "ordering" variables and a new formulation is such that the set of feasible solutions of its linear programming relaxation has integer extreme points.

In the paper we denote by $F(P)$ the set of feasible solutions of a given model P with optimal value $v(P)$ and denote by P_L its Linear Programming relaxation.

2 The CSTP and the cardinality constrained shortest path problem

In this section we show that the CSTP can be modelled as a cardinality constrained shortest path problem in an adequate graph. Then, by following [4] we obtain a formulation whose linear programming relaxation is integer.

Consider a digraph $G = (V', A)$ where $V' = \{0, \ldots, n+1\}$, $n+1$ is a destination node (dummy node) and A is the set of arcs. There is an arc $(i, j) \in A$, for all $i, j \in V$ if and only if node j can be included in the subtree solution immediately after node i, following the DFO. There is an arc $(j, n+1) \in A$ for each node $j \in V$ indicating whether node j is the last node (following the DFO) in the subtree solution.

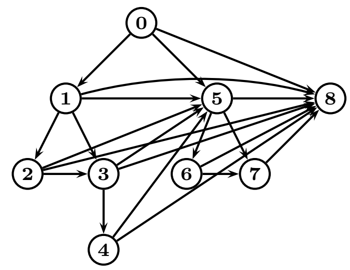

Figure 1: Example of the digraph G corresponding to the tree of Figure 2.

See Figure 1. More formally, let $P(\ell) = \{0, \ell_1, \ell_2, \ldots, \ell\}$ denote the unique path in tree T connecting the root node to node ℓ. Then node j is a successor of node i in G if and only if $P(j) \setminus \{j\} \subseteq P(i)$. We denote by $S(i)$ the set of successors of i in G.

It is easy to verify that for each subtree T' of T there exists only one path in G from node 0 to node $n+1$ that traverses all the nodes in T' and only those. Conversely, for a set P of nodes traversed by a path in G, from node 0 to node $n+1$, corresponds a unique subtree of T exactly with the nodes from P.

Following this discussion, we can follow [4] and model the CSTP as a cardinality shortest path problem over the digraph G.

To built this shortest path model consider the binary variables y_{ij}^k, for all $(i,j) \in A$ and $k \in \{1, \ldots, K\}$, indicating whether arc (i,j) is in the path in position k. In terms of the CSTP, y_{ij}^k indicates whether, in the original tree T, node i is in the solution with order k and node j is in the solution with order $k+1$. As before, assume $y_{ij}^k = 0$ if either the length of the unique path in T from node 0 to node i is greater or equal to k or $k > \min\{j, K-1\}, j < n+1$. For simplicity we omit these constraints from the model. The new formulation is given as follows.

$$
\begin{aligned}
(\text{SP}) \qquad \max \quad & \sum_{(i,j)\in A} \sum_{k=1}^{K} f_i y_{ij}^k \\
s.t. \quad & \sum_{j \in S(0)} y_{0j}^1 = 1 && (5) \\
& \sum_{i=p_j}^{j-1} y_{ij}^{k-1} = \sum_{t \in S(j)} y_{jt}^k, \ j = 1, \ldots, n, \ k = 2, \ldots, K && (6) \\
& \sum_{i \in V} \sum_{k=1}^{K} y_{i,n+1}^k = 1 && (7) \\
& y_{ij}^k \in \{0,1\}, \ (i,j) \in A, k \in \{1, \ldots, K\} && (8)
\end{aligned}
$$

Constraints (5) indicate that a feasible solution must have node 0 with order 1. Constraints (6) are the flow conservation constraints for a node j with order k, for $k = 2, \ldots, K$ and indicate that in a feasible solution a node has order k only if one of its predecessors in G has order $k-1$. Constraints (7) indicate that a feasible solution must have node $n+1$ with order from 2 to $K+1$.

Let SP_L denote the linear programming relaxation of model SP obtained by replacing constraints (8) with the corresponding nonnegativity constraints

$$
y_{ij}^k \geq 0, \quad (i,j) \in A, k \in \{1, \ldots, K\} \tag{9}
$$

Following, Gouveia [4] we have

Theorem 2.1 *The extreme points of the polyhedron defined by the set of feasible solutions of SP_L are integer valued.*

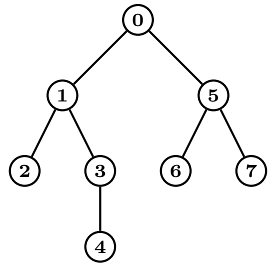

Figure 2: Example of an undirected tree where the nodes are numbered by the DFO.

3 A Reformulation with Ordering Variables

In this section we introduce the concept of "ordering" of a node in a subtree solution, introduce new variables which also give information on the ordering of the corresponding node in the subtree and derive a compact formulation with $\mathcal{O}(nK)$ variables and $\mathcal{O}(nK)$ constraints.

We assume that the nodes of the rooted tree T are numbered by the Depth First Ordering (DFO), from left to right. DFO numbers a node as far as possible along each branch before backtracking, see Figure 2.

The concept of ordering of a node in a subtree solution is similar to the DFO numbering of the nodes in the original tree. However, in the context of a solution, the ordering is applied only to the nodes of the solution. For simplicity we assume that the "ordering" starts on value 1 and that the ordering of the root node, node 0, is always one since we have assumed it is always included in every subtree solution. For instance, in the example of Figure 2 consider the subtree $T' = \{0, 1, 3, 5, 7\}$. Following this ordering, node 0 has order 1, node 1 has order 2, node 3 has order 3, node 5 has order 4 and node 7 has order 5.

To introduce the new model consider the binary variables x_i^k, for all $i \in V$ and $k \in \{1, \ldots, K\}$, indicating whether, following the DFO, node i is in the solution with order equal to k. The variables of the original model and the new variables can be related as follows:

$$x_i = \sum_{k=1}^{K} x_i^k, \text{ for all } i \in V. \tag{10}$$

We note that several of these new variables can be set to zero. First notice that the root node 0 is always selected, $x_0 = 1$, and we assume it will be selected in position 1 leading to

$$x_0^1 = 1. \tag{11}$$

If the length of the unique path in T from node 0 to node i is equal to $l_i - 1$, then, since, in any feasible subtree T', it is impossible that node i either has an ordering less or equal to $l_i - 1$ or has an ordering greater than $\min\{i+1, K\}$, we must have $x_i^k = 0,\ i = 1, \ldots, n,\ k < l_i$ or $k > \min\{i+1, K\}$. To ease the notation, we consider these variables in the model and assume they are set to zero in the formulations presented next.

Using the "ordering" variables the CSTP can be formulated as follows.

$$
\begin{aligned}
\text{(O-CST)} \quad & \max \quad \sum_{j \in V} \sum_{k=1}^{K} f_j x_j^k \\
& s.t. \quad (1), (2), (3), (4), (10), (11) \\
& \qquad x_i^k \leq \sum_{j=p_i}^{i-1} x_j^{k-1}, i = 1, \dots, n, \ k = 2, \dots, K \qquad (12) \\
& \qquad \sum_{j \in V \setminus \{0\}} x_j^k \leq 1, \ k = 2, \dots, K \qquad (13) \\
& \qquad x_j^k \in \{0, 1\}, \ j = 1, \dots, n, \ k = 1, \dots, K \qquad (14)
\end{aligned}
$$

Inequalities (11) - (14) give the required interpretation of the new variables. Constraints (12) state that if the order of node i is equal to k and the order of node j is equal to $k-1$, then j belongs to the set $\{p_i, \dots, i-1\}$, since the predecessor of node i, p_i, must also be included in the solution. Constraints (13) state that at most one node has a given order k, $k = 2, \dots, K$. Constraints (14) are the integrality constraints on the "ordering" variables.

Clearly, O-CST is a valid formulation for the problem and since the constraints (1) - (4) of the original model are included in the extended formulation, the following result can be easily established.

Proposition 3.1 *$v(CST_L) \geq v(O\text{-}CST_L)$.*

To show that the LP bound of the new model can strictly dominate the LP bound of the original model it suffices to note that for the example depicted in Figure 2 with $K = 4$ and $P = [3\ 2\ 1\ 4\ 1\ 1\ 4\ 1]$ we have $v(\text{CST}_L)$=11.5 and $v(\text{O-CST}_L)$=10.

The "key" inequalities for obtaining a tight model are based on the observation that for each value k at most one variable x_i^k can assume value 1 as stated by constraints (13). Thus, inequalities (12) can be lifted into the stronger inequalities

$$\sum_{j=p_i+1}^{i} x_j^k \leq \sum_{j=p_i}^{i-1} x_j^{k-1}, i = 1, \dots, n, \ k = 2, \dots, K. \qquad (15)$$

Proposition 3.2 *Constraints* (15) *are valid for F(O-CST).*

Constraints (10), (11), (14) and (15) imply constraints (2), (3), (12) and (13). Thus, after including (15) in model O-CST we can remove constraints (2), (3), (12) and (13). This new model, with (1), (4), (10), (11), (14), (15) will be denoted by SO-CST (Strengthened Ordering Cardinality Subtree of a Tree). We have the following result.

Proposition 3.3 *F(SO-CST) = F(O-CST).*

In order to present the main result of the paper, we consider the linear programming relaxation of SO-CST. For simplicity we write SO-CST$_L$ using only the "ordering variables" x_i^k. We use the linking constraints (10) to eliminate the x_i variables and, clearly, (1) and (4) can be dropped. We replace constraints (14) by constraints

$$x_j^k \geq 0, \ j = 1, \dots, n, \ k = 1, \dots, K. \qquad (16)$$

For completeness, the model SO-CST$_L$ rewritten with variables x_i^k alone is as follows

$$\max \sum_{j \in V} \sum_{k=1}^{K} f_j x_j^k \ \text{ subject to } (11), (15), (16).$$

The main result of our paper is stated next.

Theorem 3.4 *The extreme points of the polyhedron defined by the set of feasible solutions of SO-CST$_L$ are integer valued.*

This result can be proved by using projection techniques to show that the set of feasible solutions of the linear programming relaxation of the shortest path formulation can be projected into the set of feasible solutions of the linear programming relaxation of the "ordering" formulation SO-CST.

We conclude the paper by pointing out that Theorem 3.4 follows from Theorem 2.1 and a projection result.

References

[1] E.H. Aghezzaf, T.L. Magnanti, and L.A. Wolsey. Optimizing constrained subtrees of trees. *Mathematical Programming*, 71:113–126, 1995.

[2] M. Constantino and L. Gouveia. Reformulation by discretization: Application to economic lot sizing. *Operations Research Letters*, 35(5): 645-650, 2007.

[3] I. Correia, L. Gouveia and F. Saldanha-da-Gama, Discretized formulations for capacitated location problems with modular distribution costs Working Paper 7-2007, CIO, 2007.

[4] L. Gouveia. Using variable redefinition for computing lower bounds for minimum spanning and Steiner trees with hop constraints. *INFORMS Journal on Computing*, 10(2):180–188, 1998.

[5] D.S. Johnson and K.A. Niemi. On knapsacks, partitions, and new dynamic programming technique for trees. *Mathematics of Operations Research*, 8:1–14, 1983.

[6] T. Magnanti and L. Wolsey. Optimal trees. In M.O. Ball, T.L. Magnanti, C.L. Monma, and G.L. Nemhauser, editors, *Network Models, Handbooks in Operations Research and Management Science, Vol. 7.* Elsevier Science Publishers, North-Holland, 1995.

[7] G. Nemhauser, L. Wolsey, *Integer and Combinatorial Optimization*, Wiley, 1988.

[8] D.X. Shaw and G. Cho. A depth-first dynamic programming algorithm for the tree knapsack problem. *INFORMS Journal on Computing*, 9:431–438, 1997.

[9] D.J. van der Merwe and J.M. Hattingh. Tree knapsack approaches for local access network design. *Operations Research*, 38:1968–1978, 1990.