On the Complexity of the Robust Spanning Tree Problem with Interval Data

Ionut Aron and Pascal Van Hentenryck

Department of Computer Science
Brown University
Providence, Rhode Island 02912

CS-03-04
March 2003

On the Complexity of the Robust Spanning Tree Problem with Interval Data

Ionuţ D. Aron* and Pascal Van Hentenryck
Brown University,
Box 1910,
Providence RI 02912

December 11, 2002

Abstract

This paper studies the complexity of the robust spanning tree problem with interval data (RSTID). It settles the conjecture of [9] by proving that the problem is NP-complete. The paper also shows that the problem remains NP-complete for (1) complete graphs and (2) when the cost intervals are all $[0, 1]$. The proofs of the results show that the difficulty of the RSTID problem resides in two distinct aspects of the problem: the topology of the graph and the numerical properties of the cost intervals. As a consequence, the results suggest new directions for improving and evaluating existing search algorithms [20, 2, 15] for the RSTID problem, since they have only focused on one of these aspects.

Keywords: Robust spanning tree, robust optimization, uncertainty, interval data.

1 Introduction

Many practical applications, particularly in telecommunication networks, can be formulated as *robust optimization* problems. The goal in robust optimization is to find a solution that hedges against the worst possible scenario. Several robustness criteria may be used when choosing among solutions, depending on the goal and specifics of the problem. Kouvelis and Yu [9] discuss three important such criteria, namely *absolute robustness*, *robust deviation* and *relative robustness*. In many cases, under these criteria, the robust equivalent of a polynomially solvable problem becomes **NP**-hard. A weaker criterion for robustness, which allows controlling the degree of conservatism of the solution, was proposed by [3]. As opposed to the criteria dicussed in [9], where the target is a solution whose worst-case *value* is optimal, the goal in [3] is to find a solution whose *variance* is optimal – the actual value (cost) of the solution can in fact be quite poor. This approach enables robust formulations of polynomially solvable (α-approximable) problems to remain polynomially solvable (α-approximable).

Our study is focused on the robust spanning tree problem in graphs where the edge costs are given by intervals (RSTID), under the *robust deviation* framework[1]. The goal of the RSTID problem is to find a spanning tree that minimizes the maximum deviation of its cost from the minimum spanning trees obtained for all possible realizations of the edge costs (i.e., costs within the given intervals). The problem has attracted considerable attention recently ([10, 9, 20, 2, 15]), particularly because of its importance in communication networks. For instance, reference [9] discusses the design of communication

*On leave at Carnegie Mellon University.

[1] It is interesting to remark that under the *absolute robustness* framework, the problem is known to have a polynomial time algorithm.

networks where the routing delays on the edges are uncertain, since they depend on the network traffic. A robust spanning tree hedges against the worst possible delays and is desirable in this problem.

Attempts to solve the problem include a mixed integer programming (MIP) formulation by [20], a constraint satisfaction approach by [2], and a branch and bound approach by [15]. However, no efficient algorithm was discovered and the problem was conjectured to be NP-hard by [9].

This paper closes this conjecture by showing that the RSTID problem is at least as hard as the central tree (CT) problem introduced by [5] and proved NP-complete by [4]. The paper also indicates that the RSTID problem remains NP-complete even in the case of complete graphs. This is in contrast to the central tree problem which admits efficient solutions for complete graphs [11]. These results also shed a new light on the algorithms which have been proposed for the RSTID problem [20, 2, 15]. These algorithms have focused almost exclusively on the cost structure and have ignored the topological properties of the graph. Our results show that both the cost structure and the topology of the graph contribute to the complexity of the problem. As a consequence, they provide new directions to improve these algorithms and to evaluate them experimentally.

The rest of the paper is organized as follows. Section 2 defines the RSTID problem. Section 3 summarizes prior work on central trees which dates back to 1966. Section 4 proves that the special case of the RSTID problem where all the cost intervals are $[0, 1]$ is NP-complete. Section 5 shows that the RSTID problem remains NP-complete for complete graphs. Section 6 concludes the paper.

2 The RSTID problem

The RSTID problem is defined on undirected connected graphs. Given such a graph $G = (V, E)$, the RSTID problem also assumes that each edge e is associated with a cost interval $[\underline{c_e}, \overline{c_e}]$. A **scenario** is a particular assignment of costs to edges from their corresponding intervals. The **deviation** of a spanning tree T under a given scenario s, denoted by $\Delta_T(s)$, is the difference between the cost of T and the cost of a minimum spanning tree under scenario s, where the cost of a tree is simply the sum of its edge costs:

$$\Delta_T(s) = cost(T, s) - cost(MST_s, s).$$

The scenario s_T^* under which this difference is maximized is called the **worst case scenario** for the spanning tree T and the corresponding deviation is the **robust deviation** of T, which is denoted by Δ_T^*. In symbols,

$$s_T^* = arg \max_s \Delta_T(s)$$

and

$$\Delta_T^* = \Delta_T(s_T^*).$$

The following result gives a characterization of the worst case scenario for a spanning tree T:

Lemma 2.1 ([20]). *The robust deviation of a spanning tree T is achieved when edges in T are at the highest cost and edges in the complement of T in G are at the lowest cost (i.e. $c_e = \overline{c_e}, \forall\ e \in T$ and $c_e = \underline{c_e}, \forall\ e \in E \setminus T$).*

A **robust spanning tree** is a spanning tree whose robust deviation is minimum. The **robust spanning tree problem with interval data** is defined as the following decision problem: *Given an undirected and connected graph G with interval edge costs and a value $D \in \mathbf{R}$, is there a spanning tree T of G whose robust deviation $\Delta_T^* \leq D$?*

3 Central Trees

The **distance** between two spanning trees of a graph is defined as the number of edges which are in one tree but not in the other ([19]):[2]

$$d(T_1, T_2) = |T_1 - T_2| = |T_2 - T_1|.$$

Recall that the cospanning tree of a spanning tree T is the subgraph of G having all the vertices of G and

[2]The authors of [4] and [5] use slightly different formulations for the notion of distance between trees, but it is easy to see that they are equivalent, since the notion refers only to complete (i.e., spanning) trees.

exactly those edges of G not in T. Recall also that the rank $\rho(G)$ of a graph G with n vertices and k connected components is $n-k$.

A **central tree** ([5]) of a graph $G=(V,E)$ is a tree T_0 such that the rank r of its cospanning tree T_0^* is minimum, i.e.,

$$r=\rho(T_0^*)\leq\rho(T_i^*),\ \text{for every tree } T_i\in G.$$

Deo [5] pointed out that, if r is the rank of the cospanning tree T^* of a spanning tree T, then there is no tree in G at a distance greater than r from T and there is at least one tree in G at distance exactly r from T.[3] Thus, we have the following characterization (and hence the name) of central trees:

Lemma 3.1 ([5]). *A spanning tree T_0 is a central tree of G if and only if the largest distance from T_0 to any other tree in G is minimum, i.e.,*

$$\max_i d(T_0,T_i)\leq\max_i d(T_j,T_i),\ \textit{for every tree } T_j\in G.$$

Central trees were studied intensively in the literature (e.g., [12, 16, 18, 17]), primarily due to their importance in circuit analysis. Deo's initial study of the problem was motivated by the intuition that a central tree would provide a better starting point for tree generation methods, such as those proposed in [6, 14]. Unfortunately, no good algorithm for constructing a central tree was found for almost three decades,[4] although there are some interesting similarities with problems for which efficient algorithms exist. The dual of the central tree problem, for instance, which asks for a pair (T_1,T_2) of **maximally distant** trees (i.e., $d(T_1,T_2)\geq d(T_i,T_j)$ for any spanning trees $T_i,T_j\in G$), can be solved in polynomial time [8]. Also, as pointed out in [13], the distances between tree pairs of a graph G are in a one-to-one correspondence with the distances between the corresponding vertex pairs in the **tree-graph** $\mathcal{T}_G$ (recall that the vertices of a tree-graph $\mathcal{T}_G$ correspond to spanning trees in G, and two vertices are connected if the original spanning trees are at distance 1 of each other in G). Thus, finding a central tree in G is equivalent to finding a **central vertex** in $\mathcal{T}_G$. However, while the central vertex problem is known to have a polynomial time algorithm (in the number of vertices), such an algorithm cannot be used to efficiently find a central tree, since the number of vertices in $\mathcal{T}_G$ can be exponential. The search for an efficient algorithm for the central tree problem concluded with the following result, due to [4]:

Theorem 3.2 ([4]). *The* **Central Tree** *problem is NP-complete.*

4 The Zero-One RSTID

Consider the special case of the RSTID problem where the edges of G take their costs in the interval $[0,1]$. We refer to this problem as the Zero-One Robust Spanning Tree Problem (ZO-RSTID). More precisely, given a graph $G=(V,E)$, the RSTID problem whose underlying graph is G and whose cost intervals are all in $[0,1]$ is called the ZO-RSTID problem associated with G.

The next theorem is the main result of the paper. It shows that robust trees are central trees of the underlying graph and vice-versa.

Theorem 4.1. *Let $G=(V,E)$ be a connected and undirected graph and let $\mathcal{P}$ be the ZO-RSTID problem associated with G. T is a robust tree of $\mathcal{P}$ iff it is a central tree of G.*

Proof. [$\Rightarrow$] Let T be a robust tree for $\mathcal{P}$ with robust deviation D, i.e.,

$$cost(T,s_T^*)-cost(MST_{s_T^*},s_T^*)=D.$$

We first show that the cospanning tree of T has rank D. Let s be the worst-case scenario for T. By Lemma 2.1, this scenario assigns cost 1 to all edges in T and cost 0 to all the other edges. Let MST_s be the MST for scenario s. Since $cost(T,s)=|V|-1$, it follows that $cost(MST_s,s)=|V|-1-D$. Hence, T and MST_s must have $k=|V|-1-D$ edges in common. Let $I=\{e_1,\ldots,e_k\}$ be these edges and take such an edge $e_i=(v_i,w_i)$. By definition of an MST, vertices

[3] It can be shown that the maximum number of edges without a circuit in a graph G equals the rank of G.

[4] Initial attempts to solve the central tree problem resulted in a polynomial time algorithm proposed by [1], but a gap in this algorithm was uncovered by [7].

v_i and w_i are not connected in $G = (V, E \setminus \{e_i\})$. As a consequence, the graph $G = (V, E \setminus I)$ has $k+1$ connected components. By definition of a spanning tree, all the $k+1$ components are connected and, since $T \cap MST_s = I$, each of these compoents can be spanned by using edges in MST_s. Hence, $G = (V, E \setminus T)$ has $k+1$ components and the rank of the cospanning tree of T is thus $|V| - (k+1) = D$.

We now show that all cospanning trees of G have ranks at least D. Assume that there exists another spanning tree T' whose cospanning tree has rank $D' < D$. This means that the removal of T' from G creates $k = n - D'$ connected components $C_1, C_2, \ldots, C_k$. Let $t_1, t_2, \ldots, t_k$ be arbitrary spanning trees of these components. Since G is connected, there exist $k-1$ uniquely defined edges $e_1, e_2, \ldots, e_{k-1}$ in T' such that $\bigcup_{i=1}^{k} t_i \cup \bigcup_{j=1}^{k-1} e_j$ form a spanning tree T'' of G (note that this tree may or may not be the same as T', but it shares $e_1, e_2, \ldots, e_{k-1}$ with T'). Let s'' be the scenario in which all edges of T'' have cost 1 and all other edges have cost 0. By Lemma 2.1, s'' is the worst case scenario for T''. Since $e_1, e_2, \ldots, e_{k-1}$ are the only ways to connect $C_1, C_2, \ldots, C_k$, it follows that $MST_{s''}$ must contain all of them (otherwise it would not span G). Hence, the robust deviation of T'' satisfies

$$cost(T'', s'') - cost(MST_{s''}, s'') \leq |V| - 1 - (k-1) < D$$

which contradicts the fact that T is a robust spanning tree. The result follows.

[$\Leftarrow$] Let T be a central tree for G, whose cospanning tree has rank D. The cospanning tree of T has $k = |V| - D$ connected components. We first show that T has robust deviation D for $\mathcal{P}$. By Lemma 2.1, the worst-case scenario s for T assigns cost 1 to all edges in T and 0 to all other edges. Since the cospanning tree of T has rank D, it has $|V| - D$ connected components. Hence, an MST for scenario s must select $|V| - D - 1$ edges from T to connect these components and select edges not in T to connect the components themselves. As a consequence,

$$cost(T, s) - cost(MST_s, s) = |V| - 1 - (|V| - D - 1) = D.$$

We now show that no tree can have a robust deviation smaller than D. Assume that there exists a tree T_r with a robust deviation $D' < D$. This means that there exists a scenario s_r such that

$$cost(T_r, s_r) - cost(MST_{s_r}, s_r) = D'.$$

Since $cost(T_r, s_r) = |V| - 1$, we have that $cost(MST_{s_r}, s_r) = |V| - 1 - D'$ and T_r and MST_{s_r} have k' edges in common where

$$k' = |V| - 1 - D'.$$

Since $D' < D$, we have $k' \geq k$. Let $I = \{e_1, \ldots, e_{k'}\}$ be these edges. Removing I from G produces $k'+1$ connected components. As a consequence, any tree including I would have at least $k'+1$ connected components in its cospanning tree, which would have rank

$$|V| - (k'+1) < |V| - k = D.$$

This contradicts the hypothesis that T is a central tree. □

The NP-completeness of the robust spanning tree with interval data follows directly from Theorem 4.1.

Corollary 4.2. *The ZO-RSTID problem is NP-complete.*

Corollary 4.3. *The RSTID problem is NP-complete.*

5 RSTID on Complete Graphs

In the case of complete graphs, the central tree problem is known to have a polynomial time algorithm [11]. This section shows that this is not the case for the RSTID problem. This result is interesting, since it shows that the topology of the graph is not the only reason why the RSTID problem is hard (as Theorem 4.1 may seem to suggest).

Yaman et al. [20] have shown that, in a graph G with interval edge costs, there exists a well defined set of trees whose union contains all the solutions of the RSTID problem. These trees are called **weak trees** and they are essentially the minimum spanning trees of G with respect to the scenario set. More precisely, a spanning tree T is weak if there exists at least one scenario under which T is a minimum spanning tree of G. The following result relates the solutions of the RSTID problem to the weak trees of G:

Lemma 5.1 ([20]). *A robust spanning tree is a weak tree.*

Edges appearing on weak trees are called **weak edges** and are characterized by the following property.

Lemma 5.2 ([20]). *An edge e is weak if and only if there exists a minimum spanning tree using e when its cost is at the lowest bound and the costs of the remaining edges are at their highest bounds.*

Note that this property, together with Lemma 5.1, can be used to reduce the size (i.e., number of edges) of G before attempting to solve RSTID.[5] However, this does not change the complexity of the problem, as can be seen from the following result:

Theorem 5.3. *The RSTID problem remains NP-complete on complete graphs.*

Proof. Let $G = (V, E)$ be an arbitrary undirected and connected graph. We show how to construct a RSTID problem $\mathcal{P}$ over a complete graph, whose robust trees are the robust trees of the ZO-RSTID associated with G. Consider the complete graph $G' = (V, E \cup E')$ and assign a cost interval $[0, 1]$ to all edges in E and a cost interval $[1 + \epsilon, 1 + 2\epsilon]$ ($\epsilon > 0$) for all edges in E'. By construction, the intersection of any minimum spanning tree MST_s of G' and E' is empty, regardless of the scenario s. That is, all minimum spanning trees of G' (with respect to the set of scenarios) take edges only from E. By Lemma 5.2, a robust tree for $\mathcal{P}$ only takes edges from E. As a consequence, the robust trees of $\mathcal{P}$ are the same as the robust trees for the ZO-RSTID problem associated with G. The result follows. □

[5] The MIP approach proposed by [20] achieves considerable speedup by exploiting these properties to preprocess the graph. An enhanced algorithm for detecting non-weak edges was proposed in [2] and made it possible to use preprocessing for subproblems, which further reduced the time to find a solution.

6 Conclusion

This paper reconsidered the RSTID problem, which has attracted increasing attention in recent years, due to its particular importance in telecommunication networks. It closed the conjecture of Kouvelis and Yu [9], by showing that the RSTID problem is a least as hard as the Central Tree problem [5, 19] and therefore NP-Complete. Furthermore, it showed that the RSTID problem remains hard even on complete graphs, although the Central Tree problem admits polynomial time solutions on such graphs [5]. These results shed a new light on the algorithms that have been proposed for the RSTID problem [20, 2, 15]. Indeed, these algorithms have focused almost exclusively on the cost structure and have ignored the topological properties of the graph. Our results show that both aspects contribute to the complexity of the problem and, as a consequence, they suggest new directions to improve these algorithms and to evaluate them experimentally.

Acknowledgements

Many thanks to Shabbir Ahmed, Ismael de Farias, Michel Goemans, George Nemhauser, and Eli Upfal for interesting discussions on robust spanning trees. This work is partially supported by by NSF ITR Awards DMI-0121495 and ACI-0121497.

References

[1] A. Amoia and G. Cottafava. Invariance properties of central trees. In *IEEE Transactions on Circuit Theory*, volume CT-18, pages 465–467, 1971.

[2] I. Aron and P. Van Hentenryck. A constraint satisfaction approach to the robust spanning tree problem with interval data. In *Proc. 18th Conf. UAI*, Edmonton, Canada, August 1-4 2002.

[3] D. Bertsimas and M. Sim. Robust discrete optimization. Working Paper, Operations Research Center, MIT, January 2002.

[4] S. Bezrukov, F. Kaderali, and W. Poguntke. On central spanning trees of a graph. In *Lecture Notes in Computer Science*, volume 1120, pages 53–58. Springer Verlag, 1996.

[5] N. Deo. A central tree. In *IEEE Transactions on Circuit Theory*, volume CT-13, pages 439–440, December 1966. Correspondence.

[6] S. L. Hakimi. On trees of a graph and their generation. In *J. Franklin Institute*, volume 272, pages 347–359, November 1961.

[7] F. Kaderali. A counterexample to the algorithm of Amoia and Cottafava for finding central trees. Technical Report FB 19, TH Darmstadt, June 1973. preprint.

[8] G. Kishi and Y. Kajitani. Maximally distant trees and principal partition of a linear graph. In *IEEE Transactions on Circuit Theory*, volume CT-16, pages 323–330, August 1969.

[9] P. Kouvelis and G. Yu. *Robust Discrete Optimization and Its Applications*. Kluwer Academic Publishers, 1997.

[10] G. Kozina and V. Perepelista. Interval spanning tree problem: Solvability and computational complexity. *Interval Computing*, 1:42–50, 1994.

[11] Y. Y. Lee. Use of compound matrices in topological analysis. Master's thesis, Kansas State University, Manhatten, KS, 1968.

[12] N. Malik. On deo's central tree concept. In *IEEE Transactions on Circuit Theory*, volume CT-15, pages 283–284, September 1968. Correspondence.

[13] N. Malik and Y. Y. Lee. Finding trees and signed tree pairs by the compound method. In *Tenth Midwest Symposium on Circuit Theory*, pages VI–5–1–VI–5–11, 1967.

[14] W. Mayeda and S. Seshu. Generation of trees without duplications. In *IEEE Transactions on Circuit Theory*, volume CT-12, pages 181–185, June 1965.

[15] R. Montemanni, L. Gambardella, A. Rizzoli, and A. Donati. An branch and bound algorithm for the robust shortest path problem with interval data. Technical report, Istituto Dalle Molle di Studi sull'Intelligenza 20 Artificiale., 2002.

[16] S. Shinoda and T. Kawamoto. Central trees and critical sets. In D.E. Kirk, editor, *Proc. 14th Asilomar Conference on Circuit, Systems and Comp.*, volume 108, pages 183–187, Pacific Grove, California, 1980.

[17] S. Shinoda and T. Kawamoto. On central trees of a graph. In *Lecture Notes in Computer Science*, volume 108, pages 137–151. Springer Verlag, 1981.

[18] S. Shinoda and K. Saishu. Conditions for an incidence set to be a central tree. Technical Report CAS80-6, Inst. Elec. Comm. Eng., Japan, 1980.

[19] K. Thulasiraman and M.N.S Swamy. *Graphs: Theory And Algorithms*. John Wiley & Sons, Inc., 1992. ISBN 0-471-51356-3.

[20] H. Yaman, O. Karasan, and M. Pinar. The robust spanning tree problem with interval data. *Operations Research Letters*, 29:31–40, 2001.