Steiner tree problems with profits

Alysson M. Costa, Jean-François Cordeau and Gilbert Laporte

Centre for Research in Transportation and

Canada Research Chair in Distribution Management, HEC Montréal

3000 chemin de la Côte-Sainte-Catherine, Montréal, Canada H3T 2A7

{alysson, cordeau, gilbert}@crt.umontreal.ca

April 6, 2005

Abstract

This is a survey of the Steiner tree problem with profits, a variation of the classical Steiner problem where, besides the costs associated with edges, there are also revenues associated with vertices. The relationships between these costs and revenues are taken into consideration when deciding which vertices should be spanned by the solution tree. The survey contains a classification of the problems falling within this category and an overview of the methods developed to solve them. It also lists several graph preprocessing procedures and analyzes their validity for the different variants of the problem. Finally, a brief comparison is made between the profit versions of the Steiner tree problem and of the travelling salesman problem.

Keywords: Prize collecting, network design, Steiner tree problem, quota, budget, reduction tests.

1 Introduction

Let $G = (V, E)$ be a graph with vertex set $V = \{1, \dots, n\}$ and edge set $E = \{e = (i, j) : i, j \in V, i < j\}$, where each edge $e \in E$ has an associated cost c_e. The problem of determining a minimum cost network spanning all vertices V of G is known as the *Minimum Spanning Tree Problem* (MSTP). The *Steiner Tree Problem* (STP) is a very similar problem arising when some vertices need not be spanned, but may be used if their inclusion reduces the solution cost (see, for instance, Hwang, Richards and Winter, 1992). Unlike the MSTP, the STP is NP-hard.

Steiner Tree Problems with Profits (STPP) are an important variation of the classical STP.

In the STPP, in addition to the costs associated with the edges, there are also revenues r_i, associated with the vertices i of the graph. The goal is to determine a subtree minimizing cost or maximizing revenue (or profit), subject to constraints. The exact criteria guiding the optimization vary for the different versions of the STPP. In some particular problems, both cost and revenue are combined in the objective function, while in others, limits on either the cost or the revenue will appear as constraints. Therefore, four basic criteria can be used to distinguish between the variants: the cost of edges, the revenue of vertices, the minimum collected revenue (quota), and the maximum allowed cost (budget). Depending on the problem, some of these criteria can be ignored or combined.

Our aim is to review the main applications of the STPP as well as the models and algorithms proposed for its resolution. The remainder of this paper is organized as follows. Applications are described in Section 2, followed, in Section 3, by a more detailed description of the different problems and by an overview of the existing methods. An important component of modern algorithms is the application of reduction tests to eliminate some vertices and edges that are absent from at least one optimal solution. In Section 4, these preprocessing tests are reviewed and their validity is analyzed for the different variants of the STPP. Then, in Section 5, a brief comparison between the STPP and the travelling salesman problem with profits is reported. The paper ends with some conclusions in Section 6.

2 Applications

Several network design problems can be appropriately modelled as an STPP. One traditional application is the design of telecommunications local access networks. Here, the goal is to create or expand a local access network to offer service to new customers. Each new consumer represents a potential revenue for the company but there also exists a connection cost associated with the network to be constructed. The problem is clearly modelled in a graph, where customers are represented by vertices and physical links between them are represented by edges. The cost associated with an edge is the cost of laying down the optical fiber, while the revenue of a vertex is derived from customers. Ljubić *et al.* (2004) have reported a similar application in the planning of heating networks, where customers have an estimated heat demand and the street network provides the underlying graph where pipes can be laid down.

The STPP may also appear as a subproblem of more general problems. Engevall, Göthe-Lundgren and Värbrand (1998) have solved an STPP as a subproblem within a constraint generation algorithm for the problem of cost allocation in minimum 1-trees (in a graph G, a 1-tree is a spanning tree on $G = (V\backslash\{1\}, E)$ connected to vertex 1 by two edges). Chawla *et al.* (2003) have used an STPP as part of a mechanism for the extended multicasting game in which vertices and edges are selfish agents trying to maximize their own profit. This mechanism obtains a fraction of the market profit or demonstrates that no profitable solution exists. Finally, Friedman and Parkes (2003) have suggested the use of an STPP as part of an online mechanism for the dynamic pricing of WiFi service in an internet cafe. The idea is to recompute the optimal STPP tree at each period, based on the current clients, and to maximize the profit equal to the revenues on the connections, minus the cost of maintaining the links.

3 Classification of the Problems

Several variants of the STPP can be derived depending on how the costs and revenues are considered in the optimization problem. In the sequel, we present the most important variants.

3.1 Node-Weighted and Prize-Collecting Steiner Tree Problems

In the node-weighted and prize-collecting STP, both costs and revenues are present in the objective function. The idea is not to obtain a non-dominated solution, as would be the case in a multi-objective optimization, but to optimize a linear combination of cost and revenue.

Several mathematical formulations have been developed for this case. The following formulation makes use of a generalization of the classical subtour elimination constraints (SEC) introduced by Dantzig, Fulkerson and Johnson (1954) for the *Travelling Salesman Problem* (TSP). Let x_{ij} be real-valued variables associated with edges $(i,j) \in E$ and y_i be binary variables associated with vertices $i \in V$. Variable y_i is equal to 1 if vertex i belongs to the solution, and 0 otherwise. For $S \subseteq V$, define $E(S)$ as the set of edges with both endpoints in S. Finally, let T be the set of vertices that must be present in a solution. These vertices are called *mandatory* or *terminal* vertices. With these definitions, the problem can be written as:

$$\text{Minimize} \quad \sum_{(i,j)\in E} c_{ij}x_{ij} + \sum_{i\in V} r_i(1-y_i) \tag{1}$$

subject to

$$\begin{aligned}
\sum_{(i,j)\in E} x_{ij} &= \sum_{i\in V} y_i - 1 && && (2)\\
\sum_{(i,j)\in E(S)} x_{ij} &\leq \sum_{i\in S\setminus\{k\}} y_i && k \in S \subseteq V,\ |S| \geq 2, && (3)\\
y_i &= 1 && i \in T, && (4)\\
0 \leq x_{ij} &\leq 1 && (i,j) \in E, && (5)\\
y_i &\in \{0,1\} && i \in V. && (6)
\end{aligned}$$

The objective function minimizes the cost of the opened edges and the lost revenue of the unreached vertices. Constraint (2) forces the presence of $s-1$ edges, where s is the number of spanned vertices. Constraints (3) are the generalized subtour elimination constraints. These are stronger than the classical SEC used in the TSP formulation in which the right-hand side is $|S|-1$. Note that for subsets S formed only by spanned vertices, constraints (3) reduce to the classical SEC. Constraints (4) force the presence of the vertices of T in the solution.

Model (1)–(6) was used by Lucena and Resende (2004) (without constraints (4)) and follows from an extended formulation for the Steiner tree problem proposed by Lucena (1991), Goemans (1994) and Margot, Prodon and Liebling (1994). It is interesting to observe that the x variables associated with edges need not to be declared as integer. Indeed, once the y variables are equal to 0 or 1, constraints (2) and (3) define the convex hull of the characteristic vectors of the spanning trees on the subgraph of G formed by the selected vertices (see Margot, Prodon and Liebling, 1994).

Goemans (1994) has observed that in the presence of mandatory vertices, some redundant generalized subtour elimination constraints can be eliminated. Let $N = V \backslash T$ be the set of non-mandatory vertices. Constraints (3) then reduce to

$$\sum_{(i,j)\in E(S)} x_{ij} \leq \sum_{i\in S\cap N} y_i + |S\cap T| - 1 \qquad S\subseteq V, S\cap T\neq\emptyset, |S|\geq 2 \tag{7}$$

and

$$\sum_{(i,j)\in E(S)} x_{ij} \leq \sum_{i\in S\backslash\{k\}} y_i \qquad k\in S\subseteq N, |S|\geq 2. \tag{8}$$

The objective function (1) can be equivalently written as:

$$\text{Maximize} \sum_{i\in V} r_i y_i - \sum_{(i,j)\in E} c_{ij}x_{ij}, \tag{9}$$

i.e., the goal is to maximize the profit, equal to the revenue of the spanned vertices minus the cost of the selected edges. When the objective function (1) is used, the problem is sometimes referred to as the Goemans-Williamson minimization problem (Goemans and Williamson, 1995). When the "real profit" of objective function (9) is used, the problem is known as the Net-Worth maximization problem. While these are equivalent from an optimization point of view, they are not equivalent for the computation of the worst case performance ratio of approximation algorithms. Later in this section, we present some approximation algorithms for the Goemans-Williamson problem. However, no such algorithm exists for the Net-Worth version. Feigenbaum, Papadimitriou and Shenker (2001) have indeed proved that it is NP-hard to derive an approximation to a constant factor for this problem.

Both the Goemans-Williamson and the Net-Worth problems are known under the more general names of *Node-Weighted Steiner Tree Problem* (NWSTP) and *Prize-Collecting Steiner Tree Problem* (PCSTP). The difference between the NWSTP and the PCSTP is that in the second case, the set of mandatory vertices T is empty. In practice, these mandatory vertices often correspond to facilities (like a telecommunications central) that must be present in the network in order to supply the service.

The NWSTP was first proposed by Segev (1987) who developed three simple heuristic solution methods for a special case of the problem in which a single vertex is mandatory. This author also described two different Lagrangian relaxation bounding procedures. Both the heuristics and the bounding procedures were tested in complete graphs of up to 40 vertices.

Bienstock *et al.* (1993) were the first to propose a solution method for the PCSTP. They worked primarily with the *TSP with Profits* (TSPP), a version of the TSP in which a vertex has a revenue but need not be visited (see Section 5). They developed a heuristic based on Christofides's heuristic for the TSP (Christofides, 1976). This method was later adapted to the PCSTP (with the Goemans and Williamson objective function), yielding a heuristic with a worst-case performance ratio of 3. Goemans and Williamson (1995) have later improved this worst-case performance with a $(2-1/(|V|-1))$-approximation algorithm. The authors deal with the rooted version of the problem (where one vertex is a root and must be present in the final tree, as first proposed by Segev). Their procedure is an adaptation of a more general method which can be applied to a large number of graph optimization problems called *Constrained forest problems* (including the shortest path problem and the *Generalized Steiner Tree Problem*, a variant of the STP in which clusters of vertices are defined and the tree must include at least one vertex from each cluster). The method maintains a set of components. Initially, each vertex is a component. Each component has an associated surplus (initially, the vertex revenue) and is declared active if its surplus is positive. The edges have associated deficits, initially equal to their costs. The deficits and surpluses are dynamically reduced until the deficit of an edge connecting two active components reaches zero, in which case the two components are merged together, or until an active component becomes inactive. The process ends when no active components are left (see Johnson, Minkoff and Phillips, 2000). As suggested by the authors, the method can be viewed as a greedy algorithm in which, at each iteration, a minimum reduced cost edge is selected. Through the execution, the reduced costs are updated, giving rise to a primal-dual structure for the algorithm. In a final phase, unnecessary edges are pruned.

The Goemans and Williamson algorithm was recently revisited by Cole *et al.* (2001) who proposed a faster implementation. Their algorithm runs in $O(k(|V|+|E|)\log^2|V|)$ time in comparison to $O(|V|^2\log|V|)$ for the original algorithm, at the expense of an additive degradation of $|V|^{-k}$ in the approximation factor, for any constant k. Johnson, Minkoff and Phillips (2000) have worked specifically with the STPP and proposed a new strategy for the pruning phase. The authors have compared the results in graphs of up to 25,600 vertices and their tests have shown a slightly better performance of the new pruning strategy, which has the additional advantage of being conceptually much simpler. Johnson, Minkoff and Phillips have also proposed modifications to the core of the algorithm which leads to a $(2 - 1/n)$-approximation for the unrooted problem (PCSTP). The authors also extended the methodology to the quota and budget versions of the STPP, as will be seen in Subsections 3.2 and 3.3.

The most recent works on the PCSTP and NWSTP have explored two main streams of research: a) the obtention of good and reasonably fast methods for real-size problems and b) the development of exact algorithms and lower bounding procedures. Regarding the first stream, the tendency has been to use modern heuristic methods. A good example is the work by Canuto, Resende and Ribeiro (2001) who have developed and tested a multi-start local search method on instances with up to 1000 vertices and 25,000 edges. The different restarts are conducted in a GRASP fashion: the initial solutions are given by the Goemans and Williamson algorithm feeded with perturbed vertex revenues. The final solution of each restart goes through a path-relinking procedure, with one of the solutions maintained in a *pool* of the best solutions found so far. Finally, in the last phase of the algorithm, the best known solution is used within a variable neighborhood search procedure (Mladenović and Hansen, 1997). Optimal solutions were obtained for most of the tested graphs.

Klau *et al.* (2004) have reached similar results but with less computational time. Their approach uses a memetic algorithm which works with a population of trees. Each tree goes through an exact algorithm to eliminate non-optimal branches. The entire procedure is followed by an exact method. The idea is to use the memetic algorithm to eliminate edges that are probably not present in the optimal solution and then apply the exact method, based on a separation procedure, on the reduced instance.

Cunha *et al.* (2003) formulate the PCSTP as a restricted minimum forest problem. The authors then use a Lagrangian relaxation which dualizes some constraints only when they first become violated. A constraint is dropped from the objective function when the associated multiplier is zero. A heuristic procedure based on the work of Johnson, Minkoff and Phillips (2000) is used to obtain upper bounds. Results obtained on instances of up to 400 vertices and 1507 edges are rather modest, an outcome that the authors attribute to probable cycling in the procedure.

Besides heuristics, a fair amount of research has been conducted on exact algorithms and lower bounding procedures. Engevall, Göthe-Lundgren and Värbrand (1998) have proposed a

Lagrangian-based lower bounding procedure using a formulation based on a graph transformation which enables the NWSTP to be treated as an MSTP. The resulting bounds were stronger than those obtained by Segev (1987).

Lucena and Resende (2004) have also developed a new lower bounding method. The authors have used formulation (1)-(6), and proposed a cutting-plane algorithm based on a separation procedure that identifies violated subtour elimination constraints. The separation procedure has been refined to introduce the idea of orthogonal cuts (cuts generated from subtour elimination constraints with no common vertices). A total of 114 instances with up to 1000 vertices and 25,000 edges were solved and the bounds were proved to be optimal in 99 of these.

Recently, Ljubić *et al.* (2004) have improved these results with a branch-and-cut algorithm, applied to a formulation that uses connectivity inequalities. These inequalities are generated efficiently via a separation procedure based on a maximum flow algorithm. The authors have solved instances of up to 2500 vertices and 62,500 edges, including the Lucena and Resende instances, for which similar results were obtained with a decrease of two orders of magnitude in the computation time.

3.2 The quota STPP

In the quota version of the STPP the goal is to minimize the costs while guaranteeing a minimum value for the collected revenues:

$$\text{Minimize} \sum_{(i,j)\in E} c_{ij}x_{ij} \tag{10}$$

subject to (2)–(6) and

$$\sum_{i\in V} r_i y_i \geq Q, \tag{11}$$

where Q is the quota, i.e., the minimum acceptable collected revenue.

The quota version of the STPP has received almost no attention in the literature. To our knowledge, this problem is mentioned in only one article: Johnson, Minkoff and Phillips (2000) have observed that the quota STPP is a generalization of the k-MSTP, in which one tries to determine a minimum cost spanning tree containing at least k vertices (note that if the revenues of all vertices are set to 1 and the quota is set to k, then the PCSTP becomes a k-MSTP). The authors use this fact to obtain a 2.5-approximation algorithm for the resulting quota STPP, based on an algorithm for the unrooted k-MSTP (Arya and Ramesh, 1998). The resulting procedure is, however, not likely to be efficient in practice.

A more practical algorithm has also been developed. The idea is to multiply the vertex revenues by a factor α and then use the Goemans and Williamson algorithm on the modified instance. The larger the value of α, the more revenues should be collected (since these will easily counterbalance the costs). The value of α is increased until a solution with the minimum quota has been reached. If the total collected revenue exceeds the quota, then one can try to reduce α and still find a total revenue that will satisfy (11). The authors use binary search to find the minimum value α yielding an acceptable quota. By varying α, they also construct tradeoff curves between the total cost and total revenue.

3.3 The budget STPP

The budget STPP is the counterpart of the quota problem. Here, one wants to maximize the total collected revenues under the restriction that the total cost is limited by a budget:

$$\text{Maximize} \sum_{i \in V} r_i y_i \tag{12}$$

subject to (2)–(6) and

$$\sum_{(i,j) \in E} c_{ij} x_{ij} \leq B, \tag{13}$$

where B is the allowed budget.

Almost no research has been devoted to the budget STPP. Johnson, Minkoff and Phillips (2000) have proved that by repeatedly running a 3-approximation algorithm for the quota problem, one can find a $(5+\epsilon)$-approximation algorithm for the unrooted budget STPP, where $\epsilon > 0$. Again, this approximation algorithm is not likely to be useful in practice. For a more efficient procedure, the authors propose applying the same strategy as for the quota STPP, i.e., multiplying the costs by a factor α (this time, in order to try to find a solution that satisfies the budget constraint) and applying a binary search procedure on the values of α.

3.4 The fractional STPP

A non-linear version of the STPP maximizes the revenue-to-cost ratio. In this version of the problem, the value of a solution associated with a solution tree $T = (V_T, E_T)$, where $V_T \subseteq V$ is the set of selected vertices and $E_T \subseteq E$ is the set of selected edges, is given by

$$v(T) = \frac{\sum_{v \in V_T} r_v}{c_0 + \sum_{e \in E_T} c_e}, \tag{14}$$

where c_0 is a fixed cost, and the goal is to determine a tree T maximizing $v(T)$. The fractional STPP enables the consideration of a fixed cost for the construction of the network.

Klau *et al.* (2003) have dealt with a variation of this problem where the underlying graph is already a tree. The goal is, therefore, to identify an optimal subtree containing the root. The authors have developed three algorithms based on a parametric formulation: a binary search, a Newton method, and an algorithm based on Megiddo's parametric search (Megiddo, 1979). Tests were conducted on trees of up to 10,000 vertices and showed that the Newton method had a better performance. One could probably consider adding budget or quota constraints to this problem. However, to our knowledge, not even the basic version of the fractional STPP on graphs has been the subject of any research.

4 Reduction Tests

When solving the Steiner tree problem and its variants, several authors make use of reduction tests. These tests attempt to find vertices and edges that are absent from at least one optimal solution, or to find sets of vertices and edges that can be treated as a single vertex. Preprocessing procedures often iterate between one or more different tests until no more reductions can be made.

The field of graph preprocessing for Steiner tree problems is rooted in the work of Beasley (1984) and Balakrishnan and Patel (1987). A classical set of tests was established by Duin and Volgenant (1989). With the appearance of the profit versions of the problem, some of these tests were promptly adapted to the new problems (Duin and Volgenant, 1987) and combinations of these started being used as a preprocessing phase in many solution procedures. These preprocessing phases have become an important part of the more recent algorithms. Since they can significantly reduce the size of the graph that must effectively be solved, they enable the resolution of larger instances. Ljubić *et al.* (2004) have studied the efficiency of a few simple reduction tests. In about 100 seconds of computational time, their largest instances of 2500 vertices and 62,500 edges could be slightly reduced in terms of the number vertices, while the number of their edges was more than halved. In one extreme case, the final number of vertices and edges after preprocessing were 861 and 3881, a reduction of 65% and 94%, respectively.

4.1 Description of the reduction tests

As part of this survey on the STPP, we present an overview of the most common reduction tests. These tests are adapted from the classical STP to the PCSTP and to the NWSTP. Their validity is not always clear for other variants of the STPP, like the quota or budget versions, and is therefore analyzed in this section. For all tests presented in the sequel, we assume all costs c_{ij} and revenues r_i to be non-negative. Moreover, note that one can eliminate any edge (i, j) with cost $c_{ij} = 0$ by merging vertices i and j into a single vertex.

Test 1 (Degree one test): *(a) A vertex i that is only adjacent to one vertex j can be eliminated if $r_i < c_{ij}$. (b) If $r_i > c_{ij}$ and $r_j > c_{ij}$, vertices i and j are merged into a single vertex v of revenue $r_v = r_j + r_i - c_{ij}$.*

Test 2 (Degree two test): *A vertex i that is only adjacent to vertices j and k cannot have degree 1 in the optimal solution if $r_i < min\{c_{ij}, c_{ik}\}$. In this case, the vertex can be eliminated and an edge (j, k) of cost $c_{jk} = c_{ij} + c_{ik} - r_i$ is created or, if edge (j, k) exists, its cost is redefined as $min\{c_{jk}, c_{ij} + c_{ik} - r_i\}$.*

Test 3 (General degree test): *A generalization of tests 1 and 2 can be applied to vertices with degree superior to two. Consider a vertex i with $r_i = 0$ and let $Ad(i) = \{v_1, v_2, \ldots, v_n\}$ be the set of vertices adjacent to it. For a subset of vertices $V' \subseteq V$, let MST(V') be the cost of the minimum spanning tree problem in the complete graph formed by the nodes in V', where the cost of each edge (v_i, v_j) is given by the shortest path distance between v_i and v_j in G. We know that if*

$$MST(V') \leq \sum_{v \in V'} c_{iv}, \qquad \forall V' \subseteq Ad(i), \quad |V'| \geq 3, \tag{15}$$

then the degree of i in the optimal PCSTP solution must be zero or two. Therefore, vertex i can be eliminated and each pair of edges $(v_i, i), (i, v_j)$ can be replaced by an edge (v_i, v_j) of cost $c_{v_i v_j} = c_{v_i i} + c_{i v_j}$ or, if this edge already exists, of cost $c_{v_i v_j} = min\{c_{v_i v_j}, c_{v_i i} + c_{i v_j}\}$.

Test 4 (Minimum adjacency or V\K reduction test): *Adjacent vertices i and j can be merged into a single vertex v if c_{ij} is the smallest cost of all edges incident to vertex i (or j), and c_{ij} is smaller than each revenue r_i and r_j. The revenue r_v of the merged vertex is equal to $r_i + r_j - c_{ij}$.*

Test 5 (Shortest-path or least-cost test): *An edge (i, j) can be eliminated if there exists a shortest path between i and j (considering the costs c_{ij}) not containing (i, j).*

Test 6 (Nearest vertex test): *In the NWSTP, let T be the set of mandatory vertices and d_{ij} be the shortest path distance between vertices i and j, considering the edge costs. Suppose $k \in T$ and (i, k) is a minimum-weighted edge incident to k, i.e., $c_{ik} = min\{c_{jk} | j \in V, j \neq k\}$. If a vertex $l \in T \backslash \{k\}$ exists with*

$$d_{li} + c_{ik} \leq min\{c_{jk} | j \in V, j \neq i, k\},$$

then edge (i, k) is part of an optimal solution and vertices k and i can be merged.

Recently some new tests have been proposed by Uchoa (2005) who has revisited the tests for the classical Steiner problem and, in particular, two tests using what Duin and Volgenant have called *bottleneck distances*. Let $\mathcal{P}(i,j)$ be the set of all paths connecting vertices i and j. The bottleneck distance between two vertices i and j is defined as

$$B(i,j) = \min\{SD(P) | P \in \mathcal{P}(i,j)\}, \tag{16}$$

where $SD(P)$ is the *Steiner distance*, i.e., the maximal distance between two terminal vertices in path P. Uchoa proposes a counterpart of the bottleneck distances for problems with revenues. Consider $P \in \mathcal{P}(i,j)$, and let $e(P)$ be the set of edges $(k,\ell) \in E$ such that k and ℓ are consecutive in P, and $v(P)$ be the set of vertices $v \in V$ appearing in P. Also let S_{xy} be the subpath of P between x and y, with $x, y \in v(P)$. Then define the Steiner distance associated with this subpath as

$$SD(S_{xy}) = \sum_{e \in e(S_{xy})} c_e - \sum_{i \in v(S_{xy}) \setminus \{x,y\}} r_i, \tag{17}$$

and the Steiner distance associated with the whole path as

$$SD(P) = \max_{x,y \in v(P)} SD(S_{xy}). \tag{18}$$

This new definition of the Steiner distance can be used in the bottleneck distance formula (16), and the tests originally proposed for the Steiner Problem in the article of Duin and Volgenant (1989) become valid for the PCSTP. In these tests, reproduced below, $B(i,j)^{-e}$ is the bottleneck distance without passing through a given edge e, and is defined as:

$$B(i,j)^{-e} = \min\{SD(P) | P \in \mathcal{P}(i,j); e \notin e(P)\}. \tag{19}$$

Test 7 (Special distance): *Let $(i,j) \in E$. If $B(i,j)^{-(i,j)} \leq c_{ij}$, then edge (i,j) can be removed.*

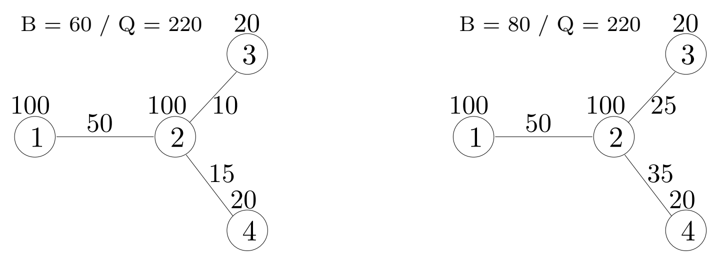
Test 8 (Non-terminal degree 3): *Let i be a vertex with $r_i = 0$, adjacent to vertices v, w and z. If*

$$min\{B(v,w) + B(v,z), B(w,v) + B(w,z), B(z,v) + B(z,w)\} \leq c_{iv} + c_{iw} + c_{iz},$$

then there exists an optimal solution in which the degree of i is 0 or 2. Therefore, i and its three adjacent edges can be replaced by the following three edges: (v,w) with cost $c_{iv} + c_{iw}$, (v,z) with cost $c_{iv} + c_{iz}$, and (w,z) with cost $c_{iw} + c_{iz}$.

4.2 Validity for the quota and budget STPP

The structure of the objective function for the PCSTP and the NWSTP is quite convenient for the development of reduction tests since it combines additive edge costs and vertex revenues. In the cases of the quota and budget STPP, however, the objective function contains only one of these terms (either revenues or costs) and therefore, all tests based on a cost analysis using both the costs and revenues become invalid. This fact is even more pronounced in the presence of the additional constraints. Indeed, the quota constraint might force the presence of a vertex that was considered unprofitable for the PCSTP or the NWSTP (since the cost to reach it was larger than its revenue), while the budget constraint might avoid the inclusion of a profitable vertex. Therefore, all tests based on a combination of costs and revenues become invalid for the quota and budget STPP. These include tests 1, 2 and 4. In Figure 1, the representative case of degree-one test (test 1 in the list) is analyzed through two simple examples.



(a) (b)

Figure 1: Invalidity of test 1 for the budget and quota STPP.

In Figure 1, the values on the vertices represent their revenues, while those on the edges represent their costs. Consider the application of test 1(a) to the graph of Figure 1a. Vertices 3 and 4 should be merged with vertex 2. However, in the optimal solution of the budget STPP, vertex 2 is present along with vertices 1 and 3, but vertex 4 is not, due to the budget constraint. In the quota STPP, vertex 4 is not present either, since its presence would increase the objective function and the presence of vertices 1, 2 and 3 is sufficient to satisfy the quota constraint. For the case of test 1(b), consider the graph of Figure 1b. According to the test, vertices 3 and 4 should be deleted from the graph. However, vertex 3 belongs to the optimal solution of the budget STPP along with vertices 1 and 2 since its addition increases the objective function value and does not violate the budget constraint. For the quota STPP, vertex 3 is also present in the final solution, along with vertices 1 and 2, since it is necessary to satisfy the quota constraint.

Like tests 1, 2 and 4, tests 7 and 8 make use of calculations mixing revenues and edge costs (see the definition of the special distance for one subpath in (17)). They are therefore invalid for the quota and budget STPP, since the presence of either budget or quota constraints, or the absence of either costs or revenues in the objective function might invalidate the conclusions. Even for the case of the PCSTP and NWSTP, Uchoa (2005) reports that the computation of the new bottleneck distances is NP-hard. However, the author uses a heuristic method to calculate these distances and implements a preprocessing package including classical preprocessing tests as well as tests 7 and 8. A comparison of the results with those of Ljubić *et al.* (2004) has revealed that significant extra reductions were obtained with the additional tests, especially for instances containing a large number of mandatory vertices.

4.3 Validity for the fractional STPP

Although no author has considered the fractional STPP in general graphs, it is interesting to analyze the validity of the reduction tests for this problem. One can easily conclude that the presence of the quotient in the objective function forbids any local consideration using both the costs and revenues. Therefore, as for the quota and budget STPP, tests 1, 2 and 4 become invalid. Tests 3, 5 and 6 make considerations based on the edge costs, acting only on the denominator of (14) and are, therefore, valid. Like tests 1, 2 and 4, tests 7 and 8 are invalid for the fractional STPP since they are based on local considerations using both the costs and the revenues.

4.4 New tests for the budget STPP

As a last note on the reduction tests, one can clearly see that the introduction of the quota and budget constraints invalidates many of the tests developed for the NWSTP and for the PCSTP. However, these constraints may also represent an opportunity. Indeed, some new reduction tests specifically adapted to the quota and budget versions of the problem can be derived because of the introduction of the new constraints. The two simple tests presented below for the case of the budget STPP are illustrative examples.

Test 9 (Distance to terminal): *In the budget STPP with a mandatory vertex set T, a vertex $i \notin T$ can be eliminated if the cost of the shortest path linking i and any vertex in T is larger than B.*

Test 10 (Minimum distance to terminal tree): *Let T be the set of mandatory nodes and $STP_G(T)$ be the cost of the STP in G. A vertex $i \notin T$ with a single incident edge (i, k) can be eliminated if $STP_G(T) + c_{ik} > B$.*

4.5 Summary of the reduction tests

Table 1 presents a summary of the reduction tests presented in this section. For each test, it is possible to see its applicability to the different problems, the original reference in which the test is described and the articles where it has been used as part of a preprocessing algorithm.

Test	NWSTP/ PCSTP	Quota STPP	Budget STPP	Fractional STPP	Proposed by	Used in
1	yes	no	no	no	Duin and Volgenant (1987)	Cunha *et al.* (2003), Lucena and Resende (2004) Klau *et al.* (2004), Lujbic *et al.* (2004)
2	yes	no	no	no	Duin and Volgenant (1987)	Cunha *et al.* (2003), Lucena and Resende (2004), Klau *et al.* (2004), Lujbic *et al.* (2004)
3	yes	yes	yes	yes	Duin and Volgenant (1989) [a]	Cunha *et al.* (2003) Lucena and Resende (2004), Klau *et al.* (2004), Lujbic *et al.* (2004)
4	yes	no	no	no	Duin and Volgenant (1987)	Cunha *et al.* (2003), Klau *et al.* (2004), Lujbic *et al.* (2004)
5	yes	yes	yes	yes	Duin and Volgenant (1987)	Cunha *et al.* (2003), Lucena and Resende (2004), Klau *et al.* (2004), Lujbic *et al.* (2004)
6[b]	yes	yes	yes	yes	Duin and Volgenant (1989)	–
7	yes	no	no	no	Uchoa (2005)[c]	–
8	yes	no	no	no	Uchoa (2005)[c]	–
9	no	no	yes[d]	no	this article	–
10	no	no	yes[d]	no	this article	–

[a] Proposed for the classical STP.
[b] The problem must contain at least two mandatory vertices.
[c] Adapted from Duin and Volgenant (1989).
[d] For the version with mandatory vertices.

Table 1: Summary of the reduction tests.

5 A comparison with the Travelling Salesman Problem with Profits

The *Travelling Salesman Problem with Profits* (TSPP) is a variation of the classical TSP where revenues are associated with vertices. It has received considerable attention as witnessed by the review articles of Feillet, Dejax and Gendreau (2005) and of Laporte and Rodríguez-Martín (2005). Several versions of the TSPP have been proposed, as in the case of the STPP. Indeed, the similarities between the two problems are numerous. Concerning the different versions, the TSPP also exists in a variation where both revenues and costs are considered in the objective function (the *Profitable Tour Problem* (PTP)), and there exist variants where limits on either the revenues or the costs are imposed as constraints: these are the *Prize-Collecting Travelling Salesman Problem* (PCTSP) and the *Orienteering Problem* (OP), respectively. Concerning the solution methodologies for the two families of problems, there are also some common features. As seen in Section 3, the approximation algorithm of Bienstock, Goemans, Simchi-Levi and Williamson (1993) for the PCSTP has been adapted from an algorithm for the PTP. Moreover, the general framework of Goemans and Williamson (1995) has been used to develop approximation algorithms both for the PCSTP and for the PTP.

Due to the closeness of the problems, we found that it could be insightful to present a comparison between the methods proposed both for the STPP and for the TSPP. In Table 2, we present a summary of these methods. For the STPP, we use Section 3 of this article, while for the TSPP, we use the articles surveyed in Feillet, Dejax and Gendreau (2005) and in Laporte and Rodríguez-Martín (2005).

Table 2 shows that the TSPP has received considerably more attention than the STPP. More interesting, and perhaps less justifiable, is the fact that the research efforts for both problems have been distributed differently. While for the TSPP, most research has concentrated on the OP (where the costs are introduced in the constraints of the model), for the STPP almost all research has been devoted to the NWSTP and the PCSTP.

6 Conclusions

We have studied the Steiner Tree Problem with Profits, and we have presented an overview of the available algorithms for several variants of this problem. We have also presented an overview of the existing reduction tests to eliminate or agglomerate vertices and edges of the original graph. We have observed that the research efforts to solve the STPP are concentrated on the PCSTP and NWSTP variants. Indeed, a single article mentions the quota and the budget versions of the STPP, and no author seems to have considered the fractional STPP on general graphs. While this may not come as a surprise for the fractional STPP (since it is hard to associate its objective function with a profit), this is less justifiable for the other problems. In the particular case of the budget STPP, the absence of any study is even more remarkable considering that most of the research conducted on the similar TSPP has been devoted to the OP.

The question that arises is: are the budget variants only important in the context of the TSP and not in the case of the STPP ? There are at least two indications that the answer to this question is negative. First, budget constraints seem to be relevant in practice in several design problems. It is quite natural to conceive that they are also present in the major projets associated with the expansion of telecommunications or heating networks, two important applications of the STPP. Secondly, the PCSTP and NWSTP variants may present optimal solutions with the same value but with very different practical implications. It is easy to construct an example along the lines of Johnson, Minkoff and Phillips (2000). Consider two different solutions, both with a profit of \$100. The PCSTP and the NWSTP are insensitive to the fact that this solution could have come from a situation where the sum of the costs are \$10 and the sum of the revenues are \$110, or from a network with construction costs of \$10,000 and revenues of \$10,100. The budget STPP as well as the quota STPP are more robust to this kind of situation, since only the revenues (or the costs) are present in the objective function.

Even if the budget STPP also has its own flaws (as seen in the example of Figure 1b, where the unprofitable vertex 3 is unnecessarily spanned), this version of the problem seems to deserve more attention than it has received so far in the literature. The existing knowledge on the PCSTP and NWSTP versions can certainly be useful in developing methods that deal with the particularities of this problem. An interesting study from a practical point of view is the analysis of cost *vs* revenue tradeoff curves, as suggested by Johnson, Minkoff and Phillips (2000).

Concerning the preprocessing phases, Section 4 has presented an overview of the most common reduction tests. These tests have been developed for the NWSTP and PCSTP and are in general invalid for the other variants of the problem. Since the preprocessing phase which is based on these reduction tests has become an important part of most algorithms, the need for additional research for the different versions of the STPP is, again, justified. We hope that this survey will encourage such research efforts and prove a helpful source of ideas and references.

Acknowledgments

This work was supported by the Brazilian Federal Agency for Post-Graduate Education - CAPES, under grant BEX 1097/01-6 and by the Canadian Natural Science and Engineering Research Council under grants 227837-00 and OGP0039682. This support is gratefully acknowledged.

Type	Steiner Tree Problem with Profits	Travelling Salesman Problem with Profits
Costs and profits in the objective function	**Node-Weighted STP or Prize-Collecting STP** Segev (1987): Simple heuristics and Lagrangian bounding procedures. Bienstock *et al.* (1993): Heuristic with worth case performance of 3. Goemans and Williamson (1995): $(2 - 1/(\|V\| - 1))$-approximation algorithm. Johnson, Minkoff and Phillips (2000): $(2 - 1/\|V\|)$-approximation algorithm for the unrooted problem. Cole *et al.* (2001): Faster implementation of the Goemans-Williamson algorithm with a small loss in the approximation. Canuto, Resende and Ribeiro (2001): Multi-Start with local search and VNS post-optimization phase. Cunha *et al.* (2003): Lagrangian relaxation with dynamically generated cutting planes. Lucena and Resende (2004): Lower bounding procedure based on a cutting planes algorithm. Klau *et al.* (2004): Hybrid Exact-Memetic algorithm. Ljubić *et al.* (2004): Branch and cut algorithm based on connectivity inequalities.	**Profitable Tour Problem** Mittenthal and Noon (1992): Insertion/deletion heuristic procedure. Bienstock *et al.* (1993): Heuristic with worth case performance of 2.5. Goemans and Williamson (1995): $(2 - 1/(\|V\| - 1))$-approximation algorithm. Dell'Amico, Maffioli and Sciomachen (1998): Branch-and-bound on the problem transformed into an asymmetric TSP.
Costs in the objective function, profits as constraints	**Quota STPP** Johnson, Minkoff and Phillips (2000): Theoretical 2.5-approximation algorithm and a more efficient heuristic based on the work of Goemans and Williamson (1995).	**Prize-Collecting Travelling Salesman Problem** Fischetti and Toth (1988): Branch-and-bound with additive bounding procedures. Dell'Amico, Maffioli and Sciomachen (1998): Heuristic based on Lagrangian relaxation and improvement moves. Awerbuch *et al.* (1998): Heuristic based on an approximation algorithm for the k-MST problem.

Table 2: Summary and comparison between the methods for the STPP and for the TSPP.

Type	Steiner Tree Problem with Profits	Travelling Salesman Problem with Profits
	Budget STPP	**Orienteering Travelling Salesman Problem**
Profits in the objective function, costs as constraints	Johnson, Minkoff and Phillips (2000): Theoretical 2.5-approximation algorithm and a more efficient heuristic based on the work of Goemans and Williamson (1995).	Tsiligirides (1984): Two approximate algorithms, one based on a former vehicle routing procedure and another based on Monte Carlo techniques.
		Golden, Levy and Vohra (1987): Heuristic based on the idea of center of gravity.
		Golden, Wang and Liu (1988): Heuristic based on the idea of center of gravity improved by a learning measure.
		Kataoka and Morito (1988): Branch-and-bound algorithm.
		Keller (1989): Two-phase algorithm consisting of a construction procedure followed by an improvement phase.
		Laporte and Martello (1990): Branch-and-bound algorithm.
		Ramesh and Brown (1991): Four-phase algorithm incorporating a tabu list.
		Ramesh, Yoon and Karwan (1992): Lagrangian relaxation based branch-and-bound algorithm.
		Leifer and Rosenwein (1994): Cutting planes algorithm.
		Wang, Sun, Golden and Jia (1995): Neural network algorithm.
		Chao, Golden and Wasil (1996): Construction algorithm followed by an deterministic annealing algorithm.
		Fischetti, Salazar-González and Toth (1998): Branch-and-cut algorithm.
		Gendreau, Laporte and Semet (1998a): Branch-and-cut algorithm.
		Gendreau, Laporte and Semet (1998b): Tabu search with neighborhoods based on clusters.
		Tasgetiren and Smith (2000): Genetic algorithm.

Table 2: (continued) Summary and comparison between the methods for the STPP and for the TSPP.

References

Arya, S. and Ramesh, H. (1998). A 2.5 factor approximation algorithm for the k-MST problem, *Information Processing Letters* **65**: 117–118.

Awerbuch, B., Azar, Y., Blum, A. and Vempala, S. (1998). New approximation guarantees for minimum-weight k-trees and prize-collecting salesmen, *SIAM Journal on Computing* **28**: 254–262.

Balakrishnan, A. and Patel, N. (1987). Problem reduction methods and a tree generation algorithm for the Steiner network problem, *Networks* **17**: 65–85.

Beasley, J. E. (1984). An algorithm for the Steiner problem in graphs, *Networks* **14**: 147–159.

Bienstock, D., Goemans, M. X., Simchi-Levi, D. and Williamson, D. (1993). A note on the prize collecting traveling salesman problem, *Mathematical Programming* **59**: 413–420.

Canuto, S. A., Resende, M. G. C. and Ribeiro, C. C. (2001). Local search with perturbations for the prize-collecting Steiner tree problem in graphs, *Networks* **38**: 50–58.

Chao, I.-M., Golden, B. L. and Wasil, E. A. (1996). A fast and effective heuristic for the orienteering problem, *European Journal of Operational Research* **88**: 475–489.

Chawla, S., Kitchin, D., Rajan, U., Ravi, R. and Sinha, A. (2003). Profit maximizing mechanisms for the extended multicasting game, *Proceedings of the 4th ACM Conference on Eletronic Commerce* : 190–191.

Christofides, N. (1976). Worst-case analysis of a new heuristic for the travelling salesman problem, *Technical Report 388 - Graduate School of Industrial Administration, Carnegie Mellon University* .

Cole, R., Hariharan, R., Lewenstein, M. and Porat, E. (2001). A faster implementation of the Goemans-Williamson clustering algorithm, *SODA'01: Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms* : 17–25.

Cunha, A. S., Lucena, A., Maculan, N. and Resende, M. G. C. (2003). A relax and cut algorithm for the prize collecting Steiner problem in graphs, *Proceeding of Mathematical Programming in Rio, Búzios - Brazil* : 72–78.

Dantzig, G. B., Fulkerson, D. R. and Johnson, S. M. (1954). Solution of a large-scale traveling salesman problem, *Operations Research* **2**: 393–410.

Dell'Amico, M., Maffioli, F. and Sciomachen, A. (1998). A Lagrangian heuristic for the prize-collecting travelling salesman problem, *Annals of Operations Research* **81**: 289–305.

Duin, C. W. and Volgenant, A. (1987). Some generalizations of the Steiner problem in graphs, *Networks* **17**: 353–364.

Duin, C. W. and Volgenant, A. (1989). Reduction tests for the Steiner problem in graphs, *Networks* **19**: 549–567.

Engevall, S., Göthe-Lundgren, M. and Värbrand, P. (1998). A stronger lower bound for the node weighted Steiner tree problem, *Networks* **31**: 11–17.

Feigenbaum, J., Papadimitriou, C. H. and Shenker, S. (2001). Sharing the cost of multicast transmissions, *Journal of Computer and System Sciences* **63**: 21–41.

Feillet, D., Dejax, P. and Gendreau, M. (2005). Traveling salesman problems with profits, *Transportation Science* . Forthcoming.

Fischetti, M., Salazar-González, J. J. and Toth, P. (1998). Solving the orienteering problem through branch-and-cut, *INFORMS Journal on Computing* **10**: 133–148.

Fischetti, M. and Toth, P. (1988). An additive approach for the optimal solution of the prize-collecting travelling salesman problem, *in* B. L. Golden and A. A. Assad (eds), *Vehicle Routing: Methods and Studies*, North-Holland, Amsterdam.

Friedman, E. J. and Parkes, D. C. (2003). Pricing WiFi at Starbucks: issues in online mechanism design, *Proceedings of the 4th ACM Conference on Electronic Commerce* : 240–241.

Gendreau, M., Laporte, G. and Semet, F. (1998a). A branch-and-cut algorithm for the undirected selective travelling salesman problem, *Networks* **32**: 263–273.

Gendreau, M., Laporte, G. and Semet, F. (1998b). A tabu search heuristic for the undirected selective travelling salesman problem, *European Journal of Operational Research* **106**: 539–545.

Goemans, M. X. (1994). The Steiner tree polytope and related polyhedra, *Mathematical Programming* **63**: 157–182.

Goemans, M. X. and Williamson, D. P. (1995). A general approximation technique for constrained forest problems, *SIAM Journal on Computing* **24**: 296–317.

Golden, B. L., Levy, L. and Vohra, R. (1987). The orienteering problem, *Naval Research Logistics* **34**: 307–318.

Golden, B. L., Wang, Q. and Liu, L. (1988). A multifaceted heuristic for the orienteering problem, *Naval Research Logistics* **35**: 359–366.

Hwang, F. K., Richards, D. S. and Winter, P. (1992). *The Steiner Tree Problem*, North-Holland, Amsterdan.

Johnson, D. S., Minkoff, M. and Phillips, S. (2000). The prize collecting Steiner tree problem: theory and practice, *Proceedings of the 11th ACM-SIAM Symposium on Discrete Algorithms, San Francisco* : 760–769.

Kataoka, S. and Morito, S. (1988). An algorithm for the single constraint maximum collection problem, *Journal of the Operations Research Society of Japan* **31**: 515–530.

Keller, C. P. (1989). Algorithms to solve the orienteering problem: a comparison, *European Journal of Operational Research* **41**: 224–231.

Klau, G. W., Ljubić, I., Moser, A., Mutzel, P., Neuner, P., Pferschy, U., Raidl, G. and Weiskircher, R. (2004). Combining a memetic algorithm with integer programming to solve the prize-collecting Steiner tree problem, *Springer Lecture Notes in Computer Science* **3102**: 1304–1315.

Klau, G. W., Ljubić, I., Mutzel, P., Pferschy, U. and Weiskircher, R. (2003). The fractional prize-collecting Steiner tree problem on trees, *Technical Report TR-186-1-03-01, Institute of Computer Graphics and Algorithms, Vienna University of Technology* .

Laporte, G. and Martello, S. (1990). The selective travelling salesman problem, *Discrete Applied Mathematics* **26**: 193–207.

Laporte, G. and Rodríguez-Martín, I. (2005). Locating cycles in transportation and telecommunications networks. Unpublished working paper.

Leifer, A. C. and Rosenwein, M. (1994). Strong linear programming relaxations for the orienteering problem, *European Journal of Operational Research* **73**: 517–523.

Ljubić, I., Weiskircher, R., Pferschy, U., Klau, G., Mutzel, P. and Fischetti, M. (2004). Solving the prize-collecting Steiner tree problem to optimality, *Technical Report TR 186-1-04-01, Institute of Computer Graphics and Algorithms, Vienna University of Technology* .

Lucena, A. (1991). Tight bounds for the Steiner problem in graphs, *Presentation made at the TIMS XXX - SOBRAPO XXIII Joint International Meeting, Rio de Janeiro, Brazil* .

Lucena, A. and Resende, M. (2004). Strong lower bounds for the prize collecting Steiner problem in graphs, *Discrete Applied Mathematics* **141**: 277–294.

Margot, F., Prodon, A. and Liebling, T. M. (1994). Tree polyhedron on 2-tree, *Mathematical Programming* **63**: 183–192.

Megiddo, N. (1979). Combinatorial optimization with rational objective functions, *Mathematics of Operations Research* **4**: 414–424.

Mittenthal, J. and Noon, C. E. (1992). Insert/delete heuristic for the travelling salesman subset-tour problem with one additional constraint, *Journal of the Operational Research Society* **43**: 277–283.

Mladenović, N. and Hansen, P. (1997). Variable neighborhood search, *Computers & Operations Research* **24**: 1097–1100.

Ramesh, R. and Brown, K. M. (1991). An efficient four-phase heuristic for the generalized orienteering problem, *Computers & Operations Research* **18**: 151–165.

Ramesh, R., Yoon, Y.-S. and Karwan, M. H. (1992). An optimal algorithm for the orienteering tour problem, *ORSA Journal on Computing* **4**: 155–165.

Segev, A. (1987). The node-weighted Steiner tree problem, *Networks* **17**: 1–17.

Tasgetiren, F. M. and Smith, A. E. (2000). A genetic algorithm for the orienteering problem, *In Proceedings of the 2000 Congress on Evolutionary Computation, San Diego, USA* : 1190–1195.

Tsiligirides, T. (1984). Heuristic methods applied to orienteering, *Journal of the Operational Research Society* **35**: 797–809.

Uchoa, E. (2005). Reduction tests for the prize-collecting Steiner problem, *Operations Research Letters* . Forthcoming.

Wang, Q., Sun, X., Golden, B. L. and Jia, J. (1995). Using artificial neural networks to solve the orienteering problem, *Annals of Operations Research* **61**: 111–120.