Survivable networks, linear programming relaxations and the parsimonious property

Michel X. Goemans

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA

Dimitris J. Bertsimas

Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA

We consider the survivable network design problem — the problem of designing, at minimum cost, a network with edge-connectivity requirements. As special cases, this problem encompasses the Steiner tree problem, the traveling salesman problem and the k-edge-connected network design problem. We establish a property, referred to as the parsimonious property, of the linear programming (LP) relaxation of a classical formulation for the problem. The parsimonious property has numerous consequences. For example, we derive various structural properties of these LP relaxations, we present some algorithmic improvements and we perform tight worst-case analyses of two heuristics for the survivable network design problem.

Key words: Network design, LP relaxations, worst-case analysis, heuristics.

1. Introduction

In this paper, we study from several perspectives the linear programming (LP) relaxations of a class of network design problems. This class includes a number of classical combinatorial optimization problems as special cases, including the Steiner tree problem, the traveling salesman and the k-edge-connected network design problem. The central problem we consider can be described as follows. Given a complete undirected network $G = (V, E)$ and a cost c_{ij} associated with each edge $(i, j) \in E$ we want to select a set of (possibly multiple) edges at minimum cost, so that the resulting network satisfies certain connectivity properties. In particular, if we associate to vertex i a *connectivity type* r_i representing the importance of communication from and to vertex i, we call a network *survivable* if it has at least $r_{ij} = \min(r_i, r_j)$

edge-disjoint paths between any pair of vertices i and j. In a survivable network, the loss or failure of any k edges still allows communication between vertices whose connectivity type is greater than k. An example of a survivable network is given in Figure 1. Gomory and Hu [14] show that the analysis problem of checking whether a given network is survivable can be solved by means of n maximum flow problems, where $n = |V|$. In this paper, we consider *the problem of designing a minimum cost survivable network.* This problem is also known as the multiterminal synthesis problem [14, 11] or the generalized Steiner problem [37].

The *survivable network design problem* (SNDP) is of particular importance in the design of communication or transportation systems in which the lack of communication or connectivity between parts of the network might be catastrophic. For example, this issue is particularly relevant in the design of communication systems using fiber optic links (see Monma et al. [29, 27] for a detailed description).

The Steiner tree problem can be modeled as an SNDP by letting the connectivity types r_i be 1 for a set S of terminals and 0 for optional vertices, also called Steiner vertices. Consequently, the minimum spanning tree $(S = V)$ and the shortest path problems $(|S| = 2)$ are also special cases. When the requirements are uniform, say equal to k, the SNDP reduces to the problem of designing a minimum cost k-edge-connected network. As we'll see when presenting the parsimonious property, it is very natural to allow the addition of degree constraints to the problem formulation. This extended version captures the well celebrated traveling salesman problem (TSP).

When all edges have a unit cost, the SNDP can be solved in polynomial time by an algorithm due to Sridhar and Chandrasekaran [34] — an adaptation of the classical Gomory and Hu's algorithm [14] producing a network with possibly half edges. For arbitrary costs, the SNDP, generalizing the Steiner tree problem, is NP-hard. Typically, such a negative result from complexity theory leads researchers to obtain approximate rather than optimal solutions. In fact, heuristic algorithms based on local search have been proposed by Steiglitz et al. [35] for the SNDP in its full generality, by Monma and Ko [27] for the k-edge-connected network design problem ($r_i = k$ for all i) and by Monma and Shallcross [29] for the case where $r_i \in \{1, 2\}$ for
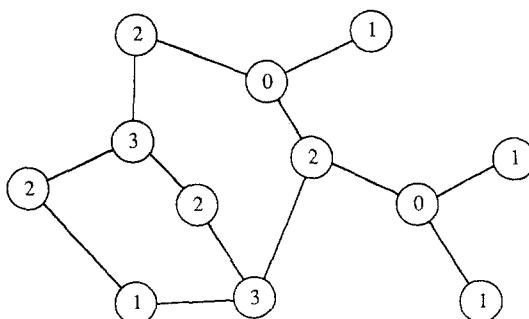

Fig. 1. Survivable network: The connectivity types are indicated inside earth vertex.

all *i*. Moreover, a wide variety of heuristics have been proposed for the Steiner tree problem (for a survey, see Winter [37]) and for the TSP (see Lawler et al. [23]).

The foundation of this paper is the derivation of a structural property, *the parsimonious property*, of the LP relaxation of a classical formulation of the SNDP. The variety of consequences makes the parsimonious property particularly important.

First, it gives strong relations between different relaxations and different combinatorial optimization problems which, in turn, have algorithmic consequences. For example, we prove the surprising result that Steiner vertices (i.e. vertices with $r_i = 0$) are unnecessary when solving the LP relaxation of the Steiner tree problem when the triangle inequality holds. This reduces the size of the problems to be solved. The simplest relation that we derive from the parsimonious property is the fact that the 1-tree relaxation with Lagrangean objective function for the TSP, also referred to as the Held–Karp lower bound [15, 16], is also a lower bound on the cost of the minimum cost 2-connected subgraph. Another corollary is the surprising result that the LP relaxations under consideration of the Steiner tree problem, the *k*-edge-connected network design problem and the TSP are essentially *identical* under the triangle inequality, the value of the first being exactly equal to the value of the second divided by *k* or half the value of the third. This can be used to compute the corresponding LP bounds efficiently. In a companion paper [12], a new formulation for the Held–Karp lower bound, which follows from the parsimonious property, is used to perform a probabilistic analysis of the bound when the vertices are identically and independently distributed in some Euclidean space.

The parsimonious property is also crucial in the analysis of two heuristics for the SNDP. Our first heuristic, referred to as the *tree heuristic*, is based on the computation of minimum spanning trees and reduces to previously known heuristics when applied to the Steiner tree problem. The second heuristic, called the *improved tree heuristic*, reduces to Christofides' heuristic when applied to the 2-connected network design problem. Among other results, our worst-case analysis shows that the ratio between the value of either heuristic and the LP relaxation bound is always less than twice the number of distinct nonzero connectivity types and also less than twice the logarithm of the largest connectivity type. Moreover, these bounds are tight. For the Steiner tree problem, our analysis strengthens results due to Takahashi and Matsuyama [36], Kou et al. [21], Plesnik [30] and El-Arbi [10]. Moreover, whenever $r_i \in \{0, 1, 2\}$, $r_i \in \{0, 1, 3\}$ or $r_i \in \{0, k\}$ for all *i*, we show that the improved tree heuristic is within twice the value of the optimal network. This also generalizes the result on the Steiner tree problem.

The remainder of the paper is structured as follows. In Section 2, we present and prove the parsimonious property. Consequences for the Held–Karp lower bound are investigated in Section 3. In Section 4, we present several algorithmic implications of the parsimonious property. Section 5 contains the description of the tree heuristics for the SNDP and their worst-case analyses. Finally, we conclude with some possible extensions.

2. The parsimonious property

Let $G = (V, E)$ be the complete undirected graph with vertex set V. For any pair (i, j) of vertices, let r_{ij} be the connectivity requirement between i and j (r_{ij} is assumed to be symmetric, i.e. $r_{ij} = r_{ji}$). Although we concentrate our attention on the typical case in which $r_{ij} = \min(r_i, r_j)$ for some set $\{r_i\}$ of *connectivity types*, we will state the parsimonious property in its full generality. By abuse of notation, r will denote either the set $\{r_{ij}\}$ or, if applicable, the set $\{r_i\}$. Call a network *survivable* if it has at least r_{ij} edge-disjoint paths between any pair (i, j) of vertices. If some edge, say e, is selected in a network, we incur a fixed cost c_e. In this paper, we assume that any edge may be chosen repeatedly. The *survivable network design problem* (SNDP) consists in finding the minimum cost survivable network. This problem may be formalized by the following integer program:

$$(\text{IP}_\emptyset(r)) \quad \text{IZ}_\emptyset(r) \;=\; \text{Min} \sum_{e \in E} c_e x_e$$

$$\text{subject to} \quad \sum_{e \in \delta(S)} x_e \geq \max_{(i,j) \in \delta(S)} r_{ij}, \quad S \subset V, \ S \neq \emptyset, \tag{1}$$

$$0 \leq x_e, \quad e \in E,$$
$$x_e \text{ integral}, \quad e \in E,$$

where $\delta(S)$ represents the set of edges connecting S to $V \setminus S$. Indeed, constraints (1) insure that the value of a minimum cut separating i from j is at least r_{ij}, implying that there are at least r_{ij} edge-disjoint paths between i and j. We denote by $(\text{IP}_\emptyset(r))$ the above integer program and by $\text{IZ}_\emptyset(r)$ its optimal value. Let $(\text{P}_\emptyset(r))$ denote the LP relaxation of $(\text{IP}_\emptyset(r))$ obtained by dropping the integrality restrictions and let $Z_\emptyset(r)$ be its optimal value. Clearly $Z_\emptyset(r)$ is a lower bound on $\text{IZ}_\emptyset(r)$. The meaning of the symbol $\emptyset$ in this notation will become clear shortly.

Although the linear program $(\text{P}_\emptyset(r))$ has an exponential number of constraints, the value $Z_\emptyset(r)$ can be computed in polynomial time either using the ellipsoid algorithm since the separation problem over $(\text{P}_\emptyset(r))$ can be solved by Gomory and Hu's algorithm [14] or using Karmarkar's algorithm since $(\text{P}_\emptyset(r))$ can be reformulated compactly using flow variables. However, these computational approaches are not satisfactory in practice and in fact, so far, no efficient and practical algorithm to compute $Z_\emptyset(r)$ exists.

As noticed in the introduction, the SNDP has some interesting special cases. For example, the Steiner tree problem — the problem of connecting at minimum cost a subset S of terminals possibly using some Steiner vertices in $V \setminus S$ — can be formulated as $(\text{IP}_\emptyset(\mathbf{1}_S))$ where $(\mathbf{1}_S)_{ij} = 1$ if $i, j \in S$ and 0 otherwise (or, $(\mathbf{1}_S)_i = 1$ if $i \in S$ and 0 otherwise). The formulation $(\text{IP}_\emptyset(\mathbf{1}_S))$ is known as the set covering formulation [2]. When $r_{ij} = k$ for all $i, j \in V$ we obtain the minimum-cost k-edge-conected network design problem.

For any feasible solution x either to $(\text{IP}_\emptyset(r))$ or to $(\text{P}_\emptyset(r))$, the degree of vertex i, defined by $d_x(i) \triangleq \sum_{e \in \delta(\{i\})} x_e$, is at least equal to $\max_{j \in V \setminus \{i\}} r_{ij}$ because of constraints

(1) for $S = \{i\}$. If $d_x(i) = \max_{j \in V \setminus \{i\}} r_{ij}$ then we say that x is *parsimonious* at vertex i. If we impose that the solution x be parsimonious at all vertices of a set $D \subseteq V$ we get some interesting variations of $(IP_{\emptyset}(r))$ and $(P_{\emptyset}(r))$, denoted by $(IP_D(r))$ and $(P_D(r))$, respectively. The most interesting special case is the traveling salesman problem. Indeed, when $r_{ij} = 2$ for all $i,j \in V$, the feasible solutions to $(IP_V(2))$ (2 denotes the vector of 2's) correspond to Hamiltonian tours. The formulation of $(IP_D(r))$ as an integer program is:

$(IP_D(r))$ $\quad IZ_D(r) = \text{Min} \sum_{e \in E} c_e x_e$

$$\text{subject to} \quad \sum_{e \in \delta(S)} x_e \geq \max_{(i,j) \in \delta(S)} r_{ij}, \quad S \subset V, \ S \neq \emptyset,$$

$$\sum_{e \in \delta(\{i\})} x_e = \max_{j \in V \setminus \{i\}} r_{ij}, \quad i \in D,$$

$$0 \leq x_e, \quad e \in E,$$

$$x_e \text{ integral}, \quad e \in E.$$

When we have integrality restrictions, the problem is clearly altered by the introduction of parsimonious constraints. For example, the TSP and the minimum-cost 2-connected problem have the same edge connectivity requirements but with different parsimonious constraints. Another illustration is given by the Steiner tree problem and the minimum spanning tree problem on S. However, when the integrality restrictions are relaxed, the value of the LP relaxation is not affected by the introduction of parsimonious constraints when the costs satisfy the triangle inequality, i.e. when $c_{ij} + c_{jk} \geq c_{ik}$ for all $i, j, k \in V$. This somewhat surprising result, which we refer to as the parsimonious property, constitutes the foundation of this paper.

Theorem 1 (the parsimonious property). *If the costs $\{c_e\}$ satisfy the triangle inequality then $Z_{\emptyset}(r) = Z_D(r)$ for all subsets $D \subseteq V$.*

The proof of this theorem is based on Lemma 2 which is a stronger version of a result due to Lovász [24] on connectivity properties of Eulerian multigraphs. The proof of this version, similar to the proof given by Lovász, is given in Appendix 1. The use of Lovász' lemma in a similar context also appears in Monma et al. [29].

Lemma 2. *Let $G = (V, E)$ be an Eulerian multigraph. Let $c_G(i, j)$ $(i, j \in V)$ denote the maximum number of edge-disjoint paths between i and j. Let x be any vertex of G and let u be any neighbor of x. Then there exists another neighbor of x, say v, such that, by splitting (x, u) and (x, v) i.e., removing the edges (x, u) and (x, v) and adding the edge (u, v), we obtain a multigraph G' satisfying the following two conditions:*
  1. *$c_{G'}(i, j) = c_G(i, j)$ for all $i, j \in V \setminus \{x\}$, and*
  2. *$c_{G'}(x, j) = \min(c_G(x, j), d_G(x) - 2)$ for all $j \in V \setminus \{x\}$, where $d_G(x)$ represents the degree of vertex x in G.*

Condition 2, which does not appear in Lovász's result, states that the splitting operation can be performed while maintaining most connectivity requirements involving vertex x.

Proof of Theorem 1. Clearly $Z_\emptyset(r) \leq Z_D(r)$ since $(P_D(r))$ is more constrained than $(P_\emptyset(r))$. In order to prove that $Z_\emptyset(r) \geq Z_D(r)$ we consider an optimal solution, say x, to $(P_\emptyset(r))$. We shall construct a feasible solution y to $(P_D(r))$ whose cost is at most equal to the cost of x. Since all data is rational, we may assume that all components of x are rational. Hence, there exists some integer k such that kx_e and kr_{ij} are even integers for all $e = (i, j) \in E$. Let $\tilde{G} = (V, \tilde{E})$ be the Eulerian multigraph which has kx_e copies of edge e. By the max-flow–min-cut theorem, $c_{\tilde{G}}(i, j) \geq kr_{ij}$ for all $(i, j) \in E$. As a result, by applying Lemma 2 repeatedly with x chosen among the vertices in D, we will eventually obtain a multigraph G' such that

  (i) $c_{G'}(i, j) \geq kr_{ij}$ $\forall (i, j) \in E$,

  (ii) $d_{G'}(i) = \max\limits_{j \in V \setminus \{i\}} kr_{ij}$ $\forall i \in D$.

Therefore, if we let y_e $(e \in E)$ be equal to the number of copies of edge e in G' divided by k, we obtain a feasible solution to $(P_D(r))$. Moreover, since the costs satisfy the triangle inequality, each time we perform a splitting operation the cost of the solution does not increase which implies that $\sum_{e \in E} c_e x_e \geq \sum_{e \in E} c_e y_e$. Since D was arbitrary, this completes the proof of Theorem 1.   □

In general, when the costs do not satisfy the triangle inequality, the parsimonious property does not hold. Nevertheless, this is not a restriction for the survivable network design problem and its special cases, such as the Steiner tree problem or the k-edge-connected network design problem. Indeed, let us consider an instance of the SNDP with arbitrary costs $\{c_e\}$. Define c'_e $(e = (i, j))$ to be the length of the shortest path between i and j with respect to the lengths $\{c_e\}$. Clearly, $\{c'_e\}$ satisfy the triangle inequality. Theorem 3 states that we can replace c_e by c'_e without affecting $IZ_\emptyset(r)$ or $Z_\emptyset(r)$.

Theorem 3. *For any set $\{c_e\}$ of costs, $IZ_\emptyset(r) = IZ'_\emptyset(r)$ and $Z_\emptyset(r) = Z'_\emptyset(r)$, where $IZ'(\cdot)$ and $Z'(\cdot)$ refer to the costs $\{c'_e\}$.*

Proof. Since $c'_e \leq c_e$ for all $e \in E$, $IZ'_\emptyset(r) \leq IZ_\emptyset(r)$ and $Z'_\emptyset(r) \leq Z_\emptyset(r)$. Now, consider an optimal solution x^* to $(IP_\emptyset(r))$ (resp. to $(P_\emptyset(r))$) with respect to the costs $\{c'_e\}$. In order to construct an optimal solution with respect to the costs $\{c_e\}$, we perform the following transformation. If some edge $e = (i, j)$ with $c'_e < c_e$ has some nonzero weight x^*_e, then we decrease x^*_e to 0 and increase by x^*_e the weights on the edges of a shortest path from i to j. Notice that this maintains feasibility and optimality of the solution. By repeating this operation, we obtain an optimal solution $\hat{x}$ to

$(IP_\emptyset(r))$ (resp. to $(P_\emptyset(r))$) with respect to the costs $\{c'_e\}$ such that $\hat{x}_e = 0$ whenever $c'_e < c_e$. As a result, the cost of this solution remains unchanged if we replace c'_e by c_e. This and the fact that $IZ'_\emptyset(r) \leqslant IZ_\emptyset(r)$ (resp. $Z'_\emptyset(r) \leqslant Z_\emptyset(r)$) imply that $\hat{x}$ is also optimal with respect to $\{c_e\}$. This completes the proof of Theorem 3. $\square$

The above transformation gives a generic transformation to convert a survivable network of total cost C' with respect to $\{c'_e\}$ into a survivable network of the same cost but with respect to $\{c_e\}$.

In the rest of this paper, we consider implications of the parsimonious property.

3. Structural properties of the Held–Karp bound

In this section, we consider the Held–Karp lower bound [15, 16] for the traveling salesman problem. This bound has been successfully used by several researchers to solve instances of the TSP by branch and bound methods (see Balas and Toth [4]). Moreover, in a striking computational study, Johnson [17, 18] estimates the degree of suboptimality of heuristic solutions by computing the Held–Karp lower bound and, as a result, he is able to show that the solutions he generates are within 1% of optimality for instances with as many as 100 000 vertices. The Held–Karp lower bound can be formulated in several equivalent ways, the most classical being in terms of the 1-tree relaxation with Lagrangean objective function [16]. As a linear program, it can be expressed by:

(P_1) $Z_{HK} = \text{Min} \sum_{e \in E} c_e x_e$

subject to $\sum_{e \in \delta(S)} x_e \geqslant 2, \quad S \subset V, \ S \neq \emptyset,$

$\sum_{e \in \delta(\{i\})} x_e = 2, \quad i \in V,$

$0 \leqslant x_e, \quad e \in E.$

Z_{HK} is thus precisely $Z_V(2)$. By the parsimonious property, we have that $Z_{HK} = Z_\emptyset(2)$ under the triangle inequality, i.e.

(P_2) $Z_{HK} = \text{Min} \sum_{e \in E} c_e x_e$

subject to $\sum_{e \in \delta(S)} x_e \geqslant 2, \quad S \subset V, \ S \neq \emptyset,$

$0 \leqslant x_e, \quad e \in E.$

This new formulation is very helpful in order to derive properties of the bound. For example, we have:

Theorem 4 (Cunningham [28]). *Under the triangle inequality, the Held–Karp lower bound Z_{HK} is a lower bound on the value of the minimum-cost 2-connected netweork design problem.*

Proof. The theorem follows from the fact that by adding integrality constraints to (P_2) we obtain an integer programming formulation of the 2-connected network design problem. □

By the same argument, the Held–Karp lower bound is also a lower bound on the Steiner version of the minimum 2-connected network design problem in which optional vertices are allowed.

In a companion paper [12], formulation (P_2) is used to perform a probabilistic analysis of the bound when the vertices are identically and independently distributed in some Euclidean space. Under this probabilistic model, the Held–Karp lower bound is proved to have a similar asymptotic behavior as the traveling salesman problem or the 2-connected network design problem. More specifically, the bound is almost surely and asymptotically a fraction of the optimal value, this fraction being empirically evaluated to be greater than 99% [17, 18]. This probabilistic analysis uses several structural properties of the bound, such as its subadditivity, monotonicity and upperlinearity. This last property seems hard to prove without referring to the new formulation (P_2). In the next theorem, we show that the monotonicity of the Held–Karp lower bound is an immediate corollary of the parsimonious property.

Theorem 5. *Let $Z_{HK}(S)$ be the Held–Karp lower bound for the subgraph G_S induced by the subset S of vertices. If $\{c_e\}$ satisfy the triangle inequality then $Z_{HK}(S) \leq Z_{HK}(V)$ for all $S \subseteq V$.*

Proof. By definition, $Z_{HK}(S) = Z_V(2_S)$, where $(2_S)_{ij} = 2$ if $i, j \in S$ and 0 otherwise. Moreover, by the parsimonious property, $Z_V(2_S)$ is equal to $Z_\emptyset(2_S)$. Since $(P_\emptyset(2_S))$ is a relaxation of $(P_\emptyset(2_V))$, we have that $Z_\emptyset(2_S) \leq Z_\emptyset(2_V)$. Using again the parsimonious property, we obtain $Z_\emptyset(2_V) = Z_V(2_V) = Z_{HK}(V)$. This proves that $Z_{HK}(S) \leq Z_{HK}(V)$. □

The monotonicity of the bound is not only useful for its probabilistic analysis but also for its worst-case analysis. Shmoys and Williamson [33] use this monotonicity property to give another proof of a result due to Wolsey [38] stating that the ratio between Christofides' heuristic and the Held–Karp lower bound is bounded by $\frac{3}{2}$ under the triangle inequality. In Section 5, a monotonicity property similar to Theorem 5 will be the basis of our worst-case analysis of the tree heuristic for the survivable network design problem.

4. Algorithmic implications

The parsimonious property also has algorithmic implications in order to compute LP relaxation bounds efficiently. Throughout this section, we assume without loss of generality (see Theorem 3), that the costs satisfy the triangle inequality. By the parsimonious property, any vertex with $r_i = 0$ can be deleted when solving the linear program $(P_\emptyset(r))$. This reduces the size of the problems to be solved. An even more astonishing result relates the LP relaxations of the Steiner tree problem and the TSP.

Theorem 6. *Let $Z_{SP}(S)$ be the optimal value of the linear program $(P_\emptyset(1_S))$. Then $Z_{SP}(S) = \frac{1}{2}Z_{HK}(S)$, where $Z_{HK}(S)$ is defined in Theorem 5.*

Proof. Using the parsimonious property and our concise notation, we have that $Z_{SP}(S) = Z_\emptyset(1_S) = Z_V(1_S)$. By linearity, this last quantity is equal to $\frac{1}{2}Z_V(2_S) = \frac{1}{2}Z_{HK}(S)$. $\square$

The relation expressed in Theorem 6 leads to an algorithm to compute $Z_{SP}(S)$. Indeed, Held and Karp [15, 16] show the equivalence between the relaxation $P_V(2)$ and the 1-tree relaxation with Lagrangean objective function and show that $Z_{HK}(S)$ can be obtained by solving a sequence of minimum spanning tree problems. Moreover, their algorithm can be implemented efficiently to obtain very close approximations of $Z_{HK}(S)$ for instances with as many as 100 000 vertices [18]. The reader is referred to the original paper [16] or to [4] for a detailed presentation of Held–Karp's algorithm.

Similarly, the LP relaxation bound for the k-edge-connected network design problem can be related to the Held–Karp lower bound. In fact, under the triangle inequality, $Z_\emptyset(k) = \frac{1}{2}kZ_{HK}$. Therefore, Held and Karp's approach can also be used to obtain a lower bound on the cost of a k-edge-connected network. Moreover, given the experimental observation and some theoretical explanation that Z_{HK} is very close to the cost of the optimal tour [17, 18, 4, 12], we can assert that $Z_\emptyset(k)$ is a very good lower bound in order to assess the quality of a k-edge-connected heuristic network.

The subadditivity and the upperlinearity (for details see [12]) also have algorithmic consequences. These properties justify the use of partitioning schemes a la Karp [20] for obtaining LP relaxation bounds corresponding to Euclidean problems. For example, suppose we would like to compute the Held–Karp lower bound defined by a set of vertices in the unit square. If $\{Q_i: i = 1, \ldots, m\}$ is a partition of the unit square into subregions of finite perimeter then the upperlinearity implies that $\sum_{i=1}^{m} Z_{HK}(V \cap Q_i) - 2(c - 2) \leq Z_{HK}(V)$ where c is the total length of the boundary of the partition. As a result, we can obtain a lower bound by solving m smaller subproblems. Moreover, using the subadditivity of the Held–Karp lower bound we can construct a feasible solution to (P_2) whose cost is at most $\sum_{i=1}^{m} Z_{HK}(V \cap Q_i) + \sqrt{2m}$. When the partition consists of $m = k^2$ similar subsquares, this partitioning

scheme gives an upper bound and a lower bound on the Held–Karp lower bound which differ by at most $(4+\sqrt{2})k$ units. Moreover, the ratio of this difference to the Held–Karp lower bound is almost surely equal to 0 when the vertices are independently and identically distributed in the unit square and $m = o(n)$.

5, The tree heuristics and their worst-case analyses

In this section, we consider the instances of the SNDP for which the requirements are of the form $r_{ij} = \min(r_i, r_j)$. As previously mentioned, this is the most typical case. We introduce two heuristics, the *tree heuristic* and the *improved tree heuristic*, and we show that they have some interesting worst-case guarantees.

The tree heuristic consists in constructing a survivable network as a union of trees. More precisely, in the kth iteration, we construct a minimum cost tree spanning all vertices for which $r_i \geqslant k$. The resulting network is survivable since, at iteration k, we have at least 1 additional path from i to j if both r_i and r_j are greater or equal to k. The implementation of this heuristic can be made more efficient by noticing that several iterations might have the same vertex set.

Tree heuristic.
   Step 1. Compute the shortest path lengths $\{c'_e\}$.
   Step 2. Prepare a sorted list $L = \{\rho_0 = 0 < \rho_1 < \rho_2 < \cdots < \rho_p\}$ consisting of all distinct connectivity types.
   Step 3. $x := 0$.
   For $k = 1$ to p do:
   ● Let $V_k := \{i \in V : r_i \geqslant \rho_k\}$.
   ● Compute $T_k = (V_k, E_k)$, the minimum spanning tree with respect to $\{c'_e\}$ of the complete graph induced by V_k.
   ● Let $x_e := x_e + (\rho_k - \rho_{k-1})$ for all $e \in E_k$.
   Step 4. Use the transformation described in Theorem 3 to obtain a survivable network whose total cost with respect to $\{c_e\}$ is equal to the cost of x with respect to $\{c'_e\}$.
   Step 5. Apply some local improvement heuristic.

Step 1 and 4 reduce the instance into one in which the costs satisfy the triangle inequality. The tree heuristic is a construction heuristic: it constructs piece by piece a survivable network. Step 5, which is optional, allows to combine the tree heuristic with an improvement heuristic — a heuristic which starts from a feasible solution and iteratively performs some local transformation in order to obtain a solution with smaller total cost. Improvement heuristics were proposed by Steiglitz et al. [35] for the general SNDP, by Monma and Ko [27] for the k-edge-connected network design problem ($r_i = k$ for all i) and by Monma and Shallcross [29] for the case where $r_i \in \{1, 2\}$ for all i. If the original costs satisfy the triangle inequality,

the three heuristic can be implemented in $O(pn^2)$ time where n represents the number of vertices. Otherwise, Step 1 is the bottleneck operation in the above description. However, although convenient for the analysis, Step 1 is not really needed. The minimum spanning tree heuristic for the Steiner tree problem can indeed be implemented in $O(n^2)$ time even when the costs do not obey the triangle inequality (see Mehlhorn [26]). This results in an implementation of the tree heuristic running in $O(pn^2)$ time.

When applied to the Steiner tree problem, the heuristic reduces to the minimum spanning tree heuristic proposed in slightly different versions by Kou et al. [21], Plesnik [30], El-Arbi [10], and Takahashi and Matsuyama [36] (see also [26]). The tree heuristic is also a generalization of the double spanning tree heuristic for the 2-connected network design problem or the TSP.

Before analyzing the tree heuristic in its full generality, we consider the worst-case analysis of the minimum spanning tree heuristic for the Steiner tree problem. Kou et al. [21], Plesnik [30], El-Arbi [10], and Takahashi and Matsuyama [36] show that the ratio between the value of the minimum spanning tree heuristic and the optimal value of the Steiner tree problem is bounded by $2 - 2/|S|$, where S denotes the set of terminals. In fact, we can prove the following stronger result.

Theorem 7. *For any set $\{c_e\}$ of costs, $Z_{\text{tree}}(S)/Z_{\text{SP}}(S) \leq 2 - 2/|S|$, where $Z_{\text{tree}}(S)$ denotes the value of the tree heuristic when applied to a Steiner tree problem in which S is the set of terminals and where $Z_{\text{SP}}(S)$ is the LP relaxation bound $Z_\emptyset(1_S)$. Furthermore, the bound is symmetrically tight in the sense that, for any set S, there exist instances for which $IZ_\emptyset(1_S)/Z_{\text{SP}}(S) = 2 - 2/|S|$, and other instances for which $Z_{\text{tree}}(S)/IZ_\emptyset(1_S) = 2 - 2/|S|$.*

Proof. Using Theorem 3, its proof and the structure of the tree heuristic, we can restrict our attention to instances which satisfy the triangle inequality. In this case, the tree heuristic reduces to taking a minimum spanning tree over S. Let x^* be the optimal solution to $(P_V(1_S))$. By the parsimonious property, the value of x^* is precisely $Z_{\text{SP}}(S)$. The value of the tree heuristic can be expressed as the optimal value of a linear program by using Edmonds' complete characterization [9] of the spanning tree polytope.

$$(\text{P}_{\text{MST}}) \quad Z_{\text{tree}}(S) = \text{Min} \sum_{e \in E} c_e x_e$$

$$\text{subject to} \quad \sum_{e \in E(T)} x_e \leq |T| - 1, \quad T \subset S, \ T \neq \emptyset,$$

$$\sum_{e \in E(S)} x_e = |S| - 1,$$

$$0 \leq x_e, \quad e \in E(S).$$

It is easy to verify that $(2 - 2/|S|)x^*$ satisfies all the constraints of (P_{MST}). Therefore, $(2 - 2/|S|)Z_{\text{SP}}(S) = (2 - 2/|S|)Z_V(1_S) \geq Z_{\text{tree}}(S)$.

The heuristic attains the worst-case bound when there is one Steiner vertex linked to all other vertices by edges of cost 1, while all other edges have cost 2 (see Figure 2). The heuristic solution has value $2(|S|-1)$ while the optimal Steiner tree has value $|S|$.

In order to show that the cost of the optimum Steiner tree can be $(2-2/|S|)$ times the value of its LP relaxation value, consider the minimum spanning tree problem on S with $c_e = 2$ for all $e \in E$ (see Figure 3). Clearly, the optimal Steiner tree has value $2(|S|-1)$, while the optimum solution of the LP relaxation is obtained by setting $x_e = 0.5$ along some Hamiltonian cycle, resulting in a total cost of $|S|$. □

We can use Theorem 7 to perform a worst-case analysis of the tree heuristic for the general survivable network design problem.

Theorem 8. *For any set* $\{c_e\}$ *of costs,*

$$\frac{Z_{\text{tree}}(r)}{Z_0(r)} \leq \left(2 - \frac{2}{|V_1|}\right)\left(\sum_{k=1}^{p} \frac{\rho_k - \rho_{k-1}}{\rho_k}\right).$$

where $Z_{\text{tree}}(r)$ *denotes the value of the tree heuristic when the connectivity types are given by the vector* r, *and* ρ_i *and* V_1 *are defined in Steps 2 and 3 of the tree heuristic.*
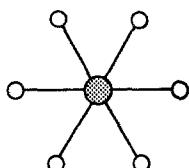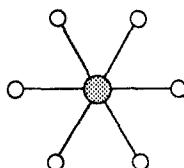
(a) Problem instance	(b) Optimal Steiner tree	(c) Heuristic solution

⬤ Steiner vertex

⬤——○ edge of cost 1

○   ○ edge of cost 2
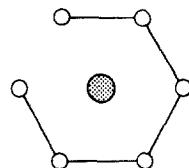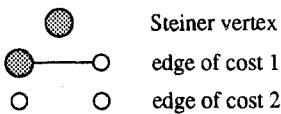
Fig. 2. Worst-case instance.
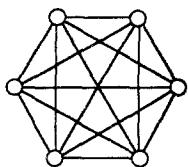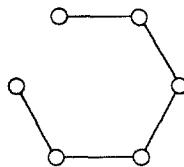
(a) Problem instance	(b) Optimal Steiner tree	(c) LP optimal solution

○———○ edge of cost 2   ○-----○ weight of 0.5
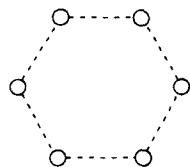
Fig. 3. Worst-case instance.

Proof. As in Theorem 7, we assume without loss of generality that the costs satisfy the triangle inequality. Therefore, $Z_{\text{tree}}(r)$ can be expressed by

$$Z_{\text{tree}}(r) = \sum_{k=1}^{p} (\rho_k - \rho_{k-1}) Z_{\text{MST}}(V_k) \tag{2}$$

where $Z_{\text{MST}}(V_k)$ denotes the cost of the minimum spanning tree over V_k. From Theorem 7, we know that

$$Z_{\text{MST}}(V_k) \leq \left(2 - \frac{2}{|V_k|}\right) Z_{\text{SP}}(V_k) \leq \left(2 - \frac{2}{|V_1|}\right) Z_{\text{SP}}(V_k).$$

But,

$$Z_{\text{SP}}(V_k) = Z_{\emptyset}(\mathbf{1}_{V_k}) = \frac{1}{\rho_k} Z_{\emptyset}(\rho_k \mathbf{1}_{V_k}). \tag{3}$$

Combining (2) and (3), we obtain

$$Z_{\text{tree}}(r) \leq \left(2 - \frac{2}{|V_1|}\right) \sum_{k=1}^{p} \frac{\rho_k - \rho_{k-1}}{\rho_k} Z_{\emptyset}(\rho_k \mathbf{1}_{V_k}). \tag{4}$$

Since, by definition of V_k, the requirements in $\rho_k \mathbf{1}_{V_k}$ are less or equal to the requirements in r, $Z_{\emptyset}(\rho_k \mathbf{1}_{V_k}) \leq Z_{\emptyset}(r)$. Therefore,

$$Z_{\text{tree}}(r) \leq \left(2 - \frac{2}{|V_1|}\right) \sum_{k=1}^{p} \frac{\rho_k - \rho_{k-1}}{\rho_k} Z_{\emptyset}(r)$$

which proves the theorem. $\square$

We now consider the question whether the bound in Theorem 8 is tight. For a given set $L = \{\rho_0 = 0, \rho_1, \rho_2, \ldots, \rho_p\}$ of distinct connectivity types, let

$$f(L) = \sup_{I:r_i \in L \;\forall i \in V} \frac{Z_{\text{tree}}(r)}{Z_{\emptyset}(r)}$$

where the supremum is taken over all instances whose connectivity types are within L. Notice that $f(\{0, \rho_1, \rho_2, \ldots, \rho_p\}) = f(\{\rho_1, \rho_2, \ldots, \rho_p\})$ since vertices whose connectivity type is 0 affect neither the heuristic nor the LP relaxation. Theorem 8 implies that

$$f(\{\rho_0 = 0, \rho_1, \rho_2, \ldots, \rho_p\}) \leq 2 \sum_{k=1}^{p} \frac{\rho_k - \rho_{k-1}}{\rho_k}.$$

In the next theorem, we show that this bound can be achieved in some cases.

Theorem 9. *If $\rho_2/\rho_1, \rho_3/\rho_2, \ldots, \rho_{p-1}/\rho_{p-2}$ and $2\rho_p/\rho_{p-1}$ are integers then*

$$f(\{\rho_0 = 0, \rho_1, \rho_2, \ldots, \rho_p\}) = 2 \sum_{k=1}^{p} \frac{\rho_k - \rho_{k-1}}{\rho_k}.$$

The proof of this result is lengthy and technical, and is sketched in Appendix 2. From Theorem 9 we can derive the following result.

Corollary 10. 1. $f(0, 2^0, 2^1, \ldots, 2^{p-1}) = p + 1$ *for any* $p \geq 1$,

2. $f(0, 1, 2, \ldots, r_{max}) = \Theta(\log r_{max})$,

3. $\sup_{\mathscr{I}_p} f(0, \rho_1, \ldots, \rho_p) = \Theta(p)$, *where* $\mathscr{I}_p$ *denotes the set of instances with at most* p *distinct connectivity types.*

Proof. 1 follows directly from Theorem 9.

By Theorem 8,

$$f(0, 1, 2, \ldots, r_{max}) \leq 2 \sum_{k=1}^{p} \frac{1}{k} = O(\log r_{max}). \tag{5}$$

Moreover, by definition of $f(\cdot)$, $f(L) \geq f(L')$ whenever $L \supseteq L'$. Therefore, by 1,

$$f(0, 1, 2, \ldots, r_{max}) \geq f(0, 2^0, 2^1, \ldots, 2^{\lfloor \log_2 r_{max} \rfloor})$$

$$= \lfloor \log_2 r_{max} \rfloor + 2 = \Omega(\log r_{max}). \tag{6}$$

Combining (5) and (6), we get 2.

By Theorem 8,

$$f(0, \rho_1, \ldots, \rho_p) \leq 2 \sum_{k=1}^{p} \frac{\rho_k - \rho_{k-1}}{\rho_k} < 2p.$$

Moreover, using 1,

$$\sup_{\mathscr{I}_p} f(0, \rho_1, \ldots, \rho_p) \geq f(0, 2^0, 2^1, \ldots, 2^{p-1}) = p + 1.$$

This proves 3. $\square$

In the next theorem, we compare the value of the tree heuristic to the optimal value rather than the LP relaxation bound. We show that when some vertex has a connectivity type of 1, we can lower the constant of Theorem 8 by one unit. Although LP relaxations do not appear in the statement of Theorem 11, they play a privileged role in its proof.

Theorem 11. *If* $\rho_1 = 1$ *then*

$$\frac{Z_{tree}(r)}{IZ_0(r)} \leq \left(2 - \frac{2}{|V_1|} \right) \left(\sum_{k=1}^{p} \frac{\rho_k - \rho_{k-1}}{\rho_k} - \frac{1}{2} \right).$$

where $V_1 = \{i \in V : r_i \geq 1\}$.

Proof. Consider the optimal solution x^* to $(IP_0(r))$. This optimal network consists of a maximal 2-connected block to which trees are attached (see Figure 4). Moreover, this block spans a set B of vertices containing V_2 (the set of vertices whose
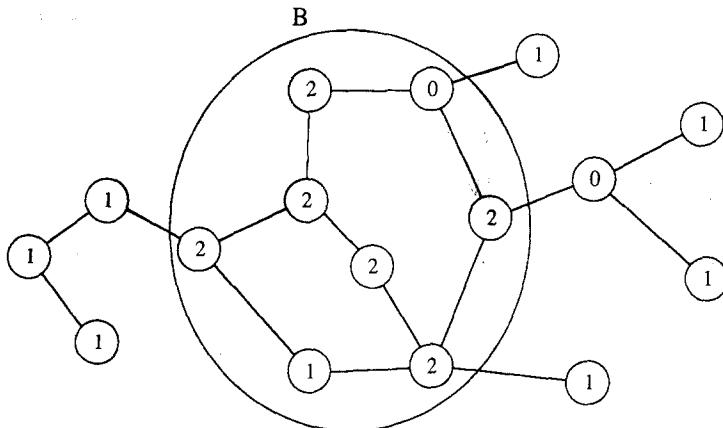
Fig. 4. Example in proof of Theorem 11.

connectivity types is at least ρ_2). We decompose x^* into the sum of two vectors y and z where

$$y_e = \begin{cases} \frac{1}{2}x_e^* & \text{if } e \in E(B), \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad z_e = \begin{cases} \frac{1}{2}x_e^* & \text{if } e \in E(B), \\ x_e^* & \text{otherwise.} \end{cases}$$

Clearly, $x^* = y + z$. The crucial observation is that, by definition of B, $2y$ is a feasible solution to $(P_\emptyset(\rho_2 \mathbf{1}_{V_2}))$ while z is a feasible solution to $(P_\emptyset(\mathbf{1}_{V_1}))$. Hence,

$$IZ_\emptyset(r) = \sum_{e \in E} c_e x_e^* = \sum_{e \in E} c_e y_e + \sum_{e \in E} c_e z_e \geq \tfrac{1}{2} Z_\emptyset(\rho_2 \mathbf{1}_{V_2}) + Z_\emptyset(\mathbf{1}_{V_1}). \tag{7}$$

Combining inequality (4) from the proof of Theorem 8, in equality (7) and the fact that $Z_\emptyset(\rho_k \mathbf{1}_{V_k}) \leq IZ_\emptyset(r)$, we obtain

$$Z_{\text{tree}}(r) \leq \left(2 - \frac{2}{|V_1|}\right) \sum_{k=1}^p \frac{\rho_k - \rho_{k-1}}{\rho_k} Z_\emptyset(\rho_k \mathbf{1}_{V_k})$$

$$= \left(2 - \frac{2}{|V_1|}\right)\left(Z_\emptyset(\mathbf{1}_{V_1}) + \sum_{k=2}^p \frac{\rho_k - \rho_{k-1}}{\rho_k} Z_\emptyset(\rho_k \mathbf{1}_{V_k})\right)$$

$$\leq \left(2 - \frac{2}{|V_1|}\right)\left(IZ_\emptyset(r) + \left(\frac{\rho_2 - 1}{\rho_2} - \frac{1}{2}\right) Z_\emptyset(\rho_2 \mathbf{1}_{V_2})\right.$$

$$\left. + \sum_{k=3}^p \frac{\rho_k - \rho_{k-1}}{\rho_k} Z_\emptyset(\rho_k \mathbf{1}_{V_k})\right)$$

$$\leq \left(2 - \frac{2}{|V_1|}\right)\left(\frac{1}{2} + \sum_{k=2}^p \frac{\rho_k - \rho_{k-1}}{\rho_k}\right) IZ_\emptyset(r)$$

$$= \left(2 - \frac{2}{|V_1|}\right)\left(\sum_{k=1}^p \frac{\rho_k - \rho_{k-1}}{\rho_k} - \frac{1}{2}\right) IZ_\emptyset(r), \tag{8}$$

where we have used the fact that $(\rho_2 - 1)/\rho_2 - \frac{1}{2} \geq 0$ since $\rho_2 \geq 2$. $\quad \square$

As a corollary, we obtain that the tree heuristic is within twice the value of the optimal solution not only for the Steiner tree problem but also for the case in which $r_i \in \{0, 1, 2\}$ for all $i \in V$.

Corollary 12. *If $r_i \in \{0, 1, 2\}$ for all $i \in V$ then*

$$Z_{\text{tree}}(r)/IZ_\emptyset(r) \le 2 - 2/|V_1|. \qquad \square$$

Our second heuristic, *the improved tree heuristic*, improves upon the tree heuristic in the worst-case when there is some gap in the sequence $\{\rho_0 = 0 < \rho_1 < \rho_2 < \cdots < \rho_p\}$. The improved tree heuristic generalizes Christofides' heuristic [6] for the traveling salesman problem in the same way as the tree heuristic generalizes the minimum spanning tree heuristic for the Steiner tree problem [10, 21, 30, 36]. The improved tree heuristic can be described as follows:

Improved tree heuristic.
  Step 1. Compute the shortest path lengths $\{c'_e\}$.
  Step 2. Prepare a sorted list $L = \{\rho_0 = 0 < \rho_1 < \rho_2 < \cdots < \rho_p\}$ consisting of all distinct connectivity types.
  Step 3. $x := 0$.
  For $k = 1$ to p do:
  • Let $V_k := \{i \in V: r_i \ge \rho_k\}$.
  • Compute $T_k = (V_k, E_k)$, the minimum spanning tree with respect to $\{c'_e\}$ of the complete graph induced by V_k.
  • If $\rho_k = \rho_{k-1} + 1$ then $x_e := x_e + 1$ for all $e \in E_k$.
    Else
      – Let O_k be the vertices of odd degree in T_k.
      – Compute $M_k = (O_k, E'_k)$, the minimum weight matching with respect to $\{c'_e\}$ of the complete graph induced by O_k.
      – Let $x_e := x_e + \lceil \frac{1}{2}(\rho_k - \rho_{k-1}) \rceil$ for all $e \in E_k$.
      – Let $x_e := x_e + \lfloor \frac{1}{2}(\rho_k - \rho_{k-1}) \rfloor$ for all $e \in E'_k$.
  Step 4. Use the transformation described in Theorem 3 to obtain a survivable network whose total cost with respect to $\{c_e\}$ is equal to the cost of x with respect to $\{c'_e\}$.
  Step 5. Apply some local improvement heuristic.

In other words, whenever we would like to increase the edge-connectivity between vertices in V_k by 2 units, we add to the current solution a minimum spanning tree T_k as well as a minimum weight matching M_k on the odd degree vertices of T_k. Since the union of T_k and M_k is Eulerian and, hence, 2-edge connected, the resulting network has at least 2 more edge-disjoint paths between any pair of vertices in V_k. If there are no gaps in the sequence L, the improved tree heuristic reduces to the tree heuristic. Otherwise, its overall time complexity is $O(rn^3)$ where r denotes the

number of gaps in L. In the next theorem, we present a worst-case analysis of the improved tree heuristic.

Theorem 13. *For any set* $\{c_e\}$ *of costs,*

$$\frac{Z_{\text{imp}}(r)}{Z_\emptyset(r)} < \sum_{k=1}^{p} \frac{f(\rho_k - \rho_{k-1})}{\rho_k},$$

where $Z_{\text{imp}}(r)$ *denotes the value of the improved tree heuristic when the connectivity types are given by the vector* r *and* $f(l) = \frac{3}{2}l$ *if l is even and* $f(l) = \frac{1}{2}(3l+1)$ *if l is odd.*

Proof. The value $Z_{\text{imp}}(r)$ of the tree heuristic is given by

$$Z_{\text{imp}}(r) = \sum_{k=1}^{p} \left\lceil \frac{\rho_k - \rho_{k-1}}{2} \right\rceil Z_{\text{MST}}(V_k) + \sum_{k=1}^{p} \left\lfloor \frac{\rho_k - \rho_{k-1}}{2} \right\rfloor Z_{\text{M}}(O_k), \tag{9}$$

where $Z_{\text{M}}(O_k)$ represents the cost of the minimum cost matching on O_k. From Theorem 8 (see equation (3)), we know that

$$Z_{\text{MST}}(V_k) < \frac{2}{\rho_k} Z_\emptyset(\rho_k \mathbf{1}_{V_k}), \tag{10}$$

the strict inequality arising from the fact that we have replaced $2 - 2/|V_k|$ by 2. Moreover,

$$Z_{\text{M}}(O_k) \leqslant Z_{O_k}(\mathbf{1}_{O_k}) = Z_\emptyset(\mathbf{1}_{O_k}) \leqslant Z_\emptyset(\mathbf{1}_{V_k}) = \frac{1}{\rho_k} Z_\emptyset(\rho_k \mathbf{1}_{V_k}), \tag{11}$$

where the first inequality follows from the complete description of the perfect matching polytope due to Edmonds [8] and the first equality follows from the parsimonious property. Combining equations (9), (10) and (11) and the fact that $Z_\emptyset(\rho_k \mathbf{1}_{V_k}) \leqslant Z_\emptyset(r)$ we obtain

$$\frac{Z_{\text{imp}}(r)}{Z_\emptyset(r)} < \sum_{k=1}^{p} \left(\frac{2}{\rho_k} \left\lceil \frac{\rho_k - \rho_{k-1}}{2} \right\rceil + \frac{1}{\rho_k} \left\lfloor \frac{\rho_k - \rho_{k-1}}{2} \right\rfloor \right) = \sum_{k=1}^{p} \frac{f(\rho_k - \rho_{k-1})}{\rho_k}. \quad \square$$

For the Steiner k-edge-connected network design problem, we obtain:

Corollary 14. *If* $r_i \in \{0, k\}$ *for all* $i \in V$ *and some k then*

$$\frac{Z_{\text{imp}}(r)}{1Z_\emptyset(r)} < \begin{cases} \frac{3}{2} & \text{for } k \text{ even,} \\ \frac{3}{2} + 1/(2k) & \text{for } k \text{ odd.} \end{cases} \quad \square$$

Using the same technique as in Theorem 11, we can improve the constant in Theorem 13 when $\rho_1 = 1$.

Theorem 15. *If $\rho_1 = 1$ then*

$$\frac{Z_{\text{imp}}(r)}{IZ_\emptyset(r)} < \left(\sum_{k=1}^{p} \frac{f(\rho_k - \rho_{k-1})}{\rho_k} - 1 \right).$$

Proof. The proof is identical to the proof of Theorem 11 except that equation (8) becomes

$$Z_{\text{imp}}(r) < \sum_{k=1}^{p} \frac{f(\rho_k - \rho_{k-1})}{\rho_k} Z_\emptyset(\rho_k 1_{V_k})$$

$$= \left(2 Z_\emptyset(1_{V_1}) + \sum_{k=2}^{p} \frac{f(\rho_k - \rho_{k-1})}{\rho_k} Z_\emptyset(\rho_k 1_{V_k}) \right)$$

$$\leq \left(2 IZ_\emptyset(r) + \left(\frac{f(\rho_2 - 1)}{\rho_2} - 1 \right) Z_\emptyset(\rho_2 1_{V_2}) + \sum_{k=3}^{p} \frac{f(\rho_k - \rho_{k-1})}{\rho_k} Z_\emptyset(\rho_k 1_{V_k}) \right)$$

$$\leq \left(1 + \sum_{k=2}^{p} \frac{f(\rho_k - \rho_{k-1})}{\rho_k} \right) IZ_\emptyset(r)$$

$$= \left(\sum_{k=1}^{p} \frac{f(\rho_k - \rho_{k-1})}{\rho_k} - 1 \right) IZ_\emptyset(r),$$

where we have used the fact that $f(\rho_2 - 1)/\rho_2 - 1 \geq 0$ since $\rho_2 \geq 2$. □

Corollary 16. *If $r_i \in \{0, 1, 3\}$ for all $i \in V$ then $Z_{\text{imp}}(r)/IZ_\emptyset(r) < 2$.* □

This corollary also generalizes the result on the worst-case analysis of the Steiner tree problem.

6. Concluding remarks and extensions

We conclude by mentioning some generalizations of the parsimonious property. The property still holds if we have additional degree constraints of the form $\sum_{e \in \delta(\{i\})} x_e = a_i$ for all i in some subset T of vertices. In that case, D has to be a subset of $V \backslash T$. This generalization allows one to consider other combinatorial optimization problems such as the k-TSP. Moreover, the parsimonious property remains valid if we impose the cutset constraints (1) only for those subsets S of odd cardinality. Again, the class of problems that fit into this framework becomes richer and it now encompasses matching-type problems. It would be interesting to investigate whether the parsimonious property has important consequences for this broader class of problems.

Very recently, Agrawal, Klein and Ravi [1] (see also Goemans and Williamson [13]) generalized the results of Section 5 to instances for which r_{ij} is not necessarily equal to $\min(r_i, r_j)$. Their approach is based upon a heuristic algorithm for the case where $r_{ij} \in \{0, 1\}$ and upon the techniques developed in Section 5.

Finally, we would also like to mention that Theorem 7 was recently used by Bienstock, Goemans, Simchi-Levi and Williamson [5] to propose a constant guaranteed heuristic for the prize collecting traveling salesman problem without a reward constraint (see also Goemans and Williamson [13] for an improved heuristic).

Appendix 1: Proof of Lemma 2

As previously mentioned, Lovász proves a slightly weaker version of Lemma 2 in which condition 2 is not present. His elegant proof proceeds along the following lines.

1. There exists at most one set S satisfying:
   (a) $x \in S, u \notin S$,
   (b) $|\delta(S)| = c_G(i, j)$ for some $i, j \in V$ with $i \in S, j \notin S$ and $i \neq x$, and
   (c) S is minimal with respect to the above two conditions.
2. If there is no such S, then any neighbor v of x can be used for the splitting operation.
3. If such an S exists then there exists at least one neighbor of x in S. Moreover, any neighbor of x in S can be used for the splitting operation.

Our proof of Lemma 2 directly rests upon Lovász's proof.

If $d_G(x) = 2$ then the result follows directly from Lovász's result. Hence, assume that $d_G(x) \geqslant 4$. Let $\hat{G} = (\hat{V}, \hat{E})$ be obtained from G by adding a new vertex $\hat{x}$ and by linking that vertex to x through $d_G(x) - 2$ edges. Clearly, $\hat{G}$ is Eulerian. Moreover,

$$c_{\hat{G}}(i, j) = c_G(i, j) \quad \forall i, j \in V \setminus \{x\},$$

$$c_{\hat{G}}(\hat{x}, j) = \min(c_G(x, j), d_G(x) - 2) \ \forall j \in V \setminus \{x\}.$$

Using the first step of Lovász's proof, we obtain that there exists at most one set $S \subset V \cup \{\hat{x}\}$ satisfying

   (a) $x \in S, u \notin S$,
   (b) $|\delta(S)| = c_{\hat{G}}(i, j)$ for some $i, j \in V \cup \{\hat{x}\}$ with $i \in S, j \notin S$ and $i \neq x$,
   (c) S is minimal with respect to the above two conditions.

If such a set S exists we see that $\hat{x} \in S$. Indeed, if $\hat{x} \notin S$ then $|\delta(S)| \geqslant d_G(x)$ since both u and $\hat{x} \notin S$, and $\hat{G}$ is Eulerian. Moreover, $|\delta(S)|$ would be equal to $c_{\hat{G}}(i, \hat{x})$ for some $i \in S \setminus \{x\}$. This follows by (b) and the fact that $|\delta(S \cup \{\hat{x}\})| < |\delta(S)|$ which implies that there is no i and j in $V \setminus \{x\}$ with $i \in S$ and $j \notin S$ such that $|\delta(S)| = c_{\hat{G}}(i, j)$. This leads to a contradiction since $d_G(x) \leqslant |\delta(S)| = c_{\hat{G}}(i, \hat{x}) \leqslant d_{\hat{G}}(\hat{x}) < d_G(x)$.

Moreover, if a set S satisfying (a)–(c) exists then there must exist some $v \in S$ with $v \neq \hat{x}$ such that v is a neighbor of x. Indeed, if S does not contain any neighbor of x other than $\hat{x}$ then $|\delta(S)| \geqslant d_G(x)$ which certainly dominates $c_{\hat{G}}(\hat{x}, j)$. Hence, there would exist some $i \in S, x \neq i \neq \hat{x}$ and $j \notin S$ such that $\delta(S) = c_{\hat{G}}(i, j)$. This leads to a contradiction since $S \setminus \{x, \hat{x}\}$ would separate i from j and $|\delta(S \setminus \{x, \hat{x}\})| < |\delta(S)|$. As a consequence, using the last part of Lovász's proof we see that there exists some $v \neq \hat{x}$ such that by splitting off (x, u) and (x, v) we obtain a graph $\tilde{G}$ with $c_{\hat{G}}(i, j) = c_{\tilde{G}}(i, j) \ \forall i, j \in V \cup \{\hat{x}\} \setminus \{x\}$. By removing $\hat{x}$, we get a graph G' which could have been

obtained from G by splitting off (x, u) and (x, v). G' satisfies $c_{G'}(i, j) = c_{\hat{G}}(i, j) = c_{\hat{G}}(i, j) = c_G(i, j)$ for all $i, j \in V \setminus \{x\}$ and

$$c_{G'}(x, j) = c_{\hat{G}}(x, j) \geq c_{\hat{G}}(\hat{x}, j) = c_{\hat{G}}(\hat{x}, j) = \min(c_G(x, j), d_G(x) - 2)$$

for all $j \in V \setminus \{x\}$. Moreover, due to the splitting operation, $c_{G'}(x, j) \leq c_G(x, j)$ and $c_{G'}(x, j) \leq d_G(x) - 2$. This completes the proof of Lemma 2. □

Appendix 2: Proof of Theorem 9

Using Theorem 8, we need to construct a family of instances for which the ratio $Z_{\text{tree}}(r)/Z_\emptyset(r)$ is arbitrarily close to $2(\sum_{k=1}^{p} (\rho_k - \rho_{k-1})/\rho_k)$.

Let $\alpha_k = \rho_{k+1}/\rho_k$ for $k = 1, \ldots, p-1$. The following procedure constructs an instance whose connectivity types are in $\{\rho_1, \ldots, \rho_p\}$.

Step 1. Take a $2\alpha_{p-1}$-connected $2\alpha_{p-1}$-regular graph G^p with np vertices and $n_p\alpha_{p-1}$ edges. The existence of such a graph can be shown by induction. Assign each vertex of G^p a connectivity type of ρ_p.

Step 2. Replace each edge of G^p by a path with n_{p-1} internal vertices. Each internal vertex is assigned a connectivity type of ρ_{p-1}. Denote the resulting graph by G^{p-1}.

Step 3. For $i = p - 2$ down to 1 do:
- Replace all edges of G^{i+1} by α_i parallel paths, each containing n_i internal vertices of connectivity type ρ_i. Denote the resulting graph by G^i.

Step 4. Let c_e $(e = (i, j))$ be the number of edges in the shortest path in G^1 from i to j.

The cost of the heuristic solution for this instance can be shown to be equal to

$$Z_{\text{tree}}(r) = \sum_{k=1}^{p} \left[(\rho_k - \rho_{k-1}) \left(\prod_{i=1}^{p} n_i \right) \left(\prod_{i=k}^{p-1} \alpha_i \right) \right] + o\left(\prod_{i=1}^{p} n_i \right)$$

$$= \rho_p \left(\prod_{i=1}^{p} n_i \right) \sum_{k=1}^{p} \frac{\rho_k - \rho_{k-1}}{\rho_k} + o\left(\prod_{i=1}^{p} n_i \right).$$

A feasible solution to the LP relaxation can be obtained by assigning a weight x_e of $\frac{1}{2}\rho_1$ to every edge of G^1 and a zero weight otherwise. By computing the cost of this solution, we obtain

$$Z_\emptyset(r) \leq \left(\prod_{i=1}^{p} n_i \right) \frac{1}{2}\rho_1 \left(\prod_{i=1}^{p-1} \alpha_i \right) + o\left(\prod_{i=1}^{p} n_i \right)$$

$$= \frac{1}{2}\rho_p \left(\prod_{i=1}^{p} n_i \right) + o\left(\prod_{i=1}^{p} n_i \right).$$

Therefore, as $n_1, \ldots, n_p \to \infty$, we have

$$\frac{Z_{\text{tree}}(r)}{Z_\theta(r)} \to 2 \sum_{k=1}^{p} \frac{\rho_k - \rho_{k-1}}{\rho_k}$$

which combined with Theorem 8 completes the proof of Theorem 9. $\square$

References

[1] A. Agrawal, P. Klein and R. Ravi, "When trees collide: An approximation algorithm for the generalized Steiner problem in networks," in: *Proceedings of the 23rd ACM Symposium on Theory of Computing* (1991) pp. 134–144.

[2] Y.P. Aneja, "An integer linear programming approach to the Steiner problem in graphs," *Networks* 10 (1980) 167–178.

[3] A. Balakrishnan, T.L. Magnanti and R.T. Wong, "A dual-ascent procedure for large-scale uncapacitated network design," *Operations Research* 37 (1989) 716–740.

[4] E. Balas and P. Toth, "Branch and Bound methods," in: E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys, eds., *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization* (Wiley, New York, 1985) pp. 361–401.

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

[6] N. Christofides, "Worst-case analysis of a new heuristic for the traveling salesman problem," Technical Report, GSIA, Carnegie-Mellon University (Pittsburgh, PA, 1976).

[7] G. Cornuejols, M.L. Fisher and G.L. Nemhauser, "Location of bank accounts to optimize float," *Management Science* 23 (1977) 789–810.

[8] J. Edmonds, "Maximum matching and a polyhedron with 0–1 vertices," *Journal of Research of the National Bureau of Standards* 69B (1965) 125–130.

[9] J. Edmonds, "Submodular functions, matroids and certain polyhedra," in: R. Guy et al., eds., *Combinatorial Structures and their Applications* (Gordon and Breach, London, 1970), pp. 69–87.

[10] C. El-Arbi, "Une heuristique pour le problème de l'arbre de Steiner," *RAIRO Operations Research* 12 (1978) 207–212.

[11] H. Frank and I.T. Frisch, *Communication, Transmission and Transportation Networks* (Addison-Wesley, Reading, MA, 1971).

[12] M.X. Goemans and D. Bertsimas, "Probabilistic analysis of the Held and Karp lower bound for the Euclidean traveling salesman problem," *Mathematics of Operations Research* 16 (1991) 72–89.

[13] M.X. Goemans and D.P. Williamson, "A general approximation technique for constrained forest problems," in: *Proceedings of the third ACM-SIAM Symposium on Discrete Algorithms* (1992) pp. 307–315.

[14] R.E. Gomory and T.C. Hu, "Synthesis of a communication network," *SIAM Journal on Applied Mathematics* 9 (1961) 551–570.

[15] M. Held and R.M. Karp, "The traveling-salesman problem and minimum spanning trees," *Operations Research* 18 (1970) 1138–1162.

[16] M. Held and R.M. Karp, "The traveling salesman problem and minimum spanning trees: Part II," *Mathematical Programming* 1 (1971) 6–25.

[17] D. Johnson, talk presented at the *Mathematical Programming Symposium, Tokyo* (1988).

[18] D. Johnson, personal communication (1989).

[19] R.M. Karp, "Reducibility among combinatorial problems", in: R.E. Miller and J.W. Thatcher, eds., *Complexity of Computer Computations* (Plenum, New York, 1972) pp. 85–103.

[20] R.M. Karp, "Probabilistic analysis of partitioning algorithms for the traveling salesman problem in the plane," *Mathematics of Operations Research* 2 (1977) 209–224.

[21] L. Kou, G. Markowsky and L. Berman, "A fast algorithm for Steiner trees," *Acta Informatica* 15 (1981) 141–145.

[22] J.B. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem," *Proceedings of the American Mathematical Society* 7 (1956) 48–50.

[23] E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys, eds., *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization* (Wiley, New York, 1985).

[24] L. Lovász, "On some connectivity properties of Eulerian graphs," *Acta Mathematica Academiae Scientiarum Hungaricae* 28 (1976) 129–138.

[25] N. Maculan, "The Steiner problem in graphs," *Annals of Discrete Mathematics* 31 (1987) 185–212.

[26] K. Mehlhorn, "A faster approximation algorithm for the Steiner problem in graphs," *Information Processing Letters* 27 (1988) 125–128.

[27] C.L. Monma and C.-W. Ko, "Methods for designing survivable communication networks," NATO Advanced Research Workshop (Denmark, 1989).

[28] C.L. Monma, B.S. Munson and W.R. Pulleyblank, "Minimum weight two-connected spanning networks," *Mathematical Programming* 46 (1990) 153–171.

[29] C.L. Monma and D.F. Shallcross, "Methods for designing communication networks with certain two-connected survivability constraints," *Operations Research* 37 (1989) 531–541.

[30] J. Plesnik, "A bound for the Steiner tree problem in graphs," *Mathematica Slovaca* 31 (1981) 155–163.

[31] R.C. Prim, "Shortest connection networks and some generalizations," *The Bell System Technical Journal* 36 (1957) 1389–1401.

[32] S. Sahni and T. Gonzalez, "P-complete approximation problems," *Journal of the Association for Computing Machinery* 23 (1976) 555–565.

[33] D. Shmoys and D. Williamson, "Analyzing the Held–Karp TSP bound: A monotonicity property with application," *Information Processing Letters* 35 (1990) 281–285.

[34] S. Sridhar and R. Chandrasekaran, "Integer solution to synthesis of communication network," Technical Report, The University of Texas (Dallas, TX, 1989).

[35] K. Steiglitz, P. Weiner and D.J. Kleitman, "The design of minimum-cost survivable networks," *IEEE Transactions on Circuit Theory* CT-16(4) (1969) 455–460.

[36] H. Takahashi and A. Matsuyama, "An approximate solution for the Steiner problem in graphs," *Mathematica Japonica* 24 (1980) 573–577.

[37] P. Winter, "Steiner problem in networks: a survey," *Networks* 17 (1987) 129–167.

[38] L.A. Wolsey, "Heuristic analysis, linear programming and Branch and Bound," *Mathematical Programming Study* 13 (1980) 121–134.

[39] R.T. Wong, "A dual ascent approach for Steiner tree problems on a directed graph," *Mathematical Programming* 28 (1984) 271–287.