Scientia Iranica D (2021) 28(3), 1479–1492

SCIENTIA IRANICA

Sharif University of Technology
Scientia Iranica
Transactions D: Computer Science & Engineering and Electrical Engineering
http://scientiairanica.sharif.edu

An approximation algorithm for the balanced capacitated minimum spanning tree problem

H. Fallah, F. Didehvar*, **and F. Rahmati**

Department of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, P.O. Box 15875-4413, Iran.

Received 27 August 2019; received in revised form 21 September 2020; accepted 18 October 2020

KEYWORDS
Capacitated minimum spanning tree problem;
Load balancing;
Approximation algorithms;
PTAS.

Abstract. Capacitated Minimum Spanning Tree Problem (CMSTP), a well-known combinatorial optimization problem, holds the central place in telecommunication network design. This problem involves finding a minimum cost spanning tree with an extra cardinality limitation on the orders of the subtrees incident to a certain root node. The Balanced Capacitated Minimum Spanning Tree Problem (BCMSTP) is a special case that aims to balance the orders of the subtrees. This problem is an NP-hard one and presents two approximation algorithms in this paper. By considering the maximum order of the subtrees Q, a $(3-\frac{1}{Q})$-approximation algorithm was provided to find a balanced solution. This result was improved to a $(2.5+\epsilon)$ approximation algorithm (for every given $\epsilon > 0$) in the $2d$-Euclidean spaces. Also, a Polynomial Time Approximation Scheme (PTAS) was presented for CMSTP.

1. Introduction

Capacitated Minimum Spanning Tree Problem (CMSTP) is a fundamental problem in the telecommunication network planning. In this problem, we are given an undirected graph with non-negative costs on its edges and non-negative weights on its nodes. Also, the given inputs include a root node r and a capacity constraint Q. The objective is to find a minimum cost spanning tree rooted at r in which the sum of the vertex weights in each rooted subtree (that indicates its load) is at most Q. In the absence of any capacity constraint, the problem is limited to finding a minimum cost spanning tree. A special case is when all the node weights are equal and known as the homogeneous demand case. This is equivalent to the case where all the node weights are units and usually referred to as CMSTP in the literature [1]. In this case, the problem is limited to finding a rooted minimum spanning tree in which each of the subtrees incident to the root contains at most Q nodes. CMSTP considers the unit demands investigated in this paper.

Many CMSTP variations are formulated depending on the type of applications (e.g., see [2–8]). A variety of CMSTPs consider additional constraints like the balance of the number of nodes in component subtrees (see [2,9]). Ali and Huang [2] tackled the spanning trees and forests with a number of balanced nodes in the component subtrees. Incel et al. [9] presented a practical application for this problem in wireless sensor networks. This is the problem we consider in this article paper and refer to it as the Balanced Capacitated Minimum Spanning Tree Problem (BCMSTP).

It can be demonstrated that BCMSTP is an NP-hard problem and provides two approximation algo-

*. *Corresponding author. Tel.: +98 21 64545665*
E-mail addresses: hfallah@aut.ac.ir (H. Fallah); didehvar@aut.ac.ir (F. Didehvar); frahmati@aut.ac.ir (F. Rahmati)

doi: 10.24200/sci.2020.54242.3663

rithms for this problem. An approximation algorithm runs in time polynomial with input size and ensures a high-quality solution [10]. The approach of Vazirani [10] was used to define an approximation algorithm and approximation guarantee. For the minimization problem, an approximation algorithm achieves the approximation ratio ρ if the ratio between the cost of the found solution by the algorithm and the cost of an optimal solution is at most ρ in every instance of the problem.

Also, a Polynomial Time Approximation Scheme (PTAS) was provided for CMSTP. A PTAS is an approximation algorithm whose ratio is $(1+\epsilon)$ for every constant $\epsilon > 0$.

Here, we confine our focus to Euclidean metrics and mention it as the Euclidean Balanced Capacitated Minimum Spanning Tree Problem (Euclidean BCMSTP). From the theoretical point of view, we seek to identify whether it is solvable and more efficient than BCMSTP in general metrics.

BCMSTP has practicable applications in converge cast wireless sensor networks [9] that aim to collect data from a set of sensors toward a common link such that the schedule length becomes a minimum. Load balancing in a CMSTP minimizes the largest load and the schedule length.

In a different scenario, the root is the central processor and the other nodes are terminals that demand traffic and must be sent to the central processor through the edges. The edge traffic is the number of traffic demands passing through it; therefore, maximum traffic appears on the gates which incident edges to the root. Balancing the loads minimizes the maximum traffic on the gates, thus maximizing the network robustness in case of an unpredictable growth of traffic demand. In fact, it reduces the likelihood of overfilling gates with additional load. Such a situation may arise when new terminals move into the network, or if the connecting network is redefined when modifying preexisting connections.

Generally, the applications of BCMSTP are in the same fields as those of CMSTP, e.g., in telecommunication network design, the local access network design in computer communication networks, the design of the local loop systems between the branch offices and final users in telephone systems, facility location planning, distribution, transportation, logistics, and telecommunications companies building fiber-optic based local access networks.

This problem appears in the design of bus routes in urban transportation systems where fair distribution of services is of importance [11]. Here, the solution is a set of Hamiltonian paths connected to the root. Accordingly, BCMSTP is a variant of the open vehicle routing problem [12] that considers a fairness issue across its distribution services.

1.1. Our results

To the best of our knowledge, we did not find any literature on an approximation algorithm for BCMSTP. The first $(3 - \frac{1}{Q})-$approximation algorithm was presented for this problem in general metrics. This result was improved to a $(2.5 + \epsilon)-$approximation (for every given $\epsilon > 0$), in $2-$dimensional ($2d$) Euclidean spaces. Also, a PTAS for CMSTP in $2d-$Euclidean spaces was provided when $Q = o(ln ln\, n)$ (n is the number of nodes). This result solved the problem of obtaining a ratio any better than 2.9 [13] for this case. As far as we are concerned, there is no previous attempt to derive a PTAS for CMSTP in $m-$dimensional Euclidean spaces for any fixed $m \geq 2$.

1.2. Related works

BCMSTP pertains to the class of balanced combinatorial optimization problems, which was first introduced by Martello et al. [14]. Many generalizations and variations of these problems have been studied by several authors, e.g. [2,15–22].

For CMSTP with unit demands, Gavish and Altinkemer in [23] presented a factor $4 - 1/(2^{\lceil logQ \rceil - 1})$ approximation algorithm. Then, Altinkemer and Gavish [24] improved this ratio to $3 - \frac{2}{Q}$ and 4 for the unit and non-unit demands, respectively. Later, Jothi and Raghavachari [13] improved the ratio of 4 to $\gamma + 2$ in the non-unit demand case, where γ is the inverse Steiner ratio. The Steiner ratio is the topmost proportion of the costs for the minimum Steiner tree against the Minimum Spanning Tree (MST) for the same instance. In graphs $\gamma = 2$ and in Euclidean and rectilinear metrics, it is $2/\sqrt{3}$ and $3/2$, respectively. Moreover, they obtained a $2.9-$approximation for the unit demand CMSTP on the L_p-metric plane, and a $2-$approximation for this problem in the general metrics when $Q \in \{3, 4\}$.

Exact solution methods for CMSTP (unit or non-unit demands) were developed by Gavish [25] and Kershenbaum and Boorstyn [26]. There are several algorithms and mathematical formulations [4,27–30] available for solving this problem. Gamvros et al. [31] studied a Multi-Level Capacitated Minimum Spanning Tree (MLCMST) problem with unit demands on tree topology networks. In their paper, the authors assume that there are multiple types of a facility (maybe a cable) with different capacities and costs which can be used to be installed between two nodes to carry the traffic. In comparison to our problem, we assume that there is a single facility type with capacity of Q. Also, we assume that the graph is complete and one facility could be installed between any two nodes at most.

The local access network design problem (see [32–34]) is one relevant problem that deals with multiple facility types. However, in this problem, the topology of the network is not a tree. A related problem dealing with multiple facility types is the Single Sink Buy-at-Bulk (SSBB) problem [35]. Hassin et al. [5] provided algorithms for the single facility SSBB problem which is a special variant of CMSTP known as the network loading problem in the literature and where multiple copies of the facility might be installed on a graph edge. Hassin et al. [5] prepared approximations of factor 2 for unit demand and 3 for non-unit demand problems. A variation of CMSTP with the connectivity constraint was considered by Jothi and Raghavachari [36].

The rest of the paper is structured as follows. Section 2 presents the notations, definitions, and important assumptions. Section 3 analyzes the complexity of the problem. Section 4 discusses the balanced solution. Section 5 addresses BCMSTP (general metrics) and provides a $(3-\frac{1}{Q})-$approximation algorithm. Section 6 addresses this problem in $2d-$Euclidean metric spaces. An improved $(2.5+\epsilon)-$approximation is prepared for every given $\epsilon > 0$. Also, a PTAS is presented for CMSTP. The conclusion section summarizes the results. An integer programing model is given for the problem in the Appendix.

2. Notations, assumptions, and problem definition

BCMSTP can be described in the following. Assume that $G = (V, E)$ is a complete graph where $V = \{r, v_1, v_2, ..., v_n\}$ is a set of nodes and E is a set of edges. Here, $r \in V$ is the root node. Each edge $(i, j) \in E$ has the weight of $c_{ij} > 0$ that denotes its length/cost. The weights on the edges are symmetric (i.e. $c_{ij} = c_{ji}$, for each $(i, j) \in E$) and follow the triangle inequality. Each node $v \in V \backslash \{r\}$ has unit demand. We will interchangeably use such terms as vertices, nodes, and terminals. Let T_v be a subtree dangling from v and the cost $C(T_v)$ indicates the sum of the edge costs incident in T_v. Moreover, the load $d(T_v)$ is the number of nodes (except root) incident in T_v that indicates its order. We seek to find a minimum cost spanning tree of G, so that order of each subtree dangling from the root does not surpass the capacity Q and orders of the subtrees are balanced. At least, $\lceil n/Q \rceil$ number of subtrees is required.

The number of subtrees may be a fixed parameter K, given by the user. However, it would be a decision variable that must be determined. When K is fixed, we seek to find K subtrees exactly. In the rest of the paper, we assume that K is fixed, except where noted.

A solution $S = \{T_1, T_2, ..., T_K\}$ of BCMSTP corresponds to a capacitated partition $P_1, P_2, ..., P_K$ which satisfies the following relations:

$$P_i \neq \emptyset, \forall i; \qquad \bigcup_{1 \leq i \leq K} P_i = V \backslash \{r\};$$

$$P_i \cap P_j = \emptyset, i \neq j; \qquad \sum\nolimits_{v \in P_i} 1 \leq Q, i \in [K].$$

Indeed, for each $T_i \in S$, we have $P_i = \{v : v \neq r, v \in T_i\}$. We say the nodes in P_i are allocated to the i^{th} subtree.

To avoid ambiguity, we notice that the notation T_v (where v is a terminal node) denotes a subtree dangling from v, while the notation like T_i (where the index i is not a terminal node) is used to denote a subtree dangling from the root r.

Let $L_S = (d(T_1), d(T_2), ..., d(T_K))$ be the loads of the subtrees in S that specify the "allocated loads" of S. In the next section, a "balanced load allocation" and a "balanced solution" are defined.

A mathematical programing model for BCMSTP is suggested in the Appendix. The model shows that BCMSTP is indeed CMSTP with extra limitations for the balance condition. The problem is NP-hard (see Section 3) and finding an approximation algorithm does not seem to be an easy task [37,38]. We prepare two approximation algorithms. The first algorithm $\Im$ relies on a tour partitioning heuristic and is valid in any metric space. The second algorithm $A(\Im)$ is applicable in the $2d-$Euclidean spaces. It separates the nodes into two parts: in the interior part, it employs $\Im$, and in the exterior part, it finds an optimal solution. The main idea is to separate the nodes such that the optimal solution in the exterior part could be found within the polynomial time with respect to n (i.e. the number of nodes). We will employ the following basic definition.

Definition 1. Let Γ be an approximation algorithm that finds a solution S_Γ with cost $C(S_\Gamma)$ to the minimization problem. Let OPT be an optimal value. The relative error of Γ is defined by the equation:

$$e^\Gamma = \frac{C(S_\Gamma) - OPT}{OPT}.$$

The approximation ratio of Γ is $\rho(\Gamma) = \frac{C(S_\Gamma)}{OPT}$.

2.1. The balance criteria

The equity measure is an indicator that evaluates the fairness of an allocated load vector. One of the accepted formulations for the notion of fairness in the load balancing domain is the range fairness that calculates the maximum distinction between the loads (e.g., see [39]). Another equity measure is the ratio fairness that calculates the maximum ratio between the loads [40,21].

Both the range and the ratio criteria comply with the weak Pigou-Dalton (PD) principle which is a widely accepted property of equity measures [41]. For an allocated load vector x and an equity function $I(x)$, let the vector x' be organized as follows: $x'_j = x_j + \delta$, $x'_i = x_i - \delta$, $x'_h = x_h$ for all $h \notin \{j, i\}$. The PD principle (weak version) declares that $I(x') \leq I(x)$ when $0 \leq \delta <$

$x_j - x_i$ is chosen. According to the strong PD principle, the inequality is strict ($I(x') < I(x)$), implying that the new allocation must be more equitable. The PD principle can be used in situations where the sum of allocation outcomes and their number are identical.

The range fairness is used to evaluate the balance of allocated loads, although the ratio fairness can also be used. Given a real parameter $\alpha \geq 0$ and the number of subtrees K, a partition of the nodes of G into K parts $\{P_i : i \in [K]\}$ is $\alpha-$balanced if and only if $j \in [K]$ for each i:

$$|Load(P_i) - Load(P_j)| \leq \alpha,$$

where $Load(P_i) = \sum_{v \in P_i} 1$ and $|x|$ is the absolute value of x. This condition results in the following inequality:

$$\max\{Load(P_i) : i \in [K]\} - \min\{Load(P_i) : i \in [K]\} \leq \alpha. \tag{1}$$

Inequality (1) is the "balanced condition". Evidently, an α_1-balanced partition is indeed α_2-balanced for each $\alpha_2 \geq \alpha_1$. The balanced range of the set $\{P_i : i \in [K]\}$ is the smallest α satisfying the balanced condition. Hence, $\max\{Load(P_i) : i \in [K]\} - \min\{Load(P_i) : i \in [K]\}$ is the balanced range of the set $\{P_i : i \in [K]\}$.

Definition 2. A solution S of BCMSTP with the related partition P is $\alpha-$balanced (for a given parameter $\alpha \geq 0$) if and only if P is $\alpha-$balanced. The solution S is balanced if the balanced range of P is the least possible among the balanced ranges of the other capacitated partitions into K parts.

When the ratio measure is used, the partition $\{P_i : i \in [K]\}$ is said to be $\beta-$balanced (for a given parameter $0 \leq \beta \leq 1$) if $j \in [K]$ for each i:

$$1 - \beta \leq \frac{Load(P_i)}{Load(P_j)} \leq 1 + \beta,$$

which simply results in the following inequality:

$$\frac{\max\{Load(P_i) : i \in [K]\}}{\min\{Load(P_i) : i \in [K]\}} \leq 1 + \beta. \tag{2}$$

Let α ($0 \leq \alpha \leq Q$) be a given load deviation and let $\beta \leq \alpha/Q$ be chosen. A $\beta-$balanced set $\{P_i : i \in [K]\}$ is also $\alpha-$balanced: we see $-\frac{\alpha}{Q} \leq -\beta \leq \frac{Load(P_i)}{Load(P_j)} - 1 \leq \beta \leq \frac{\alpha}{Q}$, so $-\alpha \leq \frac{Q}{Load(P_j)}(Load(P_i) - Load(P_j)) \leq \alpha$.

If $Load(P_i) \geq Load(P_j)$ we have:

$$-\alpha \leq 0 \leq Load(P_i) - Load(P_j) \leq \frac{Q}{Load(P_j)}(Load(P_i) - Load(P_j)) \leq \alpha.$$

If $Load(P_i) \leq Load(P_j)$, we have:

$$-\alpha \leq \frac{Q}{Load(P_j)}(Load(P_i) - Load(P_j)) \leq Load(P_i) - Load(P_j) \leq 0 \leq \alpha.$$

The role of α and β (in Definition 2) is exchanged, when the ratio measure is used.

In the rest of the paper, S denotes a set of solutions to the considered problem, S^* denotes an optimal solution, L is the load of the subtrees, L^* is the load of an optimal solution, T_i denotes a subtree, C denotes the traveling cost, and K indicates the number of subtrees.

3. Complexity analysis

It can be shown that BCMSTP is NP-hard for $Q \geq 3$ and any given $\alpha \geq 0$. Its NP-hardness can be proven by a reduction in CMSTP, which itself is NP-hard [42,43]. The complexity of CMSTP depends on the capacity Q. This problem is solvable in polynomial time if $Q = 2$ [42]. Also, it is solvable in polynomial time if vertices have 0, 1 demands and $Q = 1$ [42]. Although it is NP-hard if vertices have 0, 1 demands, $Q = 2$ and all edge weights are 0 or 1 [42]. Also, it remains NP-hard for any $Q \geq 3$ [42]. Moreover, its geometric version, in which the metric space at the edges is the Euclidean metric, remains NP-hard [43]. Camerini et al. [44,45] illustrated that many variants of this problem had the same complexity.

To demonstrate the NP-hardness of BCMSTP, its decision version is shown to be NP-complete [42]. Assume that I_α for $\alpha \geq 0$ is a decision problem that decides whether a feasible solution to the considered problem exists whose cost is at most a given bound D. Theorem 1 shows that I_α is NP-complete for each $\alpha \geq 0$ and $Q \geq 3$. The results are true when the ratio measure (Inequality (2) with $0 \leq \beta \leq 1$) is used instead of the range measure.

Theorem 1. *For every $\alpha \geq 0$ and $Q \geq 3$, I_α is NP-complete.*

Proof. We show NP-completeness of I_α by a reduction from CMSTP. As mentioned in the last section, CMSTP is NP-hard for $Q \geq 3$. Let I' be an instance of this problem (with capacity constraint Q, cost $c_{ij} > 0$ on each edge (i, j), number of terminals n, number of subtrees $K' = \left\lceil \frac{n}{Q} \right\rceil$, and a bound D) which decides whether a feasible solution exists whose cost is at most D. Clearly, I' is NP-complete.

Let $\Delta = \max\{c_{ij} : i, j \in \{0, 1, ..., n\}\}$. We define an instance I_α of BCMSTP with capacity constraint Q, number of terminals $n' = K'Q$, number of subtrees K', cost $c'_{ij} = c_{ij}$ on each edge (i, j) where $i, j \in \{0, 1, ..., n\}$ and cost $c'_{ij} = c'_{ji} = \Delta$ on each edge where $i \in \{n +$

$1,...,n'\}$, and a bound $D + \Delta(n' - n)$. Indeed, $n' - n$ auxiliary nodes are added to the set of terminals in I' that are at the distance Δ from each other and the others. Evidently, the weights on the edges obey the triangle inequality.

Let $x = \{T_1, T_2, ..., T_{K'}\}$ be a solution for I' with loads $L_1, L_2, ..., L_{K'}$ whose cost is lower than D. We add $Q - L_i$ nodes to the i^{th} subtree. Since the augmented nodes are at the distance Δ from each other and the others and the augmented length is $\Delta(n' - n)$. Therefore, a solution is obtained for I_α. Now, let x be a solution for I_α whose cost is lower than $D + \Delta(n' - n)$. The augmented nodes are removed; therefore a feasible solution is obtained for I' whose cost is lower than D. This proves the NP-completeness of I_α, for each $\alpha \geq 0$.

4. Load balancing

In BCMSTP, a set of capacitated subtrees dangling from the root is required. Thus, a solution has two portions: the number of nodes that each subtree receives and the choice of them. The number of the allocated nodes should be balanced. A solution is balanced if its balanced range is the least possible among the balanced ranges of all feasible solutions (see Definition 2). An algorithm has been constructed to find a balanced allocation of the nodes to the subtrees. We show that this allocation is the fairest in a sense. First, the fairest solution is defined.

We describe a partial order (identified with $\prec$) on the class of the allocated load vectors. For the sorted (in non-decreasing order) load vectors $l' = (l'_1, l'_2, ..., l'_K)$ and $l = (l_1, l_2, ..., l_K)$, we assume $R_{l'} = (l'_K - l'_1, ..., l'_2 - l'_1, 0)$, $R_l = (l_K - l_1, ..., l_2 - l_1, 0)$. We say $l' \prec l$, iff $R_{l'}$ is smaller than R_l in lexicographical order, that is, $R_l = R_{l'}$ or there is an index j for which $(l'_j - l'_1) < (l_j - l_1)$ and $l'_i - l'_1 = l_i - l_1$ for all $i < j$. If $l' \prec l$ and $l \prec l'$, the vectors l and l' are called equivalent. Thus, the equivalence classes admit a total order. The minimal equivalence class under $\prec$ contains the fairest allocations.

Definition 3. A solution S for BCMSTP with the allocated loads L_S is the fairest solution if and only if L_S is the fairest load vector.

Definition 3 reveals that the fairest solution is indeed balanced since its balanced range is the least possible among the balanced ranges of all the feasible solutions. However, in general, the balanced solution is not the fairest. Here, we restate the algorithm proposed by the authors in [40] to find balanced loads (see Algorithm 1).

Assume that each terminal corresponds to an object (with a volume of 1) in the set $O = \{o_1, ..., o_n\}$ and each subtree corresponds to a bin (with capacity Q) in the set $M = \{M_1, ..., M_K\}$. Each object is assigned to a bin, and the found solution $L^* = (L^*_1, ..., L^*_K)$ determines the number of objects each bin receives. Indeed, L^* is the allocated load to the K subtrees. An assignment is a function $L : O \to M$ so that L assigns each object o_j to a bin in M. In the assignment L, the degree of each bin is the number of objects assigned to it.

Lemma 1. *Algorithm 1 obtains the balanced loads in polynomial time.*

Proof. The proof is the same to the proof of Lemma 1 in [40].

In the rest of the paper, the sequence $L^*_1 \leq L^*_2 \leq ... \leq L^*_K$ denotes the balanced loads. Two possible cases are: $L^*_K = L^*_1 + 1$, or $L^*_K = L^*_1$. We suppose $L^*_K \geq 3$ and search for a set of subtrees with balanced loads.

5. BCMSTP equipped with general metrics

Our algorithm for BCMSTP relies on a method called route first-cluster second [46,24]. A special partitioning procedure that obtains the fairest solution is proposed in this case. Algorithm 2 presents this procedure in detail, referred to as $\Im$.

Let C_{BCMSTP} be the cost of an optimal solution for BCMSTP. A lower bound for C_{BCMSTP} is given in Lemma 2. Note that c_{rv} is the edge cost between r and v.

Lemma 2. $C_{BCMSTP} \geq \frac{\sum_{v \in V} c_{rv}}{L^*_K}$.

Proof. Suppose $S^* := \{T^*_1, T^*_2, ..., T^*_K\}$ is an optimal solution for BCMSTP and $T^*_i \in S^*$ is a subtree with cost $C(T^*_i)$ and let $c^{\max}_{rv} = \max\{c_{rv} : v \in T^*_i\}$. Since

---

input : A set of similar objects $O = \{o_1, ..., o_n\}$ that should be assigned to the set of bins $M = \{M_1, ..., M_K\}$.
output : Optimum assignment L^*.

Start
Assign $\lfloor \frac{n}{K} \rfloor$ objects to every bin.
Allocate an object to each bin in the set $\{M_{K-(n-\lfloor \frac{n}{K} \rfloor)+1}, ...M_{K-1}, M_K\}$.
Return degrees of the bins: $L^* = (L^*_1, ..., L^*_K)$.
End

Algorithm 1. Load balancing algorithm.

input : BCMSTP with balanced loads L^*.
output : K balanced rooted subtrees.

Start
Find a rooted minimum spanning tree $MST(V)$ of the graph $G=(V,E)$. Let $\tau := (r, v_1, v_2, ..., v_n, r)$ be an Eulerian tour of $MST(V)$.
for $i = 1$ to L_1^* **do**
Start at v_i and identify the forthcoming subtrees:
$S_i = \{T_1^i = (r, v_i, v_{i+1}, ..., v_{i+L_K^*-1}), T_2^i = (r, v_{i+L_K^*}, ..., v_{i+L_K^*+L_{K-1}^*-1}), ...,$
$T_K^i = (r, v_{i-L_1^*}, ..., v_{i-1})\}$.
Find the total cost $C(S_i) = \sum_{j=1}^{j=K} C(T_j^i)$.
end for
Restore the solution $S_p := \{T_1^p, T_2^p, ..., T_K^p\}$, $1 \le p \le L_1^*$ having the smallest total cost $C(S_p)$.
End

Algorithm 2. Approximation algorithm for the balanced capacitated minimum spanning tree problem $\Im$.

$d(T_i^*) = \sum_{v(\neq r)\in T_i^*} 1 \le L_K^*$, we have:

$$C(T_i^*) \ge c_{rv}^{\max} = \frac{\sum_{v(\neq r)\in T_i^*} 1}{\sum_{v(\neq r)\in T_i^*} 1} c_{rv}^{\max}$$

$$\ge \frac{\sum_{v(\neq r)\in T_i^*} 1}{L_K^*} c_{rv}^{\max} \ge \frac{\sum_{v\in T_i^*} c_{rv}}{L_K^*}.$$

Summing over all the subtrees in S^*, we obtain:

$$C_{BCMSTP} = \sum_{i=1}^{i=K} C(T_i^*) \ge \frac{\sum_{v\in V} c_{rv}}{L_K^*}.$$

Theorem 2. *The total traveling cost of S_p satisfies $C(S_p) \le (3 - \frac{1}{L_1^*})C_{BCMSTP}$.*

Proof. Assume that $C(\tau)$ is the cost of the Eulerian tour τ of the MST. We observe that each vertex $v \in V\backslash\{r\}$ emerges at most once as the initial node of a subtree in all the iterations. Thus, every edge (r, v) emerges at most once during the iterations. When $v \in \tau$ emerges as the initial node of a subtree, the edge $(u, v) \in \tau$ does not appear in that solution. If $v \in \tau$ is not the initial node of any subtree, $(u, v) \in \tau$ appears in all the iterations. An upper bound for the total cost of the found solutions is:

$$\sum_{j=1}^{L_1^*} \sum_{i=1}^{K} C(T_i^j) \le (L_1^* - 1)C(\tau) + \sum_{v\in V} c_{rv}. \tag{3}$$

The right side of Inequality (3) is an upper bound for the total cost since we have added $c_{ru} + c_{rv} - c_{uv}$ to the right-hand side when the edge $(u, v) \in \tau$ emerges in all the iterations. Based on the triangle inequality $c_{ru} + c_{rv} - c_{uv} \ge 0$.

Since S_p is the cheapest among the others, we obtain:

$$L_1^* C(S_p) \le \sum_{j=1}^{L_1^*} \sum_{i=1}^{K} C(T_i^j) \le (L_1^* - 1)C(\tau) + \sum_{v\in V} c_{rv}.$$

Suppose that $L_K^* = L_1^*$. Thus, we have:

$$\begin{aligned} C(S_p) &\le (1 - \tfrac{1}{L_1^*})C(\tau) + \tfrac{\sum_{v\in V} c_{rv}}{L_1^*} \\ &\le (1 - \tfrac{1}{L_1^*})C(\tau) + \tfrac{\sum_{v\in V} c_{rv}}{L_K^*} \\ &\le (2 - \tfrac{2}{L_1^*} + 1)C_{BCMSTP} \\ &= (3 - \tfrac{2}{L_1^*})C_{BCMSTP}. \end{aligned}$$

The third inequality was derived from Lemma 2.

Now, suppose $L_K^* = L_1^* + 1$. Thus:

$$\begin{aligned} C(S_p) &\le (1 - \tfrac{1}{L_1^*})C(\tau) + \tfrac{\sum_{v\in V} c_{rv}}{L_1^*} \\ &\le (2 - \tfrac{2}{L_1^*})C_{BCMSTP} + \tfrac{L_K^*}{L_1^*}(\tfrac{\sum_{v\in V} c_{rv}}{L_K^*}) \\ &\le (3 - \tfrac{1}{L_1^*})C_{BCMSTP}. \end{aligned}$$

where Lemma 2 was used.□

In addition, $3 - \frac{1}{L_1^*} \le 3 - \frac{1}{Q}$ since $L_1^* \le Q$. Therefore, the proposed algorithm provides $3 - \frac{1}{Q}$ factor of approximation for BCMSTP. In the following, BCMSTP was considered in the $2d-$Euclidean metric spaces and a better approximation algorithm was produced.

6. BCMSTP equipped with $2d-$Euclidean metrics

First, CMSTP is studied in the plane and a PTAS is provided to solve this problem. Then, a similar technique is used to provide a factor $2 + \frac{1}{L_1^*} + \epsilon$ approximation algorithm for BCMSTP for every given $\epsilon > 0$. To prove the performance ratios of the algorithms, it is required to find an upper bound for the MST in the plane. This matter is elaborated in the next section.

6.1. Approximation of MST in $\mathbb{R}^2$

A technique similar to that of given in [47] has been used to find an upper bound for the MST. Let $C(MST(V))$ be the cost of the MST of a complete graph defined on the set V, $c_{rv}^{\max} = \max\{c_{rv} : v \in V\}$, and $\bar{C}_{rv} = \frac{\sum_{v\in V} c_{rv}}{n}$; thus, we have the following theorem.

Theorem 3. $C(MST(V)) \le 2\sqrt{\pi n c_{rv}^{\max} \bar{C}_{rv}}$.

Proof. We partition the circle of radius $c_{rv}^{\max}$ into $4h$ equal sectors. The boundaries of the sectors are used to construct two star-shaped trees partitioning the circle (see Figure 1). Each tree is converted into a spanning tree by a double connection of minimal length from each point; the sum of these double connections is less than $2(2\pi/4h)c_{rv_i} = \pi c_{rv_i}/h$ due to $a \leq c_{rv_i}\sin\theta \leq c_{rv_i}\theta$ (see Figure 1). Hence, the sum of the costs of the trees is less than:

$$\pi n\bar{C}_{rv}\frac{1}{h} + 4hc_{rv}^{\max}$$

and we conclude that:

$$C(MST(V)) \leq \pi n\bar{C}_{rv}\frac{1}{2h} + 2hc_{rv}^{\max}.$$

By taking h equal to the minimized value on the right-hand side and rounding it up, i.e., $h = \left\lceil\sqrt{\pi n\bar{C}_{rv}/4c_{rv}^{\max}}\right\rceil$, the desired result can be achieved.□

6.2. PTAS for CMSTP equipped with 2d−Euclidean metrics

The tour partitioning heuristic proposed was used in [24] (see Lemma 3), as represented below by Γ.

Lemma 3. *There exists an approximation algorithm for CMSTP with performance ratio of* $3-\frac{2}{Q}$.

Proof. See the proof in [24].

Using the algorithm Γ, we construct PTAS for the $2d-$Euclidean CMSTP and refer to it as $A(\Gamma)$. The main objective of the algorithm is to separate the nodes into two parts: in the interior part, it employs Γ and in the exterior part, it finds an optimal solution. The idea is to separate the nodes such that the optimal solution in the exterior part could be found in polynomial time with respect to n (number of nodes). Khachay and Dubinin [48] used a similar technique to find PTAS for the capacitated vehicle routing problem in the Euclidean space.

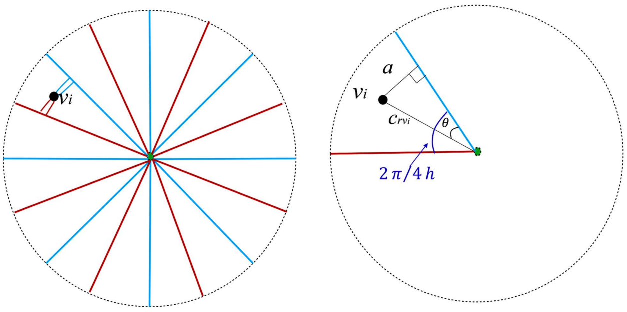

Figure 1. Approximation of Minimum Spanning Tree (MST) in R^2.

Let $S^*_{CMST} := \{T_1^{*CMST}, T_2^{*CMST}, ..., T_k^{*CMST}\}$ be a set of optimal subtrees for CMSTP with optimal cost $C(S^*_{CMST})$. Haimovich and Rinnooy Kan [47] proved that $\frac{n}{Q}\bar{C}_{rv} = \frac{\sum_{v\in V}c_{rv}}{Q}$ was a lower bound for the optimal cost: $\frac{n}{Q}\bar{C}_{rv} \leq C(S^*_{CMST})$. Altinkemer and Gavish [24] developed an upper bound of $2C(MST(V)) + \frac{n}{Q}\bar{C}_{rv}$ for the optimal cost and derived $(3-\frac{2}{Q})-$approximation ratio using this upper bound. Based on the results of the authors in [24,47] the following inequalities were achieved:

$$\frac{n}{Q}\bar{C}_{rv} \leq C(S^*_{CMST}) = \sum_{i=1}^{k} C(T_i^{*CMST})$$
$$\leq 2C(MST(V)) + \frac{n}{Q}\bar{C}_{rv}. \quad (4)$$

The algorithm $A(\Gamma)$ is presented in Algorithm 3.

Theorem 4. *The algorithm $A(\Gamma)$ is a PTAS for the Euclidean CMSTP.*

Proof. It can be illustrated that for any $\epsilon > 0$, the relative error $e^{A(\Gamma)}(V)$ satisfies $e^{A(\Gamma)}(V) \leq \epsilon$. First, we find an upper bound for $e^{A(\Gamma)}(V)$.

Let I be a given instance of CMSTP with an optimal solution $S^*_{CMST}(V)$. Consider the circle with radius $c_l = c_{rv_l}$ centered at the root (we will later determine l). We connect the nodes in $V(l)$ incident to the edges between $V(l)$ and $V\backslash V(l)$ directly to the root (see Figure 2). Let $S^*_{CMST}(V(l)), S^*_{CMST}(V\backslash V(l))$ be the optimal solutions of the problems defined on the

---

input : An instance of CMSTP.
output : A feasible solution for CMSTP.

Start
Enumerate terminals by decreasing their distance from the root

$$c_{rv_1} \geq c_{rv_2} \geq ... \geq c_{rv_n}.$$

Take the set $V(l) = \{v_1, v_2, ..., v_{l-1}\}$ of outside nodes. We will later determine the value of l, and will show that it is independent of n.
Apply Γ to the set of inside nodes $V\backslash V(l)$. Denote the obtained solution by $S_{CMST}(V\backslash V(l))$.
Find the optimal solution for CMSTP defined on the set of outside nodes $V(l)$ and the same root. Denote the found solution by $S^*_{CMST}(V(l))$.
Return $S_{CMST}(V) := S_{CMST}(V\backslash V(l)) \cup S^*_{CMST}(V(l))$.
End

Algorithm 3. PTAS for the capacitated minimum spanning tree problem $A(\Gamma)$.

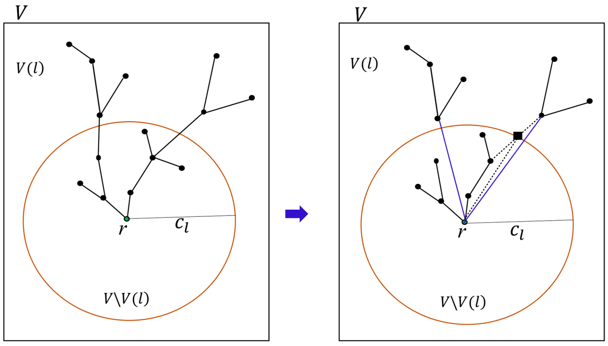

Figure 2. An optimal solution of the Capacitated Minimum Spanning Tree Problem (CMSTP) defined on the set of nodes V which can be transformed into a solution for CMSTP defined on the set of nodes $V(l)\cup\{r\}$ and a solution for CMSTP defined on the set $V/V(l)$ by deleting and adding some edges.

sets $V(l)\cup\{r\}, V\backslash V(l)\cup\{r\}$, respectively, and let $C(S^*_{CMST}(V(l))), C(S^*_{CMST}(V\backslash V(l)))$ be their respective costs. We have:

$$C(S^*_{CMST}(V(l)))+C(S^*_{CMST}(V\backslash V(l)))$$

$$\leq C(S^*_{CMST}(V))+(l-1)c_l.$$

From the $A(\Gamma)$ algorithm:

$$C(S_{CMST}(V))=C(S^*_{CMST}(V(l)))$$

$$+C(S_{CMST}(V\backslash V(l))).$$

Since $C(S^*_{CMST}(V\backslash V(l)))\geq \sum_{i=l}^{n}\frac{c_{rv_i}}{Q}$ (see Inequality (4)), we get:

$$C(S_{CMST}(V))\leq C(S^*_{CMST}(V(l)))$$

$$+C(S^*_{CMST}(V\backslash V(l)))$$

$$+\left(C(S_{CMST}(V\backslash V(l)))-\sum_{i=l}^{n}\frac{c_{rv_i}}{Q}\right).$$

Thus, we obtain the equation shown in Box I. From Inequality (4) and Theorem 3, we have:

$$C(S^*_{CMST}(V))\geq \frac{n}{Q}\bar{C}_{rv},$$

and:

$$C(S_{CMST}(V\backslash V(l)))\leq 2C(MST(V\backslash V(l)))$$

$$+\frac{n}{Q}\bar{C}_{rv}(V\backslash V(l))\leq 4\sqrt{\pi n c_{rv}^{\max}\bar{C}_{rv}}$$

$$+\frac{\sum_{i=l}^{n}c_{rv_i}}{Q}.$$

Thus, we obtain:

$$e^{A(\Gamma)}(V)\leq \frac{(l-1)c_l}{\frac{n}{Q}\bar{C}_{rv}}$$

$$+\frac{C(S_{CMST}(V\backslash V(l)))-\frac{\sum_{i=l}^{n}c_{rv_i}}{Q}}{\frac{n}{Q}\bar{C}_{rv}}$$

$$\leq \frac{Q(l-1)c_l}{\sum_{i=1}^{n}c_{rv_i}}$$

$$+Q\left(\frac{4\sqrt{\pi n c_{rv}^{\max}\bar{C}_{rv}}+\frac{\sum_{i=l}^{n}c_{rv_i}}{Q}-\frac{\sum_{i=l}^{n}c_{rv_i}}{Q}}{\sum_{i=1}^{n}c_{rv_i}}\right)$$

$$\leq \frac{Q(l-1)c_l}{\sum_{i=1}^{n}c_{rv_i}}+Q\frac{4\sqrt{\pi n c_l\bar{C}_{rv}}}{\sum_{i=1}^{n}c_{rv_i}}$$

$$\leq Ql\frac{c_l}{\sum_{i=1}^{n}c_{rv_i}}+4Q\sqrt{\pi}\sqrt{\frac{c_l}{\sum_{i=1}^{n}c_{rv_i}}}. \quad (5)$$

Now, we choose l such that Inequality (5) is less than ϵ. For large values of l, the right-hand side of Inequality (5) is smaller than ϵ since the algorithm finds an optimal solution in the exterior part. However, its running time is exponential concerning l. We need to choose l such that CMSTP defined on the set of outside nodes could be solved in polynomial time concerning n. Suppose that l is selected to be the smallest number (from n to 1) for which Inequality (5) is less than ϵ,

$$e^{A(\Gamma)}(V)=\frac{C(S_{CMST}(V))-C(S^*_{CMST}(V))}{C(S^*_{CMST}(V))}$$

$$\leq \frac{C(S^*_{CMST}(V(l)))+C(S^*_{CMST}(V\backslash V(l)))+C(S_{CMST}(V\backslash V(l)))-\frac{\sum_{i=l}^{n}c_{rv_i}}{Q}-C(S^*_{CMST}(V))}{C(S^*_{CMST}(V))}$$

$$\leq \frac{(l-1)c_l+C(S_{CMST}(V\backslash V(l)))-\frac{\sum_{i=l}^{n}c_{rv_i}}{Q}}{C(S^*_{CMST}(V))}.$$

Box I

for some fixed $\epsilon > 0$ or, in the other direction, without loss of generality, assume l is selected to be the largest number (from 1 to n) for which Inequality (5) is larger than ϵ. We obtain an upper bound on l independent of n. To do so, we put $s_h = \sqrt{\frac{c_l}{\sum_{i=1}^{n} c_{rv_i}}}$, $A = Q$, $2B = 4Q\sqrt{\pi}$ and investigate the lower bound of inequality solutions:

$$Ahs_h^2 + 2Bs_h - \varepsilon > 0, \qquad (h = 1, ..., l-1). \qquad (6)$$

Hence, s_h must be larger than the positive root of the quadratic equation defined by the left-hand side of Inequality (6) for $h = 1, ..., l-1$:

$$s_h > \frac{-2B + \sqrt{4B^2 + 4Ah\varepsilon}}{2Ah}, \quad (h = 1, ..., l-1).$$

Thus:

$$s_h^2 > \left(\frac{-2B + \sqrt{4B^2 + 4Ah\varepsilon}}{2Ah}\right)^2, \quad (h = 1, ..., l-1),$$

$$s_h^2 = \frac{c_l}{\sum_{i=1}^{n} c_{rv_i}} = \left(\frac{-2B + \sqrt{4B^2 + 4Ah\varepsilon}}{2Ah}\right)^2$$

$$\geq \frac{\varepsilon}{Ah} + 2\frac{B^2}{A^2h^2} - 4\frac{B\sqrt{4B^2 + 4Ah\varepsilon}}{4A^2h^2}$$

$$\geq \frac{\varepsilon}{Ah} - \frac{4B\sqrt{\varepsilon}}{2\sqrt{A^3h^3}}, \qquad (h = 1, ..., l-1).$$

Consequently:

$$1 \geq \sum_{h=1}^{l-1} s_h^2 = \sum_{h=1}^{l-1} \frac{c_{rv_h}}{\sum_{i=1}^{n} c_{rv_i}} \geq \varepsilon \sum\nolimits_{h=1}^{l-1} \frac{1}{Ah}$$

$$-\frac{4B\sqrt{\varepsilon}}{2} \sum\nolimits_{h=1}^{l-1} \frac{1}{\sqrt{A^3h^3}}.$$

Note that:

$$\sum_{h=1}^{l-1} \frac{1}{h} > \int_{1}^{l-1} \frac{1}{z} dz = \ln(l-1),$$

and:

$$\sum_{h=1}^{l-1} \frac{1}{\sqrt{A^3h^3}} < \frac{1}{A^{3/2}} \int_{1}^{l} \frac{1}{z^{3/2}} dz < \frac{2}{A^{3/2}}.$$

We conclude that:

$$\frac{\varepsilon}{A} \ln(l-1) - \frac{4B\sqrt{\varepsilon}}{A^{3/2}} < 1,$$

i.e.:

$$l < e^{\frac{A}{\varepsilon}\left(1 + \frac{4B\sqrt{\varepsilon}}{A^{3/2}}\right)} + 1.$$

It follows that the computational effort for seeking an optimal set of subtrees for the $l-1$ outside nodes does not rely on n. Moreover, the other steps of the algorithm can be done in polynomial time. Thus, we prove that $A(\Gamma)$ is PTAS for CMSTP. Its running time relies on algorithm solving MST.□

Since A and B are $\Theta(Q)$, the algorithm $A(\Gamma)$ is PTAS for CMSTP and $Q = o(ln ln\, n)$. Indeed, the running time of the algorithm is exponential with respect to l. Since l is exponential with respect to Q, the running time will be polynomial (concerning n) for $Q = o(ln ln\, n)$.

6.3. BCMSTP equipped with 2d−Euclidean metrics

In this section, a similar technique is used to find a factor $2 + \frac{1}{L_1^*} + \epsilon$ approximation algorithm for BCMSTP for every given $\epsilon > 0$. The algorithm given in Section 5 is used and represented by $\Im$. As we have seen, $\Im$ provides a solution S_p with an approximation factor of $3 - \frac{1}{L_1^*}$. We see that:

$$C(S_p) = \sum_{i=1}^{i=K} C(T_i^H) \leq 2C(MST(V)) + C_{rv}$$

$$= 2C(MST(V)) + \frac{\sum_{i=1}^{i=n} c_{rv_i}}{L_1^*}$$

$$= 2C(MST(V)) + \frac{n}{L_1^*}\bar{C}_{rv},$$

where $\bar{C}_{rv} = \frac{\sum_{i=1}^{i=n} c_{rv_i}}{n}$, and n is the number of terminal nodes. Furthermore, it follows from Lemma 2 that each optimal solution $S_V^* = \{T_1^*, T_2^*, ..., T_K^*\}$ of BCMSTP satisfies:

$$C(S_V^*) = \sum_{i=1}^{i=K} C(T_i^*) \geq \frac{\sum_{i=1}^{i=n} c_{rv_i}}{L_K^*} = \frac{n}{L_K^*}\bar{C}_{rv}.$$

Thus, S_V^* satisfies the following relation:

$$\frac{n}{L_K^*}\bar{C}_{rv} \leq C(S_V^*) \leq 2C(MST(V)) + \frac{n}{L_1^*}\bar{C}_{rv}.$$

These inequalities are valid in any metric space, especially in Euclidean metrics. Algorithm 4 presents our algorithm referred to as $A(\Im)$ below.

Theorem 5. *The algorithm $A(\Im)$ achieves $2 + \frac{1}{L_1^*} + \epsilon$ factor of approximation for the Euclidean BCMSTP.*

Proof. To prove the theorem, we show for any $\epsilon > 0$, the relative error $e^{A(\Im)}(V)$ of $A(\Im)$ satisfies the inequality $e^{A(\Im)}(V) \leq 1 + \frac{1}{L_1^*} + \epsilon$. First, we find an upper bound for $e^{A(\Im)}(V)$.

We provide a solution for BCMSTP on the set $V(l') \cup \{r\}$ using the loads $L_K^*, L_{K-1}^*, ..., L_k^*$, and also obtain a solution to this problem on the set $V \backslash V(l') \cup \{r\}$ using the loads $L_1^*, L_2^*, ..., L_{k-1}^*$. Let I be an

input : An instance of BCMSTP.
output : A feasible solution for BCMSTP.

Start
Enumerate terminals by decreasing their distance from the root

$$c_{rv_1} \geq c_{rv_2} \geq ... \geq c_{rv_n}.$$

Take the set $V(l') = \{v_1, v_2, ..., v_l, v_{l+1}, ..., v_{2l}, ..., v_{l'}\}$, $l' < 3l$, of outside nodes. We will later determine the values of l and l', and will show that they are independent of n.
Apply $\Im$ to the set of inside nodes $V \backslash V(l')$. Use the loads $L_1^*, L_2^*, ..., L_{k-1}^*$, where $1 \leq k \leq K$ and $\sum_{i=1}^{k-1} L_i^* = n - l'$. Denote the solution by $S_{V \backslash V(l')}$.
Find the optimal solution for BCMSTP defined on the set of outside nodes $V(l')$ and the same root. Use the loads $L_K^*, L_{K-1}^*, ..., L_k^*$ for the subtrees. Denote the solution by $S_{V(l')}^*$.
Return $S_V = S_{V \backslash V(l')} \cup S_{V(l')}^*$.
End

Algorithm 4. Approximation algorithm for the balanced capacitated minimum spanning tree problem in Euclidean metrics $A(\Im)$.

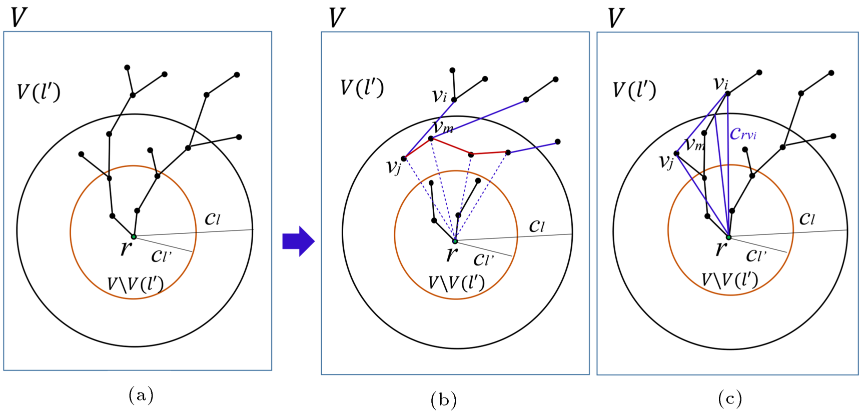

Figure 3. An optimal solution of the Balanced Capacitated Minimum Spanning Tree Problem (BCMSTP) defined on the set of nodes V which can be transformed into a solution for BCMSTP defined on the set of nodes $V(l') \cup \{r\}$.

instance of BCMSTP and S_V^* be an optimal solution. Consider the circle with radius $c_l = c_{rv_l}$ centered at the root (we will later determine l).

To provide a solution for BCMSTP on the set $V(l') \cup \{r\}$, the following steps need to be taken:

1. Remove the edges between the nodes in $V(l) = \{v_1, v_2, ..., v_l\}$ and $V \backslash V(l)$.
2. $S'_{V(l')} = \emptyset$. If the load/order of a subtree T' (in $V(l)$) is L_K^*, take $S'_{V(l')} = S'_{V(l')} \cup T'$. Without loss of generality, we assume loads of the subtrees in $V(l)$ are at most L_1^*.
3. Connect the nodes in $V(l)$ incident to the removed edges (i.e., the edges between $V(l)$ and $V \backslash V(l)$) to the i^{th} node in the set $V'(l') = \{v_{l+1}, v_{l+2}, ..., v_{2l}, ..., v_{l'}\}$. Some of the nodes in $V'(l')$ may not be connected to any subtree in $V(l)$ (see Figure 3(b)). We see $c_{v_i v_j} \leq c_{rv_i} + c_{rv_j} \leq c_{rv_i} + c_l$ for $v_i \in \{v_1, ..., v_l\}$ and $v_j \in \{v_{l+1}, v_{l+2}, ..., v_{l'}\}$. Since $c_{rv_i} \leq c_{v_i v_m} + c_l$ (see Figure 3(c)), we get $c_{v_i v_j} \leq c_{v_i v_m} + 2c_l$.
4. Connect the nodes in $V'(l')$ to each other, as shown in Figure 3(b). We see $c_{v_g v_h} \leq 2c_l$ for $g, h \in \{l+1, l+2, ..., l'\}$.
5. Find an Eulerian tour $\tau_{V(l')}$ spanning the vertices in the found tree by doubling and shortcutting the edges. Note that, there are smaller than L_1^* nodes of $V(l)$ between two consecutive nodes of $V'(l')$ in $\tau_{V(l')}$, since loads of the subtrees in $V(l)$ are at most L_1^*.
6. Let $(v'_1, v'_2, ..., v'_{l'})$ be an order of the nodes in $\tau_{V(l')}$; we obtain subtrees with loads $L_K^*, L_{K-1}^*, ..., L_k^*$ as follows:

$$T_1^{V(l')} := \{v'_1, v'_2, ..., v'_{L_K^*}\}, T_2^{V(l')} :$$

$$= \{v'_{L_K^*+1}, v'_{L_K^*+2}, ..., v'_{L_K^*+L_{K-1}^*}\}, ...,$$

$$T_{K-k+1}^{V(l')} := \{v'_{l'-L_k^*+1}, ..., v'_{l'}\}.$$

7. Connect each subtree $T_i^{V(l')}$ to the root using one of the nodes in $V'(l')$ and incident on it. This can be done, since there are smaller than L_1^* nodes (of

$V(l)$) between two consecutive nodes of $V'(l')$ in $\tau_{V(l')}$.

Thus, we obtain a solution $S'_{V(l')}$ of the problem on the set $V(l')$ whose cost is at most $2C(S^*(V))|_{V(l')}+9lc_l$.

To provide a solution for the problem on the set $V\backslash V(l')$, remove the edges between $V(l')$ and $V\backslash V(l')$. There are at most K subtrees dangling from the root and inside the circle with radius $c_{l'}$ whose loads are $L_1, L_2, ..., L_K$. Without loss of generality, we assume that $L_1 \geq L_2 \geq ... \geq L_K$. To construct a solution to the problem on $V\backslash V(l')$, we act in the following manner. We choose the $k-1$ largest subtrees and connect the nodes of the other trees to the i^{th} subtree $(1 \leq i \leq k-1)$ until its load becomes L_i^* (see Figure 4). The sum of the length of these edges is at most $6lc_l$, since $l' < 3l$. Let $S^*_{V(l')}$, $S^*_{V\backslash V(l')}$ be the optimal solutions for the problems defined on the sets $V(l'), V\backslash V(l')$, respectively and let $C(S^*_{V(l')})$, $C(S^*_{V\backslash V(l')})$ be their respective costs. We have:

$$C(S^*_{V(l')}) + C(S^*_{V\backslash V(l')}) \leq 2C(S^*_V) + 15lc_l.$$

Let $C(S_V)$, $C(S_{V\backslash V(l')})$ be the costs of the solutions S_V, $S_{V\backslash V(l')}$, respectively. By construction, for any

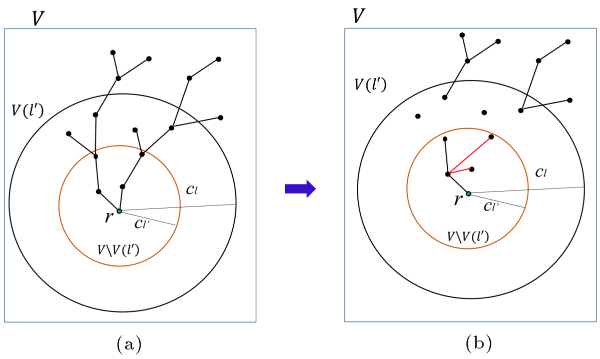

Figure 4. An optimal solution of the Balanced Capacitated Minimum Spanning Tree Problem (BCMSTP) defined on the set of nodes V which can be transformed into a solution for Balanced Capacitated Minimum Spanning Tree Problem (BCMSTP) defined on the set of nodes $V/V(l')$.

$A(\Im)$-based approximation algorithm:

$$C(S_V) = C(S^*_{V(l')}) + C(S_{V\backslash V(l')}).$$

Since $C(S^*_{V\backslash V(l')}) \geq \sum_{i=l'+1}^{n} \frac{c_{rv_i}}{L_K^*}$ (from Lemma 2), we get:

$$C(S_V) \leq C(S^*_{V(l')}) + C(S^*_{V\backslash V(l')}) + \left(C(S_{V\backslash V(l')}) - \sum_{i=l'+1}^{n} \frac{c_{rv_i}}{L_K^*} \right).$$

Thus, we have the equation shown in Box II. Since $C(S_V^*) \geq \frac{n}{L_K^*}\bar{C}_{rv}$, $C(S_{V\backslash V(l')}) \leq 2C(MST(V)) + \frac{\sum_{i=l'+1}^{n} c_{rv_i}}{L_1^*}$, and $C(MST(V)) \leq 2\sqrt{\pi n c_{rv}^{\max} \bar{C}_{rv}}$, we obtain Eqs. (7) and (8) as shown in Box III. Due to $L_K^* = L_1^* + 1$, Inequality (7) holds. In the case that $L_K^* = L_1^*$, the term $\frac{1}{L_1^*}$ would be removed from Eq. (8).

Now, we choose l such that Eq. (8) is less than $1 + \frac{1}{L_1^*} + \epsilon$. We obtain an upper bound on l independent of n. To do so, we put $s_h = \sqrt{\frac{c_l}{\sum_{i=1}^{n} c_{rv_i}}}$, $A = 15L_K^*$, $2B = 4L_K^*\sqrt{\pi}$ and investigate the lower bound of inequality solutions:

$$Ahs_h^2 + 2Bs_h - \varepsilon > 0, \qquad (h = 1, ..., l).$$

A similar method as it is given in the proof of Theorem 4 shows that:

$$l < e^{\frac{A}{\varepsilon}\left(1+\frac{4B\sqrt{\varepsilon}}{A^{3/2}}\right)}.$$

We choose the smallest l', $2l \leq l' < 3l$, so that $\sum_{i=k}^{K} L_i^* = l'$, $1 \leq k \leq K$. It follows that finding an optimal set of subtrees for the l' outside nodes does not depend on n. Moreover, the other steps of $A(\Im)$ can be done in polynomial time concerning n. We conclude that the heuristic $A(\Im)$ is a polynomial time approximation algorithm for BCMSTP.□

Since $L_K^* \leq Q$ and A and B are $\Theta(L_K^*)$, the algorithm $A(\Im)$ is an approximation algorithm for BCMSTP and $Q = o(\ln \ln n)$.

7. Conclusion

In this paper, the Balanced Capacitated Minimum Spanning Tree Problem (BCMSTP) was considered and an attempt was made to design two approximation

$$e^{A(\Im)}(V) = \frac{C(S_V) - C(S_V^*)}{C(S_V^*)} \leq \frac{C(S^*_{V(l')}) + C(S^*_{V\backslash V(l')}) + C(S_{V\backslash V(l')}) - \frac{\sum_{i=l'+1}^{n} c_{rv_i}}{L_K^*} - C(S_V^*)}{C(S_V^*)}$$

$$\leq \frac{2C(S_V^*) + 15lc_l + C(S_{V\backslash V(l')}) - \frac{\sum_{i=l'+1}^{n} c_{rv_i}}{L_K^*} - C(S_V^*)}{C(S_V^*)} \leq \frac{C(S_V^*) + 15lc_l + C(S_{V\backslash V(l')}) - \frac{\sum_{i=l'+1}^{n} c_{rv_i}}{L_K^*}}{C(S_V^*)}.$$

Box II

$$e^{A(\Im)}(V) \leq 1 + \frac{15lc_l}{\frac{n}{L_K^*}\bar{C}_{rv}} + \frac{C(S_{V\setminus V(l')}) - \frac{\sum_{i=l'+1}^{n} c_{rv_i}}{L_K^*}}{\frac{n}{L_K^*}\bar{C}_{rv}} \leq 1 + \frac{L_K^* 15lc_l}{\sum_{i=1}^{n} c_{rv_i}}$$

$$+L_K^* \left(\frac{4\sqrt{\pi n c_{rv}^{\max}\bar{C}_{rv}} + \frac{\sum_{i=l'+1}^{n} c_{rv_i}}{L_1^*} - \frac{\sum_{i=l'+1}^{n} c_{rv_i}}{L_K^*}}{\sum_{i=1}^{n} c_{rv_i}} \right) \leq 1 + \frac{L_K^* 15lc_l}{\sum_{i=1}^{n} c_{rv_i}} + L_K^* \left(\frac{4\sqrt{\pi n c_l \bar{C}_{rv}}}{\sum_{i=1}^{n} c_{rv_i}} \right) + \frac{1}{L_1^*} \quad (7)$$

$$= 1 + \frac{1}{L_1^*} + L_K^* 15l \frac{c_l}{\sum_{i=1}^{n} c_{rv_i}} + 4L_K^* \sqrt{\pi} \sqrt{\frac{c_l}{\sum_{i=1}^{n} c_{rv_i}}}. \quad (8)$$

Box III

algorithms. A factor $3 - \frac{1}{L_1^*}$ approximation algorithm that could find the fairest solution was proposed. In the Euclidean metrics, we provided an improved algorithm that achieved $2 + \frac{1}{L_1^*} + \epsilon$ factor of approximation. In addition to its applications, BCMSTP on Euclidean metrics is compelling theoretically. Most of the geometric problems accept Polynomial Time Approximation Scheme (PTAS) in Euclidean metrics; therefore, an interesting question is whether the Euclidean Balanced Capacitated Minimum Spanning Tree Problem (Euclidean BCMSTP) has PTAS, which remains an open problem. This paper presented PTAS for the $2d$-Euclidean CMSTP. Future work could be to improve this algorithm.

Acknowledgements

The authors thank Amirkabir University of Technology for facilities.

References

1. Oncan, T. and Altinel, I.K. "Parametric enhancements of the Esau-Williams heuristic for the capacitated minimum spanning tree problem", *Journal of the Operational Research Society*, **60**, pp. 259–267 (2009).
2. Ali, A.I. and Huang, C.H. "Balanced spanning forests and trees", *Networks*, **21**(6), pp. 667–687 (1991).
3. Arkin, E.M., Guttmann-Beck, N., and Hassin, R. "The (K,k)-capacitated spanning tree problem", *International Conference on Algorithmic Applications in Management, Algorithmic Aspects in Information and Management*, pp. 25–34 (2010).
4. Gavish, B. "Topological design of centralized computer networks-formulations and algorithms", *Networks*, **12**, pp. 355–377 (1982).
5. Hassin, R., Ravi, R., and Salman, F.S. "Approximation algorithms for capacitated network design problems", *APPROX*, pp. 167–176 (2000).
6. Kritikos, M. and Ioannou, G. "A greedy heuristic for the capacitated minimum spanning tree problem", *Journal of the Operational Research Society*, **68**, pp. 1223–1235 (2017).
7. Ruiza, E., Albareda-Sambola, M., Fernndez, E., et al. "A biased random-key genetic algorithm for the capacitated minimum spanning tree problem", *Journal Computers and Operations Research*, **57**, pp. 95–108 (2015).
8. Sharma, R.L. "Design of an economical multidrop network topology with capacity constraints", *IEEE Transactions on Communications*, **31**, pp. 590–591 (1983).
9. Incel, O.D., Ghosh, A., Krishnamachari, B., et al. "Fast data collection in tree-based wireless sensor networks", *IEEE Transactions on Mobile Computing*, **11**(1), pp. 86–99 (2012).
10. Vazirani, V.V., *Approximation Algorithms*, Springer (2001).
11. Bowerman, R., Hall, B., and Calamai, P. "A multi objective optimization approach to urban school bus routing: formulation and solution method", *Transportation Research Part A: Policy and Practice*, **29**, pp. 107–123 (1995).
12. Atefi, R., Salari, M., Coelho, L.C., et al. "The open vehicle routing problem with decoupling points", *European Journal of Operational Research*, **265**(1), pp. 316–327 (2018).
13. Jothi, R. and Raghavachari, B. "Approximation algorithms for the capacitated minimum spanning tree problem and its variants in network design", *ACM Transactions on Algorithms*, **1**, pp. 265–282 (2005).
14. Martello, S., Pulleyblank, W.R., Toth, P., et al. "Balanced optimization problems", *Operations Research Letters*, **3**, pp. 275–278 (1984).
15. Bassetto, T. and Mason, F. "The 2-period balanced traveling salesman problem", Working Papers, **154**, Department of applied mathematics, Universitá Cá Foscari Venezia (2007).
16. Cappanera, P. and Scutella, M.G. "Balanced paths in a cyclic networks: tractable cases and related approaches", *Networks*, **45**, pp. 104–111 (2005).

17. da Cunha, A.S. "Formulation and branch-and-cut algorithm for the minimum cardinality balanced and connected clustering problem", In *Proceedings of the International Network Optimization Conference (INOC)*, Avignon, France, June 12–14 (2019).
18. Feldmann, A.E. and Foschini, L. "Balanced partitions of trees and applications", *Algorithmica*, **71**, pp. 354–376 (2015).
19. Ficker, A., Spieksma, F.C., and Woeginger, G.J. "Balanced vector optimization", KUL, Research Report (2016).
20. Kinable, J., Smeulders, B., Delcour, E., et al. "Exact algorithms for the equitable traveling salesman problem", *European Journal of Operational Research*, **261**, pp. 475–485 (2017).
21. Yousefikhoshbakht, M., Didehvar, F., and Rahmati, F. "An effective rank based ant system algorithm for solving the balanced vehicle routing problem", *International Journal of Industrial Engineering*, **23**(1), pp. 13–25 (2016).
22. Zhang, H. "Balanced graph partitioning: optimizing graph cut based on label swapping", *International Conference on Behavioral, Economic and Socio-cultural Computing (BESC)*, Nanjing, China, 30 Oct. – 1 Nov. (2015).
23. Gavish, B. and Altinkemer, K. "A parallel savings heuristic for the topological design of local access tree networks", *Proceedings IEEE-INFOCOM'86*, pp. 130–139 (1986).
24. Altinkemer, K. and Gavish, B. "Heuristics with constant error guarantees for the design of tree networks", *Management Science*, **34**, pp. 331–341 (1988).
25. Gavish, B. "Augmented lagrangean based algorithms for centralized network design", *IEEE Transactions on Communications*, **33**, pp. 1247–1257 (1985).
26. Kershenbaum, A. and Boorstyn, R. "Centralized teleprocessing network design", *Networks*, **13**, pp. 279–293 (1983).
27. Ahuja, R., Orlin, J., and Sharma, D. "A composite neighborhood search algorithm for the capacitated minimum spanning tree problem", *Operations Research Letters*, **31**, pp. 185–194 (2003).
28. Chandy, K.M. and Lo, T. "The capacitated minimum tree", *Networks*, **3**, pp. 173–182 (1973).
29. Gavish, B. "Topological design of telecommunication networks local access design methods", *Annals of Operations Research*, **33**, pp. 17–71 (1991).
30. Gouveia, L. "A comparison of directed formulations for the capacitated minimal spanning tree problem", *Telecommunication Systems*, **1**, pp. 51–76 (1993).
31. Gamvros, I., Golden, B., and Raghavan, S. "The multilevel capacitated minimum spanning tree problem", *INFORMS Journal on Computing*, **18**, pp. 348–365 (2006).
32. Berger, D., Gendron, B., Potvin, J.Y., et al. "Tabu search for a network loading problem with multiple facilities", *Journal of Heuristics*, **6**, pp. 253–267 (2000).
33. Salman, F.S., Ravi, R., and Hooker, J. "Solving the local access network design problem", *Working paper*, Krannert Graduate School of Management, Purdue University, West Lafayette (2001).
34. Rothfarb, B. and Goldstein, M.C. "The one-terminal Telpak problem", *Operations Research*, **19**, pp. 156–169 (1971).
35. Salman, F.S., Cheriyan, J., Ravi, R., et al. "Buy-at-bulk network design: Approximating the single-sink edge installation problem", *Proceedings 8th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 619–628 (1997).
36. Jothi, R. and Raghavachari, B. "Survivable network design: The capacitated minimum spanning network problem", *Information Processing Letters*, **91**, pp. 183–190 (2004).
37. Goemans, M.X. and Williamson, D.P. "A general approximation technique for constrained forest problems", *SIAM Journal on Computing*, **24**(2), pp. 296–317 (1995).
38. Ruiz, H.E.R. "The capacitated minimum spanning tree problem", Ph.D. Thesis, Universitat Politecnica de Catalunya, BarcelonaTech (2013).
39. Chen, J. and Chen, S. "Optimization of vehicle routing problem with load balancing and time windows in distribution", *4th International Conference on Wireless Communications, Networking and Mobile Computing* (2008). DOI: 10.1109/WiCom.2008.1527
40. Fallah, H., Didehvar, F., and Rahmati, F. "Approximation algorithms for the load-balanced capacitated vehicle routing problem", *Bulletin of the Iranian Mathematical Society* (2020). DOI: 10.1007/s41980-020-00440-3
41. Cowell, F.A. "Measurement of inequality", In *Handbook of Income Distribution*, A.B. Atkinson, and F. Bourguignon, Ed., North Holland, Amsterdam (2000).
42. Garey, M.R. and Johnson, D.S. "Computers and intractability: a guide to the theory of NP-completeness" Freeman, W.H. and Co. subs. of scientific American, New York, United States, 338 pages (1979).
43. Papadimitriou, C.H. "The Complexity of the capacitated tree problem", *Networks*, **8**, pp. 217–230 (1978).
44. Camerini, P.M., Galbiatai, G., and Maffoli, F. "Complexity of spanning tree problems: part 1", *European Journal of Operational Research*, **5**, pp. 346–352 (1980).
45. Camerini, P.M., Maffioli, F., Martello, S., et al. "Most and least uniform spanning trees", *Discrete Applied Mathematics*, **15**, pp. 181–187 (1986).
46. Beasley, J. "Route first-cluster second methods for vehicle routing", *Omega*, **11**, pp. 403–408 (1983).
47. Haimovich, M. and Rinnooy Kan, A.H.G. "Bounds and heuristics for capacitated routing problems", *Mathematics of Operations Research*, **10**, pp. 527–524 (1985).

48. Khachay, M. and Dubinin, R. "PTAS for the Euclidean capacitated vehicle routing problem in R^d", *International Conference on Discrete Optimization and Operations Research, Lecture Notes in Computer Science* 9869, pp. 193–205 (2016).

Appendix A

We present a mathematical programing model for BCMSTP. We presume that the root node corresponds to the node 0 and $N = \{1, 2, ..., n\}$ is the set of terminals. Moreover, we assume that demand of the root is zero. Let x_{ij}^k be the binary variables on the edges $(i, j) \in E$ that decide whether the edge (i, j) is presented in the k^{th} subtree. An edge (i, j) is presented in the k^{th} subtree, iff $x_{ij}^k = 1$ and is not presented, otherwise. Let y_{ij}^k be the quantity that is carrying through i to j along the k^{th} subtree. The following is an integer programing formulation of BCMSTP:

$$C_{opt} = \min \sum_{k=1}^{K} \sum_{i=0}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}^k,$$

$$\sum\nolimits_{k=1}^{K} \sum\nolimits_{i=0}^{n} x_{ij}^k = 1 \quad j := 1, \ldots, n, \tag{A.1}$$

$$\sum\nolimits_{k=1}^{K} \sum\nolimits_{i=0}^{n} y_{ij}^k - \sum\nolimits_{k=1}^{K} \sum\nolimits_{i=1}^{n} y_{ji}^k = 1 \quad j := 1, \ldots, n, \tag{A.2}$$

$$\left| \sum_{j=1}^{n} y_{0j}^k - \sum_{j=1}^{n} y_{0j}^l \right| < \alpha \quad k, l := 1, \ldots, K, \tag{A.3}$$

$$x_{ij}^k \leq y_{ij}^k \leq Q x_{ij}^k \quad j := 1, \ldots, n, \quad i := 0, \ldots, n, \quad k := 1, \ldots, K \tag{A.4}$$

$$\sum_{j=1}^{n} x_{0j}^k = 1, \quad k := 1, \ldots, K, \tag{A.5}$$

$$y_{ij}^k \geq 0, x_{ij}^k \in \{0, 1\} \quad j := 1, \ldots, n, \quad i := 0, \ldots, n, \quad k := 1, \ldots, K. \tag{A.6}$$

The constraints of Eq. (A.1) ensure that each node $j \in N$ is sourced by exactly one edge (i, j) from some node $i \in N \cup \{0\}$. Constraint set (A.2) implies that the cumulative flow going into every node j is one unit more than the cumulative flow coming out of that node. The loads of the subtrees should satisfy the balanced condition (1) (or (2)) for a parameter α (or β), specified by the user. This is guaranteed by the constraints (A.3). Constraint set (A.4) implies that the flow on an activated (or used) edge will not exceed the capacity Q. In this formulation, for a certain k, there can be more than one subtree dangling from the root. In other words, what the model considers as a subtree dangling from the root would be the union of several such trees. Constraint set (A.5) is to fix this bug. Constraint set (A.6) implies that flows on all edges are nonnegative and it can be implied that an edge is either used or not. The cost of using an edge is fixed regardless of the volume of flow on the edge.

Biographies

Haniyeh Fallah is a PhD graduate student in Amirkabir University of Technology, Tehran, Iran. Her research interests include approximation algorithms, vehicle routing problems, operations research, applied dynamical systems and neuroscience. She received her MS degree in Dynamical Systems of Neurons from Sharif University of Technology. She has published papers in journals such as Biological Cybernetics, International Journal of Bifurcation and Chaos among others.

Farzad Didehvar is an Assistant Professor of Amirkabir University of Technology. His fields of interests are in the scope of theoretical computer science and computer Science, more exactly algorithm and theory of computation. He trained many students and authored several articles in these areas.

Farhad Rahmati is an Associate Professor of Amirkabir University of Technology, Tehran, Iran. His research interests include algebra and its applications in computer science, design and analysis of algorithms and operations research. He has published papers in journals such as Journal of Algebra and Its Applications, IEEE Access, The International Journal of Industrial Engineering: Theory, Applications and Practice, Journal of Engineering, Communications in Algebra, The Electronic Journal of Combinatorics, Graphs and Combinatorics, International Journal of Production Research among others.