Operations Research Letters 41 (2013) 138–141

Contents lists available at SciVerse ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

The hardness and approximation of the star p-hub center problem

Hongyu Liang *

Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 7 October 2012
Received in revised form
10 December 2012
Accepted 10 December 2012
Available online 27 December 2012

Keywords:
Star p-hub center problem
NP-hardness
Approximation algorithm
Hardness of approximation

ABSTRACT

We consider the star p-hub center problem recently introduced by Yaman and Elloumi [H. Yaman and S. Elloumi. Star p-hub center problem and star p-hub median problem with bounded path lengths, Comput. Oper. Res., 39 (11) (2012) 2725–2732]. We first show that the problem does not admit a $(1.25 - \epsilon)$-approximation algorithm for any $\epsilon > 0$ unless P = NP. In particular this gives the first strong NP-hardness result for the problem in a metric space. We also present, complementing the inapproximability result, a purely combinatorial 3.5-approximation algorithm for the star p-hub center problem.

1. Introduction

In this work, we study the *star p-hub center problem* (abbreviated as SpHCP) introduced by Yaman and Elloumi [15]. There are a set of demand nodes (users) located on a metric space, each of which wants to communicate with all the others through a two-level hub network specified as follows. There is a given fixed central hub, and we can choose another p hubs among the set of demand nodes and connect each of them to the central hub by a directed link. Then we connect each of the remaining demand nodes to exactly one of the p chosen hubs. The resulting network is a tree of depth 2 (or a star–star network), in which the root is the central hub, the depth-1 nodes are the hubs, and the depth-2 nodes (i.e., leaves) are the children of their assigned hubs. (See Fig. 1 for an example.) The length of each edge equals the distance between its two endpoints in the metric space. The *path length* between two demand nodes is the length of the unique path connecting them in this tree. The goal of SpHCP is to find p hubs and a corresponding hub assignment so as to minimize the maximum path length between any pair of demand nodes.

The star p-hub center problem can be used to model, e.g., the optimization of the poorest service quality in a two-level telecommunications network [15]. Yaman and Elloumi [15] proved the NP-hardness of SpHCP and proposed several formulations of SpHCP using integer programs. They also performed computational studies of the problem through extensive experiments. However, to the best of our knowledge, no other theoretical results are known for SpHCP. Furthermore, in their NP-hardness proof of SpHCP (Theorem 1 in [15]), the distance function between the demand nodes is not a metric, which violates the definition of SpHCP. Thus the complexity of SpHCP (on a metric space) still remains open.

In this work, we fill in the blank of theoretical aspects of the star p-hub center problem by investigating its polynomial-time approximability. We prove that there is no $(1.25 - \epsilon)$-approximation algorithm for SpHCP on a metric space, unless P = NP, by a reduction from the dominating set problem. This in particular implies the (strong) NP-hardness of SpHCP. We then present a 3.5-approximation algorithm for SpHCP. Our algorithm is purely combinatorial and easy to implement.

Related work. Various hub location problems have been well studied in the literature because of their important applications in transportation and telecommunication systems; here hubs are intermediate transshipment points for traffic between the members of origin–destination pairs. See [1,4] for two recent surveys on network hub location problems. Classical hub location problems mostly assume that the hubs are fully interconnected and thus the communication between them need not go through a central hub. A problem similar to SpHCP under this assumption is the p-hub center problem introduced by Campbell [3] and O'Kelly and Miller [14], which was further studied in, e.g., [8,11,13]. Recently there has been increasing research attention to hub location problems with multi-level topologies, such as the star–star network in SpHCP (i.e., the network connecting hub nodes to the central hub is a star, and each network connecting users to a particular hub is also a star). Some problems of interest with multi-level network structures have been studied in, e.g., [2,5–7,12].

* Correspondence to: 4-609 FIT Building, Tsinghua University, Beijing 100084, China. Tel.: +86 10 62781693x1636; fax: +86 1062797331.
E-mail addresses: sjqxzlhy@gmail.com, lianghy08@mails.tsinghua.edu.cn.

0167-6377/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.orl.2012.12.007

Fig. 1. An example of the problem with $p = 2$. On the left side, the solid circle is the central hub and the empty circles are the demand nodes. The right side represents a hub assignment and the resulting network, in which the chosen hubs are redrawn as solid squares.

2. Problem definition and notation

In the star p-hub center problem, we are given a set of demand nodes D and a central hub $c \notin D$. There is a distance function $d(\cdot, \cdot)$ defined on every pair of nodes in $D \cup \{c\}$. The distance d is a metric on $D \cup \{c\}$, i.e., $d(i, i) = 0$, $d(i, j) = d(j, i)$, and $d(i, j) + d(j, k) \geq d(k, i)$ for all $i, j, k \in D \cup \{c\}$. Additionally an integer $p \geq 1$ is given. A *hub assignment* is a way of designating p nodes from D as *hubs*, connecting them to the central hub, and assigning each of the remaining (non-hub) nodes in D to exactly one of the p chosen hubs. The *hub graph* (of a hub assignment) is a tree graph with vertex set $D \cup \{c\}$ in which c is the root, the p selected hubs are the children of c, and each non-hub node is a child of its assigned hub node; in addition, each edge $\{i, j\}$ in this tree has length $d(i, j)$. The *path length* between $i, j \in D$, denoted by $l(i, j)$, is the length of the unique path connecting i and j in the hub graph. The goal of the star p-hub center problem [15] (SpHCP, for short) requires one to find a hub assignment such that the maximum path length between any pair of nodes in D is minimized.

We introduce some notation used in the design of approximation algorithms for optimization problems. Let Π be a minimization problem. An α*-approximation algorithm* for Π is a polynomial-time algorithm that, given any instance $\mathcal{I}$ of Π, returns a solution whose objective value is at most α times that of the optimal solution of $\mathcal{I}$. (Approximation algorithms for maximization problems can be analogously defined, but they will not be used in this work.)

3. Approximability of the star p-hub center problem

In this section, we study the star p-hub center problem from an approximation point of view. We first prove the following hardness result, which rules out the possibility of designing a polynomial-time approximation scheme (PTAS) for SpHCP.

Theorem 1. *There is no* $(1.25\text{-}\epsilon)$*-approximation algorithm for the star p-hub center problem for any* $\epsilon > 0$*, unless* P = NP.

Proof. We will prove that, if SpHCP can be approximated within a factor $1.25 - \epsilon$ in polynomial time for some $\epsilon > 0$, then the dominating set problem can be solved in polynomial time. This will complete the proof of the theorem, since the dominating set problem is well known to be NP-hard [9].

Let $\{G, k\}$ be an instance of the dominating set problem, where $G = (V, E)$ is the input graph and k is an integer. The goal of the problem is to decide whether G has a dominating set of size k. We assume w.l.o.g. that $1 \leq k \leq n - 2$ and that G contains no isolated vertex. Construct an instance of SpHCP as follows. The set of demand nodes is $D = \{c_v \mid v \in V\} \cup \{c'\}$; thus there is a demand node c_v corresponding to each vertex $v \in V$, and there is another demand node c' for other uses. There is a central hub $c \notin D$. The distance function d is defined as:

- $d(c, c_v) = 1$ for all $v \in V$, and $d(c, c') = 2$;
- for all distinct $u, v \in V$, $d(c_u, c_v) = 1$ if $\{u, v\} \in E$, and $d(c_u, c_v) = 2$ otherwise;
- $d(c', c_v) = 3$ for all $v \in V$.

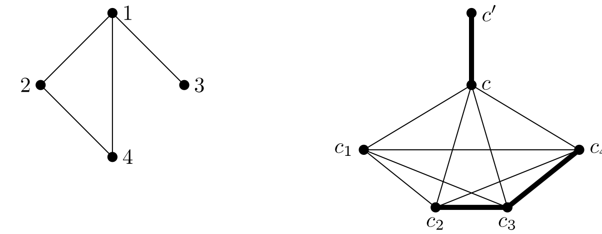

Fig. 2. An example of the reduction. The graph on the left is an instance of the dominating set problem. On the right side, $d(u, v) = 1$ if there is a normal line segment between u and v, $d(u, v) = 2$ if there is a bold line segment between u and v, and $d(u, v) = 3$ if there is no line segment connecting u and v.

It is easy to verify that d is a metric on D. (See Fig. 2 for an illustration.) Finally we set the number of allowed hubs as $p = k + 1$. This finishes the construction of the SpHCP instance, which we denote by $\mathcal{I}$. Clearly $\mathcal{I}$ can be constructed in polynomial time.

Let OPT($\mathcal{I}$) denote the objective value of the optimal solution of $\mathcal{I}$. We claim that G has a dominating set of size k if and only if OPT($\mathcal{I}$) ≤ 4. This will complete the proof, because OPT($\mathcal{I}$) > 4 implies OPT($\mathcal{I}$) ≥ 5 due to the integrality of the distance function, and thus any $(5/4 - \epsilon)$-approximation algorithm can distinguish the two cases.

First consider the "only if" part. Let $S \subseteq V$ be a dominating set of G with $|S| = k$. Consider a hub assignment of $\mathcal{I}$ obtained as follows: Choose $\{c_v \mid v \in S\} \cup \{c'\}$ as the set of $p = k + 1$ hubs. Then, for each vertex $u \in V \setminus S$, assign the demand node c_u to a hub c_v, with v satisfying that $\{u, v\} \in E$ and $v \in S$ (this can be done since S is a dominating set of G). Note that no non-hub demand node is assigned to c'. It is then easy to verify that the maximum path length between any two nodes in the underlying hub graph is at most 4. Hence, OPT($\mathcal{I}$) ≤ 4.

We now consider the "if" direction. Suppose OPT($\mathcal{I}$) ≤ 4 and consider a hub assignment that achieves the optimal objective value. We first observe that c' must be a hub node. In fact, if c' is not a hub, let c_u be the hub to which c' is assigned, and c_v be a hub other than c_u (note that $p = k + 1 \geq 2$). Then the path length between c' and c_v is $l(c', c_v) = d(c', c_u) + d(c_u, c) + d(c, c_v) = 3 + 1 + 1 = 5$, which contradicts our assumption that OPT($\mathcal{I}$) ≤ 4. Thus c' is a hub. If there is some c_v that is assigned to c', then the path length between c_v and a hub other than c' is at least $d(c_v, c') + d(c', c) + 1 = 6$, a contradiction. Therefore no demand node is assigned to c'.

Let H be the set of demand nodes other than c'. We have $|H| = p - 1 = k$. Let $S = \{v \in V \mid c_v \in H\}$. We claim that S is a dominating set of G. Assume to the contrary that S is not a dominating set; then there exists $v \in V \setminus S$ such that S contains no neighbors of v. Let $c_u \in H$ be the hub node that c_v is assigned to. Then the path length between c_v and c' is $l(c_v, c') = d(c_v, c_u) + d(c_u, c) + d(c, c') = 2 + 1 + 2 = 5$, contradicting our assumption again. Therefore, S is a dominating set of G of size k. This completes the reduction from the dominating set problem to SpHCP, and hence Theorem 1 follows. □

In the above hardness proof, the distance between two demand points can only be 1, 2, or 3. Thus the following corollary is immediate from Theorem 1.

Corollary 1. *The star p-hub center problem is strongly NP-hard.*

Next, to match the inapproximability result, we present a constant factor approximation algorithm for SpHCP.

Theorem 2. *There is a* 3.5*-approximation algorithm for the star p-hub center problem.*

Proof. Let $\mathcal{I} = (D \cup \{c\}, d, p)$ be an instance of the star p-hub center problem, where D is the set demand nodes, c is the central

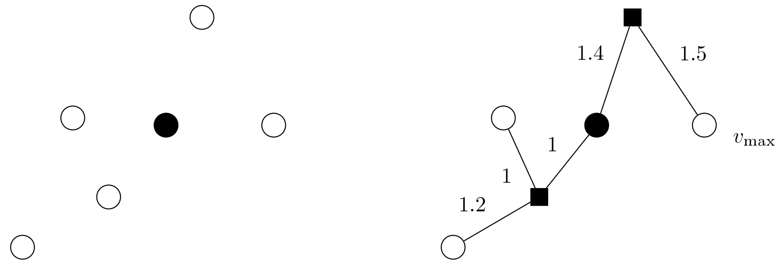

Fig. 3. An example of the parameters. On the left is an instance of SpHCP in which the solid circle is the central hub and the others are demand nodes, and $p = 2$. On the right is an optimal solution, where solid squares are selected hubs and edges represent the assignment (with the lengths). Then we have $d_0 = 1.4$, $d_1 = 1.5$, $d_2 = 1.2$, and $v_{\max}$ is the rightmost node. In this case, it is easy to see that $\mathrm{OPT} = 1.2 + 1 + 1.4 + 1.5 = 5.1$. The two lower bounds in (1) and (2) are respectively $d_1 + d_0 = 2.9$ and $d_1 + d_2 = 2.7$.

Algorithm 1: Approximation algorithm for SpHCP ($D \cup \{c\}, d, p$)

1. Guess the correct values of d_0, d_1, d_2, and $v_{\max}$. Their meanings are provided in the proof.
2. $H \leftarrow \{v \in D \mid d(v, c) \le d_0\}$.
3. Create an instance $\mathcal{J}$ of the k-center problem with forbidden centers, in which $D \setminus \{v_{\max}\}$ is the set of demand nodes, $H \setminus \{v_{\max}\}$ is the set of allowed centers, $k = p$, and the distance metric is the restriction of d to $D \setminus \{v_{\max}\}$.
4. Apply the greedy 2-approximation algorithm for the k-center problem with forbidden centers (Algorithm 2), to obtain an approximate solution of $\mathcal{J}$. Assume that $H^* \subseteq H$ is the set of centers opened in the solution.
5. **return** the solution that opens H^* as the set of p hubs and assigns each node in D to its nearest hub in H^*.

Algorithm 2: A 2-approximation for k-center with forbidden centers

//Let C be the set of demand nodes, $C' \subseteq C$ be the set of allowed centers, and d be the metric on C. Assume w.l.o.g. that $k \le |C'|$.

Guess the objective value of the optimal solution, say OPT. (Note that there are at most $|C'| \cdot |C|$ possible values for OPT.)

$R \leftarrow C$; $S \leftarrow \emptyset$.

while $R \ne \emptyset$ *and* $|S| < k$ **do**
    Choose an arbitrary node $v \in C' \cap R$.
    $B(v) \leftarrow \{u \in R \mid d(u, v) \le 2 \cdot OPT\}$.
    $R \leftarrow R \setminus B(v)$; $S \leftarrow S \cup \{v\}$.
end
return S

hub, d is the distance metric on $D \cup \{c\}$, and p is the number of allowed hubs. Let OPT denote the objective value (i.e., the maximum path length) of the optimal solution (say $\mathcal{S}$) of $\mathcal{I}$, and $D^* \subseteq D$ be the set of hub nodes opened by $\mathcal{S}$. Suppose each node $v \in D$ is assigned to the hub $f(v) \in D^*$. (A hub node is assigned to itself, i.e., $f(v) = v$ for $v \in D^*$.) Let $d_0 := \max_{v \in D^*} d(c, v)$ be the largest distance between c and any hub node in D^*. Let d_1 and d_2 be the largest distance and second largest distance (considering multiplicity) among the node–hub distances $\{d(v, f(v)) \mid v \in D \setminus D^*\}$, respectively. (For example, if the collection of distances between nodes and their assigned hubs is $\{5, 5, 4, 3, 0, 0\}$, then we have $d_1 = d_2 = 5$, but not $d_1 = 5, d_2 = 4$.) We assume that $p \ge 2$ and $d_1 > 0$; otherwise the instance is very easy to solve. See Fig. 3 for an illustration of the parameters.

Our algorithm is presented as Algorithm 1. Line 1 of Algorithm 1 guesses the values of d_0, d_1, d_2, and the node $v_{\max} \in D$ with $d(v_{\max}, f(v_{\max})) = d_1$. (Note that $v_{\max}$ is the node with the largest distance to its assigned hub in the optimal solution $\mathcal{S}$.) We certainly do not know their exact values. However, since each of them has only polynomially many possible values, we can run the algorithm for all of their possible values and take the best solution. Therefore, in the following we assume that we know d_0, d_1, d_2, and $v_{\max}$.

We next prove that Algorithm 1 is indeed a 3.5-approximation algorithm for SpHCP. We will accomplish this by first establishing lower bounds of OPT with respect to the guessed parameters, and then showing an upper bound of ALG, the maximum path length between two demand nodes in our solution. Comparing ALG with OPT yields the desired result.

We first prove the lower bounds on OPT. Let $h, v_1, v_2 \in D$ be the nodes witnessing d_0, d_1, d_2, respectively; that is, $d(h, c) = d_0$, $d(v_1, f(v_1)) = d_1$, and $d(v_2, f(v_2)) = d_2$. If $h \ne f(v_1)$, then the path length between v_1 and h in the optimal solution $\mathcal{S}$ is $d(v_1, f(v_1)) + d(f(v_1), c) + d(c, h) \ge d_1 + d_0$. If $h = f(v_1)$, then the path length between v_1 and any $v \in D \setminus \{v_1, h\}$ is at least $d(v_1, h) + d(h, c) \ge d_1 + d_0$. Hence, we have

$$\mathrm{OPT} \ge d_1 + d_0. \tag{1}$$

We then consider the path length between v_1 and v_2 in the optimal solution. Clearly the path connecting v_1 and v_2 must contain the two edges $\{v_1, f(v_1)\}$ and $\{v_2, f(v_2)\}$, and thus we have

$$\mathrm{OPT} \ge d(v_1, f(v_1)) + d(v_2, f(v_2)) = d_1 + d_2 \ge 2d_2. \tag{2}$$

Next, we establish an upper bound on ALG by examining our algorithm. By our choice of d_0, the set H defined in line 2 contains all hub nodes in the optimal solution $\mathcal{S}$, i.e., $D^* \subseteq H$. In line 3, we create an instance $\mathcal{J}$ of the k-center problem with forbidden centers. This problem is defined as follows: The input consists of a set C of demand points in a metric space, a set $C' \subseteq C$ of allowed centers, and an integer k. The goal is to open (at most) k centers in C' such that the maximum distance between any vertex in C and its nearest center among the k opened centers is minimized. This problem is a generalization of the ordinary k-center problem (in which $C' = C$), and is a special case of the k-supplier problem (in which C' may not be a subset of C) [10]. There is a simple greedy 2-approximation algorithm for this problem, which is presented in Algorithm 2. It analysis is standard and is similar to that of the traditional k-center problem (see [10]), and thus is omitted here.

Let $\mathrm{OPT}(\mathcal{J})$ denote the objective value of the optimal solution to $\mathcal{J}$. We show that $\mathrm{OPT}(\mathcal{J}) \le d_2$. Recall that $D^* \subseteq H$ is the set of hubs opened in the optimal solution to the SpHCP instance $\mathcal{I}$. By the definition of d_1, d_2 and $v_{\max}$, we have $\max_{v \in D \setminus \{v_{\max}\}} \min_{h \in D^*} d(v, h) = d_2$. Therefore, by opening D^* as the set of centers in $\mathcal{J}$, we achieve an objective value of d_2, which implies that $\mathrm{OPT}(\mathcal{J}) \le d_2$. Hence, by applying the 2-approximation algorithm (Algorithm 2) to implement line 4 of Algorithm 1, we obtain a solution H^* of $\mathcal{J}$ with objective value at most $2\mathrm{OPT}(\mathcal{J}) \le 2d_2$, that is,

$$\max_{v\in D\setminus\{v_{\max}\}} \min_{h\in H^*} d(v,h) \le 2d_2. \quad (3)$$

Since we assume that $d_1 > 0$, we have $f(v_{\max}) \ne v_{\max}$ and thus $f(v_{\max}) \in D \setminus \{v_{\max}\}$ (recall that $f(v_{\max})$ is the hub assigned to $v_{\max}$ in the optimal solution of $\mathcal{I}$). By inequality (3), there exists $h \in H^*$ such that $d(f(v_{\max}), h) \le 2d_2$, which implies that

$$d(v_{\max}, h) \le d(v_{\max}, f(v_{\max})) + d(f(v_{\max}), h) \le d_1 + 2d_2. \quad (4)$$

Line 5 returns a solution that opens H^* as the set of p hubs. For each $v \in D$, let $f'(v) := \mathrm{argmin}_{h\in H^*} d(v,h)$; i.e., $f'(v)$ is the hub in H^* assigned to v in the solution returned by the algorithm. Let d_1' and d_2' be the largest value and second-largest value in the multi-set $\{d(v, f'(v)) \mid v \in D\}$. Combining inequalities (3) and (4), we have

$$d_1' \le d_1 + 2d_2 \quad \text{and} \quad d_2' \le 2d_2. \quad (5)$$

Let $v_1, v_2 \in D$ be the nodes achieving the maximum path length in our solution, i.e., $l(v_1, v_2) = \mathrm{ALG}$. It suffices to show that $\mathrm{ALG} \le 3.5 \cdot \mathrm{OPT}$. If $f'(v_1) = f'(v_2)$, then $\mathrm{ALG} = d(v_1, f'(v_1)) + d(v_2, f'(v_2)) \le d_1' + d_2'$. If $f'(v_1) \ne f'(v_2)$, then $\mathrm{ALG} = d(v_1, f'(v_1)) + d(f'(v_1), c) + f(f'(v_2), c) + d(v_2, f'(v_2)) \le d_1' + 2d_0 + d_2'$, where we use $d(h,c) \le d_0$ for all $h \in H$ by our choice of H. Using inequality (5), we always have

$$\begin{aligned}\mathrm{ALG} &\le 2d_0 + d_1' + d_2' \le 2d_0 + (d_1 + 2d_2) + 2d_2 \\ &= 2d_0 + d_1 + 4d_2. \quad (6)\end{aligned}$$

By inequalities (6), (1) and (2), we obtain

$$\begin{aligned}\mathrm{ALG} &\le 2d_0 + d_1 + 4d_2 \\ &\le 2(d_0 + d_1) + 3d_2 \quad (\text{using } d_2 \le d_1) \\ &\le 2\mathrm{OPT} + 1.5 \cdot \mathrm{OPT} \\ &= 3.5 \cdot \mathrm{OPT},\end{aligned}$$

which indicates that Algorithm 1 is a 3.5-approximation algorithm for S*p*HCP. Hence we have the theorem. □

4. Concluding remarks

In this work, we have studied the star p-hub center problem from an approximation viewpoint. An interesting open problem is that of bridging the gap between the lower bound $1.25 - \epsilon$ and the upper bound 3.5 on the approximability of S*p*HCP. We conjecture that a better approximation ratio may be achievable by considering the two-level assignment integrally, or by a proper rounding of some linear programming relaxations of S*p*HCP. It is also of interest to explore whether S*p*HCP can be solved efficiently in some special metric spaces.

Acknowledgments

This work was supported in part by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301, and the National Natural Science Foundation of China Grant 61033001, 61061130540, 61073174.

References

[1] S.A. Alumur, B.Y. Kara, Network hub location problems: the state of the art, European J. Oper. Res. 190 (1) (2008) 1–21.
[2] J. Byrka, B. Rybicki, Improved LP-rounding approximation algorithm for k-level uncapacitated facility location, in: Proc. 39th ICALP, 2012.
[3] J.F. Campbell, Integer programming formulations of discrete hub location problems, European J. Oper. Res. 72 (2) (1994) 387–405.
[4] J.F. Campbell, A.T. Ernst, Hub location problems, in: Z. Drezner, H.W. Hamacher (Eds.), Facility Location: Applications and Theory, Springer, Berlin, 2002, pp. 373–407.
[5] P. Chardaire, J.L. Lutton, A. Sutter, Upper and lower bounds for the two-level simple plant location problem, Ann. Oper. Res. 86 (0) (1999) 117–140.
[6] I. Contreras, E. Fernández, A. Marín, Tight bounds from a path based formulation for the tree of hub location problem, Comput. Oper. Res. 36 (12) (2009) 3117–3127.
[7] I. Contreras, E. Fernández, A. Marín, The tree of hubs location problem, European J. Oper. Res. 202 (2) (2010) 390–400.
[8] A.T. Ernst, H. Hamacher, H. Jiang, M. Krishnamoorthy, G. Woeginger, Uncapacitated single and multiple allocation p-hub center problems, Comput. Oper. Res. 36 (7) (2009) 2230–2241.
[9] M. Garey, D. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W. H. Freeman, 1979.
[10] D.S. Hochbaum, D.B. Shmoys, A unified approach to approximation algorithms for bottleneck problems, J. ACM 33 (3) (1986) 533–550.
[11] B.Y. Kara, B.Ç Tansel, On the single-assignment p-hub center problem, European J. Oper. Res. 125 (3) (2000) 648–655.
[12] M. Labbé, H. Yaman, Solving the hub location problem in a star–star network, Networks 51 (1) (2008) 19–33.
[13] T. Meyer, A. Ernst, M. Krishnamoorthy, A 2-phase algorithm for solving the single allocation p-hub center problem, Comput. Oper. Res. 36 (12) (2009) 3143–3151.
[14] M.E. O'Kelly, H.J. Miller, Solution strategies for the single facility minimax hub location problem, Pap. Reg. Sci. 70 (4) (1991) 367–380.
[15] H. Yaman, S. Elloumi, Star p-hub center problem and star p-hub median problem with bounded path lengths, Comput. Oper. Res. 39 (11) (2012) 2725–2732.