An approximation algorithm for single allocation k-hub routing problem*

Li-Hsuan Chen[1]†, Sun-Yuan Hsieh[1], Ling-Ju Hung[1]‡, Bang Ye Wu[2]
[1]Department of Computer Science and Information Engineering
National Cheng Kung University, Tainan 701, Taiwan
clh100p@cs.ccu.edu.tw
hsiehsy@mail.ncku.edu.tw
hunglc@cs.ccu.edu.tw
[2]Department of Computer Science and Information Engineering
National Chung Cheng University, Chiayi 42102, Taiwan
bangye@cs.ccu.edu.tw

Abstract

Given a metric graph $G = (V, E, w)$ and a positive integer k, the Single Allocation k-Hub Routing *problem is to find a spanning subgraph H^* of G such that (i) $C^* \subset V$ is a clique of size k in H^*; (ii) $V \setminus C^*$ forms an independent set in H^*; (iii) each $v \in V \setminus C^*$ is adjacent to exactly one vertex in C^*; and (iv) the routing cost $r(H^*) = \sum_{u,v \in V} d_{H^*}(u, v)$ is minimized where $d_{H^*}(u, v)$ is the distance of vertices u and v in H^*. The vertices selected in C^* are called* hubs *and the rest of vertices are called* non-hubs. *In this paper, we show that the* Single Allocation k-Hub Routing *problem is NP-hard in metric graph and we give a 2-approximation algorithm to solve this problem running in $O(n^2)$ time where n is the number of vertices in the input graph.*

1 Introduction

The design of hub-and-spoke networks is a key issue with applications on transportation, *e.g.,* airline [8] and cargo delivery systems [14]. The major concern to design a hub-and-spoke network with high quality is to connect a large amount of origin/destination (O/D) pairs by using a small number of links. The usage of hub facilities helps to reduce the connections between all nodes. To locate k hubs in hub networks in order to route the traffic between origin/destination pairs with minimum cost is the classical hub location problem called the k-Hub Median problem [13, 15]. Notice that the general k-hub median problem considers that each pair of origin/destination has different unit traffic (flow) cost. We call a hub location problem *multi-allocation*, if a demand node can be served by several hubs. If each demand node can be served by exactly one hub, the hub location problem is *single-allocation*. The k-hub median problem is NP-hard. Many linear programming-based and heuristic algorithms were proposed to solved the k-hub median problem and its variants problems (see the survey papers [1, 2, 12]).

Another hub location problem, the Single Allocation k-Hub Center problem, is to choose a fixed number k of vertices as hubs and to assign each non-hub vertex to exactly one of the chosen hubs in such a way that the maximum distance/cost between origin-destination pairs is minimized [3, 14]. Unlike the k-Hub Median problem to minimize the total cost of all origin-destination pairs, the Single Allocation k-Hub Center problem is to minimize the poorest service quality. Chen *et al.* [4] proved that for any $\epsilon > 0$, it is NP-hard to approximate the Single Allocation k-Hub Center problem to a ratio $\frac{4}{3} - \epsilon$ and gave a $\frac{5}{3}$-approximation algorithm running in time $O(kn^3)$ to solve the same problem. The Star k-Hub Center problem is another hub location

*Parts of this work were supported by the Ministry of Science and Technology of Taiwan under grants MOST 105–2221–E–006–164–MY3 and MOST 103–2221–E–006–135–MY3

†Li-Hsuan Chen is supported by the Ministry of Science and Technology of Taiwan under grant MOST 106–2811–E–006–008

‡Ling-Ju Hung is supported by the Ministry of Science and Technology of Taiwan under grant MOST 105–2811–E–006–046.

problem with min-max criterion. It is to pick k nodes as hubs among the set of demand nodes connecting with the central given hub c and to connect each of the remaining demand nodes to exactly one of the k chosen hubs such that the longest path in the tree structure network is minimized. Chen *et al.* [5] showed that for any $\epsilon > 0$, to approximate the Star k-Hub Center problem to a ratio $1.5 - \epsilon$ is NP-hard and give a $\frac{5}{3}$-approximation algorithms for the same problem. Moreover, for input graphs satisfying β-triangle inequality, *i.e.*, $w(u, v) \leq \beta \cdot (w(u, x) + w(x, v))$ for all vertices u, v, x in the input graph $G = (V, E, w)$ and $\beta \geq 1/2$, it was showed that for any $\epsilon > 0$, to approximate the Star k-Hub Center problem to a ratio $g(\beta) - \epsilon$ is NP-hard and $r(\beta)$-approximation algorithms were given in the same paper where $g(\beta)$ and $r(\beta)$ are functions of β [6].

Although numerous research results on solving various hub location problems in past twenty-five years [2, 7], the design of approximation algorithms for hub location problems only made very little progress in the past two decades, especially for the k-Hub Median problem [10, 11]. In this paper, we consider a variant of the k-Hub Median problem that each pair of origin/destination has the same unit traffic (flow) cost called the Single Allocation k-Hub Routing problem. The Single Allocation k-Hub Routing problem is to choose a fixed number k of vertices as hubs and to assign each non-hub vertex to exactly one of the chosen hubs in such a way that the sum of distance/cost between all origin-destination pairs is minimized.

In this paper, we consider a graph $G = (V, E, w)$ with a distance function $w(\cdot, \cdot)$ being a metric on V such that $w(v, v) = 0$, $w(u, v) = w(v, u)$, and $w(u, v) + w(v, r) \geq w(u, r)$ for all $u, v, r \in V$. Let the routing cost $r(G) = \Sigma_{u,v} d_G(u, v)$ be the sum of all pairs distance in the graph where $d_G(u, v)$ denotes the length of shortest paths between u and v in G. We list the formal definition of the Single Allocation k-Hub Routing problem in the following.

Single Allocation k-Hub Center (SAkHC)

Input: A metric graph $G = (V, E, w)$ and a positive integer k.

Output: A spanning subgraph H^* of G such that (i) vertices (hubs) in $C^* \subset V$ form a clique of size k in H^*; (ii) vertices (non-hubs) in $V \setminus C^*$ form an independent set in H^*; (iii) each non-hub $v \in V \setminus C^*$ is adjacent to exactly one hub in C^*; and (iv) the routing cost $r(H^*)$ is minimized.

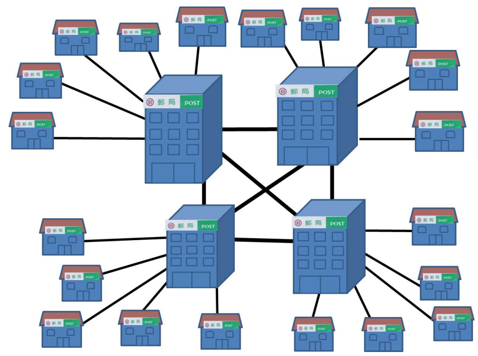

Figure 1: An example of a single allocation k-hub routing network with $k = 4$ where the hubs are the major post offices and the non-hubs are the other small post offices.

In this paper, we investigate the approximability of the Single Allocation k-Hub Routing problem.

The paper is organized as follows: In Section 2, we prove that the Single Allocation k-Hub Routing problem is NP-hard. In Section 3, we give a 2-approximation algorithm running in time $O(n^2)$ for the Single Allocation k-Hub Routing problem where n is the number of vertices in the input graph.

We close this section with some notation definitions. For a vertex v in a graph H, we use $N_H(v)$ to denote the set of vertices adjacent to v and $N_H[v] = N_H(v) \cup \{v\}$. For a vertex set X, we use $N_H(X) = \bigcup_{v \in X} N_H(v) \setminus X$. For u, v in graph H, let $d_H(u, v)$ denote the distance between u and v in H. For a graph H, we use $r(H) = \Sigma_{u,v \in H} d_H(u, v)$ to denote the routing cost of H where $d_H(u, v)$ denotes the length of shortest paths between u and v in H.

2 NP-hardness

In this section, we show that the Single Allocation k-Hub Routing problem is NP-hard.

Lemma 1. *The* Single Allocation k-Hub Routing *problem is NP-hard.*

Proof. We prove the Single Allocation k-Hub Routing problem is at least as hard as the well-known NP-hard problem Exact Cover by 3-Sets (X3C) [9]. To show such a statement, we reduce the the input $(\mathcal{U}, \mathcal{S})$ of the Exact Cover by 3-Sets problem to the Single Allocation k-Hub Routing problem.

Exact Cover by 3-Sets (X3C)

Input: A universe $\mathcal{U}$ of elements, $|\mathcal{U}| = 3q$ and a collection $\mathcal{S}$ of size 3 subsets of $\mathcal{U}$, $|\mathcal{S}| = m$.

Question: Is there a subset $\mathcal{S}^* \subseteq \mathcal{S}$ with $|\mathcal{S}| = q$ such that $\bigcup_{s_i \in \mathcal{S}^*} s_i = \mathcal{U}$?

Let $(\mathcal{U}, \mathcal{S})$ be an input instance of Exact Cover by 3-Sets.

We construct a metric graph $G = (V = \mathcal{U} \cup \mathcal{S} \cup \mathcal{B}, E, w)$ of the Single Allocation k-Hub Routing problem according to $(\mathcal{U}, \mathcal{S})$. We add a vertex set $\mathcal{B}$ with $|\mathcal{B}| = m^3$ which is partitioned into m groups B_s with size $|B_s| = m^2$ and each B_s is corresponding to one vertex $s \in \mathcal{S}$. For $b \in B_s$, we say that s is the big brother of $b \in B_s$. Each vertex in $\mathcal{B}$ has exactly one big brother. We define the cost of edges as follows.

- For $u, v \in \mathcal{U}$, $w(u, v) = 2$.
- For $v \in \mathcal{U}$ and $s \in \mathcal{S}$, if $v \in s$, $w(v, s) = 1$; otherwise $w(v, s) = 2$.
- For $s_i, s_j \in \mathcal{S}$, $w(s_i, s_j) = 1$.
- For $b \in B_s$, $w(b, s) = 1$.
- For $b \notin B_s$, $w(b, s) = 2$.
- For $b \in B_s$ and $b' \in B_{s'}$, if $s = s'$, $w(b, b') = 2$; otherwise $w(b, b') = 3$.
- For $b \in \mathcal{B}$ and $u \in \mathcal{U}$, $w(b, u) = 2$.

For a feasible solution H, the routing cost will be computed in two parts: $r(H) = r_1(H) + r_2(H)$, where $r_1(H)$ is the total pairwise distance between vertices in $\mathcal{B}$ and $r_2(H)$ is the remaining distance of other pairs $r_2(H) = r(H) - r_1(H)$.

It is not hard to see that G is a metric graph. Let H^* be an optimal solution of the Single Allocation k-Hub Routing problem in G with $k = m$ and C^* be the set of k hubs in H^*.

Suppose that $\mathcal{S}^* \subset \mathcal{S}$ is a solution of X3C. We now construct a spanning subgraph H by choosing vertices in $\mathcal{S}$, *i.e.,* $C = \mathcal{S}$. For each $v \in \mathcal{U}$, connect v to a set $s \in \mathcal{S}^*$ such that $v \in s$ and $w(v, s) = 1$. All vertices in $\mathcal{B}$ are connected to their big brothers.

Notice that there is exact one such set $s \in S^*$ that contains v since $\mathcal{S}^*$ is a solution of X3C. We see that

$$\begin{aligned}
r(H) &= r_{X3C} = \sum_{u,v \in V} d_H(u, v) \\
&= \sum_{u,v \in C} d_H(u, v) + \sum_{u \in V \setminus C, v \in C} d_H(u, v) \\
&\quad + \sum_{u,v \in V \setminus C} d_H(u, v) \\
&= \frac{m \cdot (m-1)}{2} \\
&\quad + 3q \cdot (1 + 2(m-1)) \\
&\quad + m^3 \cdot (1 + 2(m-1)) \\
&\quad + 3 \cdot \frac{3q \cdot (3q-1)}{2} - 3 \cdot q \\
&\quad + 3 \cdot \frac{m^3 \cdot (m^3 - m^2)}{2} \\
&\quad + 2 \cdot \frac{m^3 \cdot (m^2 - 1)}{2} \\
&\quad + 3q \left(3 \cdot (m^3 - m^2) + 2 \cdot m^2\right) \\
&= m^5 \left(\frac{3m-1}{2}\right) + o(m^5).
\end{aligned}$$

It follows that $r(H^*) \leq r_{X3C}$.

We then show that if there exists an algorithm that finds an optima solution H^* of the Single Allocation k-Hub Routing problem in G such that $r(H^*) \leq r_{X3C}$, then there exists a solution $\mathcal{S}' \subseteq \mathcal{S}$ of the X3C problem. Supposed that H^* is a solution of the Single Allocation k-Hub Routing problem on graph G with $r(H^*) \leq r_{X3C}$.

Claim 1. *There are less than $\frac{m^2}{2} + m$ many vertices from $\mathcal{S}$ not connected to their big brothers in H^*.*

Proof. If there are $\frac{m^2}{2} + m$ vertices from $\mathcal{S}$ that are not connected to their big brothers. Since there are only m hubs allowed, at least $\frac{m^2}{2}$ of them are non-hubs and we call these non-hubs B'. The total pairwise distance between vertices in $\mathcal{B}$ is at least

$$\begin{aligned}
r_1(H^*) &= \sum_{u,v \in \mathcal{B}} d_H(u, v) \\
&= \sum_{u,v \in \mathcal{B} \setminus B'} d_G(u, v) \\
&\quad + \sum_{u \in \mathcal{B} \setminus B', v \in B'} d_H(u, v) \\
&\quad + \sum_{u,v \in B'} d_H(u, v).
\end{aligned}$$

Since the distance on graph H can not be smaller than the distance on G for all pairs of vertices, we have

$$\begin{aligned} r_1(H^*) &\geq \sum_{u,v\in\mathcal{B}\setminus B'} d_G(u,v) \\ &\quad + \sum_{u\in\mathcal{B}\setminus B', v\in B'} (1 + d_G(u,v)) \\ &\quad + \sum_{u,v\in B'} 4 \\ &\geq m^5(\frac{3m}{2}) + o(m^5) \\ &> r_{X3C}. \end{aligned}$$

This contradicts to the assumption that $r(H^*) \leq r_{X3C}$. This completes the proof. $\square$

Claim 2. *All $s \in \mathcal{S}$ must be hubs and all elements from $\mathcal{B}$ and $\mathcal{U}$ are non-hubs in H^*.*

Proof. If there is a non-hub $s \in \mathcal{S}$ in H^*, there are at least $m^2 - m$ vertices in $\mathcal{B}$ that are non-hubs and not connected to their big brother s. This contradicts to the fact of Claim 1. Therefore, $\mathcal{S}$ is the hub set in H^*. Since the size of the hub set is m, we see that all elements in $\mathcal{B}$ and $\mathcal{U}$ are non-hubs. This completes the proof. $\square$

Claim 3. *For all non-hubs $v \in \mathcal{B} \cup \mathcal{U}$, if u is the hub connected to v in H, we have $d_{H^*}(u,v) = 1$.*

Proof. Suppose that there is a non-hub $v \in \mathcal{B}\cup\mathcal{U}$ connected to a hub u with $d_{H^*}(u,v) > 1$. By Claim 2 the the hub set of graph H^* is $\mathcal{S}$, we can always find another hub u' with $d_H(u',v) = 1$. Since $|B_s| = m^2$, by the help of Claim 1, there are at least $\frac{m^2}{2} - m$ non-hub in $\mathcal{B}$ connected to u'. Let H' be the graph obtained by changing the connected hub and making each non-hub in H^* connected to a hub with distance 1. We have $r(H^*) \geq r(H') + \frac{m^2}{2} - m$. This is contradict to the assumption that $r(H) \leq r_{X3C}$. This completes the proof. $\square$

Claim 4. *Each $s \in C$ either is not adjacent to any non-hub in $\mathcal{U}$ or is adjacent to exactly three non-hubs in $\mathcal{U}$.*

Proof. In graph H^*, if there are x_1 hubs connected to three non-hubs in $\mathcal{U}$, x_2 hubs connected to two non-hubs in $\mathcal{U}$ and x_3 hubs connected to one non-hub in $\mathcal{U}$ with $3x_1 + 2x_2 + x_3 = 3q$ and $x_1 < q$, the pairwise distance between vertices in $\mathcal{U}$ is at least

$$\begin{aligned} & 3 \cdot \frac{3q \cdot (3q-1)}{2} - 3 \cdot x_1 - x_2 \\ \geq\ & 3 \cdot \frac{3q \cdot (3q-1)}{2} - 3 \cdot q. \end{aligned}$$

We have $r(H^*) > r_{X3C}$. This contradicts to the assumption that $r(H^*) \leq r_{X3C}$. This completes the proof.

$\square$

According to Claims 1–4, the collection of hubs that are connecting to non-hubs in $\mathcal{U}$ is a solution to the (X3C) problem. Therefore the Single Allocation k-Hub Routing problem is NP-hard. $\square$

3 A 2-approximation algorithm

In this section, we give a 2-approximation algorithm for the Single Allocation k-Hub Routing problem.

Let H^* be an optimal solution of Single Allocation k-Hub Routing problem in G and $r(H^*)$ is the routing cost of H^*. We have the following lemma.

Lemma 2. *Let G be a connected weight graph. If G' is a spanning subgraph of G, the routing cost of G' is not smaller than the routing cost of G, i.e., $r(G') \geq r(G)$*

Proof. Since G' is a subgraph of G and G is a metric graph, we have $d_{G'}(u,v) \geq d_G(u,v)$. By definition, routing cost is the sum of all pairs distance in the graph, we have

$$\begin{aligned} r(G') &= \sum_{u,v\in V} d_{G'}(u,v) \\ &\geq \sum_{u,v\in V} d_G(u,v) = r(G). \end{aligned}$$

Therefore, we have $r(G') \geq r(G)$ and the lemma is proved. $\square$

Let $z = \arg\min_{v\in V} \sum_{u\in V} d_G(u,v)$ and Z is a spanning star G with center z, we have the following lemma.

Lemma 3. *The routing cost of Z is not greater than two times the sum of all edge cost in G, i.e., $r(Z) \leq 2 \cdot \sum_{u,v\in V} d_G(u,v)$.*

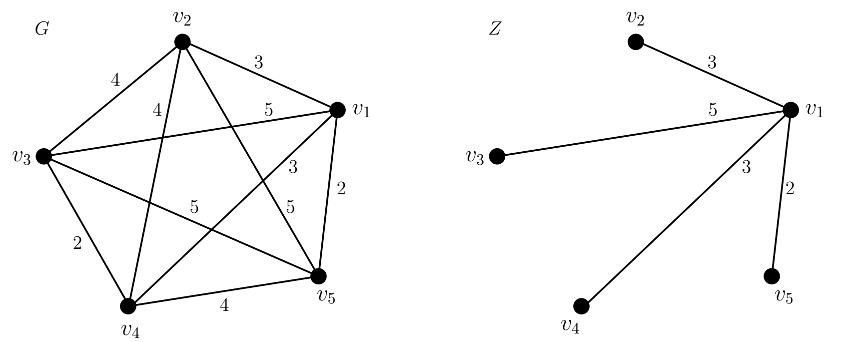

Figure 2: An example of a spanning star Z of graph G with $v_1 = \arg\min_{v\in V} \sum_{u\in V} d_G(u, v)$.

Proof. Let $\mathcal{S}$ be the set of all spanning stars of G, then the weight sum of all spanning stars in $\mathcal{S}$ is equal to adding all edge weight in G twice. In other words,

$$\sum_{s\in V}\sum_{u\in V} w(u, s) = 2 \cdot \sum_{u,v\in V} w(u, v).$$

Since $z = \arg\min_{v\in V} \sum_{u\in V} w(u, v)$, we have

$$\begin{aligned}\sum_{u\in V} w(u, z) &\leq \frac{\sum_{s\in V}\sum_{u\in V} w(u, s)}{n}\\ &= \frac{2\cdot\sum_{u,v\in V} w(u, v)}{n}.\end{aligned}$$

By definition,

$$\begin{aligned} r(Z) &= \sum_{u,v\in V} d_Z(u, v)\\ &= (n-1)\cdot\sum_{u\in V} w(u, z)\\ &\leq \frac{2(n-1)\cdot\sum_{u,v\in V} w(u, v)}{n}\\ &\leq 2\cdot\sum_{u,v\in V} w(u, v).\end{aligned}$$

Therefore, we have $r(Z) \leq 2\cdot \sum_{u,v\in V} w(u, v)$ and the lemma is correct. □

Algorithm BasicAPX$_{SAkHR}$
Let $U := V$. Initially, $C = \emptyset$. Construct a spanning subgraph H of G by the following steps.

Step 1: Find $v_1 = \arg\min_{v\in V} \sum_{u\in V} w(u, v)$

Step 2: Pick $k-1$ vertices $\{v_2, \ldots, v_k\}$ in U. Let $C := C \cup \{v_1, v_2, \ldots, v_k\}$ and $U := U \setminus \{v_1, v_2, \ldots, v_k\}$.

Step 3: Connect all vertices in U to v_1.

Step 4: Return H.

Theorem 1. *There is a 2-approximation algorithm for the* Single Allocation k-Hub Routing *problem running in time $O(n^2)$.*

Proof. It is easy to see that in time $O(n^2)$ Algorithm BasicAPX$_{SAkHR}$ return a spanning subgraph of G satisfying that C is a clique of size k in H; $V \setminus C$ forms an independent set in H; and each vertex in $V \setminus C$ is adjacent to exactly one vertex in C.

We now show that H is a 2-approximate solution. Let H^* denote an optimal solution of the Single Allocation k-Hub Routing problem and $r(H^*)$ is the routing cost of H^*. Since H^* is a spanning subgraph of G, according to Lemma 2, we have $r(G) \leq r(H^*)$. By Lemma 3, we have $r(Z) \leq 2\cdot r(G) \leq 2\cdot r(H^*)$. Since Z is a subgraph of H, by Lemma 2, we have $r(H) \leq r(Z) \leq 2\cdot r(H^*)$. Thus,

$$r(H) \leq 2\cdot r(H^*).$$

This completes the proof. □

4 Concluding remarks

In this paper, we give a 2-approximation of the Single Allocation k-Hub Routing problem. For the future work, it is interesting to see whether the gap between lower and upper bounds can be reduced. One possibility is to show that for any $\epsilon > 0$, it is NP-hard to approximate the Single Allocation k-Hub Routing problem to a ratio $\alpha - \epsilon$ where $\alpha > 0$. The other possibility is to design a γ-approximation algorithm for the Single Allocation k-Hub Routing problem and $\gamma < 2$.

References

[1] S. Alumur and B.Y. Kara, Network hub location problems: the state of the art, *European Journal of Operational Research*, **190**, pp. 1-21, 2008.

[2] J. F. Campbell and M. E. O'Kelly, Twenty-five years of hub location research, *Transportation Science*, **46**, pp. 153–169, 2012.

[3] J. F. Campbell, Integer programming formulations of discrete hub location problems, *European Journal of Operational Research*, **72** (1994), 387–405.

[4] L.-H. Chen, D.-W. Cheng, S.-Y. Hsieh, L.-J. Hung, C.-W. Lee, and B. Y. Wu, Approximation algorithms for single allocation k-hub center problem, *Proceedings of the 33rd Workshop on Combinatorial Mathematics and Computation Theory (CMCT 2016)*, pp. 13–18, 2016.

[5] L.-H. Chen, D.-W. Cheng, S.-Y. Hsieh, L.-J. Hung, C.-W. Lee, B. Y. Wu, Approximation algorithms for the star k-hub center problem in metric graphs, *Proceeding of the 22nd Annual International Computing and Combinatorics Conference (COCOON 2016)*, LNCS 9797, pp. 222-234,2016.

[6] L.-H. Chen, S.-Y. Hsieh, L.-J. Hung, R. Klasing, C.-W. Lee, B. Y. Wu, On the complexity of the star p-hub center problem with parameterized triangle inequality, *The 10th International Conference on Algorithms and Complexity (CIAC 2017)*, accepted.

[7] I. Contreras, Hub location problems, in G. Laporte, S. Nickel, and F. Saldanha da Gama, editors, *Location Science*, Springer, 2015.

[8] A. T. Ernst, H. Hamacher, H. Jiang, M. Krishnamoorthy, and G. Woeginger, Uncapacitated single and multiple allocation p-hub center problems, *Computer & Operations Research*, **36** (2009), 2230–2241.

[9] M.R. Garey, D.S. Johnson: *Computers and Intractability: A guide to the theory of NP-completeness*. W. H. Freeman and Company, San Francisco, 1979.

[10] M. Iwasa, H. Saito, T. Matsui, Approximation algorithms for the single allocation problem in hub-and-spoke networks and related metric labeling problems, *Discrete Applied Mathematics*, **157**, pp. 2078–2088, 2009.

[11] Y. Kuroki and T. Matsui, Approximation algorithms for hub location problems, *The 9th Annual Meeting of Asian Association for Algorithms and Computation (AAAC 2016)*, accepted.

[12] N. Mladenović, J. Brimberg, P. Hansen, J.A. Moreno-Pérez, The p-median problem: a survey of metaheuristic approaches, *European Journal of Operational Research*, **179**, pp. 927–939, 2007.

[13] M. E. O'Kelly, A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research*, Vol. 32, pp. 393–404, 1987.

[14] M. E. O'Kelly and H. J. Miller, Solution strategies for the single facility minimax hub location problem, *Papers in Regional Science*, **70** (1991), 367–380.

[15] R. Todosijević, D. Urošević, N. Mladenović, and S. Hanafi, A general variable neighborhood search for solving the uncapacitated r-allocation p-hub median problem, *Optimization Letters*, Online. Available: DOI 10.1007/s11590-015-0867-6