An Approximation Algorithm for Star p-Hub Routing Cost Problem

Sun-Yuan Hsieh(✉), Li-Hsuan Chen, and Wei Lu

Department of Computer Science and Information Engineering,
National Cheng Kung University, Tainan 701, Taiwan
{hsiehsy,p76041166}@mail.ncku.edu.tw,
clh100p@cs.ccu.edu.tw

Abstract. Given a metric graph $G = (V, E, w)$, a center $c \in V$, and an integer p, we discuss the Star p-Hub Routing Cost Problem in this paper. We want obtain a depth-2 tree which has a root and root is adjacent to p vertices called hubs. We call that this tree is a star p-hub tree and let the sum of distance in tree between all pairs of vertices be minimum. We prove the Star p-Hub Routing Cost Problem is NP-hard by reducing the Exact Cover by 3-Sets Problem to it. The Exact Cover by 3-Sets Problem is a variation of set cover problem and known NP-hard problem. After proving the Star p-Hub Routing Cost Problem is NP-hard, we present a 4-approximation algorithm running in polynomial time $O(n^2)$ for the Star p-Hub Routing Cost Problem.

Keywords: Graph theory · Algorithm · NP-hard · Approximation · Hub location · Star · Routing cost

1 Introduction

Among the available forms of interconnecting network, the intense exchange of entities among different areas impels the desire for more efficient architecture. Instead of directly routing flows from their origin node to the destination, a hub-and-spoke network is the architecture which can arise to better routing performance of nodes exchange flows. Hub-and-spoke network design consists of locating hubs from a set of candidate nodes, deciding which hub arcs will be installed and granted scale economies, and allocating non-hub nodes to hubs such that a given objective is optimized (O'Kelly and Miller 1994) [13]. The objective of this is the total cost to construct the connected network which usually composed of fixed setup costs for hubs and hub arcs, and of transportation costs. Hub location problems constitute an important class of problems in logistics which numerous applications in passenger/cargo transportation, postal services, telecommunications, etc. According to the definition of hub location problems, hubs are special facilities of the network responsible for switching, aggregating, transshipment and sorting. In many-to-many distributed systems, these hubs routing the flows either to their final destinations or to other hubs. Instead of serving each pair of service directly, hubs concentrate flows to take advantage of economies of scales. Hub facilities are inter-connected by hub arcs forming a hub level network, since the demand flows are routed in bulks through this sub-network, high-capacity

C.-Y. Chang et al. (Eds.): ICS 2018, CCIS 1013, pp. 524–531, 2019.
https://doi.org/10.1007/978-981-13-9190-3_58

carriers can be used on those hub arcs so that lower unitary transportation cost can be attained. A complete network between the installed hubs equipped with efficient means of transport that allow a flow- independent discount factor to be applied to the inter-hub transportation costs.

Hub location problem is used in many situations in reality, including telecommunication system, traffic system, or medical system, etc. The target is different in different situation. Hub location problem has many different target functions. The k-hub median problem minimizes the total cost of output and the k-hub center problem minimize the diameter of output which had been proved that it is NP-hard and presented 4/3-inapproxibility and a 5/3-approximation algorithm by Chen et al. [3].

In this paper, we discuss the Star p-Hub Routing Cost Problem(SpHRP) that output is a 2-level rooted tree and routing cost is minimum. We prove the problem SpHRP is NP-hard by reducing a NP-hard problem to SpHRP and present a approximation algorithm for SpHRP.

The rest of the paper is organized as follows. In the Sect. 2, we formally define the problem SpHRP. In the Sect. 3, we prove that SpHRP is NP-hard. In the Sect. 4, we present a approximation algorithm for SpHRP. We conclude the paper in Sect. 5.

2 Preliminaries

Let u, v be two vertices. $d_G(u, v)$ denotes distance between u, v in graph G and $d_T(u, v)$ denotes distance between u, v in tree T. Define $C(T) = \sum_{u,v \in V} d_T(u, v)$ called the routing cost of tree T and $c_s = \mathrm{argmin}_{u \in V} \sum_{u,v \in V} d_T(u, v)$ called the center of minimum star. The input of the problem SpHRP is a metric, undirected, and complete graph, a vertex c and a parameter p. We want to obtain a tree that root is c, p vertices called hubs are adjacent to c and each of remaining vertices called terminal is adjacent to a hub. Our target is that minimizing routing cost. Routing cost is sum of distance between all pairs of vertices.

Star p-Hub Routing Cost Problem

Input: A undirected metric graph $G = (V, E, w)$, a vertex $c \in V$, and positive integer p.
Output: A depth-2 spanning tree T^* rooted by c called the central hub such that c has exactly p children (called hubs) and the routing cost of T^*, $C(T^*)$, is minimized.

To prove SpHRP is NP-hard, we need to reduce a NP-hard problem to SpHRP. Exact Cover by 3-Sets Problem (X3C) is a variation of set cover problem and a well known NP-hard problem [9]. Original set cover problem is that give a universal set and many subset of universal set and let union of the least subsets be universal set. In the problem X3C, the amount of member of universal set is multiple of three, every subsets are 3-sets, and each member of universal set appears in the union of chosen subsets once.

3 NP-Hardness

In this section, we prove that the problem SpHRP is a NP-hard problem with the problem X3C, which is one of NP-hard problem.

Exact Cover by 3-Sets Problem

Input: Given a set U, with $|U| = n = 3q$ (so, the size of U is a multiple of 3), and a collection S of 3-element subsets of U.
Output: Is there a subset S' of S where every element of U occurs in exactly one member of S'?

Theorem 1. *If the* Star p-Hub Routing Cost Problem *can be solved in polynomial time, then* Exact Cover by 3-Sets Problem *is polynomial time solvable.*

Proof. Let $(\mathcal{U}, \mathcal{S})$ be an input instance of Exact Cover by 3-Sets Problem. $\mathcal{U}$ is universal set and $\mathcal{S}$ is a collection of 3-element subset of $\mathcal{U}$. We construct a metric graph $G = (V \cup S \cup L \cup \{c\}, E, w)$ of the Star p-Hub Routing Cost Problem according to $(\mathcal{U}, \mathcal{S})$. First, we define the vertex set of G. c is a specified center. V corresponds to $\mathcal{U}$. For each subset $S_i \in \mathcal{S}$, we create a vertex s_i which belongs to S, and a vertex set L_i which contains b vertices and belongs to L.

It is obviously that any three vertices u, v, r in G satisfy $w(u, v) + w(v, r) \geq w(u, r)$. Thus G is a metric graph and we reduce input of X3C to input of SpHRP. Let T^* be an optimal solution of the Star p-Hub Routing Cost Problem with input graph G. We can construct T^* on condition that all vertex s in S is hub, all vertex l_i in L is terminal and is adjacent to s_i in S which creates l_i, all v_i in U is terminal and is adjacent to s_i which covers v_i, and v in U must be adjacent to the same vertex in S with the others two vertices in U. The routing cost in T^* is $C(T^*)$, where $C(T^*) = (4m^2 - 2m)b^2 + (6m^2 + 24mq - 2m - 12q)b + (2m^2 + 18mq + 36q^2 - 24q)$.

In the following, we will prove T^* is an optimal solution of the Star p-Hub Routing Cost Problem with input graph G. First, we prove that all vertex s in S is hub.

Claim 3.1. *In the T^*, all vertices in S are hubs and adjacent to c.*

Proof. Let T_1 be solution of SpHRP and T_1^* be optimal solution in T_1. We support that a vertex s belonging S is not a hub or adjacent to c in T_1. If the lower bound of the routing cost of T_1 is bigger than the routing cost of T^*, T_1 can't be T^*. When we calculating the lower bound of the routing cost of T_1, we just need to calculate cost between L instead of V because the lower bound won't increase when amount of vertices which are calculated are decrease.

We can sum up all minimum distance between each pair of vertices when calculating the lower bound of routing cost of T_1 by observing the above equation. A vertex l in L just has five possible location in solution. The first location is hub, and the other four possible location are terminals and distance to the hub is 1, 2, 3, 4, respectively. In the following, we will discuss sum of minimum distance from l to the other vertices when l in each possible location.

The first possible location is that l is a hub and adjacent to c. Up to m members in L can be put in this location.

(The remaining detailed proofs are omitted due to page limit.)

With Claim 3.1, we will prove the second condition of T^*.

Claim 3.2. *In the T^*, all vertices in L are adjacent to the vertex in S which creates it.*

Proof. Let T_2 be solution of SpHRP and T_2^* be optimal solution in T_2. We support that a vertex l in L is not adjacent to s in S which creates l in T_2. The vertex set V is consist of S, U, L', $\{l\}$, and $\{c\}$. Among them, L' is L excluding l.

$$\sum_{x,y\in A\cup B} d(x,y) = \sum_{x,y\in A} d(x,y) + 2\times \sum_{x\in A,y\in B} d(x,y) + \sum_{x,y\in B} d(x,y)$$

Using the above equation, we simplify the calculation of the routing cost of T_2.

$$\begin{aligned} C(T_2) &= \sum_{x,y\in V} d_{T_2}(x,y) \\ &= \sum_{x\in L'}\sum_{y\in L'} d_{T_2}(x,y) + 2\sum_{x\in L'}\sum_{y\in S\cup\{c\}} d_{T_2}(x,y) + \sum_{x\in S\cup\{c\}}\sum_{y\in S\cup\{c\}} d_{T_2}(x,y) \\ &\quad + 2\sum_{y\in L'\cup S\cup\{c\}} d_{T_2}(l,y) + 2\sum_{x\in U}\sum_{y\in L'} d_{T_2}(x,y) + 2\sum_{x\in U}\sum_{y\in S\cup\{c\}} d_{T_2}(x,y) \\ &\quad + 2\sum_{x\in U} d_{T_2}(x,l) + \sum_{x\in U}\sum_{y\in U} d_{T_2}(x,y) \end{aligned}$$

First, we calculate the lower bound of routing cost of T_2 between vertices in L'.

(Claims 3.2.1 to 3.2.7 are omitted due to page limit.)

In the third part, we will prove the third condition of T^* with Claim 3.1 and Claim 3.2.

Claim 3.3. *In the T^*, all vertices in U are adjacent to the vertex in S which covers it.*

Proof. Let T_3 be solution of SpHRP and T_3^* be optimal solution in T_3. We support that a vertex v in U is not adjacent to s in S which covers v in T_3. The vertex set V is consist of S, U', L, $\{v\}$, and $\{c\}$. Among them, U' is U excluding v. Using the Eq. 1, we simplify the calculation of the routing cost of T_3.

$$\begin{aligned} C(T_3) &= \sum_{x,y\in V} d_{T_3}(x,y) \\ &= \sum_{x\in L}\sum_{y\in L} d_{T_3}(x,y) + 2\sum_{x\in L}\sum_{y\in S\cup\{c\}} d_{T_3}(x,y) + \sum_{x\in S\cup\{c\}}\sum_{y\in S\cup\{c\}} d_{T_3}(x,y) \\ &\quad + 2\sum_{y\in L\cup S\cup\{c\}} d_{T_3}(v,y) + 2\sum_{x\in U'} d_{T_3}(x,v) \\ &\quad + 2\sum_{x\in U'}\sum_{y\in L\cup S\cup\{c\}} d_{T_3}(x,y) + \sum_{x\in U'}\sum_{y\in U'} d_{T_3}(x,y) \end{aligned}$$

First, we calculate the lower bound of routing cost of T_3 between vertices in L.

(Claims 3.3.1 to 3.3.6 with their proofs are omitted due to page limit.)

In the last part, we will prove the fourth condition of T^* with Claim 3.1, Claim 3.2, and Claim 3.3.

Claim 3.4. *In the* T^*, v_i *in* U *must be adjacent to the same vertex in* S *with the others two vertices in* U.

Proof. Let T_4 be solution of S*p*HRP and T_4^* be optimal solution in T_4. We support that a vertex v in U is not adjacent to the same vertex in S with the others two vertices in U in T_4. The vertex set V is consist of S, U', L, $\{v\}$, and $\{c\}$. Among them, U' is U excluding v. Using the Eq. 1, we simplify the calculation of the routing cost of T_4 of T_4.

$$\begin{aligned} C(T_4) &= \sum_{x,y\in V} d_{T_4}(x,y) \\ &= \sum_{x\in L}\sum_{y\in L} d_{T_4}(x,y) + 2\sum_{x\in L}\sum_{y\in S\cup\{c\}} d_{T_4}(x,y) + \sum_{x\in S\cup\{c\}}\sum_{y\in S\cup\{c\}} d_{T_4}(x,y) \\ &\quad + 2\sum_{x\in U}\sum_{y\in L\cup S\cup\{c\}} d_{T_4}(x,y) + \sum_{x\in U'}\sum_{y\in U} d_{T_4}(x,y) + \sum_{y\in U} d_{T_4}(v,y) \end{aligned}$$

Next, we calculate the lower bound of routing cost of T_4 between vertices in U and vertices in $L \cup S \cup \{c\}$.

Claim 3.4.1. $\sum_{x\in U}\sum_{y\in L\cup S\cup\{c\}} d_{T_4}(x,y) \geq (12mq - 6q)b + 9mq.$

Next, we will calculate the lower bound of routing cost of T_4 between vertices in U' and vertices in U.

Claim 3.4.2. $\sum_{x\in U'}\sum_{y\in U} d_{T_4}(x,y) \geq 36q^2 - 36q + 8.$

At last, we will calculate the lower bound of routing cost of T_4 between v which is not adjacent to the same vertex in S with the others two vertices in U and vertices in U.

Claim 3.4.3. $\sum_{y\in U} d_{T_4}(v,y) \geq 12q - 6.$

According above four parts, we can construct T^* on condition that all vertices in S are hubs, all vertices in L are terminals and adjacent to the vertex in S which creates it, all vertices in U are terminal and adjacent to the vertex in S which covers it, and each vertex v in U must be adjacent to the same vertex in S with the others two vertices in U.

When we know the structure of output of S*p*HRP, we can get output of X3C from output of S*p*HRP. Each hub in S which is adjacent to three vertices in U covers adjacent vertices in U. Obviously, all vertices in U are adjacent to a vertex in S. It means that all elements in U are covered and the hubs, vertices in S, which are adjacent to three vertices in U are the output of X3C. We prove that we can obtain the input of S*p*HRP from the input of X3C in polynomial time and obtain the output of X3C from the output of S*p*HRP in polynomial time. If we can solve S*p*HRP in polynomial time, we can solve X3C in polynomial time by obtaining the input of S*p*HRP from the input of X3C, solving S*p*HRP and obtaining the output of X3C from the output of S*p*HRP in polynomial time.

4 Approximation Algorithm

Next we will present a 4-approximation algorithm for the Star p-Hub Routing Cost Problem. Let c_s be the center of minimum star in G, $c_s = \text{argmin}_{v \in V} \sum_{u \in V} (u, v)$, and c_n be the nearest vertex of vertex c_s, $c_n = \text{argmin}_{v \in V} d_G(c_s, v)$. Let H^* be hubs in optimal solution T^* and $h_1^*, h_2^*, \ldots, h_p^*$ present each hub in H^*. H be hubs in our algorithm solution T and $h_1, h_2, \ldots, h_p$ present each hub in H.

We can get the following lemma.

Lemma 1. $C(T^*) \geq C(G)$.

Proof. Because input graph G is a metric graph, complete graph and satisfied triangle inequality, the minimum distance between two vertices is weight of edge between two vertices in G.

$$d_{T^*}(u, v) \geq d_G(u, v)$$

$$C(T^*) = \sum_{u \in V} \sum_{v \in V} d_{T^*}(u, v) \geq \sum_{u \in V} \sum_{v \in V} d_G(u, v) = C(G)$$

The routing cost of optimal solution T^* is bigger or equal than the routing cost of input graph G.

Lemma 2. $C(T^*) \geq n \sum_{v \in V} d_G(c_s, v)$.

Proof. $s_c = \arg\min_{v \in V} \sum_{u \in V} (u, v)$

$$C(T^*) \geq C(G) = \sum_{u \in V} \sum_{v \in V} d(u, v) = n \sum_{v \in V} d_G(c_s, v)$$

(The proofs of Lemmas 3 and 4 are omitted due to page limit.)

Lemma 3. $C(T^*) \geq n(n-1) d_G(c_s, c_n)$.

Lemma 4. $C(T^*) \geq 2(n-1) \sum_{i=2}^{p} d_G(h_i, c)$.

Algorithm APX

Step 1: Find c_s, center of minimum star. $c_s = \text{argmin}_{u \in V} \sum_{v \in V} (u, v)$.
Step 2: Find c_n, the nearest vertex of vertex c_s. $c_n = \text{argmin}_{v \in V} d_G(c_s, v)$.
Step 3: If vertex c_s is different from vertex c, let vertex c_s be h_1. Otherwise, let vertex c_n be h_1.
Step 4: Connect vertex c and vertex h_1.
Step 5: Pick $p-1$ vertices $\{h_2, h_3, \ldots, h_p\}$ closest to c from $V \backslash \{c, h_1\}$ and connect them to vertex c. Let $N_T(c) = \{h_1, h_2, h_3, \ldots, h_p\}$.
Step 6: Connect all vertices in $V \backslash \{c, h_1, h_2, \ldots, h_p\}$ to h_1 in T.
Step 7: Return the tree T.

Theorem 2. There is a 4-approximation algorithm for the Star p-Hub Routing Cost Problem running in time $O(n^2)$ where n is the number of vertices in the input graph.

Proof. There are two kinds of the output of algorithm. One is T_s that h_1 is vertex c_s, the other is T_n that h_1 is vertex c_n. We will prove that both routing cost of T_s and T_n are less or equal than four times of routing cost of optimal solution.

(The proofs of Lemmas 5 and 6 are omitted due to page limit.)

Lemma 5. $C(T_s) \leq 4C(T^*)$.

Lemma 6. $C(T_n) \leq 4C(T^*)$.

By Lemmas 5 and 6, we prove the routing cost of algorithm solution is less or equal than four times of the routing cost of optimal solution. Next, we will prove that the algorithm run in time $O(n^2)$. Obviously, finding center of the minimum induced star run in time $O(n^2)$, finding the nearest vertex of center of the minimum induced star run in time $O(n)$, picking $p-1$ vertices closet to c run in time $O(n)$, and connecting two vertices run in time $O(1)$. The proposed algorithm is a 4-approximation algorithm running in time $O(n^2)$.

5 Conclusion

In this paper, we formally define the Star p-Hub Routing Cost Problem. Then, we prove the Star p-Hub Routing Cost Problem is NP-hard, by reducing a known NP-hard problem, the Exact cover by 3-sets problem, to our problem. In the last, we present a 4-approximation algorithm using minimum star to find hub to solve the Star p-Hub Routing Cost Problem. For the future work, it is interesting to see whether there exists an α-approximation algorithm and $\alpha < 4$ or to prove that for any $\varepsilon > 0$, it is NP-hard to approximate the Star p-Hub Routing Cost Problem to a ratio $4 - \varepsilon$.

References

1. Campos, R., Ricardo, M.: A fast algorithm for computing minimum routing cost spanning trees. Comput. Netw. **52**(17), 3229–3247 (2008)
2. Chen, L.H., Cheng, D.W., Hsieh, S.Y., Hung, L.J., Lee, C.W., Wu, B.Y.: Approximation algorithms for single allocation k-hub center problem. In: Proceedings of the 33rd Workshop on Combinatorial Mathematics and Computation Theory (CMCT 2016), pp. 13–18 (2016)
3. Chen, L.-H., Cheng, D.-W., Hsieh, S.-Y., Hung, L.-J., Lee, C.-W., Wu, B.Y.: Approximation algorithms for the star k-hub center problem in metric graphs. In: Dinh, T.N., Thai, M.T. (eds.) COCOON 2016. LNCS, vol. 9797, pp. 222–234. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42634-1_18
4. Chen, L.-H., Hsieh, S.-Y., Hung, L.-J., Klasing, R., Lee, C.-W., Wu, B.Y.: On the complexity of the star p-hub center problem with parameterized triangle inequality. In: Fotakis, D., Pagourtzis, A., Paschos, V.T. (eds.) CIAC 2017. LNCS, vol. 10236, pp. 152–163. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57586-5_14

5. Chen, L.-H., Hsieh, S.-Y., Hung, L.-J., Klasing, R.: The Approximability of the p-hub center problem with parameterized triangle inequality. In: Cao, Y., Chen, J. (eds.) COCOON 2017. LNCS, vol. 10392, pp. 112–123. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62389-4_10
6. Wu, Y., et al.: Approximability and inapproximability of the star p-hub center problem with parameterized triangle inequality. J. Comput. Syst. Sci. **92**, 92–112 (2018)
7. Dionne, R., Florian, M.: Exact and approximate algorithms for optimal network design. Networks **9**(1), 37–59 (1979)
8. Fischetti, M., Lancia, G., Serafini, P.: Exact algorithms for minimum routing cost trees. Networks **39**(3), 161–173 (2002)
9. Garey, M.R., Johnson, D.S.: Computers and Intractability: A guide to the theory of NP-completeness. W. H. Freeman and Company, San Francisco (1979)
10. Hu, T.C.: Optimum communication spanning trees. SIAM J. Comput. **3**(3), 188–195 (1974)
11. Hochuli, A., Holzer, S., Wattenhofer, R.: Distributed approximation of minimum routing cost trees. In: Halldórsson, M.M. (ed.) SIROCCO 2014. LNCS, vol. 8576, pp. 121–136. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-09620-9_11
12. Lin, W.C., Wu, B.Y.: A 2-approximation algorithm for the clustered minimum routing cost tree problem. In: Intelligent Systems and Applications: Proceedings of the International Computer Symposium (ICS) Held at Taichung, Taiwan, 12–14 December 2014, vol. 274, p. 3. IOS Press, April 2015
13. O'Kelly, M.E., Miller, H.J.: The hub network design problem: a review and synthesis. J. Transp. Geograph. **2**(1), 31–40 (1994)
14. Reddy, K.R.U.K.: A survey of the all-pairs shortest paths problem and its variants in graphs. Acta Universitatis Sapientiae, Informatica **8**(1), 16–40 (2016)
15. Wu, B.Y., Chao, K.M., Tang, C.Y.: Approximation algorithms for some optimum communication spanning tree problems. Discrete Appl. Math. **102**(3), 245–266 (2000)
16. Wu, B.Y., Lancia, G., Bafna, V., Chao, K.M., Ravi, R., Tang, C.Y.: A polynomial-time approximation scheme for minimum routing cost spanning trees. SIAM J. Comput. **29**(3), 761–778 (2000)
17. Yaman, H., Elloumi, S.: Star p-hub center problem and star p-hub median problem with bounded path lengths. Comput. Oper. Res. **39**(11), 2725–2732 (2012)