On the Complexity of the Uncapacitated Single Allocation p-Hub Median Problem with Equal Weights

Steffen Wolf
Distributed Algorithms Group
University of Kaiserslautern
wolf@informatik.uni-kl.de

Internal Report
No. 363/07

July 2007

Abstract

The Super-Peer Selection Problem is an optimization problem in network topology construction. It may be cast as a special case of a Hub Location Problem, more exactly an Uncapacitated Single Allocation p-Hub Median Problem with equal weights. We show that this problem is still NP-hard by reduction from Max Clique.

1 Problem Description

The Super-Peer Selection Problem (SPSP) is the problem of finding a Super-Peer topology in a Peer-to-Peer network with minimal total communication cost. It may be cast as a special case of a well known NP-hard Hub Location Problem, the Uncapacitated Single Allocation p-Hub Median Problem (USApHMP) [3, 5, 6, 8].

In both problems, p nodes are to be selected out of n nodes to serve as hubs (or super-peers). The remaining nodes are assigned to one hub, respectively. Hubs are assumed to be assigned to themselves. Denoting the assignment of node i to hub j by a binary variable $x_{ij} = 1$, the transportation costs for the link between i and j by d_{ij}, and the amount of flow from i to j by w_{ij}, the cost of such a hub location can be defined as

$$C(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \left(d_{ik} + d_{kl} + d_{jl}\right) \cdot w_{ij} \cdot x_{ik} \cdot x_{jl}. \tag{1}$$

This cost is to be minimized. There are a number of constraints:

$$x_{ij} \leq x_{jj} \qquad i, j = 1, \ldots, n \tag{2}$$

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad i = 1, \ldots, n \tag{3}$$

$$\sum_{j=1}^{n} x_{jj} = p \tag{4}$$

$$x_{ij} \in \{0, 1\} \qquad i, j = 1, \ldots, n \tag{5}$$

The set of constraints (2) ensures that nodes are assigned only to hubs, while (3) enforces the allocation of a node to exactly one hub. Due to constraint (4), there will be exactly p hubs. Now, the SPSP is the special case of the USApHMP, where all flow is $w_{ij} = 1$.

The general USApHMP is known to be NP-hard. In fact, even when the set of hubs is fixed, the remaining assignment problem is still NP-hard, as soon as $p \geq 3$ [8], but polynomial time algorithms exist for $p = 2$ or $p = 1$ [7]. The proof for NP-hardness in [8] reduces the NP-hard 3-Terminal Cut to USApHMP. It needs the weights to be arbitrarily selectable from $w_{ij} \in \{0, 1\}$. Since SPSP fixes the weights to $w_{ij} = 1$, the proof cannot be reused, and it is still unclear whether SPSP is NP-Hard.

2 NP-Hardness

The SPSP is NP-hard. We show this by reducing Max Clique [4, problem GT19] to a special SPSP.

Max Clique problem: Given a graph $G = (V, E)$ and a number $k > 0$, does there exist a subset $V' \subseteq V$ with cardinality $|V'| \geq k$, such that every two vertices in $i, j \in V'$ are joined by an edge $(i, j) \in E$? Max Clique is NP-hard [4].

Max Clique can be transformed to an SPSP by introducing a new vertex $a \notin V$, setting $V' := V \cup \{a\}$, the number of hubs to $p := k + 1$, and the distance function d_{ij} to:

$$d_{ij} = \begin{cases} 0 & \text{if } i = a \vee j = a \vee i = j \\ 1 & \text{else, if } (i, j) \in E \\ 2 & \text{else} \end{cases} \tag{6}$$

Theorem 2.1. *The decision problem whether there is a super-peer selection X with cost $C(X) \leq Z = (p-1) \cdot (p-2)$ in the weighted graph $G' = (V', V' \times V', d)$ is equivalent to the Max Clique decision problem in the original graph $G = (V, E)$.*

Proof. We first show that a solution for the Max Clique Problem is also a solution for the SPSP. We then show that every solution for the SPSP is a solution for Max Clique.

Consider a solution for the Max Clique problem, i. e. a subset $V' \subset V$, $|V'| = k$, such that every two vertices $i, j \in V'$ are joined by and edge $(i, j) \in E$. The super-peer selection $X_a(V')$ that chooses all $p = k + 1$ nodes from the set $V' \cup \{a\}$ to be hubs, and assigns the remaining nodes to hub a, only creates costs on the inter-hub links. Since all distances to and from node a are zero, its links to the non-hub nodes and the other $p - 1$ hub nodes again adds nothing to the total cost. The cost for the super-peer selection $X_a(V')$ is just the sum of the remaining inter-hub links between the $p - 1$ hubs V':

$$C(X_a(V')) = \sum_{k \in V'} \sum_{l \in V'} d_{kl} \cdot x_{kk} \cdot x_{ll} = \sum_{k \in V'} \sum_{l \in V' \setminus \{k\}} 1 = (p-1) \cdot (p-2) \leq Z \qquad (7)$$

Thus, every solution for the Max Clique Problem is a solution for the SPSP.

Consider an arbitrary assignment X_b that selects p hubs and assigns the non-hub nodes to hubs. We will consider the following non-disjoint cases, and show that in each case the assignment X_b does not give a solution to the SPSP, i. e. its cost is larger than Z:

1. Node a is not selected as a hub in X_b.
2. A non-hub node is not assigned to a in X_b.
3. A set of nodes V' is selected as the hub-set in X_b, such that two of them $i, j \in V'$ are not joined by an edge $(i, j) \in E$.

If node a is not selected as a hub, all inter-hub links between the p hubs will have distances $d_{ij} \geq 1$, creating a cost of $C(X_b) \geq p \cdot (p-1) > Z$. Thus, every solution to the SPSP must select a as a hub. Since all other distances are $d_{ij} \geq 1$, the remaining inter-hub links create a cost of $H \geq (p-1) \cdot (p-2) = Z$.

If a non-hub node i is not assigned to a, its link to the hub $j \neq a$ creates additional cost of at least $d_{ij} \geq 1$. The total cost is then $C(X_b) \geq H + d_{ij} > Z$. Thus, every solution to the SPSP must assign all non-hub nodes to hub a.

If two of the selected hubs $i, j \in V'$ are not joined by an edge $(i, j) \in E$, the corresponding distance is $d_{ij} = 2$. This increases the cost for the inter-hub links and thus the total cost to $C(X_b) \geq H + 1 > Z$.

In all cases 1, 2 and 3, the assignment X_b is not a solution to the SPSP, thus, every solution to the SPSP must select node a as a hub, assign all non-hub nodes to a, and select a fully connected set of nodes V' as hubs, which is also a solution to the Max Clique problem. □

Note, that if X_b selects a different set of fully connected nodes than X_a as hub-set, it only constitutes a solution to the SPSP if node a is also selected as a hub and all non-hub nodes are assigned to a. However, the chosen hubset minus node a is again a solution to the Max Clique problem.

3 Other Results

The proof can be reused to show the NP-hardness of the p-hub center problem with equal weights ($\min C(X) = \max_{i,j,k,l \in V}(d_{ik} + d_{kl} + d_{jl}) \cdot x_{ik} \cdot x_{jl}$, s. t. (2), (3), (4), (5)), by using the same distance function d_{ij} as in (6) and setting $Z = 1$.

The Max Clique problem is also known to be NP-hard if $k = r \cdot |V|$, for any fixed r with $0 < r < 1$ [4]. Thus, the SPSP with $p = r \cdot |V| + 1$ is also NP-hard.

For the special case of $p \in \mathcal{O}(1)$, however, the remaining assignment problem when all hubs are fixed becomes solvable in polynomial time for the SPSP, but is still NP-hard for a constant $p \geq 3$ for the general USApHMP [8]. Since there are only $\binom{n}{p} \in \mathcal{O}(n^p)$ different selections of hubs, the full SPSP becomes solvable in polynomial time in this case.

The SPSP uses equal weights $w_{ij} = 1$, thus the weight on the inter-hub link (i, j) only depends on the number of nodes assigned to i and j, but not on the actual node assignment. Let $E_i = \{j \mid x_{ji} = 1\}$ denote the set of nodes assigned to hub i. Now, the cost for the inter-hub link (i, j) is $c_{ij} = 2 \cdot |E_i| \cdot |E_j| \cdot d_{ij}$, and the cost for the link between a non-hub node i and its hub j is $c_{ij} = 2 \cdot n \cdot d_{ij}$. There are only $\mathcal{O}(n^p)$ different combinations of the numbers of nodes assigned to the p hubs.

Given a set of hubs and the number of nodes assigned to each hub $|E_i|$, the remaining assignment problem can be cast as a special case of the Linear Sum Assignment Problem (LSAP) [1, 2]. The LSAP can be modelled as follows:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} \cdot x_{ij} \tag{8}$$

s. t.:

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad i = 1, \ldots, n \tag{9}$$

$$\sum_{i=1}^{n} x_{ij} = 1 \qquad j = 1, \ldots, n \tag{10}$$

$$x_{ij} \in \{0, 1\} \qquad i, j = 1, \ldots, n \tag{11}$$

The binary variables x_{ij} implicate a one-to-one assignment of the rows and columns. Consider a partition of the set $\{1, \ldots, n\}$ in p disjoint subsets F_i, such that each subset is as large as the number of nodes assigned to hub i ($|F_i| = |E_i|$), and a cost function that assigns the cost $c_{ij} = d_{ik}$ to all $j \in F_k$. Now, the solution to the LSAP minimizes the total cost for the links between non-hub nodes and their hubs.

Since all hubs and the numbers of nodes for each hub are fixed, the cost for the inter-hub link is also fixed. Thus, the total cost $C(X)$ is minimized by the LSAP. Since there exist several algorithms for the LSAP with time complexity $\mathcal{O}(n^4)$ [2], the SPSP with $p \in \mathcal{O}(1)$ fixed hubs is solvable in polynomial time, by solving the $\mathcal{O}(n^{2p})$ many LSAPs for fixed hubs and fixed numbers of nodes assigned to the hubs.

4 Acknowledgements

We would like to thank Peter Merz, Matthias Priebe, Thomas Fischer, Tanja Meyer, Johannes Kloos and Andreas T. Ernst for their comments, suggestions and thoughts on the equal weights USApHMP.

References

[1] Rainer E. Burkard and Eranda Çela. Linear assignment problems and extensions. Technical Report 127, June 1998.

[2] Rainer E. Burkard, Mauro Dell'Amico, and Silvano Martello. Linear sum assignment problem. In *Assignment Problems*, chapter 4.1. SIAM, Philadelphia, 2008. in preparation.

[3] Andreas T. Ernst and Mohan Krishnamoorthy. Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location Science*, 4(3):139–154, 1996.

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

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

[6] Morton E. O'Kelly, Deborah L. Bryan, Darko Skorin-Kapov, and Jadranka Skorin-Kapov. Hub network design with single and multiple allocation: A computational study. *Location Science*, 4(3):125–138, 1996.

[7] Jinhyeon Sohn and Sungsoo Park. A linear program for the two-hub location problem. *European Journal of Operational Research*, 100:617–622, 1997.

[8] Jinhyeon Sohn and Sungsoo Park. The single allocation problem in the interacting three-hub network. *Networks*, 35(1):17–25, 2000.