

# The design of single allocation incomplete hub networks <sup>☆</sup>

Sibel A. Alumur <sup>\*</sup>, Bahar Y. Kara, Oya E. Karasan

Department of Industrial Engineering, Bilkent University, Bilkent, 06800 Ankara, Turkey

## ARTICLE INFO

### Article history:

Received 18 July 2008

Received in revised form 9 April 2009

Accepted 9 April 2009

### Keywords:

Hub location  
Network design  
 $p$ -hub median  
Hub covering  
 $p$ -hub center

## ABSTRACT

The hub location problem deals with finding the location of hub facilities and allocating the demand nodes to these hub facilities so as to effectively route the demand between any origin–destination pair. In the extensive literature on this challenging network design problem, it has widely been assumed that the subgraph induced by the hub nodes is complete. Relaxation of this basic assumption constitutes the starting point of the present work. In this study, we provide a uniform modeling treatment to all the single allocation variants of the existing hub location problems, under the incomplete hub network design. No network structure other than connectivity is imposed on the induced hub network. Within this context, the single allocation incomplete  $p$ -hub median, the incomplete hub location with fixed costs, the incomplete hub covering, and the incomplete  $p$ -hub center network design problems are defined, and efficient mathematical formulations for these problems with  $O(n^3)$  variables are introduced. Computational analyses with these formulations are presented on the various instances of the CAB data set and on the Turkish network.

© 2009 Elsevier Ltd. All rights reserved.

## 1. Introduction

Hub facilities are used to consolidate and disseminate flow in many-to-many distribution systems. The hub location problem deals with finding the location of hub facilities and the allocation of the non-hub nodes (demand centers) to these located hub facilities. It is generally assumed that there is economy of scale generated through hub-to-hub transportations due to consolidation of flows. This economy of scale is typically estimated by using a constant cost discount factor ( $\alpha$ ) for travel on the inter-hub connections. Hub location problems arise in various application settings in telecommunication and transportation (air passenger, cargo, public) network design.

The first mathematical formulation for the hub location problem was introduced by O'Kelly (1987). He proposed a quadratic integer programming formulation for the problem of minimizing the total transportation cost for a given number of hubs to locate ( $p$ -hub median problem). The following hub location literature mostly focused on this total transportation cost objective locating both a fixed and a variable number of hubs. Such studies are O'Kelly (1992); Campbell (1996); Ernst and Krishnamoorthy (1996, 1998, 1999); Skorin-Kapov et al. (1996); Ebery (2001); Boland et al. (2004) and Marin et al. (2006). An interested reader can refer to two surveys on hub location problems by Campbell et al. (2002) and Alumur and Kara (2008a).

Campbell (1994) introduced two new hub location problems to the literature:  $p$ -hub center and hub covering problems. The  $p$ -hub center problem locates  $p$  hubs such that the maximum cost (travel time) between any origin–destination pair is minimized. Different integer programming formulations are presented in Kara and Tansel (2000) and Ernst et al. (2009) for the  $p$ -hub center problem. The hub covering problem, on the other hand, minimizes the number of hubs to establish while ensuring that the cost (travel time) between any origin–destination pair does not exceed a specified value. The  $p$ -hub center

<sup>☆</sup> This research is supported by Turkish Academy of Science.

<sup>\*</sup> Corresponding author. Tel.: +90 532 3619960; fax: +90 312 2664054.

E-mail address: [alumur@bilkent.edu.tr](mailto:alumur@bilkent.edu.tr) (S.A. Alumur).

and hub covering problems naturally arise for systems in which time is a major concern, such as ground and air cargo transportation. Kara and Tansel (2003) placed the hub covering problem in the NP-hard class and provided various linear integer programming formulations to solve it. Ernst et al. (2005) and Wagner (2008) further proposed different mathematical formulations. Hamacher and Meyer (2006) compared various formulations and presented some facet-defining valid inequalities for the hub covering problem.

In hub location studies, it is typically assumed that the hub network is complete with the presence of a direct hub link between every hub pair. However, it is sometimes possible to provide nearly the same service quality of a complete hub network, in terms of cost and/or service time, with an efficiently designed incomplete one. Thus, building complete hub networks may unnecessarily increase the total investment cost in designing hub networks. Furthermore, in reality, many less-than-truckload and telecommunication networks do not operate on a complete hub network structure.

In this study, we focus on designing hub networks that are not necessarily complete, for single allocation hub location problems. We do not impose any structure on the hub networks other than connectivity.

There are few studies in the literature relaxing the complete hub network assumption in hub location problems. O'Kelly and Miller (1994) introduced different hub network design protocols, one being the incomplete hub network design, to the literature. Nickel et al. (2001) proposed a model for the hub location problem arising in urban public transport networks. The authors minimized the total transportation cost plus the fixed costs of locating hubs and building hub links. They considered the multiple allocation structure in which a non-hub node can be allocated to more than one hub. Campbell et al. (2005a,b) proposed hub arc location problems to the literature. Such problems locate hub arcs with reduced unit costs, rather than locating hub facilities. A fixed number of hub arcs is located while minimizing the total transportation costs. The resulting hub arc network in these problems does not need to be connected unless forced. Yoon and Current (2008) studied the multiple allocation incomplete hub network design problem with fixed and variable arc costs. Their proposed model minimized the total transportation costs and the fixed costs of locating hubs and hub links. They also considered direct connections between the non-hub nodes and incurred variable arc costs associated with demand on the arcs. A study by Alumur and Kara (2008b) considered the hub covering network design problem. Inspired by the observations from the cargo sector in Turkey, the authors proposed an incomplete hub covering model in which every origin–destination pair receives service by visiting at most three hubs on a route. Recently, Calik et al. (2009) proposed a tabu-search based heuristic for the single allocation hub covering problem over incomplete hub networks.

The hub location problem has also been studied within the telecommunication network design context. Many special cases of building incomplete hub networks have been studied in this literature. Designing different hub network topologies such as star, ring, tree, and path are considered. The objective in the telecommunication network design is to minimize the total costs of building the hub networks. The reader may refer to Klinecicz (1998) for such applications.

In this work, our aim is to propose a uniform modeling treatment for the incomplete variants of all the existing single allocation hub location problems in the literature. Initially, we define the problems with the minimization of total transportation cost objective. We present efficient mathematical formulations with  $O(n^3)$  variables for the single allocation incomplete  $p$ -hub median and the hub location with fixed costs network design problems. The aim of the model with the minimization of total transportation cost objective is to find the location of hub nodes, the allocation of non-hub nodes to these hub nodes, and which hub links to establish between the hub nodes. The transportation cost model is then tested on the various instances of the Civil Aeronautics Board (CAB) data set – a benchmark data set introduced in the literature (O'Kelly, 1987) – and on the Turkish network with the optimization software CPLEX 11.2. We proceed with defining the single allocation incomplete hub covering and  $p$ -hub center network design problems and present  $O(n^3)$  mathematical formulations for these problems. We explore the performance of some valid inequalities for the hub covering model and provide some computational analysis on the CAB data set and on the Turkish network with both of the problems.

The outline of this paper is as follows. In the next section, we propose our mathematical formulations for the single allocation  $p$ -hub median and the hub location with fixed costs network design problems. The third section compiles the computational results on both the CAB data set and the Turkish network with the minimization of total transportation cost objective. In the fourth section, we present a mathematical formulation for the single allocation incomplete hub covering network design problem and in Section 5, we present some valid inequalities for this model. Section 6 presents the computational results on both the CAB data set and the Turkish network with the hub covering formulation. Section 7, we present a mathematical formulation and its computational results for the single allocation incomplete  $p$ -hub center network design problem. We present our concluding remarks in the last section of the paper.

## 2. Minimization of total transportation costs in designing incomplete hub networks

For the mathematical model, we need a given node set  $N$  consisting of  $n$  demand nodes and a potential hub set  $H$  such that  $H \subseteq N$  with  $h$  nodes. Our aim is to decide on the location of hub nodes, the allocation of non-hub nodes to these hub nodes, and which hub links to establish between the hub nodes, while minimizing total transportation costs. We aim to design our network so that any node can send flow to any other node in the network. In order to ensure this, the hub network to be established must be connected.

It may be straightforward to model the problem at hand using multi-commodity flow balance constraints with  $O(n^4)$  variables. For example, we could use  $f_{ij}^{kl}$  variables to represent the flow on hub arc  $(i,j)$  originating from node  $k$  destined to node  $l$ ,

similar to the multiple allocation formulations proposed by Nickel et al. (2001) and Yoon and Current (2008). In this paper, we developed a formulation with  $O(n^3)$  variables. Since all the flow originating from a non-hub node must visit the single hub that the non-hub node is allocated to, the calculation of the total flow from a non-hub node to its hub node is trivial. Thus, for each non-hub node, this value is exactly the total amount of flow originating from the non-hub node. Similarly, the total flow from a hub node to a non-hub node is exactly the total amount of flow destined to this non-hub node. What is challenging is the calculation of the total flow between hub nodes over an incomplete hub network with few variables. In the following we introduce one such model.

The single allocation incomplete  $p$ -hub median network design problem is to locate  $p$  hubs, to allocate each non-hub node to a single hub, and to determine which  $q$  hub links to establish between hubs such that the total transportation cost is minimized.

In order to present the mathematical formulation for the incomplete  $p$ -hub median network design problem, we first need to define some parameters:

$w_{ij}$  = flow from node  $i \in N$  to node  $j \in N$ .

$c_{ij}$  = transportation cost of a unit of flow between nodes  $i \in N$  and  $j \in N$ .

$\alpha$  = hub-to-hub transportation cost discount factor.

$p$  = number of hubs to be established.

$q$  = number of hub links to be established.

Also, let  $O_i = \sum_j w_{ij}$  be the total amount of flow originating from node  $i$  and  $D_j = \sum_i w_{ij}$  be the total amount of flow destined to node  $j$ .

We then define the decision variables of the mathematical model:

$$x_{ij} = \begin{cases} 1 & \text{if node } i \in N \text{ is allocated to hub at node } j \in H; \\ 0 & \text{otherwise.} \end{cases}$$

$$z_{ij} = \begin{cases} 1 & \text{if a hub link is established between hubs } i \in H \text{ and } j \in H; \\ 0 & \text{otherwise.} \end{cases}$$

$f_{ij}^k$  = total amount of flow originating from node  $k \in N$  to be routed on hub link  $\{i, j\}$  in the direction from  $i \in H$  to  $j \in H$ .

The single allocation incomplete  $p$ -hub median network design problem is modeled as:

$$\min \sum_{i \in N} \sum_{k \in H} c_{ik} O_i x_{ik} + \sum_{i \in H} \sum_{j \in H} \sum_{k \in N} \alpha c_{ij} f_{ij}^k + \sum_{i \in N} \sum_{k \in H} c_{ki} D_i x_{ik} \quad (1)$$

$$\text{s.t. } \sum_{j \in H} x_{ij} = 1 \quad \forall i \in N \quad (2)$$

$$x_{ij} \leq x_{ji} \quad \forall i \in N, \quad j \in H \quad (3)$$

$$z_{ij} \leq x_{ii} \quad \forall i, j \in H: \quad i < j \quad (4)$$

$$z_{ij} \leq x_{jj} \quad \forall i, j \in H: \quad i < j \quad (5)$$

$$\sum_{j \in H} x_{jj} = p \quad (6)$$

$$\sum_{i \in H} \sum_{j \in H: j > i} z_{ij} = q \quad (7)$$

$$\sum_{j \in H: j \neq i} f_{ji}^k + O_k x_{ki} = \sum_{j \in H: j \neq i} f_{ij}^k + \sum_{l \in N} w_{kl} x_{li} \quad \forall i \in H, \quad k \in N \quad (8)$$

$$f_{ij}^k + f_{ji}^k \leq O_k z_{ij} \quad \forall i, j \in H: \quad i < j, \quad k \in N \quad (9)$$

$$f_{ij}^k \geq 0 \quad \forall i, j \in H: \quad i \neq j, \quad k \in N \quad (10)$$

$$x_{ij} \in \{0, 1\} \quad \forall i \in N, \quad j \in H \quad (11)$$

$$z_{ij} \in \{0, 1\} \quad \forall i, j \in H: \quad i < j \quad (12)$$

The objective function (1) minimizes the total transportation cost. The first term in the objective function calculates the total cost of transportation from non-hub nodes to hub nodes, the second term calculates the total discounted cost of transportation in the hub network, and the third term calculates the total cost of transportation from hub nodes to non-hub nodes.

We use a single allocation structure so that every non-hub node is allocated to a single hub. This is obtained via Constraint (2). A non-hub node can only be allocated to a hub node, which is ensured by Constraint (3).

Constraints (4) and (5) guarantee that a hub link can only be established if both of the end nodes of that link are hub nodes. Note that since we are designing an undirected hub network we defined  $z_{ij}$  variables only for  $i < j$ . Constraints (6) and (7) ensure that exactly  $p$  hub facilities and  $q$  hub links are to be established.

We calculate the amount of flow to be routed on hub link  $\{i, j\}$  in either orientation from any node  $k$  in the network via Constraint (8). Note that this is the divergence equation introduced by Ernst and Krishnamoorthy (1996) for the single

allocation  $p$ -hub median problem with complete hub networks. By Constraint (8), for each node  $k$  and hub node  $i$ , the total flow originating from the node  $k$  entering into the hub node  $i$  must be equal to the outgoing flow. The first terms on the left and right-hand sides of Constraint (8) calculate the flow within the hub network ( $f$  variables), whereas the second terms correspond to the flows via the allocations. By Constraint (9), we restrict the  $f$  variables to be positive only on the established hublinks.

The rest of the constraints of the model (10)–(12) represent non-negativity and binary requirements.

In the worst case  $h = n$  and the model has  $(\frac{3}{2}n^2)$  binary variables and  $(n^3)$  real variables. The number of constraints of our model is  $(n^3 + 3n^2 + n + 2)$ . Hence, in total we have  $O(n^3)$  variables and constraints.

We name the problem in which the number of hubs and hub links to be established are taken as decision variables in the incomplete  $p$ -hub median formulation as the incomplete hub location with fixed costs network design problem. Let,

$FH_j$  = fixed cost of opening a hub at node  $j \in H$ .

$FL_{ij}$  = fixed cost of opening a hub link between hubs  $i \in H$  and  $j \in H$

The single allocation incomplete hub location with fixed costs network design problem is modeled as:

$$\min \sum_{k \in H} FH_k x_{kk} + \sum_{i \in H} \sum_{j \in H: j > i} FL_{ij} z_{ij} + \sum_{i \in N} \sum_{k \in H} C_{ik} O_i x_{ik} + \sum_{i \in H} \sum_{j \in H} \sum_{k \in H} \alpha C_{ij} f_{ij}^k + \sum_{i \in N} \sum_{k \in H} C_{ki} D_i x_{ik} \tag{13}$$

s.t. (2)–(5), (8)–(12)

In addition to the total transportation costs, the objective function (13) also includes the total cost of building hubs and the fixed costs of building the hub links. The rest of the constraints of the model comply with the incomplete  $p$ -hub median formulation, except the number of hubs and hub links to be located are now determined by the model.

Though not considered in this particular paper, one should note that both the incomplete  $p$ -hub median and the hub location with fixed costs network design models are readily extendible to take care of the capacity restrictions by adapting Constraint (9).

In the following section we present some computational analysis with the minimization of total transportation cost objective.

### 3. Computational analysis with the total transportation cost model

We tested the performance of our models on two data sets previously introduced in the literature. The first is the Civil Aeronautics Board (CAB) data set introduced in O’Kelly (1987) and based on airline passenger interactions between 25 US cities, each of which is a potential hub node. We only considered  $n = 25$  for the CAB data set. The second data set is the Turkish network (Fig. 1). There are 81 demand centers (nodes) in this network and 16 candidate hub locations over these demand centers (Yaman et al., 2007).

For both of these data sets, no real data on the fixed cost of establishing hub links has been introduced in the literature. In order to observe and compare the results for different fixed cost values, we tested three different ways for establishing hub link costs. For the first one we took the fixed costs for all hub links to be the same. In the second cost pattern, the fixed costs are directly proportional to the length of the hub link, and finally in the third one the structure from Calik et al. (2009) where the fixed costs are directly proportional to the length of the hub link and inversely proportional to  $w_{ij}$  are used.

During all of our computational analysis, we utilized the optimization software CPLEX version 11.2. We took our runs on a server with a 2.66 GHz Intel Xeon processor and 8GB of RAM.

We first tested our incomplete hub location with fixed costs network design formulation on the CAB data set with the mentioned three possible fixed cost patterns. However, since flow costs are high when compared to fixed hub link costs (under all scenarios) on the CAB data set, the model resulted in establishing complete hub networks in all of the test instances.

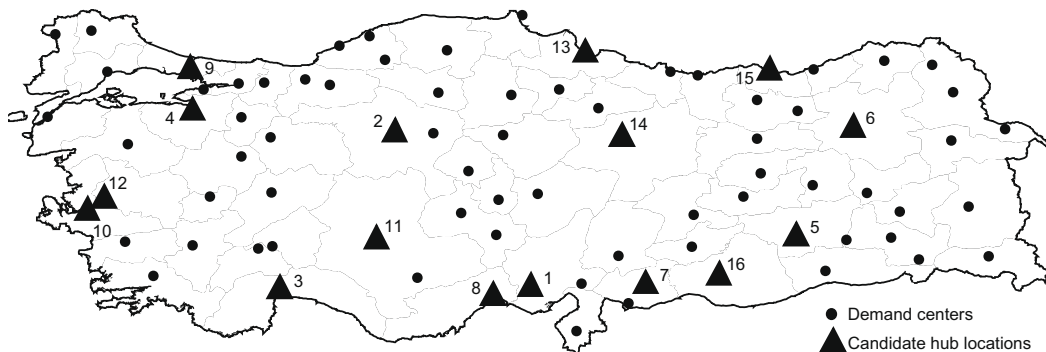


Fig. 1. Demand centers and 16 candidate hub locations on the Turkish network.

Since fixing the number of hubs and hub links to be located converts the fixed cost problem to the incomplete  $p$ -hub median problem, we do not present any computational results for the fixed cost model.

For the CAB data set with  $p$  ranging from 2 to 5 we tested differing  $q$  values for our incomplete  $p$ -hub median network design formulation. As customarily done in the literature, we took  $\alpha$  value to be 0.2, 0.4, 0.6, and 0.8.

We report our results on the CAB data set for the incomplete  $p$ -hub median problem in Table 1. For each instance, Table 1 reports the required CPU time in seconds, the locations of the hub nodes, and presents the percentage of increase in transportation costs with respect to establishing a complete hub network. In order to calculate this percentage of increase, we subtracted the total transportation cost of the complete hub network from the observed transportation cost, divided this value to the transportation cost of the complete hub network and multiplied by 100.

On the average the model is solved within 3 min of CPU time. The minimum CPU time requirement was about 1 s, whereas the maximum was just below 41 min. Observe from Table 1 that the instances with greater values of  $\alpha$  turned out to be harder. Also, the instances in which we forced the hub networks to be sparse turned out to be harder than the instances with almost complete hub networks. The hardest of these 44 instances was when  $\alpha = 0.8$ ,  $p = 5$ , and  $q = 6$  which required about 40.7 min of CPU time.

**Table 1**  
The results on the CAB data set with the incomplete  $p$ -hub median problem.

$\alpha$	$p$	$q$	CPU time (s)	Hub locations	% Increase in transportation costs
0.2	2	1	0.66	12,20	0
0.2	3	2	8.92	4,12,17	0.020
0.2	3	3	3.99	4,12,17	0
0.2	4	4	7.4	4,12,17,24	0.507
0.2	4	5	2.65	4,12,17,24	0.022
0.2	4	6	2.28	4,12,17,24	0
0.2	5	6	7.03	4,7,12,14,17	0.867
0.2	5	7	1.84	4,7,12,14,17	0.327
0.2	5	8	1.78	4,7,12,14,17	0.031
0.2	5	9	1.91	4,7,12,14,17	0.004
0.2	5	10	1.34	4,7,12,14,17	0
Average			3.62		0.162
0.4	2	1	4.35	12,20	0
0.4	3	2	32.54	4,12,18	0.082
0.4	3	3	10.5	4,12,18	0
0.4	4	4	43.98	1,4,12,17	0.866
0.4	4	5	19.23	1,4,12,17	0.036
0.4	4	6	12.99	1,4,12,17	0
0.4	5	6	40.06	4,7,12,14,17	1.209
0.4	5	7	18.51	4,7,12,14,17	0.449
0.4	5	8	8.25	4,7,12,14,17	0.047
0.4	5	9	6.39	4,7,12,14,17	0.007
0.4	5	10	6.01	4,7,12,14,17	0
Average			18.44		0.245
0.6	2	1	6.21	12,20	0
0.6	3	2	87.43	4,12,18	0.177
0.6	3	3	20.43	2,4,12	0
0.6	4	4	137.18	1,4,12,17	1.090
0.6	4	5	61.25	1,4,12,17	0.045
0.6	4	6	27.74	1,4,12,17	0
0.6	5	6	666.43	4,7,12,14,17	1.466
0.6	5	7	233.15	4,7,12,14,17	0.544
0.6	5	8	65.37	4,7,12,14,17	0.057
0.6	5	9	74.87	4,7,12,14,17	0.008
0.6	5	10	42.61	4,7,12,14,17	0
Average			129.33		0.308
0.8	2	1	18.32	12,20	0
0.8	3	2	185.66	2,4,12	0.269
0.8	3	3	85.34	2,4,12	0
0.8	4	4	970.26	1,4,12,18	1.287
0.8	4	5	432.11	1,4,12,18	0.124
0.8	4	6	122.06	1,4,12,18	0
0.8	5	6	2441.97	1,4,11,12,18	1.907
0.8	5	7	574.03	1,4,7,12,18	0.430
0.8	5	8	482.37	1,4,7,12,18	0.165
0.8	5	9	394.34	1,4,7,12,18	0.034
0.8	5	10	269.66	1,4,7,12,18	0
Average			543.28		0.383

Looking at the locations of the hub nodes in Table 1, we observe that Los Angeles (12) is always selected as a hub node and at the instances where we located three or more hub nodes, Chicago (4) is always selected. While locating four or more hubs either New York (17) or Philadelphia (18) is always present in the hub set. This is mainly due to the generation of high flow from these cities. When we rank the flow generated from the nodes in the CAB data set New York (17) has the largest flow, and Chicago (4) and Los Angeles (12) are second and third, respectively.

In general, the hub locations were insensitive to the number of hub links to be established. Except at two instances, the hub locations were exactly the same as the complete hub network solutions. At these two instances ( $\alpha = 0.6$ ,  $p = 3$ ,  $q = 2$  and  $\alpha = 0.8$ ,  $p = 5$ ,  $q = 6$ ) only one hub node differs (Philadelphia (18) and Kansas City (11), respectively) from the complete hub network solution.

With lower values of  $\alpha$ , hub nodes are located near the peripheries of the region. Since in the CAB data set  $c_{ij} = \text{Distance}_{ij}$ , by locating hub nodes near the peripheries the higher cost values get higher advantage from the economies of scale. When  $\alpha$  increases, hub locations are concentrated in the center of the region.

The percentage of increase in transportation costs is reported as 0 for the instances with complete hub networks. As expected, this percentage increases as the hub network is forced to be sparser. The highest increase we obtained at the CAB instances in Table 1 was 1.907%. We also observed from Table 1 that the percentage of increase in the transportation costs are lower when  $\alpha$  value is lower. When we took the average of percent increases for different values of  $\alpha$ , we obtain 0.162% for  $\alpha = 0.2$  and 0.245%, 0.308%, and 0.383% for  $\alpha$  equal to 0.4, 0.6, and 0.8, respectively.

Fig. 2 illustrates a sample of solutions on the CAB data set. In order to analyze the flow behavior of the designed network links (both hub links and spoke links), we explored the flow data with  $\alpha = 0.6$  and  $p = 5$  corresponding to instances a and b of Fig. 2. This is one of the common instances where hub locations and allocation decisions coincide for both sparse ( $q = 6$ ) and complete ( $q = 10$ ) hub network structures. In both of the resulting designs there are 20 spoke links. The complete design contains 10 hub links, whereas the sparse design has only six hub links. For both of the designs, we ranked the resulting flow on the links of the network in a descending order. Under both sparse and complete designs, the largest flow in the network is between hubs Chicago (4) and New York (17). For both of the networks, the spoke links Boston (3) – New York and Washington (25) – New York are in top 5. This is in sync with our expectations to have high flows on some spoke links under single allocation hub network structure. It is interesting to note that in the complete design not all of the hub links carry large flows. In particular, only 4 out of 10 hub links are among the first 10 highest flow carrying links. On the other hand, two of the remaining hub links are the two lowest flow carrying links. In contrast, in the sparse design 5 out of 6 hub links appear in the top 10 list. The remaining hub link Dallas (7) – Los Angeles (12) has the 20th largest flow.

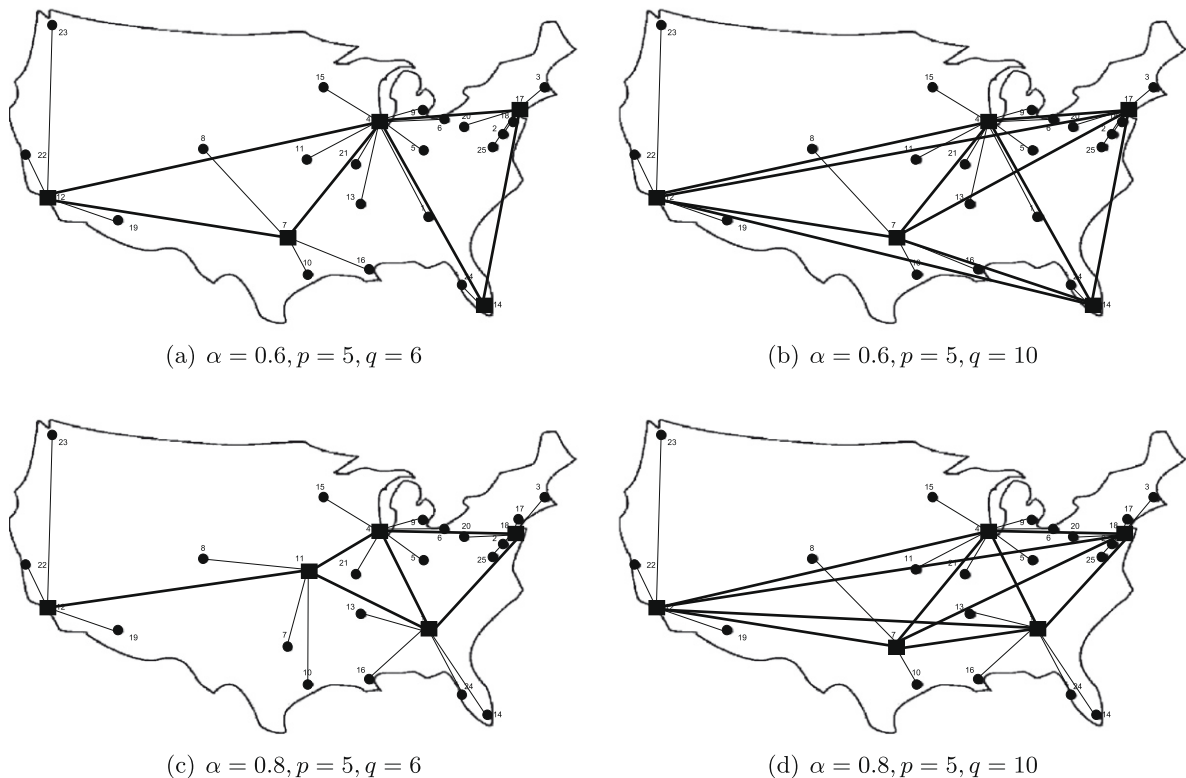


Fig. 2. CAB data set results with the transportation cost objective.

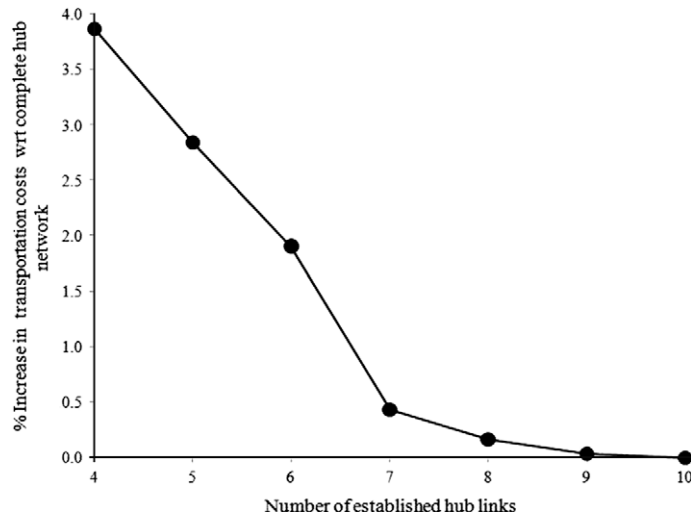


Fig. 3. Trade-off curve with  $\alpha = 0.8$  and  $p = 5$ .

The flow behavior of the above example is typical of many that we encountered during our experimentation. In the absence of a direct hub link between some hubs, the flow between these hubs is to be routed on more than one hub link. Thus, it is expected to have more flow on hub links in incomplete hub networks compared to complete ones. In hub location, it is desirable to have larger flow between hubs in order to account for the economies of scale discount factor more realistically. When compared with complete hub networks, the incomplete hub networks allow the solutions to adapt to use hub links with larger flows. However, in the incomplete hub networks since the flow between hubs is sometimes not transported directly, the total transportation cost may increase.

In order to observe the increase in transportation costs with respect to the number of established hub links, we decided to draw a trade-off curve. For the curve we analyzed the instance with  $\alpha = 0.8$  and  $p = 5$ , which is the hardest of our instances and the instance with the greatest percentage increase. For this, we calculated the percent increase with all possible number of established hub links, starting from a tree-hub network and going on to a complete one. Fig. 3 depicts the resulting trade-off curve.

In Fig. 3, when we forced the model to establish a tree-hub network with four hub links the percent increase in transportation costs was about 3.87%. This value was about 0.03% when we reduced one hub link from the complete hub network ( $q = 9$ ). Observe that, there is a steep increase in the curve below  $q = 7$ .

By analyzing the curve, the decision maker can observe the trade-off between establishing an incomplete hub network versus the increase in transportation costs. The main drawback of building incomplete hub networks is the increase in the total transportation costs. However, we observed that the increase in the total transportation costs with respect to building complete hub networks is not very significant. If the decision maker considers the fixed costs of building hub links, for example for assigning new aircrafts between two hub nodes, this increase in transportation costs can be tolerable.

In a similar fashion, we tested the incomplete  $p$ -hub median network design model with the Turkish network. We again analyzed the increase in transportation costs with respect to complete hub networks. The results are provided in Table 2.

Our findings on the CAB data set placed larger values of  $\alpha$  to be most challenging thus on the Turkish network we tested the two largest possible values for  $\alpha$ . On the Turkish network, for  $\alpha = 0.6$ , and 0.8, and for each  $p = 4, 6, 8$ , and 10, we tested three different  $q$  values, corresponding to sparse, medium, and complete hub networks. We report the CPU time requirements given by CPLEX in the fourth column of Table 2. Considering the fact that this is a decision problem of strategic nature the CPU times are reasonable.

Observe from Table 2 that some instances corresponding to sparse hub networks could not be solved to optimality within 24 h (86,400 s). At these instances, we present the gap reported by CPLEX in parenthesis and calculated the percent increase in transportation cost by using the best integer solution found by CPLEX at the end of 24 h. Thus, the percentages reported for the instances that took longer than 24 h are pessimistic.

Excluding the two instances with  $p = 4$  and  $q = 3$ , the percent increases in transportation costs with respect to complete hub networks, reported in the last column of Table 2, are all below 1%. The average of the percent increases of all the instances is 0.367%. This again shows that when building hub networks, the difference in transportation costs with respect to complete hub networks can be tolerable when the fixed costs of establishing hub links are considered.

#### 4. Model formulation for the incomplete hub covering network design problem

The aim of the single allocation incomplete hub covering network design problem is to find the location of hub nodes, the allocation of non-hub nodes to these hub nodes, and which hub links to establish between the hub nodes, while providing

**Table 2**  
Incomplete  $p$ -hub median results on the Turkish network.

$\alpha$	$p$	$q$	CPU time (s)	% Increase in transportation costs
0.6	4	3	3555.36	2.046
0.6	4	4	729.32	0.024
0.6	4	6	377.40	0
0.6	6	8	86,400 (0.33%)	0.743
0.6	6	10	541.37	0.037
0.6	6	15	302.62	0
0.6	8	15	8120.42	0.289
0.6	8	20	2139.91	0.027
0.6	8	28	564.68	0
0.6	10	20	86,400 (0.39%)	0.773
0.6	10	30	4038.39	0.058
0.6	10	45	503.44	0
0.8	4	3	6864.23	2.813
0.8	4	4	599.14	0.038
0.8	4	6	495.05	0
0.8	6	8	86,400 (0.44%)	0.797
0.8	6	10	1685.50	0.033
0.8	6	15	1137.39	0
0.8	8	15	15541.44	0.309
0.8	8	20	1611.45	0.027
0.8	8	28	697.62	0
0.8	10	20	86,400 (0.79%)	0.746
0.8	10	30	62260.15	0.060
0.8	10	45	4544.15	0
Average			5815.45	0.367

service between every origin–destination pair within the given time bound and while minimizing the fixed costs of building the hub network.

The incomplete hub covering network design problem is important for the systems where time is a major concern. As Campbell (1994) pointed out, hub covering and  $p$ -hub center type problems are important for hub systems involving perishable or time sensitive items. A major application area for hub covering problems is the cargo sector, where there is a promised service time between origin–destination pairs. In the hub covering and  $p$ -hub center type problems, the economies of scale discount factor  $\alpha$  is applied to time instead of cost. The time discount factor  $\alpha^t$  is again a number between 0 and 1, however it is most likely to be higher than the cost discount factor and it is expected to be a number closer to one. Travel time is discounted when the flow is consolidated to be transported with bigger and faster vehicles. For example, for a cargo application in Turkey, Tan and Kara (2007) found out that there is a time discount factor of 0.9 when using inter-hub connections.

For the modeling of the incomplete hub covering network design problem, we no longer require any parameters and variables associated with flow as previously presented for the incomplete  $p$ -hub median and fixed cost models. However, we need to know which hub links are used on the path from any origin to destination to calculate the travel time. Similar to the total cost models, it is possible to model the problem with  $O(n^4)$  variables. For example, we could use  $y_{ij}^{kl}$  variables and set them to one if a hub link  $\{i, j\}$  is used on the path from node  $k$  to node  $l$ . In this paper, we developed a novel formulation with  $O(n^3)$  binary variables. The idea behind our model is that, for each established hub, we would like to find a spanning tree rooted at this hub that visits every other hub in the hub network using only the established hub links. Employing these spanning trees ensures the connectivity of the underlying hub network. We calculate the travel time between all pairs of hubs, using these spanning trees. We then ensure that the travel time between all origin–destination pairs are within the given time bound. We assume throughout our model that the triangle inequality is satisfied.

The additional parameters needed for the mathematical model are as follows:

- $t_{ij}$  = travel time between nodes  $i \in N$  and  $j \in N$
- $\beta$  = maximum service time requirement
- $\alpha^t$  = hub-to-hub transportation time discount factor

In addition to the previously defined decision variables  $x$  and  $z$ , the new decision variables of the mathematical model are:

$$y_{ijk} = \begin{cases} 1 & \text{if the spanning tree rooted at hub } k \in H \text{ uses the hub link } i, j \text{ from hub } i \in H \text{ to hub } j \in H; \\ 0 & \text{otherwise.} \end{cases}$$

$r_j$  = radius of hub  $j \in H$ , i.e., the maximum travel time between hub  $j$  and the nodes that are allocated to hub  $j$ .

$d_{ij}$  = travel time from hub  $i \in H$  to hub  $j \in H$  in the hub network (not discounted).

An integer programming formulation of the single allocation incomplete hub covering network design problem defined above is as follows:



$$\min \sum_{i \in H} \sum_{j \in H: j > i} FL_{ij}Z_{ij} + \sum_{k \in H} FH_k X_{kk} \tag{14}$$

s.t. (2)–(5), (11), (12)

$$\sum_{i \in H: i \neq j} y_{ijk} \geq x_{kk} + x_{ij} - 1 \quad \forall j, k \in H: j \neq k \tag{15}$$

$$\sum_{i \in H: i \neq j} y_{ijk} \leq x_{kk} \quad \forall j, k \in H: j \neq k \tag{16}$$

$$y_{ijk} + y_{jik} \leq z_{ij} \quad \forall i, j, k \in H: i < j \tag{17}$$

$$d_{kj} \geq (d_{ki} + t_{ij})y_{ijk} \quad \forall i, j, k \in H: i \neq j \text{ and } j \neq k \tag{18}$$

$$d_{ij} = d_{ji} \quad \forall i, j \in H: i \neq j \tag{19}$$

$$d_{kk} = 0 \quad \forall k \in H \tag{20}$$

$$r_j \geq t_{ij}x_{ij} \quad \forall i \in N, j \in H \tag{21}$$

$$r_k + \alpha^t d_{kj} + r_j \leq \beta \quad \forall j, k \in H \tag{22}$$

$$y_{ijk} \in \{0, 1\} \quad \forall i, j, k \in H: i \neq j \text{ and } j \neq k \tag{23}$$

$$d_{ij} \geq 0 \quad \forall i, j \in H \tag{24}$$

The first term in the objective function (14) calculates the total cost of establishing hub links and the second term, the total cost of establishing hubs. In the objective function, we minimize the overall cost of establishing an incomplete hub network.

For the case when there will be more than one hub established in the network, Constraint (15) ensures that the degree for each hub node is at least one, so that every hub node is an end node for at least one hub link. Through this constraint, the model guarantees that the tree rooted at hub  $k$  will have an entering arc into every other hub  $j$ . Constraint (16) assures that each spanning tree rooted at hub  $k$  can have at most one entering arc into another hub node  $j$  and forces the spanning tree arcs associated with a non-hub node to take zero values. Constraint (17) causes the spanning tree arcs to be hub arcs.

Constraint (18) calculates the time needed to travel from one hub node to another using the established spanning tree arcs in the hub network. Note that the spanning tree formed for each hub node in the network does not need to be the minimum spanning tree; thus, the time calculated in Constraint (18) does not need to be the minimum traveling time. We linearized Constraint (18) with a BigM type linearization as follows:

$$d_{kj} \geq d_{ki} + t_{ij}y_{ijk} - \frac{\beta}{\alpha^t}(1 - y_{ijk}) \quad \forall i, j, k \in H: i \neq j \text{ and } j \neq k \tag{18^*}$$

For given  $i, j$  and  $k$ , when  $y_{ijk} = 1$  both of the Constraints (18) and (18\*) yield the same right-hand side. Note that the discounted maximum travel time between two hubs in the network cannot be greater than  $\beta$ . Thus, the travel time between two hub nodes in the network can be at most  $\beta/\alpha^t$ . So when  $y_{ijk} = 0$ , the right-hand side of Constraint (18) yields 0, whereas the right-hand side of Constraint (18\*) yields a number less than or equal to zero. Because  $d_{kj} \geq 0$  by Constraint (24), we conclude that Constraint (18\*) correctly linearizes Constraint (18).

Constraints (18) or (18\*) also act as subtour breaking constraints for the possible values that  $y$  variables can take. Note that the previous constraints only guarantee that each hub node has an outgoing arc as well as an incoming arc to every other hub node. We would not necessarily end up with a rooted spanning tree, were it not for these constraints.

We assume symmetric time data; thus,  $d$  variables will also be symmetric (Constraint (19)), and the distance from a node to itself will be zero (Constraint (20)).

The maximum travel time needed from a demand node to its assigned hub, the radius of a hub, is calculated in Constraint (21) for each hub. Constraint (22) ensures that the travel time between any two nodes in the network is less than the given time limit. Because  $d$  variables calculate the travel time between hubs, we need to discount this travel time by the economies of scale time discount factor  $\alpha^t$ .

Our linear integer programming formulation of the single allocation incomplete hub covering network design problem consists of the objective function (14) and the Constraints (2)–(5), (11), (12), (15)–(17), (18\*), (19)–(24). If  $h = n$ , then the model has  $(n^3 + \frac{3}{2}n^2)$  binary variables and  $(n^2 + n)$  real variables. The number of constraints of our model is  $(2n^3 + 7n^2 + 2n)$ . Hence, in total we have  $O(n^3)$  variables and constraints.

### 5. Incorporating valid inequalities

In this section, we present some inequalities that are valid for our incomplete hub covering network design model. The aim in providing these inequalities is to reduce the time needed to solve our model to optimality.

The first valid inequality that we introduce is constraint (A):

$$d_{ij} \geq t_{ij} \quad \forall i, j \in H: i \neq j \tag{A}$$

Constraint (A) states that the  $d$  variable associated with two candidate hub nodes  $i$  and  $j$  in the hub network is at least the direct travel time between these nodes.

For the second valid inequality, we need to define a second travel time parameter  $\bar{t}_{ij}$  calculated for all pairs of potential hub nodes as  $\bar{t}_{ij} = \min_{k \in H, k \neq i, j} (t_{ik} + t_{kj}) \forall i, j \in H : i \neq j$ . The second travel time parameter between nodes  $i$  and  $j$  is the minimum travel time from node  $i$  to node  $j$  using exactly one node  $k$  in between. With the definition of the second travel time parameter, we present our second valid inequality:

$$d_{ij} \geq \bar{t}_{ij}(1 - z_{ij}) \forall i, j \in H : i < j \quad (B)$$

By constraint (B), if there is not a direct hub link established between two candidate hub nodes  $i$  and  $j$ , then the  $d$  variable associated with them in the hub network is at least the second travel time between these nodes.

Our third valid inequality is based on the  $z$  variables:

$$\sum_{i \in H} \sum_{j \in H, j > i} z_{ij} \geq \sum_{k \in H} x_{kk} - 1 \quad (C)$$

Note that we are building a connected hub network in this problem, so that every node in the network can send flow to any other node. The minimum number of edges in a connected network is (number of nodes in the network - 1). By using this fact, constraint (C) states that the number of hub links to be established is at least the total number of hub nodes to be established, less one.

Our fourth valid inequality links  $d$  and  $z$  variables:

$$d_{ij} \leq t_{ij}z_{ij} + \frac{\beta}{\alpha^t}(1 - z_{ij}) \forall i, j \in H : i < j \quad (D)$$

By constraint (D) we ensure that, for a given node pair  $i$  and  $j$ , if there is a direct hub link established between these nodes, the  $d$  variable associated with them is at most the travel time between them. Otherwise, because the discounted maximum travel time between any two hubs in the network cannot be greater than  $\beta$ , we do not put any limit on the  $d$  variable.

We tested the performance of these valid inequalities (A)–(D) both individually and collectively. For this, we used both the CAB and the Turkish network data sets. Our preliminary analysis with the valid inequalities showed that the valid inequality (A) is very effective. Even though using valid inequality (A) alone did not always result in the quickest solution times, the average CPU time turned out to be the lowest. On the other hand, with respect to the average CPU times, use of valid inequalities (A) and (D) jointly has the second smallest CPU time requirement. However, neither using valid inequality (D) alone nor the addition of valid inequality (D) to other inequalities is as effective as the valid inequality (A).

With valid inequalities, we also compared the linear programming (LP) relaxations of the model and the LP relaxations at the root node reported by CPLEX. It turns out that none of the valid inequalities resulted in different LP relaxation values. However, the use of valid inequality (C) slightly increases (less than 8.5%) the LP relaxation at the root node reported by CPLEX.

In light of our observations, we decided to use our incomplete hub covering model together with valid inequality (A) alone for the rest of our computational analysis. The reader should note that different data sets may also make the other valid inequalities effective.

## 6. Computational analysis with the incomplete hub covering model

We tested the performance of our incomplete hub covering network design formulation on the CAB and the Turkish network data sets. No time data has been provided for the CAB data set, thus, similar to other hub covering studies in the literature, we took  $t_{ij} = \text{Distance}_{ij}$  for this data set. For the fixed costs of opening hubs we took  $FH_i = 100$  for all nodes (O’Kelly, 1992). We varied  $\alpha^t$  from 0.2 to 1.

On the CAB data set, we tested three different fixed cost values for opening hub links. In the first one we let  $FL_{ij} = 10$  for all  $i, j$ . For the second one we took  $FL_{ij} = \text{Distance}_{ij}$  and for the last one we used the values from Calik et al. (2009) in which they calculated the fixed costs for hub links as follows:

$$FL_{ij} = \frac{\frac{\text{Distance}_{ij}}{\text{Flow}_{ij}}}{\max_{i, j} \frac{\text{Distance}_{ij}}{\text{Flow}_{ij}}} \times 100 \forall i, j \in H, i \neq j$$

where  $\text{Distance}_{ij}$  is the distance between nodes  $i$  and  $j$ , and  $\text{Flow}_{ij}$  is the amount of flow between nodes  $i$  and  $j$ .

In what follows, we are going to exhibit the performance of our incomplete hub covering network design formulation with valid inequality (A) using CPLEX 11.2. In order to obtain the tightest possible  $\beta$  values on the CAB data set, we first solved  $p$ -hub center problems with complete hub networks (Ernst et al., 2009) with 2, 3, 4, and 5 hubs and with different possible  $\alpha^t$  values. Then we tested our model with three different fixed link cost values on these instances. The test parameters (shown in bold) and the corresponding results are provided in Table 3.

When we look at the CPU times obtained in Table 3, we observed that CPU times are in fact dependent on the test data. The average CPU times that we obtained on the CAB instances in Table 3 with three different fixed link cost values are 2.7 s, 24.2 min, and 3.0 s, respectively. In general, all of the instances are solved in reasonable CPU times. Because we used the tightest  $\beta$  values in all of these instances, it is reasonable to presume that these are among the hardest instances on this data set.

**Table 3**  
Hub covering results on the CAB data set.

$\alpha^t$	$\beta$	$FL_{ij} = 10$			$FL_{ij} = \text{Distance}_{ij}$			$FL_{ij}$ from Calik et al.		
		CPU time (s)	Hub locations	Number of hub links	CPU time (s)	Hub locations	Number of hub links	CPU time (s)	Hub locations	Number of hub links
0.2	2136	0.86	21,22	1	6.93	5,8,13	2	1.42	21,22	1
0.2	1912.8	2.69	3,13,22	2	171.41	5,13,22	2	2.4	13,17,22	2
0.2	1616.2	4.86	9,16,19,23	4	1721.51	1,5,8,22,23	4	1.86	9,16,19,23	4
0.2	1346	3.42	2,11,12,23,24	4	3609.98	1,11,13,19,20,22,23	6	6.53	11,12,17,23,24	4
0.4	2400.4	4.64	5,8	1	52.46	8,21	1	3.51	8,21	1
0.4	2098.2	2.66	1,2,8	2	282.16	1,8,18,20	3	2.17	1,8,25	2
0.4	1880.4	2.17	3,12,13,23	3	2365.12	1,7,22,23,25	4	3.24	12,13,17,23	3
0.4	1596.8	3.75	11,12,14,18,23	4	14149.08	11,12,13,17,21,23,24	6	6.48	11,12,14,18,23	4
0.6	2556.6	1.33	8,21	1	1.79	8,21	1	1.37	8,21	1
0.6	2335.2	4.26	8,16,25	3	278.69	8,20,21,24	3	4.32	8,16,25	3
0.6	2183.2	5.57	19,21,22,23	3	380.36	3,6,8,11,24	4	2.62	19,21,22,23	3
0.6	2002	4.47	13,17,19,22,23	6	3304.79	8,12,13,22,23,25	5	12.54	8,12,13,17,22,23	5
0.8	2712.8	0.83	8,21	1	3.00	8,21	1	0.93	8,21	1
0.8	2551.6	2.48	6,8,16	3	186.50	8,11,22,23	3	2.82	6,8,16	3
0.8	2456.8	4.38	19,21,22,23	5	474.77	8,9,11,13,22,23	5	2.35	19,21,22,23	5
0.8	2370.6	3.18	6,8,16,22,23	5	1709.71	8,17,21,22,23,24	5	2.19	6,8,16,22,23	5
1	2826	0.29	8,11	1	1.10	8,11	1	0.63	8,11	1
1	2762	0.61	8,11,23	3	13.43	8,11,23	3	0.55	8,11,23	3
1	2726	0.62	4,8,23,24	6	162.17	4,8,13,23,24	7	0.83	4,8,23,24	6
1	2725	1.63	7,8,9,14,23	7	176.32	4,8,13,14,23	7	1.07	4,7,8,14,23	8
Average		2.74			1452.56			2.99		

Among the three fixed cost values, the instances with  $FL_{ij} = \text{Distance}_{ij}$  turned out to be the hardest. In general, there is not a significant CPU time difference between the instances with the first and third fixed link cost values. The reason why the  $FL_{ij} = \text{Distance}_{ij}$  instances lasted longer is because  $FL_{ij}$  values are higher when compared to fixed hub costs.

Table 3 also provides the location of the hub nodes and the number of hub links in the established hub networks. Except for a few instances, the results with the second fixed link cost structure turned out to be very different than the other two sets of results. In these instances with  $FL_{ij} = \text{Distance}_{ij}$ , the model resulted in establishing more hubs to reduce the total fixed costs of establishing hub links. For example, at the first instance in Table 3 when  $\alpha^t$  is 0.2 and  $\beta$  value is 2136, both the first and the third fixed link cost values resulted in opening two hubs at nodes St. Louis (21) and San Francisco (22) with one hub link {21, 22}, whereas the second one resulted in opening three hubs at nodes Cincinnati (5), Denver (8), and Memphis (13) with two hub links {5, 13} and {8, 13}. This is because in the CAB data set ( $\text{Distance}_{21,22} = 1736$ ) > ( $\text{Distance}_{5,13} = 402$ ) + ( $\text{Distance}_{8,13} = 880$ ), and thus the difference  $FL_{21,22} - (FL_{5,13} + FL_{8,13}) > FH = 100$ . Hence the objective function value of the model with the second fixed hub link cost value is higher with opening two hubs St. Louis (21) and San Francisco (22), compared to opening hubs Cincinnati (5), Denver (8), and Memphis (13).

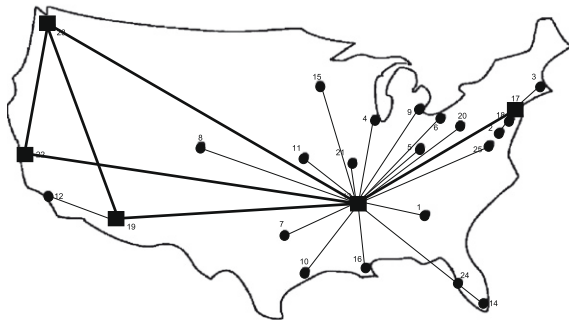
The number of established hub links reported in Table 3 indicates how sparse the hub network is. Note that 14 of the 60 instances listed in Table 3 correspond to locating two hubs, where an incomplete hub network solution is not possible. Even though we used the tightest possible  $\beta$  values, we obtained incomplete hub networks in 37 of the remaining 46 instances. Especially for the instances with  $FL_{ij} = \text{Distance}_{ij}$  all of the solutions with three or more hubs, with one exception when  $\alpha^t$  is 1 and  $\beta$  value is 2762, resulted in incomplete hub networks. For the other two fixed cost structures, the respective number of established hub links are usually the same though sometimes with the establishment of different hub nodes. In all of the instances, where the model resulted in opening four or more hubs, all of the hub networks were incomplete with one exception when  $\alpha^t = 1$  and  $\beta = 2726$ . Fig. 4 depicts incomplete hub network solutions of some instances from Table 3 on the CAB data set.

Because we are using the tightest possible time bounds on this network, in order to discount the travel time, the nodes that are located near the periphery are often selected as hub nodes (such as San Francisco (22) and Seattle (23)) especially when  $\alpha^t$  increases. In almost all of the solutions presented in Fig. 4, at least one central hub node is established, and most of the non-hub nodes are allocated to this central hub node.

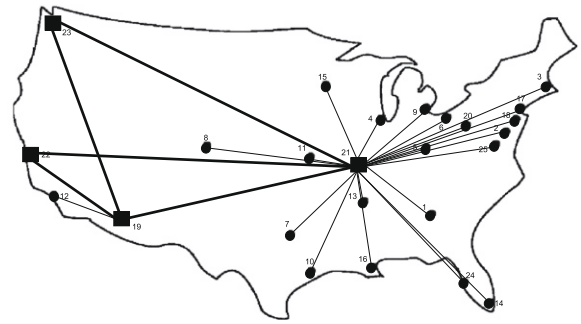
In order to observe the performance of our model on larger networks, we tested it on the Turkish network (Fig. 1). For the Turkish network data set, we took the fixed cost values for opening hubs from Tan and Kara (2007), for opening hub links from Beasley (1990), and the  $\alpha^t$  value to be 0.9 (Tan and Kara, 2007). On this network with 16 candidate hub locations, the tightest possible  $\beta$  value corresponding to  $\alpha^t = 0.9$  is 1783.1 min. In addition to the tightest possible  $\beta$ , we tested  $\beta$  values between 1800 and 2100 min at 30 min intervals. The summary of our results on the Turkish network is provided in Table 4.

All instances with the Turkish network in Table 4 are solved within 1 s due to having 16 potential hub locations. All of the solutions with four or more hubs resulted in incomplete hub networks. We illustrated two of these results in Fig. 5.

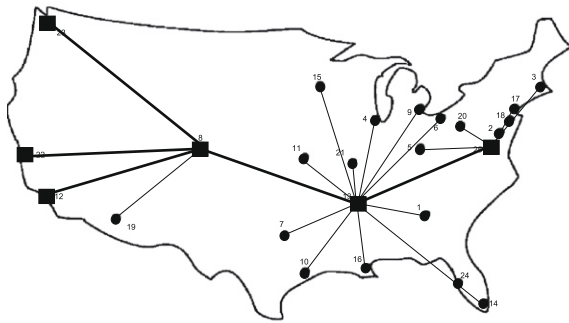
Similar to the observations from the CAB data set, we conclude from the Turkish network solutions that building complete hub networks is neither necessary nor cost efficient when building high number of hubs (for example more than three hubs) for providing service within a given time bound.



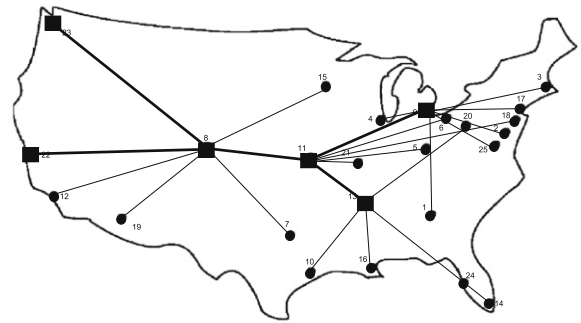
(a)  $\alpha^t = 0.6, \beta = 2002, FL_{ij} = 10$



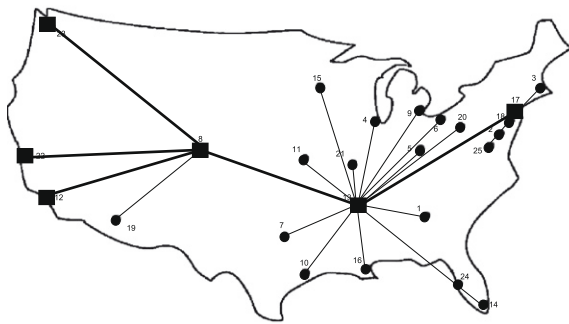
(b)  $\alpha^t = 0.8, \beta = 2456.8, FL_{ij} = 10$



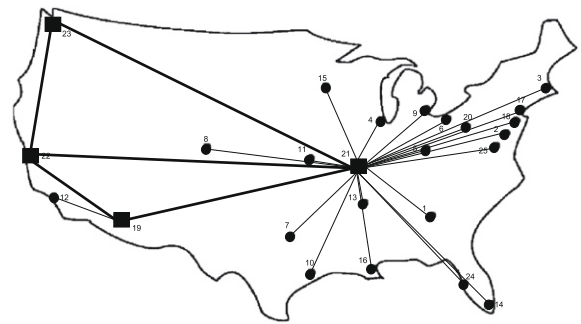
(c)  $\alpha^t = 0.6, \beta = 2002, FL_{ij} = Distance_{ij}$



(d)  $\alpha^t = 0.8, \beta = 2456.8, FL_{ij} = Distance_{ij}$



(e)  $\alpha^t = 0.6, \beta = 2002, FL_{ij}$  from Calik et al.



(f)  $\alpha^t = 0.8, \beta = 2456.8, FL_{ij}$  from Calik et al.

Fig. 4. Hub covering results with the CAB data set.

Table 4  
Hub covering results on the Turkish network.

$\alpha^t$	$\beta$	CPU time (s)	Hub locations	Number of hub links
0.9	1783.1	0.31	5,9,10,11,14	8
	1800	0.43	5,9,10,11,14	8
	1830	0.53	5,9,11,14	5
	1860	0.41	2,5,6	3
	1890	0.48	2,5,15	3
	1920	0.46	2,3,14	3
	1950	0.39	2,14	1
	1980	0.46	2,14	1
	2100	0.15	13	0
	Average		0.40	



Fig. 5. Hub covering results with the Turkish network.

**Table 5**  
Performance of the hub covering model with CPLEX on large networks.

$n$	$\alpha^t$	$\beta$	$h$	CPU time (s)
81	0.9	1800	16	0.43
			25	1.86
			35	4.17
			45	12.27
			55	24.63
			65	83.90
			75	108.47
			81	151.41

Lastly, in order to challenge the solution potential of our model with CPLEX, we fixed  $\beta$  and increased the number of candidate hub locations on the Turkish network. We report our results in Table 5. As expected, the CPU time requirement grows exponentially with the increase on the candidate number of hub nodes. With our proposed model, we were able to solve even the largest instance with 81 candidate hub locations within 2.53 min. These are the largest instances of incomplete hub network solutions solved to optimality in the literature up to now.

From our computational analysis on both the CAB data set with 25 nodes and the Turkish network with 81 nodes, we obtained optimal solutions of our model with CPLEX in few minutes, proving the effectiveness of our proposed mathematical model.

### 7. The incomplete $p$ -hub center network design problem

The  $p$ -hub center problem locates  $p$  hubs, such that the maximum travel time between any origin–destination pair is minimized. On the other hand, we define the incomplete  $p$ -hub center network design problem as one that additionally determines which  $q$  hub links to establish in the  $p$ -hub center problem. Similar to the hub covering version, it is possible to model the single allocation incomplete  $p$ -hub center network design problem with  $O(n^3)$  decision variables and constraints. The number of hubs and hub links to be located are now given, and the service time parameter introduced in the hub covering formulation is to be treated as a decision variable. The new decision variable, also named as  $\beta$ , is defined as the maximum travel time between any origin–destination pair.

More formally, we define the single allocation  $p$ -hub center network design problem as locating  $p$  hubs, allocating each non-hub node to a single hub, and determining which  $q$  hub links to establish between hubs such that the maximum travel time between any origin–destination pair is minimized. With the previously defined decision variables and parameters, the problem can be modeled as:

$$\min \beta \tag{25}$$

$$\text{s.t. } (2)–(7), (11), (12), (15)–(24),$$

$$\beta \geq 0 \tag{26}$$

For the linearization of Constraint (18), because  $\beta$  is now a decision variable, we suggest using BigM instead of using  $\frac{\beta}{\alpha^t}$  in Constraint (18\*).

Note that all valid inequalities introduced for the incomplete hub covering network design problem are also valid for the  $p$ -hub center version. However, for the valid inequality (D), we again suggest using BigM instead of using  $\frac{\beta}{\alpha^t}$ .

We again performed computational analysis with our single allocation incomplete  $p$ -hub center network design formulation with CPLEX 11.2 on the CAB data set. We included valid inequality (A) to our formulation and varied  $\alpha^t, p$  and  $q$ . We compile our results in Table 6.

We report the optimum  $\beta$  value, the CPU time requirement, and the hub locations in Table 6. Note that for a given  $\alpha^t$  and  $p$ , even though we increase the number of hub links to be established, after a while the optimum  $\beta$  value stays constant. These  $\beta$  values are the tightest possible  $\beta$  values corresponding to a given  $\alpha^t$  and  $p$ . Also note that these are the  $\beta$  values that

**Table 6**  
Incomplete  $p$ -hub center results on the CAB data set.

$\alpha^t$	$p$	$q$	Optimum $\beta$ value	CPU time (s)	Hub locations
0.2	2	1	2136	2.25	21,22
0.2	3	2	1912.8	3.86	13,18,22
0.2	3	3	1912.8	4.45	5,13,22
0.2	4	3	1648.4	13.94	11,20,22,24
0.2	4	4	1616.2	9.09	9,16,19,23
0.2	4	5	1616.2	8.82	9,16,19,23
0.2	5	4	1346	77.74	11,12,23,24,25
0.2	5	5	1346	97.17	11,12,23,24,25
0.4	2	1	2400.4	2.46	8,21
0.4	3	2	2098.2	9.44	1,8,20
0.4	3	3	2098.2	5.53	1,8,20
0.4	4	3	1880.4	97.85	2,12,13,23
0.4	4	4	1880.4	13.75	12,13,20,23
0.4	5	4	1596.8	126.01	11,12,18,23,24
0.4	5	5	1596.8	133.87	11,12,14,18,23
0.6	2	1	2556.6	2.69	8,21
0.6	3	2	2374.6	7.72	12,21,23
0.6	3	3	2335.2	5.57	8,16,20
0.6	4	3	2183.2	91.27	19,21,22,23
0.6	4	4	2183.2	24.15	19,21,22,23
0.6	5	4	2072.8	234.95	11,12,18,23,24
0.6	5	5	2032	138.95	13,18,19,22,23
0.6	5	6	2002	111.83	13,19,22,23,25
0.6	5	7	2002	93.16	13,19,22,23,25
0.8	2	1	2712.8	2.26	8,21
0.8	3	2	2648.8	217.82	8,21,23
0.8	3	3	2551.6	4.15	6,8,16
0.8	4	3	2508	68.24	11,19,22,23
0.8	4	4	2487.6	27.56	6,8,16,23
0.8	4	5	2456.8	26.60	19,21,22,23
0.8	4	6	2456.8	17.01	19,21,22,23
0.8	5	4	2456.8	319.30	11,19,21,22,23
0.8	5	5	2370.6	137.04	6,8,16,22,23
0.8	5	6	2370.6	105.10	6,8,16,22,23
1	2	1	2826	1.64	8,11
1	3	2	2826	7.72	8,11,24
1	3	3	2762	2.72	8,11,23
1	4	3	2826	335.51	8,11,18,25
1	4	4	2762	67.16	8,11,21,23
1	4	5	2739	11.56	4,8,23,24
1	4	6	2726	6.63	4,8,23,24
1	5	4	2826	9844.53	6,8,11,13,23
1	5	5	2762	1241.64	3,4,8,11,23
1	5	6	2739	1614.08	4,8,23,24,25
1	5	7	2725	103.75	4,8,14,23,24
1	5	8	2725	19.90	8,9,14,21,23
Average				336.92	

we used in our hub covering instances. We increased  $q$  in the test instances until we reach the tightest  $\beta$  value, i.e. until the optimum  $\beta$  value stays constant. Observe from Table 6 that, excluding the five instances locating two hub nodes, the model did not need to establish complete hub networks in 11 out of 15 instances (with five different  $\alpha^t$  and three different  $p$  values) in order to obtain the tightest possible  $\beta$  values.

When we compare Tables 3 and 6 we observe similarities. For example, in Table 6,  $\alpha^t = 0.8$ ,  $p = 4$  and  $q = 5$  yielded opening hub nodes Phoenix (19), St. Louis (21), San Francisco (22), and Seattle (23) with an optimum  $\beta$  value of 2456.8, and in Table 3 same hubs are established with  $\alpha^t = 0.8$  and  $\beta = 2456.8$  with both the first and third fixed link cost patterns. In almost all of the instances that required complete hub networks in the hub covering solutions, the tightest  $\beta$  values are again obtained on complete hub networks in the  $p$ -hub center solutions. On the other hand, there are some  $p$ -hub center solutions in which we obtained the tightest  $\beta$  values, i.e. the values that we used in the hub covering version, that resulted in establishing different hub nodes. For example, at the instance with  $\alpha^t = 0.6$ ,  $p = 5$  and  $q = 6$  the  $p$ -hub center model established hubs at nodes Memphis (13), Phoenix (19), San Francisco (22), Seattle (23) and Washington (25); however, we did not obtain the same solution with the hub covering formulation when  $\alpha^t = 0.6$  and  $\beta = 2002$ . This provides an example of having multiple optimal solutions with our incomplete  $p$ -hub center network design problem. In sync with the complete or location analogous versions we observed multiple optimal solutions with this model, but only reported the first optimal solution found by CPLEX in our computational analysis.

**Table 7**  
Incomplete  $p$ -hub center results on the Turkish network.

$\alpha^t$	$p$	$q$	Optimum $\beta$ value	CPU time (sec)	Hub locations
0.9	2	1	1923.7	2.96	2,14
0.9	3	2	1878.8	20.11	2,5,14
0.9	3	3	1856.3	12.91	2,5,6
0.9	4	3	1878.8	103.93	2,4,5,14
0.9	4	4	1856.3	33.00	2,5,12,14
0.9	4	5	1821.5	17.95	5,9,11,14
0.9	4	6	1821.5	21.96	4,5,11,14
0.9	5	4	1878.8	1120.34	2,5,6,8,14
0.9	5	5	1849.8	240.05	2,5,11,12,14
0.9	5	6	1805.6	169.07	5,9,11,12,14
0.9	5	7	1783.1	32.35	5,6,9,11,14
0.9	5	8	1783.1	28.22	5,6,9,11,14
0.9	6	5	1878.8	5767.56	2,5,6,14,15,16
0.9	6	6	1849.6	1488.71	2,5,9,10,11,14
0.9	6	7	1805.6	304.90	3,5,6,9,11,14
0.9	6	8	1783.1	123.29	5,6,9,11,13,14
0.9	6	9	1783.1	131.48	2,5,6,9,11,13
Average				565.81	

When we look at the CPU times on the CAB data set reported by CPLEX, the hardest instance in Table 6 lasted about 2.7 h, and the average CPU time was approximately 5.5 min. Similar to the total cost model, the instances in which we forced the model to obtain sparse hub networks turned out to be harder.

We also tested the incomplete  $p$ -hub center network design problem on the Turkish network. We provide the results in Table 7. The tightest possible  $\beta$  value in the Turkish network with 16 candidate hub locations and  $\alpha^t = 0.9$  is 1783.1, i.e. when we establish a complete hub network with 16 hubs. This  $\beta$  value can be obtained with locating five hubs and establishing seven hub links or six hubs and eight hub links. Observe from Table 7 that, in all of the instances with locating four or more hubs we obtained the tightest possible  $\beta$  values on incomplete hub networks.

The CPU time requirement with the Turkish network was higher than the CAB instances. Nonetheless, the running times reported by CPLEX are still reasonable. On the average the model is solved within 10 min, where the hardest instance lasted about 1.6 h.

## 8. Conclusion

In this paper, we relaxed the fundamental assumption in the hub location literature of building complete hub networks and studied the single allocation incomplete hub network design problems. The problems were motivated from real-life observations of many hub networks. We first defined the single allocation incomplete  $p$ -hub median and the hub location with fixed costs network design problems and presented very efficient  $O(n^3)$  mathematical formulations. The models are readily extendible to handle capacity restrictions. Through computational analysis, we were able to solve all CAB instances within 45 min of CPU time. In order to observe the performance of the model on larger networks, we presented some computational results on the Turkish network. These models are among the first single allocation incomplete hub network design models.

We also defined the single allocation incomplete hub covering and  $p$ -hub center network design problems and proposed  $O(n^3)$  mathematical formulations. To increase the exact solution potential, we proposed and compared four different valid inequalities for the hub covering model. Computational analyses with these formulations on the CAB data set and the Turkish network are also presented. All test instances were very efficiently solved with our models.

The aim of this paper is to provide a thorough treatment of the existing hub location problems under the incomplete hub network structure. In this study, we were able to show that, in some instances, the service that is provided with a complete hub network can also be provided with an efficiently designed incomplete hub network. Thus, the contribution to the quality of service with complete hub network designs may not be high enough to justify their high costs.

## Acknowledgements

The authors express their gratitude to the anonymous Referee for his/her valuable comments especially for gearing the paper towards the transportation cost models.

## References

- Alumur, S., Kara, B.Y., 2008a. Network hub location problems: the state of the art. *European Journal of Operational Research* 190 (1), 1–21.
- Alumur, S., Kara, B.Y., 2008b. A hub covering network design problem for cargo applications in Turkey. *Journal of the Operational Research Society*, doi:10.1057/jors.2008.92

- Beasley, J.E., 1990. OR-library: hub location. <<http://people.brunel.ac.uk/mastjib/jeb/orlib/phubinfo.html>> (accessed 28.01.09.). Site last updated February 2008.
- Boland, N., Krishnamoorthy, M., Ernst, A.T., Ebery, J., 2004. Preprocessing and cutting for multiple allocation hub location problems. *European Journal of Operational Research* 155 (3), 638–653.
- Calik, H., Alumur, S.A., Kara, B.Y., Karasan, O.E., 2009. A tabu-search based heuristic for the hub covering problem over incomplete hub networks. *Computers and Operations Research*. doi:10.1016/j.cor.2008.11.023.
- Campbell, J.F., 1994. Integer programming formulations of discrete hub location problems. *European Journal of Operational Research* 72 (2), 387–405.
- Campbell, J.F., 1996. Hub location and the p-hub median problem. *Operations Research* 44 (6), 923–935.
- Campbell, J.F., Ernst, A.T., Krishnamoorthy, M., 2002. Hub location problems. In: Drezner, Z., Hamacher, H. (Eds.), *Facility Location: Applications and Theory*. Springer-Verlag, Berlin.
- Campbell, J.F., Ernst, A.T., Krishnamoorthy, M., 2005a. Hub arc location problems: Part I—Introduction and Results. *Management Science* 51 (10), 1540–1555.
- Campbell, J.F., Ernst, A.T., Krishnamoorthy, M., 2005b. Hub arc location problems: Part II—formulations and optimal algorithms. *Management Science* 51 (10), 1556–1571.
- Ebery, J., 2001. Solving large single allocation p-hub problems with two or three hubs. *European Journal of Operational Research* 128 (2), 447–458.
- Ernst, A.T., Krishnamoorthy, M., 1996. Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location Science* 4 (3), 139–154.
- Ernst, A.T., Krishnamoorthy, M., 1998. Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem. *European Journal of Operational Research* 104 (1), 100–112.
- Ernst, A.T., Krishnamoorthy, M., 1999. Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research* 86 (1–4), 141–159.
- Ernst, A.T., Hamacher, H., Jiang, H., Krishnamoorthy, M., Woeginger, G., 2009. Uncapacitated single and multi allocation p-hub center problems. *Computers and Operations Research* 36 (7), 2230–2241.
- Ernst, A.T., Jiang, H., Krishnamoorthy, M., 2005. Reformulations and computational results for uncapacitated single and multiple allocation hub covering problems. Unpublished Report, CSIRO Mathematical and Information Sciences, Australia.
- Hamacher, H.W., Meyer, T., 2006. Hub cover and hub center problems. Working paper. Department of Mathematics, University of Kaiserslautern, Gottlieb-Daimler-Strasse, 67663 Kaiserslautern, Germany.
- Kara, B.Y., Tansel, B.C., 2000. On the single assignment p-hub center problem. *European Journal of Operational Research* 125 (3), 648–655.
- Kara, B.Y., Tansel, B.C., 2003. The single assignment hub covering problem: models and linearizations. *Journal of the Operational Research Society* 54 (1), 59–64.
- Klincewicz, J.G., 1998. Hub location in backbone/tributary network design: a review. *Location Science* 6 (4), 307–335.
- Marin, A., Canovas, L., Landete, M., 2006. New formulations for the uncapacitated multiple allocation hub location problem. *European Journal of Operational Research* 172 (1), 274–292.
- Nickel, S., Schobel, A., Sonneborn, T., 2001. Chapter 1: hub Location problems in urban traffic networks. In: Niittymaki, J., Pursula, M. (Eds.), *Mathematics Methods and Optimization in Transportation Systems*. Kluwer Academic Publishers, pp. 1–12.
- O’Kelly, M.E., 1987. A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research* 32 (3), 393–404.
- O’Kelly, M.E., 1992. Hub facility location with fixed costs. *Papers in Regional Science* 71 (3), 293–306.
- O’Kelly, M.E., Miller, H.J., 1994. The hub network design problem: a review and synthesis. *Journal of Transport Geography* 2 (1), 31–40.
- Skorin-Kapov, D., Skorin-Kapov, J., O’Kelly, M.E., 1996. Tight linear programming relaxations of uncapacitated p-hub median problems. *European Journal of Operational Research* 94 (3), 582–593.
- Tan, P.Z., Kara, B.Y., 2007. A hub covering model for cargo delivery systems. *Networks* 49 (1), 28–39.
- Wagner, B., 2008. Model formulations for hub covering problems. *Journal of the Operational Research Society* 59 (7), 932–938.
- Yaman, H., Kara, B.Y., Tansel, B.C., 2007. The latest arrival hub location problem for cargo delivery systems with stopovers. *Transportation Research Part B* 41 (8), 906–919.
- Yoon, M.G., Current, J., 2008. The hub location and network design problem with fixed and variable arc costs: formulation and dual-based solution heuristic. *Journal of the Operational Research Society* 59 (1), 80–89.