On Hub Location Models

Darko Skorin–Kapov and Jadranka Skorin–Kapov

W. A. Harriman School for Management and Policy, Stony Brook, USA

The study of hub location models involves designing communication networks where some of the nodes serve as focal points (i.e. hubs) and other nodes are connected to those hubs. Possible applications include airline traffic flow, telecommunications, and mail delivery networks. In this paper we present an overview of recent results on solvability of some hub location models. The overview includes a heuristic approach based on tabu search, lower bounds for cases where distances satisfy triangular inequality, tight linear programming relaxations, and a linkage between optimal and heuristic solutions. As a result of those studies the range of optimally solvable instances of NP-hard hub location problems was extended. In particular, well known and heavily used bench-mark data set of real world problems (Civil Aeronautics Board (CAB) data set), that has resisted efficient solutions for more than a decade, has been solved to optimality. The paper concludes with the discussion of some avenues for future research.

Keywords: Hub location; Linear programming; Integer programming; Tabu search.

1. Introduction

Hub networks play an important role in modeling transportation and telecommunication systems. In this paper we address the p-hub location problem (p-HLP) that can be informally described as follows. For a given set of nodes (origins and/or destinations and potential hub locations), and given flow requirements between pairs of origin-destination nodes, the strategic decision in hub models is to locate the prescribed number of hub facilities and to allocate non-hub nodes to hubs. The hubs are completely interconnected, while non-hub nodes are connected to one or more hubs. The model in which each node is allocated to exactly one hub (resp., more than one hub) will be referred to as the single (resp., multiple) allocation p-HLP. Such connectivity protocol allows the use of relatively small number of links and exploitation of economies of scale by concentrating flows, and thus resulting in efficient network investment. Although the p-HLP requires exogenous internodal interaction as input data, the volume of flow which takes place between the hubs depends on the sites chosen for the hub facilities, and the pattern of nodal allocations to the hubs. This simultaneous interdependence of location and flow distinguishes the p-HLP from the standard multifacility location problem.

Modeling of both single and multiple allocation versions of the p-HLP leads to NP-hard combinatorial problems. Specifically, quadratic integer programming formulation for the single allocation case, and more recently linear integer programming formulations for both versions of the p-HLP, received considerable attention in the literature. Before referring to particular studies we mention the well known Civil Aeronautics Board (CAB) benchmark data-set on which most of the computational testing from the literature was performed (for an extensive survey see Campbell, 1994b). This data set consists of airline passenger flow and distances among 25 US cities (see list of node names in Table 1). Subsets of these data are also generated so as to give 10x10, 15x15, 20x20 and 25x25 interaction systems.

The single allocation version of the p-HLP was formulated by O'Kelly (1987) as a quadratic integer programming problem, and a number of heuristic algorithms to solve it have been proposed, including: complete evaluation of all locational patterns with respect to allocations based on distances, such as nearest hub allocation and allocation to one of the two nearest hubs (O'Kelly, 1987); exchange and clustering heuristics (Klincewicz, 1991); tabu search and

1. Atlanta	6. Cleveland	11. Kansas City	16. New Orleans	21. St. Louis
2. Baltimore	7. Dallas–Fort W.	12. Los Angeles	17. New York	22. San Francisco
3. Boston	8. Denver	13. Memphis	18. Philadelphia	23. Seattle
4. Chicago	9. Detroit	14. Miami	19. Phoenix	24. Tampa
5. Cincinnati	10. Houston	15. Minneapolis	20. Pittsburgh	25. Washington D. C.

Table 1. List of cities in CAB data set

GRASP strategies with distance based allocations (Klincewicz, 1990); tabu search method with allocations of non-hub nodes based jointly on distances as well as on flows between the nodes (Skorin–Kapov and Skorin–Kapov, 1994).

For the CAB data set, the best known solutions for the single allocation p-HLP were obtained by the tabu search heuristic (TABUHUB) developed by Skorin–Kapov and Skorin–Kapov (1994). The quality of the above mentioned tabu search heuristic was further confirmed by obtaining good lower bounds for cases when distances satisfy the triangle inequality (O'Kelly et al., 1995). The main contribution of that work is the novel approach of using a known upper bound to improve the lower bound.

Campbell (1994a) formulated the single and multiple allocation versions of the p-HLP as mixed 0/1 linear programs. However, integrality restrictions imposed on a subset of variables, coupled with large size of formulations (for a network of size n, the number of variables is $O(n^4)$) restrict the suitability of those formulations to small instances. Since LP relaxations of Campbell's models resulted with highly fractional solutions, tighter LP relaxations were needed. Skorin–Kapov et al. (1995) have proposed new mixed 0/1 linear formulations whose linear programming relaxations often provide integral solutions. For the above CAB data set, the LP relaxations proposed in their study resulted in almost all cases with integral solutions. Where this was not the case, the LP objective function value for the multiple (resp., single) allocation case was less than 0.1% (resp., 1%) bellow the optimal integer objective function value.

Moreover, by exploiting the LP solution and the best heuristic solution together with excellent lower bounds from the LP relaxation, the integrality was achieved by adding a partial set of integrality constraints. Thus, by combining the information from optimal LP solutions and from heuristic solutions, the range of optimally solvable instances of the p-HLP was extended. Specifically, all considered instances of CAB data set were solved to optimality. Note that these problems are already large (the above LP relaxations for the case with 25 nodes and 4 hubs has 391,250 variables and 31,901 constraints), and solving higher dimensional cases optimally would require larger computer resources. However, the results of these studies suggest that TABUHUB algorithm could be used with a reasonable confidence on larger problems, since for all considered cases of CAB data it achieved optimal solutions.

The plan of the paper follows. In Section 2 we discuss quadratic mixed integer programming model for the single allocation version of the p-HLP. Therein, we summarize the key elements of the TABUHUB heuristic from Skorin–Kapov and Skorin–Kapov (1994) and the idea for the lower bound based on triangular inequality from O'Kelly et al. (1995). Section 3 contains the discussion on linear integer programming models for the multiple and single allocation version of the p-HLP. In particular we describe Skorin–Kapov et al. (1995) tight LP relaxations. Finally, in Section 4 we summarize these results, and propose some lines for further investigation.

2. Quadratic approach for the single allocation p-HLP

The single allocation p-HLP can be formulated as a quadratic integer program (O'Kelly, 1987) with a nonconvex objective function as follows. Recall that there are n nodes that should interact and p of those will be designated as hubs. For nodes i and k let x_{ik} be a 0-1 variable with the following interpretation: $x_{kk} = 1$ if a hub facility is located at node k (and k is assigned to itself), and $x_{kk} = 0$ otherwise. For $i \neq k$, $x_{ik} = 1$ if a node i is allocated to hub k, and $x_{ik} = 0$ otherwise. Note that there are no explicit variables to indicate hub linkages. The number of units of flow between nodes i and j is f_{ij}, and $f_{ii} = 0$ by assumption. The transportation cost of a unit of flow between nodes i and j is d_{ij} (distance is used as a surrogate for costs), and $\alpha \leq 1$ is the discount between hubs due to heavy traffic. The single allocation p-HLP is then:

QSA-p-HLP

$$\min Z = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}\Big(\sum_{k=1}^{n} d_{ik}x_{ik} + \sum_{m=1}^{n} d_{jm}x_{jm} + \alpha \sum_{k=1}^{n}\sum_{m=1}^{n} d_{km}x_{ik}x_{jm}\Big) \quad (1.1)$$

$$\text{s.t.} (n-p+1)x_{jj} - \sum_{i=1}^{n} x_{ij} \geq 0, \; j=1,\ldots,n, \quad (1.2)$$

$$\sum_{j=1}^{n} x_{ij} = 1, \; i=1,\ldots,n, \quad (1.3)$$

$$\sum_{j=1}^{n} x_{jj} = p, \quad (1.4)$$

$$x_{ij} \in \{0,1\}, \; i=1,\ldots,n; \; j=1,\ldots,n. \quad (1.5)$$

The set of constraints (1.2) ensures that non-hub nodes are only allocated to the hubs, the set of constraints (1.3) enforces the allocation of a non-hub node to one and only one hub, and constraint (1.4) guarantees that the initially prescribed number of hubs will be located.

2.1. Tabu search heuristic

Tabu search is an approach to overcome local optimality entrapment in optimization problems with nonconvex objective function. Its strategic principles were proposed by Glover (1989,1990). Tabu search guides the search to continue exploring the feasible region even after a local optimum has been reached, and tries to prevent falling back to the same local optimum.

In contrast to previous distance based allocations approaches for the QSA-p-HLP: "allocate the non-hub node to its nearest hub" (e.g. O'Kelly, 1987, Klincewicz, 1990,1991), Skorin–Kapov and Skorin–Kapov (1994) developed a tabu search heuristic (TABUHUB) in which equal importance was given to the locational, as well as to the allocational part of the problem. The method consists of a single exchange heuristic for hub locations, which in the evaluation of an exchange uses a single exchange heuristic for non-hub reallocations. Tabu search strategy is superimposed on both levels to deal with the local optimality problem. For completeness, we present the elements and an informal description of TABUHUB algorithm.

Locational neighborhood of x (a feasible solution to the QSA-p-HLP) consists of all feasible solutions whose set of hub locations differs in exactly one location from the set of hub locations defined by x, regardless of the allocation of non-hub nodes. Since this neighborhood is large (it consists of $[(n-p)p]^{(n-p)}$ possibilities), a strategy is designed to examine it only partially. Allocation neighborhood of a feasible solution x consists of all feasible solutions which differ from x in the allocation of exactly one non-hub node ($(n-p)(p-1)$ possibilities).

A starting solution is constructed by choosing p nodes with the largest amount of incoming and outcoming flow as hubs, and by allocating each non-hub node to its nearest hub. One *master-iteration* consists of evaluating all '*single hub location exchanges*' of the current hub set. Every such evaluation starts with the 'nearest hub' reallocations which serves as a starting point for the '*single non-hub allocation exchange*' heuristic performed to obtain a better set of allocations for a given set of hubs. After carrying on this reallocation process for a number of iterations, the best obtained objective function value is used in the evaluation of the current hub exchange.

For example, let z_c be the objective function value of the current solution. In evaluating the replacement of the hub node r with the new hub node q, first reallocate non-hub nodes according to the nearest hub rule. Let this be the starting solution for the '*single non-hub allocation exchange*' heuristic as follows: having the set of hubs fixed, in each iteration evaluate all single reallocations of non-hub nodes and perform the best one. After the prescribed number of iterations, denote the best achieved objective function value by $z_{q,r}$. The evaluation of "replace the hub r with the new hub q" is then performed on the basis of the difference $z_{q,r} - z_c$. This is done for every single exchange of hubs, and the exchange with the smallest value of the difference is actually performed and the current solution is updated accordingly.

Tabu search strategy is superimposed on both location and allocation levels with the same task to guide the search beyond local optima. The elements of the move from the current solution to its selected neighbor are recorded in the *tabu list* for the purpose of forbidding the reversal of this replacement in a number (*tabu list size*) of future iterations. (Without this assurance, the search would cycle between the first encountered local minimum and its neighbor with the smallest objective function value). The tabu list is updated circularly. The tabu status is inactive only if the move leads to a solution better than the best found so far (this is the so-called *aspiration criterion*). Due to the complexity of the problem, separate tabu lists with respective parameters are maintained for the locational and allocational part. For the locational part of the problem, when the new hub node q replaces the old hub node r, we record node r in the tabu list and forbid its inclusion in the set of hubs during the tenure of its tabu status. Similarly, in the allocational part of the problem, after the reallocation of node t from one hub to another, further reallocation of t is forbidden during its tabu tenure.

The parameters of tabu search are, in general, dependent on problem size. In TABUHUB algorithm the tabu list sizes and maximal number of iterations for the locational, as well as the allocational part of the heuristic were selected as follows. Preliminary testing suggested that the appropriate tabu list should contain approximately 25%–33% of non-hub nodes. For both parts of the search the tabu list sizes were equal and, therefore, from the interval $[n - p/4, n - p/3]$. The maximal number of iterations was a function of the respective neighborhood size. Recall that for the locational (resp., allocational) heuristic, the number of neighbors considered for a given solution equals $(n-p)p$ (resp., $(n-p)(p-1)$). For the allocational heuristic, the maximal number of iterations was set to be somewhere between 25%–33%, i.e. from the interval $[(n-p)(p-1)/4, (n-p)(p-1)/3]$. For the locational heuristic, the maximal number of iterations was set to be approximately 20% of the neighborhood size, or $(n-p)p/5$. (Recall that an iteration of the locational heuristic calls $(n-p)p$ times the allocational heuristic and is therefore computationally intensive.)

Additional experimentation has showed that the method is robust with respect to those parameter values, and that the best results do not change with relatively small changes in parameter values. In particular, the best known solutions were obtained early in the search process, therefore the reductions in the maximal number of iterations would not change the best obtained solutions. The method was tested on the CAB data sets ranging between 10 and 25 nodes, with 2, 3 and 4 hubs, and for the different values of discount hub flow parameter α. In all cases the results matched or improved the best results from the literature.

2.2. Lower bounds based on triangle inequality

In order to measure the quality of heuristic solutions to the QSA-p-HLP, in this section we present the lower bounding approach from O'Kelly et al. (1995). Their approach is applicable to instances of the QSA-p-HLP in which distances satisfy the triangle inequality. The CAB data set falls into this category. The development of their lower bounds is based on the novel approach of utilizing the information from existing heuristic solutions.

First, note that if the quadratic term in the objective function of the QSA-p-HLP is disregarded, then the remaining problem is a p-median problem for which efficient solution procedures are known, yielding a naive lower bound to the QSA-p-HLP. Instead of ignoring the quadratic term completely, one can develop a lower bound by adding an underestimate of the costs of the inter facility flows (Gavish, 1985; O'Kelly 1992). The underestimate is chosen so that the neglected quadratic terms are approximated by a linear contribution to the objective. Consider the situation shown in Figure 1.

Suppose that i and j interact and that their flows are routed from i to k, from k to m, and then from m to j. The flow between the hubs k and m

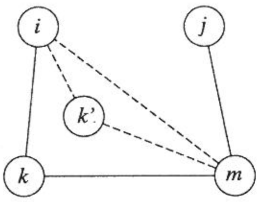

Fig. 1. Base for the underestimate of inter facility cost: $d_{ik} + d_{km} \geq d_{im}$

introduces a quadratic component to the objective. Specifically, the contribution of this flow to the objective is $\alpha x_{ik} x_{jm} f_{ij} d_{km}$, which weights the distance between hubs k and m with the volume of flow between nodes i and j, whenever x_{ik} and x_{jm} are equal to one. From Figure 1, assuming that the triangle inequality holds, note that the following is true: $d_{im} \leq d_{ik} + d_{km}$, implying that $d_{km} \geq d_{im} - d_{ik}$.

Upon substitution of $d_{im} - d_{ik}$ for d_{km} in the original objective of the QSA-p-HLP, and making use of $\sum_j x_{ij} = 1$, $\sum_j f_{ij} = O_i$, and $\sum_i f_{ij} = D_j$, it can be shown that:

$$Z \geq \sum_{i=1}^{n} \sum_{k=1}^{n} d_{ik} x_{ik} (O_i(1-\alpha) + D_i) + \alpha \Big(\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{m=1}^{n} f_{ij} d_{im} x_{jm} \Big).$$

After some relabeling, the above substitution results in a linearization of the objective function and yields a lower bound which is the optimal value of the following p-median problem.

IP1

$$LB_1 = \min \Big(\sum_{i=1}^{n} \sum_{k=1}^{n} x_{ik} ((d_{ik}(1-\alpha) O_i + D_i) + \alpha \sum_{q=1}^{n} f_{qi} d_{qk}) \Big)$$

$$s.t. \ (1.2), (1.3), (1.4) \text{ and } (1.5).$$

An effective way to solve the p-median problem is to develop a related Simple Plant Location Problem (SPLP), where the constraint on the number of facilities (1.4) is dualized, and Lagrangian relaxation is used to solve for a parameter θ which generates the correct number of facilities. The constraint that requires p facilities to be opened is relaxed, and the problem can be re-written as:

SPLP

$$\min \Big\{ \sum_i \sum_k x_{ik} \{ d_{ik} [(1-\alpha) O_i + D_i] + \alpha \sum_q f_{qi} d_{qk} \} + \theta \sum_j x_{jj} \Big\}$$

$$s.t. \ (1.2), (1.3) \text{ and } (1.5).$$

By varying the values of θ, an arbitrary number of facilities p can be generated. Subtracting θp (the artificial fixed cost) from the objective cost yields the desired transportation cost. This use of the SPLP to solve a p-median problem is well known in the literature (see Mirchandani, 1990).

O'Kelly et al. (1995) improved the above lower bound LB_1 by incorporating the knowledge from existing heuristic solutions. In computing LB_1 the part of the objective function associated with the flow f_{ij} is underestimated by the amount: $D_{ij} = (d_{ik} + d_{km} - d_{im}) f_{ij} \alpha$. The lower bound LB_2, such that $LB_2 \geq LB_1$, is constructed as follows.

The objective is to increase LB_1 by "including back" part of the above underestimate of the objective function. Namely, d_{im} in the objective function of IP_1 was replaced with $c(i,m) = d_{ik'} + d_{k'm} = \min\{d_{ik} + d_{km} : k \in N, \ k \neq i, m\}$ (see Figure 1). In such case the underestimate of the cost associated with the flow f_{ij} would be $D'_{ij} = D_{ij} - (c(i,m) - d_{im}) f_{ij} \alpha$. By the triangular inequality $D'_{ij} \leq D_{ij}$ and by construction D'_{ij} is non-negative.

The question remains how to implement the above modification of IP_1 since the hub locations, as well as the allocations of non-hub nodes, are not known in advance. Consequently, the implementation of the above modification in IP_1would require quadratic term in the objective function. To preserve the linearity of the objective function the above underestimate of the cost is applied only for certain pairs of nodes. Namely, some feasible solution to IP_1 (to be referred to as the *reference solution*) is taken, and then the above modification of the objective function of IP_1 is applied with respect to that solution. In a sense, the modification would "penalize" the choice of hubs and assignment of non-hub nodes made by the reference solution. It is shown below that if the reference solution or some "parts" of it endup in the optimal solution to the modified problem, the lower bound improves.

The choice of the reference solution was not arbitrary. It is desirable to have a reference solution to IP_1 which is likely to "survive" the above penalization, and therefore the optimal solution to IP_1 is a natural choice. Another good choice for a reference solution is the best known heuristic solution. Namely, the modification of IP_1 brings the integer linear program "closer" to the original QSA-p-HLP. The conjecture is that the best known heuristic solution is optimal or close to optimal solution of the QSA-p-HLP, and so it is likely that this solution would "survive" the above modification. The computational results confirmed the conjecture.

Below we summarize the development of improved lower bound with respect to a reference solution. Let N be the set of nodes and assume that the reference solution to QSA-p-HLP is implicitly given by a function $a : N \rightarrow N$, i.e. for a node u, $a(u) = v$ means that node u is allocated to hub v.

The above modification of the objective function of IP_1 is applied for pairs of nodes i, j for which $a(i) \neq a(j)$ and $a(i) \neq i$. Namely, in such cases $(c(i, a(j)) - d_{i,a(j)}) f_{ij} \alpha$ is added to the cost coefficient associated with the variable $x_{i,a(j)}$ in the objective function of IP_1, where $c(i, a(j)) = \min\{d_{ik} + d_{k,a(j)} : k \in N, k \neq i, a(j)\}$. This modification can be accomplished via the mixed integer programming problem:

$IP_2(a)$

$$LB_2(\mathrm{a}) = \min\Big(\mathrm{I} + \sum_{\substack{i=1 \\ a(i)\neq a(j),i}}^{n} \sum_{j=1}^{n} y_{ij}(c(i,a(j)) - d_{i,a(j)})\alpha f_{ij}\Big)$$

$$s.t.\ (1.2), (1.3), (1.4), (1.5) \text{ and,}$$

$$y_{ij} \geq x_{j,a(j)} - x_{i,a(j)} - x_{ii},\ i=1,\ldots,n; j = 1,\ldots,n$$
$$y_{ij} \geq 0,\ i = 1,\ldots,n;\ j = 1,\ldots,n,$$

where I is the part of the objective function inherited from IP_1, i.e.,

$$\mathrm{I} = \sum_{i=1}^{n}\sum_{k=1}^{n} x_{ik}((d_{ik}(1-\alpha)O_i + D_i) + \alpha \sum_{q=1}^{n} f_{qi} d_{qk}),$$

and indices $a(j)$ are obtained from the reference solution. Note that in the optimal solution to the $IP_2(a)$, y_{ij} will clearly have the following values:

$$y_{ij} = \begin{cases} 1 & \text{if } x_{j,a(j)} = 1,\ x_{i,a(j)} = 0,\ x_{ii} = 0, \\ 0 & \text{otherwise.} \end{cases}$$

O'Kelly et al. (1995) proved that $LB_2(a) \geq LB_1$, and that LB_2 is a lower bound to the QSA-p-HLP objective. Observe that in $IP_2(a)$ we lost the structure of IP_1 that enabled us to apply the Simple Plant Location Problem (SPLP) routine to compute LB_1.

However, next we will describe a compromise model and another lower bound, $LB_3(a)$, for which the SPL structure will be preserved, but $LB_3(a)$ will not necessarily be a better lower bound then LB_1.To that end, consider the following integer linear programming problem:

$IP_3(a)$

$$LB_3(a) = \min\Big(I + \sum_{\substack{i=1 \\ a(i)\neq a(j),i}}^{n} \sum_{j=1}^{n} (x_{j,a(j)} - x_{i,a(j)} - x_{ii})(c(i,a(j)) - d_{i,a(j)})\alpha f_{ij}\Big)$$

$$s.t.\ (1.2), (1.3), (1.4) \text{ and } (1.5).$$

Clearly, $LB_3(a) \leq LB_2(a)$. Hence, $LB_3(a)$ is a lower bound. However, it is plausible that in some situation $LB_3(a) < LB_1$. Indeed, since $IP_3(a)$ does not exclude cases when $x_{j,a(j)} = 0$, and $x_{i,a(j)=1}$ or $x_{ii} = 1$, the negative amount could be added to I. Nevertheless, the rationale for considering $LB_3(a)$ is that if a reference solution is robust, i.e. if it can survive the modifications in objective function coefficients, then it is likely that $LB_3(a) > LB_1$. Moreover, $LB_3(a)$ can be computed using the same SPLP routine.

Indeed, the computational results on the CAB data confirmed the usefulness of $LB_3(a)$. The lower bound (LB_1) was obtained using the approach from O'Kelly (1992) and the lower bounds resulting from the above described approach were obtained for two distinct values of reference solution: the optimal solution to LB_1 giving $LB_3(lb_1)$, and the best known heuristic solution (obtained by TABUHUB) giving

$LB_3(ts)$. The new lower bound LB_3 is computed as $LB_3 = \max(LB_3(lb_1), LB_3(ts))$.

The lower bound LB_3 was tested on the CAB data. In 83 out of 84 considered cases the best previously known lower bound was improved. As a result of that research, the average optimality gap was reduced for smaller problems (all instances with 10 and 15 nodes) to 3.3% and for larger problems (20 and 25 nodes) to 5.9%. The CPU time (on mainframe) needed to compute tabu search solutions was ranging between few seconds to a minute, while the computation of LB_1 and LB_3 required only few seconds.

3. Linear programming approach for the p-HLP

In this section we present some recent results for the p-HLP based on integer linear programming formulations. First we consider a less restricted problem, i.e. the multiple allocation p-HLP. As in Campbell (1994a), we define the following variables: x_{ijkm} = the fraction of flow from location origin i to location (destination) j, routed via hubs at locations k and m in that order; $y_k = 1$ if location k is a hub, and 0 otherwise. The cost per unit of flow between origin i and destination j, routed via hubs k and m in that order, is given by $d_{ijkm} = d_{ik} + \alpha d_{km} + d_{mj}$. We assume that $d_{ii} = 0, i = 1,\ldots,n$, so the formula for d_{ijkm} remains valid when i and/or j is a hub. Campbell 1994a formulated the multiple allocation version of the p-HLP as:

LMA-p-HLP'

$$\min \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{m=1}^{n} f_{ij} d_{ijkm} x_{ijkm} \tag{3.1}$$

$$s.t. \sum_{k=1}^{n} y_k = p, \tag{3.2}$$

$$\sum_{k=1}^{n}\sum_{m=1}^{n} x_{ijkm} = 1,\ i = 1,\ldots,n,\ j = 1,\ldots,n, \tag{3.3.}$$

$$x_{ijkm} \le y_k,\ i = 1,\ldots,n,\ j = 1,\ldots,n,\ k = 1,\ldots,n,\ m = 1,\ldots,n, \tag{3.4}$$

$$x_{ijkm} \le y_m,\ i = 1,\ldots,n,\ j = 1,\ldots,n,\ k = 1,\ldots,n,\ m = 1,\ldots,n, \tag{3.5}$$

$$y_k \in \{0,1\},\ k = 1,\ldots,n, \tag{3.6}$$

$$x_{ijkm} \ge 0,\ i = 1,\ldots,n,\ j = 1,\ldots,n,\ k = 1,\ldots,n,\ m = 1,\ldots,n. \tag{3.7}$$

The objective is to minimize the overall transportation cost subject to: having exactly p hubs (constraint 3.2); the flow between every origin-destination $(o-d)$ pair (i,j) should be routed via some hub pair (constraints 3.3) and flows can be routed only via locations that are hubs (constraints 3.4, 3.5). Variables y, serving as hub indicators, are restricted to be 0 or 1, and flow variables x are nonnegative. Due to constraints (3.3), it is clear that x variables cannot have values bigger than 1. Problem LMA-p-HLP is a very large mixed 0/1 linear problem (with $n + n^4$ variables, and $1 + n^2 + 2n^4$ constraints). Campbell (1994a), pp. 390 states: "In the absence of capacity constraints on the links, an optimal solution will have all x_{ijkm} equal to zero or one since the total flow for each $o-d$ pair should be routed via the least costly pair." Yet, due to its size, the integrality of y variables makes the problem very difficult to solve. Relaxing the integrality, however, results in a highly fractional solution.

An intuitive explanation for obtaining fractional solutions when relaxing the integrality of y variables is that the constraints (3.4) and (3.5) are not 'strong enough' with respect to hub locations. Namely, since there are no fixed costs for opening the hubs, relaxing integrality will create lots of 'partial' hubs, depending on the cheapest routes indicated via x_{ijkm} variables.

Skorin–Kapov et al. (1995) proposed a modification to Campbell's LMA-p-HLP model as follows. Replace constraints (3.4) and (3.5) by:

$$\sum_{m=1}^{n} x_{ijkm} \le y_k,\ i = 1,\ldots,n,\ j = 1,\ldots,n,\ k = 1,\ldots,n, \tag{3.8}$$

and

$$\sum_{k=1}^{n} x_{ijkm} \le y_m,\ i = 1,\ldots,n,\ j = 1,\ldots,n,\ m = 1,\ldots,n, \tag{3.9}$$

Their modified model (to be refered to as the LMA-p-HLP) is equivalent to the original LMA-p-HLP, since for a given i, j, and k (respectively., i, j, and m), only the cheapest route $i-k-m-j$ will 'survive'. However, the linear relaxation of the modified model LMA-p-HPL, is tighter than the linear relaxation of the original LMA-p-HLP.This follows because every nonnegative solution satisfying (3.8) and (3.9) satisfies (3.4) and (3.5), but not vice versa. Intuitively, we expect that y variables will have values closer to integral values. The computational results performed on the CAB data set confirmed it, and in almost all cases (58 out of 60) the LP relaxation of the LMA-p-HLP provided integral solutions. For the instances with non-integral LP solutions, the LP relaxation resulted with objective function value less than 0.1% below the optimal objective function value. Moreover, when compared to Campbell's (1994) model, the constraint set has been reduced by $2n^3(n-1)$ constraints. The CPU time (on SUN Sparc 2 station) needed to solve these relaxations for the LMA-p-HLP ranged from a minute (10 node cases) to 4 hours (25 node cases).

Solutions to the LP relaxation of LMA-p-HLP exhibit multiple allocation property. Namely, for a given $o-d$ pair (i, j), it might be best to use the link from i to the hub k, but for the $o-d$ pair (i, l), it might be best to use the link from i to some hub other than k. However, in numerous applications it is economically justified to restrict non-hub nodes to be connected to exactly one hub. Such connectivity protocol greatly reduces the number of potentially expensive links.

Campbell (1994a) also proposed a couple of mixed linear programming formulations for the single allocation version of the p-HLP. However, his linear programming relaxations are not tight and lead to fractional solutions with objective function values significantly below the optimal objective function values (for details see Skorin–Kapov et al. (1995))

Skorin–Kapov et al. (1995) proposed a new formulation for the single allocation p-HLP obtained by modifying the multiple allocation version (LMA-p-HLP). The idea was to make the allocation choice of an origin node i independent of a destination node, and vice versa. To that end, they introduce the 'allocation' variables z_{ik} where $z_{ik} = 1$ if the origin is allocated to hub k, and 0 otherwise. Note that since each hub is allocated to itself, it must be $z_{ik} \leq z_{kk}$ for all i and all k. The following mixed 0/1 formulation could then be used for the single allocation p-HLP:

LSA-p-HLP

$$\min \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{m=1}^{n} f_{ij} d_{ijkm} x_{ijkm} \tag{3.10}$$

$$s.t. \sum_{k=1}^{n} z_{kk} = p, \tag{3.11}$$

$$\sum_{k=1}^{n} z_{ik} = 1, \ i = 1, \ldots, n, \tag{3.12}$$

$$z_{ik} \leq z_{kk}, \ i = 1, \ldots, n, \quad k = 1, \ldots, n, \tag{3.13}$$

$$\sum_{m=1}^{n} x_{ijkm} = z_{ik}, \ i = 1, \ldots, n, \quad j = 1, \ldots, n, \ k = 1, \ldots, n, \tag{3.14}$$

$$\sum_{k=1}^{n} x_{ijkm} = z_{jm}, \ i = 1, \ldots, n, \quad j = 1, \ldots, n, \ m = 1, \ldots, n, \tag{3.15}$$

$$z_{ik} \in \{0, 1\}, \ i = 1, \ldots, n, \ k = 1, \ldots, n \tag{3.16}$$

$$x_{ijkm} \geq 0, \ i = 1, \ldots, n, j = 1, \ldots, n, \quad k = 1, \ldots, n, \ m = 1, \ldots, n, \tag{3.17}$$

Constraints (3.12) state that each node has to be allocated to exactly one hub, constraints (3.14) assure that for every destination j, the sum $\sum_m x_{ijkm}$ (i.e. the total flow from origin i to destination j routed via all paths using link $i-k$) will be non zero only if location i is allocated to hub k (independently of a destination). Similarly, constraints (3.15) assure that for every origin i and every hub k, a flow through the path $i-k-m-j$ is feasible only if j is allocated to hub m (independently of an origin).

The LSA-p-HLP has n^2 binary variables, n^4 continuos variables, and $1+n+n^2+2n^3$ linear constraints. For the CAB data set, the above formulation resulted in LP relaxations ranging from 10,010 to 391,250 variables and from

2,101 to 31,901 constraints, which proved to be difficult linear programs. The linear programming relaxations of the LSA-p-HLP appear to be tight: in almost all considered instances of the CAB data (57 out of 60) the LP relaxations have integral solutions. For the instances with non-integral LP solutions, the respective LP objective function values were less than 1% below the optimal objective function value. The CPU time (on SUN Sparc 2) needed to solve these LP relaxations was ranging from a few minutes (10 node cases) to 15 hours (25 node cases).

The integrality for the instances with non-integral LP solutions was achieved as follows. For the LMA-p-HLP there are n integrality constraints (i.e. n $0/1$ variables), but for each tested data instance, adding only one of those constraints (as guided by the LP solution) was sufficient to obtain an optimal integral solution. For the LSA-p-HLP there are n^2 $0/1$ variables, and it is more difficult to decide which integrality constraints should be added to obtain an optimal integral solution. In this case the best known heuristic solution for the given data instance obtained via tabu search was also used as a guidance in adding integrality constraints. Specifically, the following heuristic rule was used to determine the branching variable: branch on the fractional hub variable z_{jj} with the biggest sum of differences between values of variables z_{ij}, $i = 1, \ldots, n$ of heuristic and LP solutions. As a result, the optimality of all heuristic solutions obtained for the CAB data with TABUHUB algorithm was established.

4. Conclusions and directions for future research

Most of research concerning p-hub location problems so far was concerned with the single and multiple versions presented in this paper. More than 70 publications on the topic were classified in Campbell's (1994b) survey. Therein he indicates that many of these studies use the CAB data, thus enabling some comparison. In this paper we provided an overview of some more recent results concerning heuristic, as well as optimal solvability of the hub location problem. Integer linear programming with tight linear relaxations combined with the good heuristic solutions resulted in optimal solution for CAB data set. These results also suggest that the TABUHUB heuristic could be used with a reasonable confidence for even larger problems.

Although p-hub location models considered in this paper have lots of advantages, there are still lots of opportunities for improvements. It would be important to modify parameters in these models so as to incorporate additional practical considerations. For example, transportation costs might include some measurement for loading and unloading, waiting time, fixed costs for takeoff and landing, fixed costs of facilities, and fixed costs to establish links. Another possibility for future research is the re-evaluation of the way we model the level of discount for the transportation between hubs. Closer inspection of some of the above obtained optimal solutions on the CAB data showed that the flow between some pairs of hubs was not so large, suggesting that for those cases economies of scale were not properly utilized. We already started some experiments with the new model in which the cost of flow on a particular link is discounted only if some level of traffic is reached, the requirement on the prescribed number of hubs is removed, and the direct traffic is allowed between non-hub nodes. Note that, strictly speaking, this is not a hub location model. However, preliminary computation on small problems resulted with hub like transportation networks. This suggests that such implicit hub models might give us transportation systems with better utilization of economies of scale.

Acknowledgment

The authors would like to acknowledge that research reported here was partially supported by NSF grants DMS-9218206 and DDM-937417.

References

CAMPBELL, J. F., (1994a), "Integer programming formulations of discrete hub location problems," *European Journal of Operational Research* 72, 387–405.

CAMPBELL, J. F., (1994b), "A survey of network hub location," *Studies in Locational Analysis*, Issue 6, 31–49.

GAVISH, B., (1985), "Models for configuring large-scale distributed computing systems," AT&T *Technical Journal*, 64(2), 491–532.

KLINCEWICZ, J. G., (1991), "Heuristics for the p-hub location problem," *European Journal of Operational Research*, 53, 25–37.

KLINCEWICZ, J. G., (1992), "Avoiding local optima in the p-hub location problem using tabu search and GRASP," *Annals of Operations Research* 40, 283–302.

MIRCHANDANI, P., (1990), "The p-median problem and generalizations," Chapter 2 in "Discrete location theory", edited by P. B. Mirchandani and R. L. Francis, Wiley, New York.

O'KELLY, M. E., (1987), "A quadratic integer program for the location of interacting hub facilities," *European Journal of Operational Research*, 32, 393–404.

O'KELLY, M. E., (1992), "Hub facility location with fixed cost," *Papers on Regional Science*. 71,3, 293–306.

O'KELLY, M., D. SKORIN–KAPOV, AND J. SKORIN–KAPOV, (1995) "Lower bounds for the hub location problem," *Management Science*, Vol. 41, No. 4, 713–721.

SKORIN–KAPOV, D. AND J. SKORIN–KAPOV, (1994), "On tabu search for the location of interacting hub facilities," *European Journal of Operational Research* 73, 502–509.

SKORIN–KAPOV, D., J. SKORIN–KAPOV AND M. O'KELLY, (1995), "Tight linear programming relaxations of uncapacitated p-hub median problems," to appear in *European Journal of Operational Research*.

Received: September, 1995
Accepted: November, 1995

Contact address:

Darko Skorin–Kapov
W. A. Harriman School
for Management and Policy
State University of New York
at Stony Brook
Stony Brook, NY 11794–3775, USA
e-mail: dskorin@edge.har.sunysb.edu

Jadranka Skorin–Kapov
W. A. Harriman School
for Management and Policy
State University of new York
at Stony Brook
Stony Brook, NY 11794–3775, USA
e-mail: jskorin@frost.har.sunysb.edu

DARKO SKORIN–KAPOV is an Assistant Professor in Harriman School for Management and Policy at the State University of New York at Stony Brook. He got his undergraduate and masters degree in Mathematics and Information Science, respectively from the University of Zagreb, Croatia. He received his Ph.D. in Management Science from the University of British Columbia, Vancouver, Canada. His research is predominantly in the area of network optimization and cost allocation. He has published in Discrete Applied Mathematics, European Journal of Operational Research, SIAM Journal of Discrete Mathematics, Telecommunication Systems, Operations Research Letters, Management Science, Annals of Operations Research, Journal of Computing and Information Technology.

JADRANKA SKORIN–KAPOV is an Associate Professor in the Harriman School for Management and Policy, State University of New York at Stony Brook. She received her B.Sc. and M.Sc. in Applied Mathematics from the University of Zagreb, Croatia and her Ph.D. in Operational Research from the University of British Columbia, Canada. Her research interests are in the area of combinatorial optimization. She has published in Mathematical Programming, Operations Research Letters, ORSA Journal on Computing, Computers and Operational Research, Discrete Applied Mathematics, European Journal of Operational Research, and Annals of Operations Research.