A conditional $p$-hub location problem with attraction functions

H.A. Eiselt$^a$, Vladimir Marianov$^{b,*}$

$^a$Faculty of Business Administration, University of New Brunswick, Fredericton, NB, Canada E3B 5A3
$^b$Department of Electrical Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile

ARTICLE INFO

Available online 9 December 2008

Keywords:
Hub location
Conditional location
Gravity model

ABSTRACT

We formulate the competitive hub location problem in which customers have gravity-like utility functions. In the resulting probabilistic model, customers choose an airline depending on a combination of functions of flying time and fare. The (conditional) follower's hub location problem is solved by means of a heuristic concentration method. Computational experience is obtained using the Australian data frequently used in the literature. The results demonstrate that the proposed method is viable even for problems of realistic size, and the results appear quite robust with respect to the leader's hub locations.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

The purpose of this paper is to locate hubs in the presence of competition. While hubs are found in many transportation networks, they are most frequently associated with the airline industry. The existence of hubs can be justified not only from the supplier's point of view, but also by taking the side of the customers. Routing traffic through hubs will allow the airline industry to take advantage of larger planes that, on a basis of costs per passenger and mile, are more cost-efficient than smaller commuter airplanes. McShan and Windle [1] report in their study of the airline industry between 1977 and 1984 that each 1% increase of hub-and-spoke traffic decreased the airlines' costs by 0.11%. On the other hand, traffic from a small place to another minor destination will almost never justify direct flight, or, if it does, only very infrequently. The use of hubs allows passengers who want to travel to smaller places, to do so on a much more frequent basis as would otherwise be possible.

Given these arguments in favor of hubs, airlines have established more or less sophisticated hub systems. As a case in point, the six largest North American airlines each have between three and five hub cities, some add secondary hubs as well as focus cities. Our discussion in this paper will consider a model that locates a small number of hubs in the presence of already existing traffic networks by other carriers. As such, the model follows the work by Marianov et al. [2], except that we no longer make their assumption that customers choose an airline exclusively on the basis of cost advantage, but consider other factors as well. Furthermore, we relax their assumption that "winner takes all" and instead, we allow a more realistic market share among competitors. As such, this work brings together three main streams of research, viz:

- hub location problems,
- behavioral research that attempts to model customer choice behavior, and
- conditional location problems, in which some facilities, goods, or services already exist.

Our analysis will proceed in three phases: In the first phase we will discuss customer behavior and ways to model it. In phase 2 we formulate the model, and in phase 3 we solve some instances of our model. The sections of this paper follow these phases closely.

2. Literature review

Hub location problems were first introduced by O'Kelly [3]. Their basic structure includes origins, destinations, and hubs (or transshipment points). Customers, who are located at the origins, desire to travel to the destinations in certain numbers, so that the demand does not occur at demand points but there is a demand for trips on routes between origins and destinations, sometimes also referred to as $O-D$ pairs. Rather than connecting origins and destinations directly, which would in many cases not only be very expensive, but downright infeasible, the customer flow is directed towards and through hubs. In general, for some $O-D$ pairs a direct connection may exist, for others customers may have to travel though a single hub, while other $O-D$ pairs may require routing through two hubs. Systems with more than two hubs per route exist in some...
applications, but are generally considered too awkward in situations that involve passenger travel. The incentive for the firms to use hubs is, at least in part, due to the fact that travel between hubs is usually cheaper, as the larger numbers of passengers typically allow the use of larger and (on a per-passenger basis) more efficient aircraft.

There are many features that distinguish different hub location models. However, four features stand out. First, there is the single vs. multiple assignment rule. In case of single assignments, all traffic that begins at some origin is routed to a single hub, while in the case of multiple assignments, traffic flow from some origin i can be routed to different hubs. In airline networks, most origins are connected to more than one hub. In contrast, most mail sorting systems include single assignments.

The work by Chen [4] includes single assignments, while O’Kelly and Bryan [5] discuss models with single- and multiple assignments. Most authors, though, investigate models that allow multiple assignments; see e.g. [6–9] belong into that category.

The second distinguishing feature concerns the number of hubs that are included in the model. In p-hub location models, the number of hubs is exogenously given by the decision maker (who can solve problems with different values of p which is generally easy to do as the number of hubs is typically very small). On the other hand, rather than prespecifying the number of hubs, the model may include terms in the objective function that express the fixed costs that are incurred when hubs are established. The objective will then minimize the costs of routing the flow through the network plus the costs of establishing hubs. This would be the “hub-equivalent” of the simple plant location problem. An alternative would be to include a budget constraint in the model. The contribution by Marianov et al. [2] is an example of models that exogenously specify the number of hubs that are to be located, while Nickel et al. [10] and Bollapragada et al. [11] constrain the number of hubs in the model by virtue of a budget constraint. On the other hand, the models by Cánovas et al. [9], Rodríguez-Martín and Salazar-González [12] include the costs of the hubs in the objective function.

A third distinguishing feature is the inclusion of capacities in the model, similar to the capacitated and uncapacitated plant location problems. These capacities may either limit the throughput through a hub (as is reasonable for models involving air traffic), or they may restrict the magnitude of the flow that enters a hub from an origin, as may be appropriate in networks in which mail is sorted at hubs, so that flow that enters a hub from an origin must still be sorted, while mail that is routed to a hub from another hub does not.

Examples in the literature that deal with uncapacitated models are those by Klinkewicz [13], Hamacher et al. [14] and Chen [4], while capacitated models are discussed by Ebery et al. [6] and Yaman and Carello [15]. An example of capacity constraint that limits waiting time at hubs is presented by Marianov and Serra [16].

Finally, some models may include not only variable costs for routing traffic through a network, but also fixed costs for establishing links of the network. Such a feature is typical in network design problems. It adds another degree of difficulty. Problems of this nature are found in [17–19].

Hub location problems can be modeled in different ways. One standard formulation uses variables x_{ik} to specify the flow from origin i to destination j by way of hubs k and ℓ (where k = ℓ is possible in case the trip involves only a single hub, and where “0” is an artificial hub, so that x_{i00} indicates that there is a direct connection from origin i to destination j. This four-subscript formulation is similar to what Eiselt [20] called “path-based formulations.” Another possibility is to use an “arc-based formulation” similar to multicmodity network flows, where variables x_{ik} denote the flow from origin i to hub k, variables y_{jk} for the flow from hub k to hub ℓ that started the trip at origin i, and variables x_{ij} for the flow from hub ℓ to destination j for flow that began its journey at origin i. For simplicity assume that n denotes the number of nodes in a network that symbolize origins, destinations, and potential hub locations. Then, while arc-based formulations use O(n^2) variables, path-based formulation require O(n^4) variables. There does not appear a clear indication as to which of the two is superior. As even the most basic hub location problems are NP-hard (Sohn and Park [21] proved that the single allocation problem is NP-hard for three or more hubs), it is hardly surprising that many of the solution techniques that have been suggested for the various models are heuristics. They run the gamut from greedy heuristics [11], tabu search [22], simulated annealing [4], Teitz and Bart vertex substitution heuristics [2], and heuristic concentration [15].

Another strand of contributions uses heuristics as subalgorithms in branch and bound schemes. Cánovas et al. [9] use Erlenkotter’s [23] dual ascent techniques as also do Mayer and Wagner [7]. Another method to improve the performance of exact method involves the use of polyhedral combinatorics, such as cutting plane methods and branch and cut. Contributions in this category include works by Hamacher et al. [14], Labbé and Yaman [24,25], and Labbé et al. [26]. Finally, Rodríguez-Martín and Salazar-González [12] employ decomposition techniques to solve their capacitated hub location problem. As far as general surveys are concerned, an early such review was provided by Campbell [27]. The review by Klinkewicz [28] includes different network structures, reliability issues, and solution methods, such as decomposition techniques. Bryan and O’Kelly [29] focus on the analytical aspects of the model, while the more recent work by Campbell et al. [44] not only surveys different models but also provides a framework and taxonomy.

The history of models that use attraction functions dates back to Reilly [45]. His deterministic model was picked up and presented as a probabilistic model as the famed gravitational models by Huff [30]. The basic idea of all gravitational models is to model customer behavior by the use of attraction functions. Huff’s attraction function models the attraction a customer feels towards a facility as a ratio that includes a base attraction in the numerator (in the context of retail facilities this would typically be the size of a facility as a proxy expression for the variety of goods available), and a dissuity in the denominator, which is typically the spatial separation of customer and facility, i.e., the distance between them. This distance is usually raised to some power that expresses the attraction decay. Drezner and Drezner [32] appear to have been the first to apply attraction functions to hub location models. On the other hand, Marianov et al. [2] were the first to put hub location models in (their natural) competitive context. This paper merges these two approaches.

The remainder of this paper is organized as follows. The third section provides details on the problem and develops and states the model under consideration in this work. Section 4 describes the solution technique of choice for our model, and Section 5 reports computational evidence. Finally, Section 6 summarizes our findings and provides an outlook on potential future research directions.

3. The problem and the model

The space we have chosen for the region of interest is a network. Each node of this network represents a demand concentration (a demand node) and symbolizes a candidate location for a hub. The edges of the network represent possible connections (flights) between cities. For each combination of origin i and destination j, there exists a known, non-elastic demand h_{ij} for air traffic to node j. The region of interest is already served by one or more existing airlines. To facilitate our discussion, the subscript “o” refers to the
entering airline or US, while the subscripts \( s = 1, 2, \ldots \) refer to already existing airlines. These existing airlines organize their flights using a hub-and-spoke structure, i.e., traffic from node \( i \) is directed to a hub airport located at \( k \), where it joins the traffic originating from other nodes in the neighborhood. From the hub at \( k \), the traffic is directed either to its destination \( j \) or to a second hub airport located at node \( \ell \). If there is a second hub \( \ell \) on the route, the traffic is directed from it to its destination node \( j \). Although in other applications it could be convenient, we focus on the airline case, in which a route never uses more than two hubs. If airline \( s \) is \( 0,1,2,\ldots \) charges a fare \( c_{iks} \) for a flight between nodes \( k \) and \( \ell \), the fare is reduced by a factor of \( \pi \leq 1 \) if these nodes are hubs, representing the economies of scale (such as more efficient aircraft and a potential higher load factor) obtained by a concentration of the traffic between hubs. A total of \( r \) hubs of existing airlines are located at nodes belonging to the set \( \text{Comp} = \{ j \mid \text{there exists a competing hub at} \ j \} \).

A new airline (below referred to as US) is planning to enter the market. As a newcomer to the market, it will attempt to capture the maximum possible traffic and thus maximize its market share. In the long run, once it has gained a solid foothold in the market, the new entrant’s objective will slowly change towards profit maximization. Its network will include \( p \) hubs, at locations chosen from the set \( \text{Origin} = \{ j \mid \text{node} \ j \ is \ a \ potential \ hub \ to \ US \} \) of candidate nodes (or cities). Once located, the entering firm establishes routes connecting demand nodes \( i \) and \( j \), that visit these hubs. The entering firm also operates using a hub and spoke structure, with similar economies of scale as the incumbent airlines. Furthermore, we assume that all competitors in the market have the same correct knowledge of the demand structure.

As opposed to previous models that deal with the location of hubs for maximum passenger capture and in which each competitor had exactly one route between each pair of nodes, we assume here that each competitor can have multiple routes between nodes \( i \) and \( j \), going through different hubs or pairs of hubs. Not all passengers traveling from \( i \) to \( j \) will take the same route and/or airline. The proportion of passengers that choose a specific route and airline (or, alternatively, the proportion of times that passengers at a node choose the route/airline) depends on a relative utility that each passenger perceives for using that route. This type of model was put forward in the context of location models by Hakimi [33] and it is commonly referred to as the “proportional model.” Drezner and Dreznner [32] employ attraction models in the context of hub location. However, their model is noncompetitive, and it minimizes the total distances traveled by customers.

The utility that each passenger receives from a route from node \( i \) to node \( j \) by way of hubs \( k \) and \( \ell \) operated by airline \( s \) is \( 0,1,2,\ldots \) is denoted by \( u_{ijkts} \). This utility is assumed to depend on three factors. The first factor is the basic attractiveness \( A_{iks} \) of a pair of hubs \( (k,\ell) \) on a trip operated by airline \( s \). It is a function of the airline \( s \) (its safety record, collecting mileage points, personal space, in-flight entertainment, quality of food and service), the number of hubs on the path (1 or 2), and it also depends on the hubs themselves, in particular their location and, maybe even more importantly, the conveniences they offer to travelers. The second factor that influences a traveler’s utility is the cost of the route \( c_{iks} \) (the fare), and the third factor is the total time required by the flight, which we denote by \( t_{iks} \). Note that the flying time includes the necessary layovers, which, in turn, will depend on the congestion at the airport, which may be a function of the frequency of flights offered by airline \( s \). Grove and O’Kelly [34] estimated the travel time as a fixed time of about 30 minutes plus 0.12 multiplied by the distance (in miles) between the origin and the destination of the trip, considering layover at the hub(s). However, given the increased time required for security clearance, today’s travel times will be significantly longer.

The attraction function according to which customers choose a specific route \((i, j, k, \ell)\) and an airline \( s \) follows gravity models first put forward by Huff [30,31]. All gravity model have in common that customers choose a facility (or, as is the case here, a route and an airline) based on a ratio of a base attraction and the distance (or cost) of the trip raised to some power. The original models squared the distance. The power of the distance measures the loss of utility (or the increase in the disutility) that customers experience as distances increase. A higher power indicates that a customer considers longer distances or travel times more highly undesirable as customer whose utility function raises the distance to a lower power. Our attraction function can formally be written as

\[
u_{ijkts} = \frac{A_{iks}}{\gamma t_{iks}^\beta + (1-\gamma) c_{ists}^\lambda}
\]

In this expression, the parameters \( \beta \) and \( \lambda \) denote the attraction decay of the travel time and the cost, respectively. In particular, the exponents \( \beta \) and \( \lambda \) indicate how fast the utility drops off as travel time and travel costs increase. The parameter \( \gamma \) allows us to trade off travel time and travel costs. The travel time \( t_{iks} = t_{ik} + t_{k\ell} + t_{\ell j} \), where \( t_{ik} \) denotes the travel time from origin \( i \) to hub \( k \), and \( t_{\ell j} \) is the sum of the times of the other legs of the trip, and \( t_{k\ell} = t_{\ell j} \) are the (fixed) times spent at the respective airports, including check-in time and waiting time at the luggage belt. A similar argument is applied to the fare between origin \( i \) and destination \( j \), given that the route leads through hubs \( k \) and \( \ell \) and is offered by airline \( s \). Formally, the fare is defined as \( c_{iks} = c_{k\ell} + c_{\ell j} + c_{ij} \), where \( s \) is the discount factor that accounts for the lower fares offered between hubs, and \( c_{k\ell} \) is the fare that airline \( s \) charges for a (non hub-to-hub) trip between \( i \) and \( j \) and \( c_{ij} \) is the remaining legs of the trip. For simplicity, we will assume that air fares are proportional to the costs incurred by the airlines. Note that there are two possible routes between nodes \( i \) and \( j \) through a pair of hubs \( k \) and \( \ell \), namely the route \( i-k-\ell-j \) and the route \( i-\ell-k-j \). We assume that at most one of these is open by each airline: the one with the highest utility for customers.

Using the proportional model, the probability of a customer traveling from \( i \) to \( j \) via hubs \( k \) and \( \ell \) with airline \( s \) can then be expressed as

\[
p_{ijkts} = \frac{u_{ijkts}}{\sum_{m,n,\text{Comp}} u_{ijmns} + \sum_{i \neq j} s \sum_{m,n,\text{Comp}} u_{ijmns}}
\]

The proportion of passengers captured by a route and the airline \( o \) is then equal to this probability \( p_{ijkts} \). Thus the total capture of passengers using the route \((i, k, \ell, j)\) offered by airline \( o \) is \( h_{ij} \cdot p_{ijkts} \).

Consider now the problem the new airline faces. Its objective is to capture as large a market share as possible, by locating a fixed number \( p \) of hubs. Since this number is typically very small, we can solve the problem for, say, \( p = 1,2,\ldots,5 \). We assume that the airline is aware of customer behavior as described above. This model is an extension of the work by Marianov et al. [2], who assume that all passengers at origin \( i \) who want to travel to destination \( j \) will all choose the cheapest route. Our present work extends this research in two directions: firstly, by using attraction functions we allow factors other than price being used for the selection of the route and airline, and secondly, we do away with the “winner-take-all” assumption according to which all customers who want to travel from \( i \) to \( j \) will always choose the same route.

In order to formulate the proportional competitive hub location model, we first define binary location variables \( y_{k} \), which assume a value of one if a hub is located at node \( k \), and zero otherwise. Furthermore, we define a binary hub-pair location variable \( w_{k\ell} \), which equals 1, if hubs are located at both, nodes \( k \) and \( \ell \).
Recalling that $h_{ij}$ is the demand for air traffic from nodes $i$ to $j$, the model can then be written as

$$P: \text{Max } z = \sum_{ij} \sum_{k \in \text{Own}} h_{ij}y_{jkl}$$

$$= \sum_{ij} h_{ij} \frac{\sum_{k \in \text{Own}} W_{kl}y_{jkl}}{\sum_{j \in \text{Own}} W_{ij}y_{jkl}} \sum_{k \in \text{Own}} W_{kl}y_{jkl}$$

s.t. $w_{k\ell} \leq y_k \quad \forall k, \ell \in \text{Own}$

$$w_{k\ell} \leq y_{k\ell} \quad \forall k, \ell \in \text{Own}$$

$$\sum_k y_k = 1 \quad \forall \ell$$

$$y_k, W_{k\ell} \in [0,1] \quad \forall k, \ell$$

The objective (3) maximizes the demand captured by routes belonging to the entering airline, given that there are known competitors’ hub locations. Constraints (4) and (5) prevent two-hub variables $w_{k\ell}$ taking the value 1, unless there are hubs located at both $k$ and $\ell$. Constraint (6) establishes the number of hubs to be located and constraint (7) forces integrality of the variables.

It is worth noting that $y_k = y_{1\ell} = 1$ does not guarantee $w_{k\ell} = 1$ per se. However, as all utilities in the model are positive, increasing the variable $w_{k\ell}$ from 0 to 1, whenever possible, increases the value of the objective function as well, so that it is guaranteed that whenever $y_k = y_{1\ell} = 1$, $w_{k\ell} = 1$ will follow. Furthermore, we would also like to point out that for each $k, \ell$, constraints (4) and (5) could be combined in the single constraint $w_{k\ell} \leq \frac{1}{2}(y_k + y_{1\ell})$. Such aggregation of constraints is, however, not advantageous from a computational point of view.

4. Solution methods

The problem described in the previous section is nonlinear and integer, which precludes using standard integer programming software. Among the many options available as far as heuristic methods are concerned, heuristic concentration is a procedure that has shown good results when a fixed number of facilities is to be located and is very easy to implement. The principle of heuristic concentration was first described by Rosing and ReVelle [35] and later expanded in Rosing et al. [36].

This two-phase metaheuristic uses a low complexity method in phase 1 to restrict the number of candidate locations from the original $n$ to some (much) smaller number $m$ and applies either an exact method or an improvement heuristic in phase 2. The purpose of the concentration of the working set in phase 1 is to allow phase 2 to work with this smaller set that can much more easily be dealt with than the original larger set. Furthermore, the concentrated set generated in phase 1 is not a random set but a set of locations that includes locations that are, at least, local optimas.

Several different versions of heuristic concentration have been described in [35–40]. The specific version we use here applies a multistart procedure in phase 1 that uses the vertex substitution algorithm of Teitz and Bart [41] $q_1$ times, where $q_1$ is a prespecified number. Each of the $q_1$ runs starts with a different set of random facility locations. One by one, facilities in the set are moved to unused candidate locations and the objective is computed at every such move. Whenever such a move improves the objective function, the solution is saved as the best currently available. The process continues until none of the possible moves improves the objective. After all $q_1$ runs have been completed, the locations belonging to the best $b$ solutions are saved in a “concentration set,” where $b \leq q_1$ is a number preset by the analyst. At this point, the concentration set contains only a small subset of the original set of candidate locations.

The second phase of the heuristic performs $q_2$ runs of a two-vertex substitution heuristic. As in phase 1, each run in phase 2 starts with a randomly found initial set of $p$ locations, where $p$ is the number of hubs to be located. In this phase, the candidate locations are only those contained in the concentration set, of cardinality $m$, such that $p \leq m \leq bp$. The two-vertex substitution heuristic applied at each run is analogous to the one-vertex substitution heuristic of the first phase, but instead of moving one facility at a time, two facilities are moved at a time from their present location to any pair of unused locations. The procedure terminates when no move of any two vertices from their present locations to any currently unused locations in the concentration set results in a better value of the objective function.

As other approximation procedures, heuristic concentration does not require linearity of the objective. Even if the objective is linear, the model is reformulated in such a way that most constraints are embedded in the objective, which turns out to be nonlinear, see e.g., [40]. This formulation contains only the objective and constraint (6), making it a nonlinear knapsack problem. In our case, the inequality constraints (4) and (5) in problem $P$ are replaced by the single equation

$$w_{k\ell} = \min(y_k, y_{1\ell}) \quad \forall k, \ell \in \text{Own}$$

Using relation (8), the binary variable $w_{k\ell}$ is replaced in the objective, leading to the new formulation

$$P_1: \text{Max } z = \sum_{ij} h_{ij} \frac{\sum_{k \in \text{Own}} W_{ij} \min(y_k, y_{1\ell})}{\sum_{j \in \text{Own}} W_{ij} \min(y_k, y_{1\ell}) + \sum_{k \in \text{Own}} W_{ij} \min(y_k, y_{1\ell})}$$

s.t. $\sum_j y_j = p$

$$y_k \in [0,1] \quad \forall k$$

In both phases of the heuristic procedure, the objective (9) is evaluated for every set of locations resulting after each one facility or two-facilities move. If for a new set of locations the objective is better than the present value of the objective function, the new set is saved as the best set of locations found so far in the procedure.

5. Computational experience

The first data set used in this study is the 25-node version of the network for the Australian Post data, see, [42,43]. Given the many parameters that could be modified, we focus on customer behavior, so that we assume that there are only two competitors, them and us. We assume that they are assumed to operate between three and six hubs, while we are planning to enter the market with 2, 3, 4, 5 and 6 hubs, respectively. For the purpose of numerical testing, we will assume below that the travel times, waiting times and costs are the same for all airlines, given the same route from $i$ to $j$ via $k$ and $\ell$. The travel times over each leg were taken as proportional to distances. Also, attractivenesses are identical across all 2-hub routes. The 1-hub routes have an attractiveness that is 25% higher than 2-hub routes. Throughout the computations, we vary the value of $\omega$ (recall that $1-\omega$ expresses the discount for interhub transportation) and $\gamma$, the relative weight that measures the importance of travel time as opposed to travel costs, weighted by $(1-\omega)$. The travel time (equal for all airlines) was computed as $t_{ij} = t_{ik} + t_{k\ell} + t_{\ell j}$. The costs were simply computed as $c_{ij} = c(t_{ij} + 2t_{k\ell} + t_{\ell j})$. The same values of $\gamma$ and $x$ were used for both the entering firm and the incumbent airline (or airlines).
hub locations are optimal for the hub locations. (b) The second phase uses the heuristic concentration procedure are as follows. The first run of 10 random starts. Locations that appear at least once in any of the solutions of these runs are included in the concentration set, i.e., \( q_1 = 10 \) random starts. Locations that appear at least once in any of the solutions of these runs are included in the concentration set, i.e., \( q_1 = 10 \), the second phase uses \( q_2 = 5 \) runs, whose initial sets of locations are chosen at random from the concentration set.

The average run times were 5 seconds for three hubs, 11 seconds for four hubs, 18 seconds for five hubs and 29s for six hubs. The variance of the run times was small, and it depended on the number of hubs to be located.

First, we wanted to determine if there were any effects of the discount factor \( z \) and the parameter \( \gamma \) that measures the relative importance of the travel time (as opposed to the fare). For that purpose, we consider the case in which \( r = 5 \) competitor's hubs are already located on the market, while we also locate \( p = 5 \) facilities. Two cases were considered: in the first case, competitor's hubs are located at random, while in the second case, the competitor locates his hubs after optimally solving the \( p \)-hub median problem.

It is apparent from Figs. 1a and b that regardless of whether or not the leaders have optimized their hub locations, the follower has a strong advantage: in none of the scenarios does the follower capture less than 55% of the entire market, and our capture reaches 70% in one scenario. And this is achieved by a market strategy that has the follower locate the same number of hubs as the leader. Secondly, comparing the two figures, it becomes evident that, as expected, optimized leader locations will negatively impact on the market capture of the follower. We can also observe that if \( \gamma = 1 \) (i.e., costs are immaterial to the users, only time is considered relevant), then the value of \( z \) does not matter, as interhub discounts only affect the costs, but not time. This is the reason why the functions in Figs. 1a and b converge for all values of \( z \) at \( \gamma = 1 \). On the other hand, if only cost matters (i.e., \( \gamma = 0 \)), different values of \( z \) should result in major differences between the solution, which is shown in the two companion figures. Also, as \( \gamma \) decreases, the market capture becomes more dependent on the cost \( c_{rj} = c_{rj1} + c_{rj2} + c_{rj3} \). All the components of the cost \( (c_{rj1}, c_{rj2}, c_{rj3}) \) can be optimized by a good location, and, more so, when the value of \( z \) is large: if \( z = 1 \), there are three components to optimize and a higher degree of freedom, while if \( z = 0 \), there are only two components corresponding to the travel times of the origin and destination nodes to their closest hubs. To the contrary, when \( \gamma = 1 \), waiting times at the airports cannot be optimized by moving locations. As a consequence, in general, the entering airline should always be able to capture a larger market when \( \gamma \) is closer to 0, i.e., the curves should be decreasing with \( \gamma \). Only when \( z \) is small and for certain locations of the leader, the curves deviate from this behavior, as those in Fig. 1b for \( z = 0.2 \) and 0.4. This is due to the fact that the \( p \)-hub median solution (2.1, 8, 17, 18, 20), adopted by the leader, is very close to the optimal solution of the capture problem solved by the follower (8, 17, 18, 19, 23), when \( z \) approaches 0.

Figs. 2a and b plot the percentage of traffic captured by the entering airline’s 1-hub routes (“our 1-hub”) and the total traffic captured by 1-hub routes of all airlines (“total 1-hub”), against \( \gamma \), given values of \( z = 0.2 \) and 0.8, respectively. Fig. 2a shows that the total traffic captured by 1-hub routes increases as customers give more importance to the travel time (i.e., for increasing values of \( \gamma \)), due to the fact that 2-hub routes, regardless who offers them, require extra waiting time at a second airport and longer average flying time, making 2-hub routes a poor choice for time-conscious passengers. On the other hand, 2-hub routes may offer cheaper deals that are of obvious interest to fare-sensitive customers. Comparing Figs. 2a and b, it is evident that this tendency is stronger for larger economies of scale (i.e., for smaller values of \( z \)), because for fare-oriented customers (i.e., customer behavior that is described by small values of \( \gamma \)), large economies of scale mean a larger advantage of lower fares when choosing 2-hub routes. Again, for \( z \rightarrow 1 \), there is no interhub discount, and for Euclidean distances, virtually all traffic would normally be directed through a single hub. The fact that it is not so in this example is because we are not using a binary choice model (winner takes all). Proportional models of the type used in this study do not only model customer behavior more accurately, but also behave much more smoothly.

It is to be expected that the traffic captured by us depends not only on the locations of our hubs but also on the locations of the competitor’s hubs. Table 1 shows the differences in our captured traffic, given that we locate three hubs, for different, randomly selected, locations of their four hubs. Note that the capture differs by a factor of 15.65%. In the last row, the competitor’s locations are chosen in such a way as to result in co-location.

If the locations of the competitor’s hubs were optimally chosen by solving the leader problem (i.e., locate optimally given that a follower will locate the same number of hubs), locating the same number of hubs as the competitor reports a capture of 50% of the total traffic. In this situation, entering hubs are co-located with those already open.

For fixed values of \( z \) (economies of scale) and \( \gamma \) (weight on flying time), our captured traffic is determined by the relative numbers of competitor’s and entering airline’s hubs. Fig. 3 shows our capture as a function of the number of our and their hubs and different values.

Fig. 1. (a) Our market capture for \( r = p = 5 \), 25-node network, random competitor’s hub locations. (b) Our market capture for \( r = p = 5 \), 25-node network, competitor’s hub locations are optimal for the \( p \)-hub median problem.
The dotted line indicates the capture if \( p = r \), i.e., they locate the same number of hubs as we do.

For \( \gamma = 0.5 \), Table 2 reports the average and standard deviation of the marginal capture as an additional hub is activated by the entering airline, averaged over \( x = 0.2, 0.6 \) and 1. For example, given that they have located four hubs and we locate a third hub, our additional capture increases by 12.73%. Or, on average, when we locate a fifth hub, our marginal capture is decreasing as we located more facilities. Note the small standard deviation in all cases.

Similarly, assuming that our competitor(s) have located 4 hubs and the interhub discount factor is \( \gamma = 0.6 \), we have made runs for different values of \( \gamma \), Table 3 indicates the average increase (and standard deviation) of our capture in these cases. There is not much variation in the figures, indicating that \( \gamma \) is not much of a factor in the benefits of an additional hub.

Our second data set is the 50-node version of the network for the Australian Post data. There is only one firm in the market, with 5 hubs. The competitor’s hubs are located by solving a \( p \)-hub median problem that results in locations at nodes 4, 14, 28, 32, and 35. The number of starts of the first and second stages of the algorithm are now \( q_1 = 5 \) and \( q_2 = 3 \). Interestingly, in all cases, the solutions were the same for all starts in both phases of the algorithm. The time needed for each instance was 45 seconds, and only the combinations of \( x = 0.4 \) or 0.2 and \( \gamma = 0 \) required 58 seconds.

Our market capture is plotted in Fig. 4 for different values of \( x \) and \( \gamma \).

As Fig. 4 shows, the advantage of the follower is kept in this larger network. Note the strong similarity of Figs. 2 and 4. Different best locational strategies for different conditions can be seen in Table 4.
Table 4 shows that the solutions are very robust, and most situations are dealt with the same locations. These locations change when the values of and are small, i.e., when customers choose lower fares and when the discount in inter-hub links is important. Still, in these cases, some locations are retained (35, 36, 38).

### 6. Summary and outlook

This paper has considered the competitive hub location problem with the feature that the customer choice function is probabilistic and uses a gravity model as a utility function. The problem was formulated and a version of the principle of heuristic concentration was described for its solution. A series of tests were conducted given a well-known data set. The main results are:

- problems of realistic size can be solved,
- the results appear quite robust considering both different initial guesses for solutions, and different locations of the competitors and
- the solution technique is quite flexible so as to be able to accommodate different utility functions or other features.

Future work in this field could include different utility functions. A popular class of models is logit functions that may be investigated. It future work in this field could include different utility functions. A popular class of models is logit functions that may be investigated. It

### Acknowledgments

This research was in part supported by Grant #9160 from the Natural Sciences and Engineering Research Council of Canada, by the Instituto Milenio "Complex Engineering Systems," and by FONDECYT Grant 1070741. This support is gratefully acknowledged. The authors would also like to thank the anonymous referees for their insightful comments. Their remarks have helped to clarify a number of issues in the paper.

### References


