CrossMark

# Spatial Analysis of Single Allocation Hub Location Problems

Meltem Peker[1] · Bahar Y. Kara[1] ·
James F. Campbell[2] · Sibel A. Alumur[3]

**Abstract** Hubs are special facilities that serve as switching, transshipment and sorting nodes in many-to-many distribution systems. Flow is consolidated at hubs to exploit economies of scale and to reduce transportation costs between hubs. In this article, we first identify general features of optimal hub locations for single allocation hub location problems based on only the fundamental problem data (demand for travel and spatial locations). We then exploit this knowledge to develop a straightforward heuristic methodology based on spatial proximity of nodes, dispersion and measures of node importance to delineate subsets of nodes likely to contain optimal hubs. We then develop constraints for these subsets for use in mathematical programming formulations to solve hub location problems. Our methodology can also help narrow an organization's focus to concentrate on more detailed and qualitative analyses of promising potential hub locations. Results document the value of including both demand magnitude and centrality in measuring node importance and the relevant tradeoffs in solution quality and time.

**Keywords** Hub location problem · Single allocation · Spatial distribution · Clustering nodes

## 1 Introduction

Hubs are special facilities that act as switching, transshipment and sorting nodes in many large transportation and telecommunication networks. Rather than having direct links for each origin–destination (*o-d*) pair, hub networks use fewer links to connect the origins and destinations, and thereby concentrate flows to allow economies of scale.

✉ Bahar Y. Kara
   bkara@bilkent.edu.tr

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

2   College of Business Administration, University of Missouri-St. Louis, St. Louis, MO 63121, USA

3   Department of Management Sciences, University of Waterloo, Waterloo, ON N2L 3G1, Canada

🍃 Springer

Hub networks are designed to serve demand for movement (e.g., transportation of freight or passengers) from specified origins to specified destinations. Transportation hub location problems are concerned with locating the hub facilities, generally with the aim of minimizing the total costs for movement of all flows from origins to destinations. Solutions also require allocating demand nodes to hubs in order to route traffic between the *o-d* pairs. The first goal of our research is to better understand the characteristics of good transportation hub locations in order to predict likely optimal hub locations based only on the fundamental problem data (demand for travel and spatial locations). Thus, rather than analyzing existing hub networks and traffic flows to identify or classify hubs, we seek to identify locations (e.g., cities) likely to be optimal hubs prior to designing the network. Our second goal is then to exploit this knowledge of promising hub locations in a heuristic solution methodology to better solve hub location problems.

Much research in the past 25 years has focused on solving fundamental hub location problems, where the hub network is complete, traffic for each *o-d* pair is routed through at least one hub, and the cost between two hubs is discounted due to the consolidation of flows (see O'Kelly and Miller 1994; Campbell et al. 2002; Alumur and Kara 2008, and Campbell and O'Kelly 2012). We focus on the original hub location problem introduced by O'Kelly (1986a, 1986b, 1987): the uncapacitated single allocation *p*-hub median problem (USApHMP). In the USApHMP, *p* hub nodes must be located, each non-hub node must be assigned to one of the *p* hubs, and the objective is to minimize the total transportation cost to serve given *o-d* flows, where the cost rate for inter-hub flows is discounted by the economies of scale factor $\alpha$ ($0 \leq \alpha \leq 1$). With single allocation, the incoming and outgoing flow of each node is routed through a single hub. (In contrast, with multiple allocation problems, demand nodes can be assigned to more than one hub.)

The remainder of the paper is organized as follows: Section 2 describes the motivation and goals for the research. Section 3 describes the analysis to identify key characteristics of optimal hub locations and Section 4 explains the conversion of these into methodologies to identify subsets of nodes likely to contain hubs. Section 5 presents computational results of the methodologies over a wide range of data sets. Section 6 is a discussion and conclusion.

## 2 Motivation and Goals

The motivation for our research is the desire to better understand the factors that create optimal hub locations, with an eye towards using that knowledge to better solve hub location problems. Theoretical models of strategic hub location are well studied (e.g., Alumur and Kara 2008; Campbell and O'Kelly 2012). Many extensions to the basic hub location problems have been addressed, including fixed or flow dependent costs (e.g., O'Kelly 1992; Bryan 1998), reliability and resilience (e.g., Kim and O'Kelly 2009; Parvaresh et al. 2013; O'Kelly 2014), multiple products (Correia et al. 2014), and price sensitive demands (O'Kelly et al. 2014). Algorithmic advances have included better methods for finding optimal solutions (e.g., Contreras et al. 2011 and Sa et al. 2013) and combinations of optimization and simulation (Vidovic et al. 2011). Recent work that explored details of operations at hubs, including optimizing flows through hubs, includes Chen (2010) and O'Kelly (2014). In this paper we focus on USApHMP which lies at the heart of much hub location research.

Hub location and network design models necessarily abstract key features from the practical problems and may ignore important aspects of the underlying system (as noted in Campbell and O'Kelly 2012 and Campbell 2013). Further, optimal solutions of a model are only one factor in facility location decisions, so identifying good potential hub locations based on fundamental problem data is valuable to focus more practical and qualitative analyses. See Bowen (2012) for an interesting discussion of practical issues for FedEx and UPS hubs, including weather, trucking and access to highways, personal connections and favorable leasing arrangements. Because large hub location models lead to very difficult optimization problems, finding a small set of good potential hubs may also help speed solution of hub location models.

While the ultimate goal of our research is the design of a hub network, including the location of hubs, to serve a given demand, spatial analysis of the flows and hub locations in existing transportation networks provides useful insights. The concepts of connectivity and centrality (and related measures) have a long and rich history in analysis of social, communication and transportation networks (e.g., Garrison 1960; Kissling 1969; Freeman 1978). More recently, Rodríguez-Déniz et al. (2013) highlight the roles of traffic generation and connectivity in distinguishing types of hubs. The traffic generation capability of a node or region is often measured based on demand (e.g., passengers originating and destined for a city) or economic measures. Martin and Voltes-Dorta (2008) explore the implications of spatial concentration of demand and passenger connections on hubbing. Yu et al. (2013) consider transit hub location in China using an approach to identify candidate hub locations based on "passenger attraction", which includes the magnitude of demand within a specified distance of a node. (A similar idea of aggregating demand in the neighborhood of a node was applied to the $p$-median problem by Hillsman 1980 and Sorensen and Church 1995.) Yu et al. (2013) also provides an incentive for the separation of hubs using a procedure to eliminate overlapping service to avoid double counting demand (transit riders). More generally, the concept of dispersion in facility location modeling (Kuby 1987) is often included as one of possibly several objectives (e.g., Kim and O'Kelly 2009; Maliszewski et al. 2012).

Because in our research we seek promising hub locations prior to design of the hub network, we cannot use network measures such as flight frequencies, passenger flows, travel paths or node degrees that are calculated based on an established network and its operations. However, the concept of connectivity has been measured in a variety of ways, such as centrality, intermediacy and "betweenness" (see Rodríguez-Déniz et al. 2013) and locational attributes of these measures exist for a set of cities independent of how they are connected via a particular network (see for example, Fleming and Hayuth 1994; Bowen 2012 and Maertens et al. 2014). The discussion above highlights the important roles of the demand magnitude (e.g., city size or passenger enplanements) and geography (i.e., measures of relative node locations) for identifying good potential hub locations. Note also that proximity in hub location problems is different than in "regular" facility location problems, as the flow between two nearby nodes may actually travel a long distance (at high expense) if the assigned hub for both nodes is not close to them. This complicates finding good hub locations, as the unit of demand is $o$-$d$ pairs, rather than individual nodes. Our research takes as input only the origin–destination locations and the demand for travel between $o$-$d$ pairs. Aggregating the travel demand provides the basis for demand at a particular city. While the effect of traffic generation from the network design and hub locations can be important (see for example, O'Kelly 2010 and Rodríguez-Déniz et al. 2013), that is beyond

the scope of this research. By using fundamental properties based on the cities' demand and location, we propose a model that provides useful insights on the spatial aspects of hub location. Our specific approach is to identify general characteristics of optimal (or near-optimal) hub locations in terms of the input data, which can be exploited using parsimonious and straightforward methods to define small subsets of nodes likely to contain the optimal hubs. This differs from the heuristic concentration method presented by Rosing and ReVelle (1997) for facility location problems, where a "good" set of locations is identified as those that most frequently appear in the heuristic solutions of the problem. In contrast, our methodology identifies a set of "good" locations *prior* to solving the problem and by using only the problem data.

We use our methodology to explore the benefits of confining the search for optimal hubs to small subsets with extensive computational experiments on benchmark hub location data sets. When the size of a problem is large, obtaining optimal hub locations might not be easy; and in such cases, the methodology developed in this paper may be useful. Furthermore, because mathematical models for locating facilities necessarily can include only a subset of the key issues and objectives in any real-world problem, our methodology can help narrow an organization's focus to concentrate more detailed and qualitative analyses on promising potential hub locations. This theme in the spirit of recent multi-criteria location research (e.g., Maliszewski et al. 2012; Batta et al. 2014) that advocates modeling as a "prescriptive aid for decision makers to help generate/assess a number of good quality solutions that they can choose from" (Batta et al. 2014, p. 828).

## 3 Observations from Optimal Solutions

To explore the general characteristics of good hub locations, we first analyzed the spatial distribution of the optimal hub locations for the USApHMP with three customary data sets: (i) the CAB data set of continental-scale air passenger traffic between 25 cities in the USA (O'Kelly 1986b), (ii) the AP20 data set for postal distribution among 20 nodes in metropolitan Sydney, Australia (Ernst and Krishnamoorthy 1996), and (iii) the TR81 data set of cargo flows between 81 cities in Turkey (Tan and Kara 2007). The cities in the CAB data set are shown in Fig. 1.

We varied the number of hubs and the level of cost discount for travel between hubs. Although optimal locations for many of the instances are available in the literature, we solved them again for comparison purposes using the formulation given in Ernst and Krishnamoorthy (1996) and including the missing cuts presented in Correia et al. (2010).

The mathematical formulation of USApHMP uses a given node set $N$, representing geographic locations (e.g., cities). The decision variable $X^{ik}$ takes the value 1 if node $i$ is allocated to hub $k$ and 0 otherwise, and $Y^i_{kl}$ is the total flow that originates at node $i$ and is distributed to demand nodes after visiting hubs $k$ and $l$ respectively. Parameter $\alpha$ is the transportation rate (economies of scale discount factor) for transferring flow between two hubs, $\gamma$ is the transportation rate for collection from an origin to a hub, and $\delta$ is the transportation rate for distribution from a hub to a destination, where $\alpha \leq \delta$ and $\alpha \leq \gamma$. The given parameter $d^{ik}$ is the distance between nodes $i$ and $k$ and $w^{ij}$ is the flow from origin $i$ to destination $j$. Let $O_i = \sum_{j \in N} w_{ij}$ be the total amount flow that
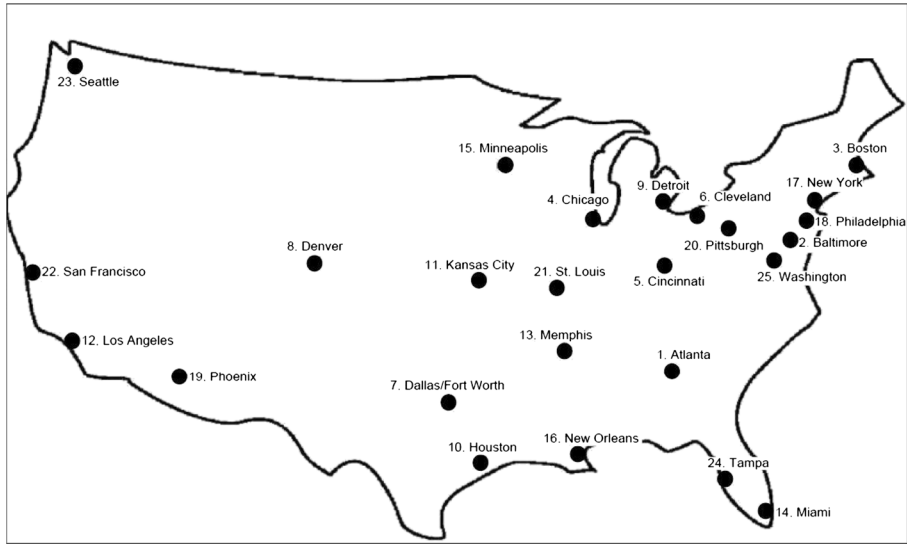
**Fig. 1** Locations of 25 cities of USA in the CAB data set

originates at node $i$ and let $D_i = \sum_{j \in N} w_{ji}$ be the total amount flow that is destined to node $i$. The number of hubs to locate is $p$. The formulation for the USApHMP is as follows:

$$\min \sum_{i \in N} \sum_{k \in N} d_{ik} X_{ik} (\gamma O_i + \delta D_i) + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \alpha d_{kl} Y_{kl}^i$$

$$\text{s.t} \tag{1}$$

$$\sum_{k \in N} X_{ik} = 1 \ \forall i \in N$$

$$\sum_{k \in N} X_{kk} = p \tag{2}$$

$$X_{ik} \leq X_{kk} \ \ \forall i, k \in N \tag{3}$$

$$\sum_{l \in N} Y_{kl}^i - \sum_{l \in N} Y_{lk}^i = O_i X_{ik} - \sum_{j \in N} w_{ij} X_{jk} \forall i, k \in N \tag{4}$$

$$\sum_{l \in N} Y_{kl}^i \leq O_i X_{ik} \ \forall i, k \in N : \ i \neq k \tag{5}$$

$$X_{ik} \in \{0, 1\} \ \forall i, k \in N \tag{6}$$

$$Y_{kl}^i \geq 0 \ \forall i, k, l \in N \tag{7}$$

The objective minimizes the total transportation cost for collection, distribution and inter-hub transfer. Constraints (1) ensure all nodes are allocated to a single hub. Constraint (2) ensures exactly $p$ hubs are located. Constraints (3) ensure a hub is opened if a node is allocated to it. Constraints (4) are the flow balance constraints. Constraints (5) link the allocation decisions to the flow variables. Constraints (6) and (7) establish the variable domains.

To identify general characteristics of the optimal hub locations, we sought general patterns from the optimum hub locations for the CAB, AP20 and TR81 data sets with $p$ ranging from 1 to 10 and $\alpha$=0.2, 0.4, 0.6 and 0.8 (with $\gamma=\delta=1$ for CAB and TR81, $\gamma=3$, $\delta=2$ for AP20). We discuss results for the CAB data set below but similar findings hold true with the AP20 and TR81 data sets. (Full results are available from the authors.) The optimum hub locations were obtained by solving the above model using CPLEX version 12.4 using a 4xAMD Opteron Interlagos 2.6GHz with 96 GB RAM. The results are shown in Table 1.

It is apparent in Table 1 that optimal solutions do not vary greatly with $\alpha$, as the same hubs are optimal for all four $\alpha$ values for $p$=1,2 and 10; and except for $p$=3 and 4, the same hubs are optimal for $\alpha$=0.2, 0.4 and 0.6. Even with the differences for $p$=3 and 4, there are strong geographical patterns to the optimal hub locations, where one of the nearby northeastern cities New York (17), Philadelphia (18) and Washington D.C. (25) is always a hub, and for $p$=4 one southern city, Atlanta (1) or Tampa (24), is a hub. Furthermore, optimal hub locations are insensitive to small changes in $\alpha$. Thus, on average with CAB, 94.4 % of the hubs stay the same with "neighboring" $\alpha$ values (e.g., compare locations for $\alpha$=0.2 and 0.4, for $\alpha$=0.4 and 0.6, for $\alpha$=0.6 and 0.8). To further investigate the influence of $\alpha$, we computed the cost for CAB with a particular value of $\alpha$, while using the optimal hub locations for the neighboring $\alpha$ value $\alpha'$, where $\alpha'=\alpha+0.2$ or $\alpha'=\alpha-0.2$. The average cost increase from using the optimal hub locations for the "neighboring" $\alpha$ value is less than 1 %. This insensitivity to $\alpha$ may be quite important in practice as the actual value of $\alpha$ is not likely to be known with certainty and may change

**Table 1**  Optimal hub locations for the CAB data with different values of the $p$ and $\alpha$

| $p$ | $\alpha$ | Optimal hub locations |
|---|---|---|
| 1 | 0.2–0.8 | 5 |
| 2 | 0.2–0.8 | 12,20 |
| 3 | 0.2 | 4,12,17 |
|  | 0.4 | 4,12,18 |
|  | 0.6–0.8 | 2,4,12 |
| 4 | 0.2 | 4,12,17,24 |
|  | 0.4–0.6 | 1,4,12,17 |
|  | 0.8 | 1,4,12,18 |
| 5 | 0.2–0.6 | 4,7,12,14,17 |
|  | 0.8 | 1,4,7,12,18 |
| 6 | 0.2–0.6 | 4,6,7,12,14,17 |
|  | 0.8 | 1,4,6,7,12,17 |
| 7 | 0.2–0.6 | 4,6,7,12,14,17,22 |
|  | 0.8 | 1,4,6,7,12,17,25 |
| 8 | 0.2–0.6 | 1,4,6,7,12,14,17,22 |
|  | 0.8 | 1,4,6,7,8,12,17,25 |
| 9 | 0.2–0.6 | 1,4,6,7,8,12,14,17,22 |
|  | 0.8 | 1,4,6,7,8,12,17,22,25 |
| 10 | 0.2–0.8 | 1,4,6,7,8,12,14,17,22,25 |

over time as the vehicle types, modes and fleet mix change. See O'Kelly and Lao (1991) for some discussion of using multiple modes in hub networks. Another observation from Table 1 (and the related results for the other data sets) is that with larger numbers of hubs, in most cases the optimal locations with $p$ hubs are also optimal locations for $p+1$ hubs. This is also common in small $p$-median problems, but is not true in general.

Another important aspect of the optimal hub locations stems from the magnitude of the demand at the nodes. With the CAB data set, the largest nodes in terms of total demand ($O_i+D_i$) are New York (17), Chicago (4), and Los Angeles (12) with 17.0, 10.0 and 7.3 % of the total demand, respectively. These cities appear as optimal hub locations in 75, 89 and 100 % of the results for $p=2$–10 in Table 1. (Note that in several solutions, Philadelphia (18) is used instead of New York, but they are never both optimal.) Interestingly, Boston, which is the fourth largest node in the CAB data set with 6.1 % of the total demand, is never a hub as it is both too peripheral and too close to the largest node at New York. Results with the AP20 and TR81 data sets are similar. The frequency of certain large cities being hubs is not perhaps surprising given the large disparity in city sizes (Nitsch 2005).

Summarizing our findings, coupled with knowledge of the geographic position of the nodes, we derived five general observations:

(1) larger demand nodes tend to be selected as hubs, especially those with greater distances to other cities, though smaller demand nodes in close proximity to large demand nodes may be chosen,
(2) among a set of "nearby" nodes, usually at most one of them becomes a hub,
(3) some cities almost never appear as hubs in optimal solutions,
(4) optimal hub locations are rather insensitive to $\alpha$ and
(5) hubs tend to be dispersed across the service region.

Translating these observations into tools for solving hub location problems is difficult as these may overlap and contradict each other for particular nodes (cities). Also, note that some of these observations reflect ideas shared with other facility location problems.

While these five observations show that optimal hub locations are strongly influenced by both the magnitude of demands at nodes and the spatial distribution of the nodes, the challenge is to capture these ideas in a algorithmic format. We chose to focus our efforts on three key areas: the importance of a node, the proximity of a node to other nodes, and the dispersion of hubs. These three areas capture the ideas in observations 1, 2 and 5, and part of observation 3 via dispersion. However, the issue of node importance requires further elaboration and analyses in the following sections, where we explore several alternative forms of node importance that integrate the magnitude of demand with concepts of betweenness and centrality.

## 4 Identifying Subsets Likely to Contain Optimal Hubs

Based on the analysis of optimal hub locations for a wide range of instances, we developed a methodology to define small subsets of nodes that seem likely to contain hubs. These subsets can be useful to identify hubs (cities) and geographic regions to be

targeted for more detailed analyses. They can also be useful to define constraints for mathematical programming formulations that require at least one hub to be selected from each set. The challenge is to define subsets that contain relatively few nodes, but are likely to include one or more hubs.

## 4.1 Clustering-Based Potential Hub Sets (CBS)

The main idea behind our methodology is to identify clusters of nodes using "hub circles" centered at important nodes and with a radius based on the proximity to other important nodes. Each hub circle then will host at least one hub. The sizes of the hub circles, and the subsets derived from the collection of hub circles, are defined using a methodology based on a large number of experiments with the data sets discussed in Section 3. The use of hub circles to identify clusters of demand nodes is similar to some aspects of the approaches in Vidovic et al. (2011) and Yu et al. (2013). A related idea is also utilized in Ernst and Krishnamoorthy (1998), where clusters of nodes are produced with a greedy merging process based on minimizing distances between nodes in a cluster. This is used to create either $p$ or $n/4$ clusters of nodes that contain hubs as a starting point for an exact solution procedure. In contrast to our approach, Ernst and Krishnamoorthy (1998) ignores the nodal demands in forming clusters. Another related work with clustering is Horner and O'Kelly (2005), which aggregates origins (destinations) into separate clusters that interact via a single origin (destination) location. In this work, the clusters are determined endogenously by solving a mathematical programming formulation of a hierarchical assignment problem.

We first describe the baseline algorithm for defining subsets, denoted "CBS" for Cluster-Based Subsets. Later we consider extensions of the CBS algorithm. The input data for our approach includes only the basic data: the number of hubs, $p$, to locate, a set of origin/destination nodes $N = \{1,2,3\ldots, n\}$, and the distances $d_{ij}$ and flows $w_{ij}$ (e.g., of passengers or freight) between nodes $i$ and $j$. The first step in CBS is to sort the nodes (potential hub locations) in decreasing order of importance, where we compare several measures of importance later in this section. The $2*p$ (abbreviated as $2p$) nodes with the largest importance form set $N^p$ and these "important nodes" are used as potential centers for hub circles. The remaining $n-2p$ nodes with the lowest importance values are termed "small nodes". The hub circles provide for a dispersion of the hubs, where the radius is the proximity measure $PM$, calculated as the average of the distances for each node in $N^p$ to the nearest of the other nodes in $N^p$. This can be interpreted as assigning each node in $N^p$ to the nearest other node in $N^p$, and taking the average distance of these assignments.

The CBS algorithm defines the subsets of nodes by iteratively considering the nodes in $N^p$ in decreasing order of importance and assigning each node in $N^p$ to either: (i) set $H_S$ if the node's hub circle contains no other nodes from $N^p$ (other than the center), (ii) set $H$ if the node's hub circle contains two or more nodes from $N^p$, or (iii) set $S_i$ if the node is within the hub circle for node $i$ of higher importance. Set $S_i$ also includes all nodes from $N$ that are within the hub circle for node $i$ to allow for less important nodes to be hubs (as long as they are near an important node in $N^p$). Formally, the CBS algorithm can be written

as follows, where the nodes are sorted in decreasing order of importance, $N^p$ is the sorted set of the *2p* most important nodes and $N^p[t]$ be the $t^{th}$ node in this ordered set.

**Algorithm CBS**:

0. $t = 1, \cdot H = \varnothing, \ H_S = \varnothing, \cdot S_i = \varnothing \ \forall \ i \in H$ and $PM = \dfrac{\sum_{i \in N^p} \left\{ \min\limits_{j \in N^p : i \neq j} d_{ij} \right\}}{2p}$

1. while ($t \leq 2p$) do {
    If ($N^p[t] \notin S_i \ \forall \ i \in H$) then {
        $S_i^0 = \{k | d_{ik} \leq PM \text{ and } k \in N^p\}$
        If ($|S_i^0| = 1$) then
            $H_S = H_S \cup \{i\}$
        If ($|S_i^0| > 1$) then
            $H = H \cup \{i\}$ and $S_i = \{k | d_{ik} \leq PM \text{ and } k \in N\}$}

2. $t = t + 1$}.

Set $H_S$ consists of isolated important nodes (i.e., there is no other node from $N^p$ within distance *PM* of these nodes) and set $S_i$ consists of a small group of neighboring nodes within distance *PM* of important node *i*. Set *H* consists of the hub circle centers for the sets $S_i$. (Note that set *H* cannot be empty and $H_S$ is empty only in the degenerate case where for every hub in $N^p$ the distance to nearest hub in $N^p$ is the same.) The subsets defined above can be used to define constraints to be added to the MILP formulation given in the previous section. Since we need to select *p* hub locations, we use one constraint derived from the set $H_S$ and (at most) *p-1* constraints for the subsets $S_i$ whose hub circle centers have the highest importance. Each constraint requires at least one hub to be selected from the associated subset. The formulation with the added constraints corresponding to the subsets can be written as follows.

$$\min \sum_{i \in N} \sum_{k \in N} d_{ik} X_{ik} (\gamma O_i + \delta D_i) + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \alpha d_{kl} Y_{kl}^i$$

s.t

(1)−(7)                                                                                          (8)

$$\sum_{k \in S_i} X_{kk} \geq 1 \ \forall i \in H$$

$$\sum_{k \in H_S} X_{kk} \geq 1$$                                                                (9)

Constraints (8) are for hub circles that contain two or more important nodes, and these help to disperse the hubs across the region. Constraint (9) also helps ensure spatial dispersion by ensuring at least one of the isolated important nodes is a hub. At most a total of *p* constraints are added. Note that solving the formulation above with constraints (1)–(9) provides a heuristic solution as the optimal solution with (1)–(9) may differ from that with constraints (1)–(7) alone.

## 4.2 Measuring Node Importance

A key aspect of the CBS algorithm is the definition of node importance, which as discussed earlier is based on both the magnitude of a node's demand and its relative geographic location. Results suggest that all things being equal, (i) nodes with greater demand are likely to be preferred as hubs, and (ii) nodes that are dispersed across the region (relative to the distribution of the demand), but not too peripheral, are likely to be preferred as hubs. Note that the demand in hub location problems is the flows between specified origin and destination nodes, so it has both a magnitude and a distance component. This is in contrast to demand in regular facility location problems (e.g., the $p$-median problem) that has only magnitude. For the magnitude of demand at node $i$ in our hub location problems we use a straightforward sum of all traffic that originates or terminates at the node: $O_i + D_i$. However, the concept of relative geographic location is more complex as it includes characteristics of centrality (and its converse "peripherality") and betweenness.

In this research, we evaluated a variety of measures of importance for a node $i$ using a straightforward approach based on the magnitude of its demand, its centrality and its "betweenness" as in Rodríguez-Déniz et al. (2013). Both centrality and betweenness are calculated from the Euclidean distances, and not based on any network. Centrality is measured by the sum of the Euclidean distances from node $i$ to all other nodes, $C_i = \sum_{j \in N} d_{ij}$, so more central nodes have smaller values of $C_i$. Our measure for betweenness is based on whether a node $k$ would be on the shortest one-stop path for traffic from origin $i$ to destination $j$ if the direct $i$-$j$ trip was not allowed. Thus, betweenness for node $k$ is the total demand that would use node $k$ as the intermediate hub (on path $i$-$k$-$j$):

$$B_k = \sum_{\substack{i,j:\text{k}=\operatorname*{argmin}_{m:\ m \neq i,j} \left\{ d_{im} + d_{mj} \right\}}} w_{ij}$$

We evaluated 14 different measures of importance, as shown in Table 2, which summarizes our experiments with the CAB data set. The second column of Table 2 is the importance measure for the nodes, $V_i$. The third column gives the percentage of the instances in which optimum hub locations are achieved using the CBS algorithm to identify subsets containing hubs. The fourth column is the percentage of instances in which there is at least one node chosen as a hub from the set of "small nodes" (i.e., not in set $N^p$). This column supports the observation in Section 2 that a smaller node close to an important node might be chosen as a hub. The fifth and sixth columns show the maximum and average gap in the objective function values (transportation cost) between the optimal solution and the solution using the CBS heuristic algorithm.

The first row of Table 2 measures node importance as the product of the magnitude of demand and centrality. The next row is similar, although the summation here is over the product of distance and demand for each node. This measure gives greater weight to large flows traveling long distances and is

**Table 2** Experiments with different measures of node importance for the CAB data set (with 16 instances where $p=3$–10 and $\alpha=0.6$ and 0.8)

| | Node importance measure, $V_i$ | Optimal solutions achieved (%) | Optimum from added rare hubs (%) | Maximum Gap (%) | Minimum Gap (%) |
|---|---|---|---|---|---|
| 1 | $(O_i+D_i)*C_i$ | 100.00 | 31.25 | 0.00 | 0.00 |
| 2 | $\sum\left[d_{ij}*(w_{ij}+w_{ji})\right]$ | 100.00 | 50.00 | 0.00 | 0.00 |
| 3 | $(O_i+D_i)/C_i$ | 50.00 | 12.50 | 14.77 | 0.69 |
| 4 | $(O_i+D_i)^2/C_i$ | 75.00 | 18.75 | 0.94 | 0.72 |
| 5 | $(O_i+D_i)/C_i^2$ | 12.50 | 0.00 | 5.44 | 0.69 |
| 6 | $C_i$ | 37.50 | 31.25 | 15.12 | 0.06 |
| 7 | $0.25\frac{(O_i+D_i)}{\sum_i(O_i+D_i)}+0.75\frac{C_i}{\sum_iC_i}$ | 37.50 | 31.25 | 15.12 | 0.06 |
| 8 | $0.5\frac{(O_i+D_i)}{\sum_i(O_i+D_i)}+0.5\frac{C_i}{\sum_iC_i}$ | 100.00 | 43.75 | 0.00 | 0.00 |
| 9 | $0.75\frac{(O_i+D_i)}{\sum_i(O_i+D_i)}+0.25\frac{C_i}{\sum_iC_i}$ | 100.00 | 18.75 | 0.00 | 0.00 |
| 10 | $O_i+D_i$ | 75.00 | 18.75 | 0.94 | 0.72 |
| 11 | Sets based on p-median results | 81.25 | – | 0.74 | 0.40 |
| 12 | $B_i$ | 12.50 | 68.75 | 7.66 | 0.00 |
| 13 | $(O_i+D_i)*B_i$ | 43.75 | 0.00 | 17.85 | 0.00 |
| 14 | $(O_i+D_i)*B_i*C_i$ | 62.50 | 18.75 | 3.80 | 0.00 |

actually the component of the objective function of hub median problems for node *i*. The next three rows of Table 2 provide reciprocal measures with the magnitude of demand in the numerator and centrality in the denominator, where squaring is used to increase emphasis on one component or the other. The next five measures can be viewed as specific cases of the weighted average of the relative magnitude of demand and the relative centrality. Row 6 of Table 2 puts all the weight on centrality, and row 10 puts all weight on the magnitude of demand. Row 11 of Table 2 provides results for comparison purposes using sets based on solving the *p*-median problem where the nodal demands are given by $O_i+D_i$ and the subsets are defined by the optimal allocation of nodes to facilities. (This is similar to the idea used in Horner and O'Kelly (2005), where all locations are considered as origins and as destinations.) The last three rows of Table 2 provide different measures using the betweenness of nodes. Every measure of importance includes the spatial influence of node locations through either centrality $C_i$ or betweenness $B_i$, except for rows 10 and 11, which rely simply on the total demand $(O_i+D_i)$.

The best results in Table 2 are achieved with the importance measures in rows 1, 2, 8 and 9, which find the optima in all instances. Measures with centrality in the denominator (rows 3–5) are not effective as they provide less dispersion and tend to favor more central hub locations. Measures that use the

magnitude of demand alone (rows 10 and 11) or betweenness alone (row 12) are ineffective, as are measures that focus primarily on centrality (rows 6 and 7). Measures with betweenness provide poor performance, most likely because betweenness in a multi-hub network is more properly a local measure influenced by the nodes allocated to each hub. The results in Table 2 show that the four measures in rows 1, 2, 8 and 9 perform best. Further testing with the CAB and AP data sets showed similar performance for the measures in rows 1, 8 and 9, with lesser performance for the measure in row 2. Therefore, we decided to conduct further experiments reported in Section 5 using the three node importance measures in rows 1, 8 and 9 of Table 2.

### 4.3 Variations of Clustering-Based Potential Hub Sets

While the CBS algorithm appeared promising in identifying good hub locations, the average reduction in CPU times was only 13.6 % (see Appendix 2). Analysis of the solutions showed that the number of subsets created by CBS is quite often strictly less than $p$, so in many cases relatively few of the hubs were being selected from the small cluster-based subsets. Therefore, to reduce the solution space (and CPU times) we developed variants of CBS that restricted the flexibility in selecting hubs. We also experimented with several variants of CBS in an effort to improve the solution quality by expanding the subsets to consider additional nodes as potential hubs. The letter "R" in an algorithm name reflects an effort to reduce CPU times, while the letters "A" and "I" indicate an effort to improve solution quality. Table 3 provides the constraints added to the mathematical programming formulation for each of the five variants of the CBS algorithm described below.

**Restricted CBS (RCBS)** The restricted variant of the CBS algorithm (RCBS) forces all $p$ hubs to be selected from the union of the subsets $S_i$ and $H_S$. This has the greatest impact when the number of constraints (8) and (9) is far less than $p$.

**Augmented CBS (ACBS) and Augmented RCBS (ARCBS)** In CBS and RCBS, the less important "small nodes" from $N \setminus N^p$ are included in the subsets $S_i$, but the "small nodes" are not included when near an isolated important node in $H_S$. Thus, for the augmented CBS (ACBS) and augmented RCBS (ARCBS) algorithms, we enlarge set $H_S$ to include all the nodes from $N$ that are within distance $PM$ of a node in $H_S$. We denote this new set as $H'_S$ and at the end of the CBS algorithm we set $H'_S = H_S \cup \{k | d_{ik} \leq PM$ and $\forall \ k \in N, \ i \in H_S\}$.

**Improved RCBS (IRCBS) and Improved Augmented RCBS (IARCBS)** Because some optimal solutions use hubs that do not have particularly high importance values, these variants of CBS incorporate $p$ additional nodes in a set $H_p$ as potential hubs. For IRCBS, $H_p$ is formed by sequentially adding the most

important nodes that are not already included in the subsets for RCBS (i.e., not in $\cup_{i \in H} S_i \cup H_S$), nor are too close to another node already added to set $H_p$ (this encourages dispersion of the hubs). For IARCBS, set $H_p$ is formed similarly by sequentially adding the most important nodes that are not already included in the subsets for ARCBS (i.e., not in $\cup_{i \in H} S_i \cup H_S'$), nor too close to a node already added to set $H_p$. The set $H_p$ is defined using only those subsets $S_i$ actually used to form constraints added to the MIP formulation and a proximity measure $PM^A$ based on **all** nodes in $N$, not just the most important nodes that comprise $N^p$. The algorithm to create set $H_p$ for IRCBS follows. This uses $N^A$ as the set of all nodes sorted in decreasing order of importance and $N^A[t]$ to denote the $t^{th}$ node in this ordered set.

**Algorithm Add $H_p$:**

0. $t = 1, \cdot q = 0, \cdot H_p = \varnothing$ and $PM^A = \dfrac{\sum_{i \in N^A} \left\{ \min\limits_{j \in N^A : i \neq j} d_{ij} \right\}}{|N^A|}$

1. while ($t \leq n$ or $q < p$) do {
      If $\left( N^A[t] \notin H_s \cup \{\cup_{i \in H} S_i\} \text{ and } d_{N^A[t]}, N^A[k] > PM^A : k < t \right)$ then
        $H_p = H_p \cup \{N^A[t]\}$ and $q = q + 1$

2. $t = t + 1$}.

At the end of Algorithm **Add $H_p$**, either the set $H_p$ has $p$ elements or there were not $p$ potential hubs that were not already in sets $S_i$ and $H_S$, and are not far enough apart. The IARCBS algorithm is identical to this, except that $H_s$ is replaced by $H_S'$.

**Table 3** The constraints for each methodology

| Methodology | Constraints from original formulation | Additional constraints | | |
|---|---|---|---|---|
| Original | (1)–(7) | – | – | – |
| CBS | (1)–(7) | $\sum_{k \in S_i} X_{kk} \geq 1 \; \forall i \in H$ | $\sum_{k \in H_s} X_{kk} \geq 1$ | – |
| RCBS | (1), (3)–(7) | $\sum_{k \in S_i} X_{kk} \geq 1 \; \forall i \in H$ | $\sum_{k \in H_s} X_{kk} \geq 1$ | $\sum_{k \in \{\cup_{i \in H} S_i\} \cup H_S} X_{kk} = p$ |
| IRCBS | (1), (3)–(7) | $\sum_{k \in S_i} X_{kk} \geq 1 \; \forall i \in H$ | $\sum_{k \in H_s} X_{kk} \geq 1$ | $\sum_{k \in \{\cup_{i \in H} S_i\} \cup H_S \cup H_p} X_{kk} = p$ |
| ACBS | (1)–(7) | $\sum_{k \in S_i} X_{kk} \geq 1 \; \forall i \in H$ | $\sum_{k \in H_s'} X_{kk} \geq 1$ | – |
| ARCBS | (1), (3)–(7) | $\sum_{k \in S_i} X_{kk} \geq 1 \; \forall i \in H$ | $\sum_{k \in H_s'} X_{kk} \geq 1$ | $\sum_{k \in (\cup_{i \in H} S_i) \cup H_S'} X_{kk} = p$ |
| IARCBS | (1), (3)–(7) | $\sum_{k \in S_i} X_{kk} \geq 1 \; \forall i \in H$ | $\sum_{k \in H_s'} X_{kk} \geq 1$ | $\sum_{k \in (\cup_{i \in H} S_i) \cup H_S' \cup H_p} X_{kk} = p$ |

In Appendix 1, we present an illustration using the CAB data set to demonstrate the formation of the subsets for the different methodologies.

By design of the methodologies, as reflected in the added constraints, the relationship of the optimal objective function values is:

$$Z^*_{opt} \leq Z^*_{ACBS} \leq Z^*_{CBS} \leq Z^*_{IRCBS} \leq Z^*_{RCBS}$$

$$Z^*_{ACBS} \leq Z^*_{IARCBS} \leq Z^*_{ARCBS} \leq Z^*_{RCBS}$$

where $Z^*$ denotes the optimal objective function value of the algorithms. Clearly, as shown in Table 3, when we add restrictions to a particular formulation, the objective function value of the resulting problem cannot improve. Graphically this is depicted in Fig. 2, where the arrows indicate a possible improvement in the objective function value.

## 5 Evaluation of Methodologies and Node Importance Measures

In this section, we evaluate the variants of the CBS algorithm by solving a variety of problem instances using four real-world data sets of varying scale, including some large problems with up to 200 nodes (40,000 o-d flows). To evaluate the effectiveness of the methodologies, we compare the hub locations, total costs and the CPU times. For comparing solution quality, the baseline is the optimal solution from the original formulation with constraints (1)–(7), and we report both the percentage of optimal solutions achieved by each methodology and the "Gap", as measured by the relative difference from the objective function value of the baseline. The average CPU time improvement is measured relative to the time for the baseline solution using the same hardware and software.

### 5.1 Comparison of Variations of CBS

We first provide results for 64 small instances with the CAB and AP data sets described earlier (using $p=3$–10 and $\alpha=0.2$, 0.4, 0.6 and 0.8), for 64 medium sized instances with the 40 node CAB25+15 data set (Campbell 2009) (using $p=3$–10 and $\alpha=0.2$, 0.4, 0.6 and 0.8),
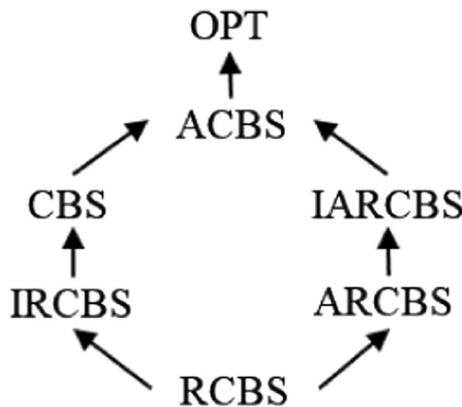


**Fig. 2** Graphical representation of the relationship between the objective function values
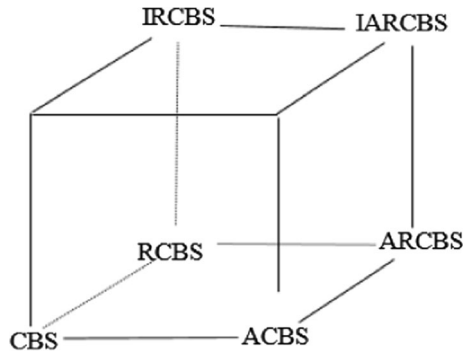
**Fig. 3** Relationship of the variants of the CBS methodology

and for the 81 node TR81 data set (Tan and Kara 2007). The math programming formulations were solved using CPLEX 12.4 for CAB, AP20 and TR81, and using Gurobi 4.5.2 for CAB25+15. All the results were obtained in a Linux environment with a 4xAMD Opteron Interlagos 2.6GHz processor and 96 GB RAM. To compare the variants of the CBS methodology, we present results using the measure of node importance from the first row of Table 2, $(O_i+D_i)*C_i$. The variants are related as shown in Fig. 3, where: (i) moving to the right in the figure to the "A" versions is an attempt to improve solution quality by allowing some of the smaller nodes to be considered as potential hubs, (ii) moving up in the figure to the "I" versions is an attempt to improve solution quality by incorporating $p$ additional important nodes as potential hubs, and (iii) moving back in the figure to the "R" versions" is an attempt to reduce CPU times by restricting the set of potential hubs. The "A" versions consider some less important nodes located close to important nodes as potential hubs, so the changes in optimal locations are likely to be on a small geographic scale. In contrast, the "I" versions consider more of the most important nodes as potential hubs, which allows a wider dispersion of hubs across the region. Finally, the "R" versions restrict the flexibility in locations to reduce solution times.

Table 4 provides a summary of results with the CAB data set, with complete results in Appendix 2. (For other data sets, we only provide a summary of the results; complete results are available from the authors.) For the small AP20 instances where optimum results are obtained very quickly (maximum is 2.79 s), comparing solution times on such small scales

**Table 4** Comparison of the methodologies for AP20 and CAB data sets

|  |  | CBS | RCBS | IRCBS | ACBS | ARCBS | IARCBS |
|---|---|---|---|---|---|---|---|
| AP20 | Average Gap (%) | 0.49 | 0.96 | 0.89 | 0.45 | 0.53 | 0.45 |
|  | Maximum Gap (%) | 4.02 | 4.02 | 4.02 | 3.90 | 3.90 | 3.90 |
|  | Optimal Solutions (%) | 84.40 | 37.50 | 50.00 | 84.40 | 68.80 | 81.30 |
| CAB | Average Gap (%) | 0.00 | 0.55 | 0.03 | 0.00 | 0.50 | 0.00 |
|  | Maximum Gap (%) | 0.00 | 5.09 | 0.98 | 0.00 | 5.09 | 0.00 |
|  | Optimal Solutions (%) | 100.00 | 68.80 | 96.90 | 100.0 | 71.90 | 100.00 |
|  | Average CPU time improvement (%) | 13.61 | 59.62 | 46.79 | 7.39 | 42.03 | 31.81 |
|  | Maximum CPU time improvement (%) | 81.60 | 94.13 | 87.64 | 39.54 | 92.99 | 79.66 |
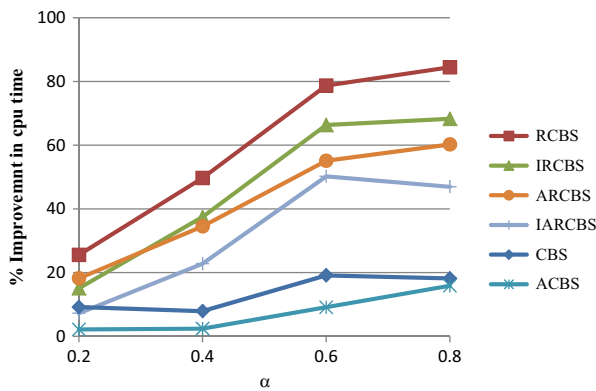
**Fig. 4** CPU time improvements for the CAB data set

will not be reliable so CPU times are not reported for AP20 in Table 4. This table shows that in terms of solution quality, the RCBS and ARCBS methodologies perform poorly and the CBS, ACBS and IARCBS methodologies perform the best. As expected, the "A" versions of the methodologies improve performance with their greater geographic flexibility, especially so for ARCBS and IARCBS. Results also show improvements with the "I" versions from considering a wider range of high importance nodes, though at the expense of additional CPU time. Table 4 shows how the "R" versions of the methodology reduce the aggregated CPU times, and Fig. 4 provides a more nuanced perspective by showing how the CPU time improvements (averaged for each value of $\alpha$; see Appendix 2) increase with $\alpha$. This figure also clearly shows the greater benefits from the "R" versions for larger values of $\alpha$ (0.6 and 0.8), where the improvements range from 47 to 85 %. (But note that the CPU times do not always improve with the various methodologies as shown in Appendix 2.) The results in Table 4 display the expected tradeoff between solution quality and CPU time improvement, and while CBS and ACBS yield good quality solutions, the CPU improvement is rather small and not nearly as good as the other four methodologies (see also Fig. 4). Therefore, to further explore faster ways to find optimal or near-optimal solutions, for the

**Table 5** Comparison of the methodologies for CAB25+15 and TR81 data sets

|  |  | RCBS | IRCBS | ARCBS | IARCBS |
|---|---|---|---|---|---|
| CAB25+15 | Average Gap (%) | 1.21 | 0.30 | 1.09 | 0.29 |
|  | Maximum Gap (%) | 5.15 | 2.30 | 5.15 | 2.30 |
|  | Optimal Solutions (%) | 37.50 | 71.90 | 46.90 | 78.10 |
|  | Average CPU time improvement (%) | 82.70 | 71.11 | 76.70 | 58.73 |
|  | Maximum CPU time improvement (%) | 98.52 | 97.08 | 92.50 | 89.02 |
| TR81 | Average Gap (%) | 3.05 | 2.61 | 1.99 | 1.74 |
|  | Maximum Gap (%) | 7.81 | 7.21 | 5.26 | 5.26 |
|  | Optimal Solutions (%) | 0.00 | 0.00 | 23.50 | 35.30 |
|  | Average CPU time improvement (%) | 98.85 | 96.91 | 92.32 | 86.77 |
|  | Maximum CPU time improvement (%) | 99.86 | 99.70 | 99.01 | 98.00 |

remaining analyses we continue with only the four fastest versions of the methodology: RCBS, IRCBS, ARCBS, and IARCBS.

Table 5 is a summary of results using the CAB25+15 and TR81 data sets. For TR81 data set optimal solutions are obtained within a 2 h limit for only 17 of the 32 instances. Therefore, in the CPU time comparisons we include only the instances where the original formulation found the optimal solution in 2 h; however, for the solution quality comparison we include all 32 instances by using the lower bound from solving the original formulation (from CPLEX) when the optimal solution is not found. The high quality of solutions with IARCBS is evident; and though it provide the least improvement in CPU times, it still averages a 58.7 % reduction for CAB25+15 and an 86.8 % reduction for TR81. The RCBS methodology is least effective in terms of solution quality, though it provides the greatest CPU time reduction. Note that the CPU time percentage improvements are greater for the more challenging data set CAB25+15, then for the CAB data set in Table 4.

Figure 5 documents the tradeoff between solution quality and CPU time improvement with the four best methodologies (RCBS, IRCBS, A
RCBS and IARCBS) for the CAB, CAB25+15, and TR81 data sets. The best solutions (high quality and fast) would be at the lower left corner of the graph. These figure shows that all of these methodologies identify good subsets for potential hubs (very near-optimal solutions are found) with considerably less CPU time than required for the original formulation. Comparison across the three data sets suggests that as the size of the problem increases (moving right to left in the figure), the improvements in CPU times increase as well, while solution quality deteriorates only a little. The figure also clearly shows that the added flexibility from an expanded set of potential hubs with the "A" or "I" versions improves the solution quality at the expense of increased CPU time.

### 5.2 Comparison of Node Importance Measures

Table 6 displays results with the IARCBS methodology (which provides the best quality solutions) for the CAB, CAB25+15 and TR81 data sets using the three best performing node importance measures identified in Section 4. These results show that all three measures perform well in terms of both solution quality and CPU time improvements, with a very slight advantage to the simpler measure $(O_i+D_i)*C_i$ from the greater CPU time improvement with CAB. The results show how having more
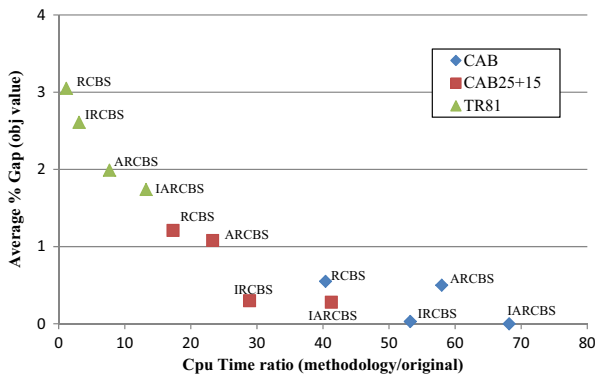


**Fig. 5** Solution quality and CPU time improvements for the CAB, CAB25+15 and TR81 data sets

**Table 6** Comparison of node importance measures for IARCBS

| | | $(O_i+D_i)*C_i$ | $\sum_j \left[ d_{ij}*(w_{ij} + w_{ji}) \right]$ | $0.75\frac{(O_i+D_i)}{\sum_i(O_i+D_i)} + 0.25\frac{C_i}{\sum_i C_i}$ |
|---|---|---|---|---|
| CAB | Average Gap (%) | 0.00 | 0.00 | 0.00 |
| | Maximum Gap (%) | 0.00 | 0.00 | 0.00 |
| | Optimal Solutions (%) | 100.00 | 100.00 | 100.00 |
| | Average CPU time improvement (%) | 31.81 | 19.79 | 22.83 |
| | Maximum CPU time improvement (%) | 79.66 | 79.14 | 79.90 |
| CAB25+15 | Average Gap (%) | 0.29 | 0.28 | 0.25 |
| | Maximum Gap (%) | 2.30 | 2.13 | 2.13 |
| | Optimal Solutions (%) | 78.1 | 78.1 | 75.0 |
| | Average CPU time improvement (%) | 58.73 | 58.30 | 54.72 |
| | Maximum CPU time improvement (%) | 89.02 | 92.24 | 86.84 |
| TR81 | Average Gap (%) | 1.74 | 1.85 | 1.72 |
| | Maximum Gap (%) | 5.26 | 4.96 | 4.63 |
| | Optimal Solutions (%) | 35.30 | 29.40 | 29.40 |
| | Average CPU time improvement (%) | 86.77 | 88.57 | 86.92 |
| | Maximum CPU time improvement (%) | 98.00 | 94.16 | 94.16 |

nodes available for locating hubs (moving from 25 nodes in CAB, to 40 nodes in CAB25+15, to 81 nodes in TR81) leads to finding fewer of the optimal hub location sets. However, the quality of the solutions is still quite good, even when the majority of solutions are non-optimal sets of hubs (29–35 % for TR81), and the CPU time savings increase with the size of the problem. This illustrates a tradeoff of solution quality and CPU time with the size of the problem.

### 5.3 AP200 Instances

To explore the use of our methodology on larger problems, we solved several instances of the full 200 node AP200 data set (Ernst and Krishnamoorthy 1996) using $p$=3–5, 10 and 15 and the given cost parameter values $\gamma$=3, $\delta$=2 and $\alpha$= 0.75. Results use CPLEX 12.4 and the hardware setup described earlier with a 3 h CPU time limit. Table 7 displays the gaps from the best known heuristic solutions for these problems (for $p$=3,4, 10 from Kratica et al. 2007 and $p$=5 and 15 from Ilic et al. 2010). Unfortunately, the optimal objective function values and optimal hub locations for the AP200 dataset have not been reported. (The earlier formulation with constraints (1)–(7) was unable to obtain even an integer feasible solution within the time limit.) As can be seen from Table 7, the gaps are greatest with the smallest instances ($p$=3) and the average gaps are

**Table 7** Gaps with best known solutions for the methodologies with the AP200 data set

| $p$ | RCBS | IRCBS | ARCBS | IARCBS |
|---|---|---|---|---|
| 3 | 7.2 % | 7.2 % | 6.9 % | 6.9 % |
| 4 | 3.5 % | 3.5 % | 3.5 % | 3.8 % |
| 5 | 2.7 % | 2.7 % | 2.5 % | 2.9 % |
| 10 | 5.5 % | 2.2 % | 5.4 % | 7.9 % |
| 15 | 5.6 % | 1.2 % | 3.3 % | 0.8 % |
| Average | 4.9 % | 3.4 % | 4.3 % | 4.5 % |

less than 5 % for all methodologies. The CPU time limit was reached for all methodologies for the problems with $p=10$ and 15 in Table 7, and for some of the smaller problems with the IRCBS, ARCBS, and IARCBS methodologies. The AP200 data set includes a large number of nodes (i.e., potential hub locations) and demand is relatively centrally concentrated (corresponding to downtown Sydney), so finding all the optimal hubs is quite challenging. However, identifying regions of likely hub locations is useful as the practical considerations would restrict the locations actually used according to a myriad of site specific factors, such as land availability, facility capacities, transportation infrastructure, labor requirements, etc. Note that the subsets provided by our methodologies can also be used to reduce the solution space in other optimal or heuristic solution procedures.

## 6 Discussion and Conclusion

In this research, we have sought to develop a better understanding of optimal and near-optimal hub locations in single allocation networks derived only from the basic data for the problem, and to use this understanding to better solve hub location problems. Using ideas from spatial analysis of real-world hub networks and optimal hub locations for benchmark hub location data sets of differing scale and scope, we first identified key characteristics of optimal hub locations. We then used these characteristics in a straightforward heuristic solution approach to delineate subsets of nodes likely to contain hubs based on spatial proximity of nodes, dispersion and measures of node importance. We developed and evaluated several variants of a basic methodology and documented the tradeoffs in terms of solution quality and CPU times. Two important aspects of the methodologies were the ability to select as hubs: (i) smaller nodes near nodes with high importance, and (ii) important isolated nodes. This highlights the key role of local spatial interactions in case (i) and wider interactions over the entire region in case (ii). Several methodologies were shown to perform very well in terms of providing near-optimal solutions and large reductions in CPU times when used with MILP formulations, though the relative benefits from different methodologies depended on the data sets.

One key contribution of this research is that simple measures of demand, centrality, and dispersion are effective in finding optimal or near-optimal hub locations. Results showed that using only the aggregate demand originating and terminating at a node ($O_i+D_i$) is not a particularly effective measure of node importance, while combining the aggregated demand with a spatial measure of node centrality does provide an effective way to identify good locations for hubs. This underscores the importance of the relative spatial locations for finding near-optimal hub locations. Another contribution was to highlight the value of centrality as an important component of node importance vs. betweenness. Although betweenness is an important concept for single hub systems (see e.g., Maertens et al. 2014), with multiple hubs in a network dispersed across the service region, a global sense of betweenness based on the entire data set may not be as important as a more restrictive "local betweenness" for those nodes allocated to a particular hub.

The computational results documented the important tradeoff between solution quality and solution speed in two dimensions. As expected, for a particular instance the variants of the CBS heuristic designed to improve solution quality required more CPU time. Interestingly, the results showed that as the problem size (number of nodes) increase, the solution quality deteriorates only a little, but the CPU time improvement increases. This suggests the heuristic approach as outlined in this paper may provide even greater CPU time savings for even larger problems.

In summary, this research provides tools to identify optimal or near-optimal subsets of nodes for locating hubs that rely on a limited amount of the input data. These subsets may be useful to help focus attention and further analyses on certain cities or geographic regions as likely locations for hubs, as well as to speed solution approaches to design good hub networks (e.g., mixed-integer linear programming models). Because hub location problems are complex combinations of facility location and network design, all in a practical setting (e.g., airlines or trucking companies), many problem features, including the input data are not likely to be known with certainty and will certainly vary over time. Thus, overreliance on specific parameter values (e.g., the transportation cost discount $\alpha$), static data sets, or on "optimal" model outputs may be too strong a simplification of the real-world problem. While hub location optimization models are certainly valuable, there is complementary value in gaining a better understanding of the general properties that make a good hub location, with the goal of using the aggregate knowledge to better solve hub location problems (Geoffrion 1976).

Some promising areas for future research include refining the methodologies and attempting to solve even larger problems, analyzing other hub location problems (e.g., multiple allocation problems, incomplete network designs, etc.), considering reliability and robustness issues (e.g., Lordan et al. 2014; O'Kelly 2014)) and analyzing the effect of the hub circle radius on the hub locations. Another area could be the direct inclusion of hub dispersion as a constraint in the MILP models, rather than as a step in the formation of clusters.

## Appendix 1. Example for the CBS Methodologies

In the following example, we demonstrate the formation of subsets for CAB dataset with the node importance measure $V_i=(O_i+D_i)*C_i$ as in row 1 of Table 2. The nodes in decreasing order of importance are as follows:

| Rank | Node # | $V_i$ | Rank | Node # | $V_i$ | Rank | Node # | $V_i$ |
|------|--------|-------|------|--------|-------|------|--------|-------|
| 1 | 17 | 8537.42 | 10 | 9 | 1707.36 | 18 | 21 | 1073.22 |
| 2 | 12 | 6020.33 | 11 | 7 | 1399.90 | 19 | 19 | 1036.16 |
| 3 | 22 | 4513.58 | 12 | 8 | 1312.30 | 20 | 24 | 984.87 |
| 4 | 4 | 3851.80 | 13 | 6 | 1191.68 | 21 | 16 | 848.00 |
| 5 | 3 | 3494.22 | 14 | 10 | 1182.99 | 22 | 11 | 796.87 |
| 6 | 14 | 3340.95 | 15 | 1 | 1163.30 | 23 | 2 | 754.70 |
| 7 | 25 | 2524.76 | 16 | 20 | 1154.96 | 24 | 5 | 574.81 |
| 8 | 23 | 1737.25 | 17 | 15 | 1126.96 | 25 | 13 | 441.10 |
| 9 | 18 | 1710.37 | | | | | | |

For $p=5$, $N^p =\{17, 12, 22, 4, 3, 14, 25, 23, 18, 9\}$ is the set of the 10 most important nodes and $PM$ is 330,71 miles and calculated as shown below:

| cities | 3 | 4 | 9 | 12 | 14 | 17 | 18 | 22 | 23 | 25 | min distance |
|--------|---|---|---|----|----|----|----|----|----|----|--------------|
| 3 | 0 | 858,33 | 621,33 | 2600,08 | 1266,85 | 190,32 | 274,31 | 2703,40 | 2503,83 | 405,79 | 190,32 |
| 4 | 858,33 | 0 | 237,07 | 1741,87 | 1186,86 | 720,47 | 675,34 | 1853,62 | 1733,13 | 592,03 | 237,07 |
| 9 | 621,33 | 237,07 | 0 | 1978,94 | 1151,87 | 489,56 | 453,26 | 2086,85 | 1936,30 | 392,90 | 237,07 |
| 12 | 2600,08 | 1741,87 | 1978,94 | 0 | 2335,82 | 2453,35 | 2396,79 | 361,54 | 986,81 | 2288,75 | 361,54 |
| 14 | 1266,85 | 1186,86 | 1151,87 | 2335,82 | 0 | 1098,28 | 1021,61 | 2591,45 | 2725,79 | 923,22 | 923,22 |
| 17 | 190,32 | 720,47 | 489,56 | 2453,35 | 1098,28 | 0 | 84,34 | 2574,08 | 2415,49 | 215,56 | 84,34 |
| 18 | 274,31 | 675,34 | 453,26 | 2396,79 | 1021,61 | 84,34 | 0 | 2526,56 | 2388,69 | 132,77 | 84,34 |
| 22 | 2703,40 | 1853,62 | 2086,85 | 361,54 | 2591,45 | 2574,08 | 2526,56 | 0 | 694,94 | 2430,27 | 361,54 |
| 23 | 2503,83 | 1733,13 | 1936,30 | 986,81 | 2725,79 | 2415,49 | 2388,69 | 694,94 | 0 | 2321,87 | 694,94 |
| 25 | 405,79 | 592,03 | 392,90 | 2288,75 | 923,22 | 215,56 | 132,77 | 2430,27 | 2321,87 | 0 | 132,77 |

average: 330,71

Table 8 shows the sets of potential hubs produced by the CBS algorithm and its variants. For example, with the CBS algorithm, the subsets are $S_{17}=\{17, 3, 18, 25, 2, 20\}$, $S_4=\{4, 9, 5, 6, 21\}$, and $H_S =\{12, 22, 14, 23\}$ and the following three constraints are added to the MILP formulation:

$$\sum_{k\in\{2,3,17,18,20,25\}} X_{kk}\geq 1 \tag{10}$$

$$\sum_{k\in\{4,5,6,9,21\}} X_{kk}\geq 1 \tag{11}$$

$$\sum_{k\in\{12,14,22,23\}} X_{kk}\geq 1. \tag{12}$$

**Table 8** Potential hub location sets with different methodologies for $p=5$ with the CAB data set

| CBS, RCBS | IRCBS | ACBS, ARCBS | IARCBS |
|---|---|---|---|
| $S_{17}=\{17, 3, 18, 25, 2, 20\}$ | | | |
| $S_4=\{4, 9, 5, 6, 21\}$ | | | |
| $H_S =\{12, 22, 14, 23\}$ | $H_S =\{12, 22, 14, 23\}$ | $H_S' =\{12, 22, 14, 23, \mathbf{24}\}$ | $H_S' =\{12, 22, 14, 23, \mathbf{24}\}$ |
| | $H_p =\{7, 8, 1, 15, 19\}$ | | $H_{p'} =\{7, 8, 1, 15, 19\}$ |

For the RCBS algorithm we also add the following constraint:

$$\sum_{k\in\{2,3,4,5,6,9,12,14,17,18,20,21,22,23,25\}} X_{kk} = 5. \tag{13}$$

As shown in Table 8, for the ACBS and ARCBS variations, the set of isolated important nodes is augmented with node 24 (Tampa) as it in within the hub circle for important node 14 (Miami), as shown in the following figure (Fig. 6):

Thus, for the ACBS algorithm, we add to the original formulation constraints (10), (11) and

$$\sum_{k\in\{12,14,22,23,24\}} X_{kk}\geq 1. \tag{14}$$

For the ARCBS algorithm, we add to the original formulation constraints (10), (11), (14) and

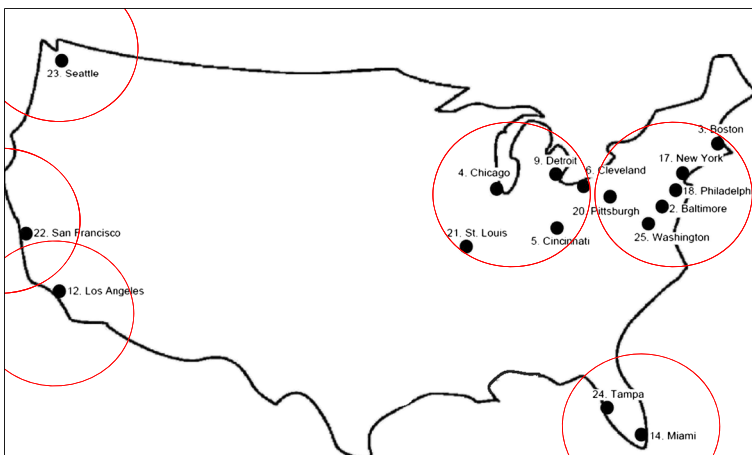$$\sum_{k\in\{2,3,4,5,6,9,12,14,17,18,20,21,22,23,24,25\}} X_{kk} = 5. \tag{15}$$



**Fig. 6** Hub circles for $p=5$ with the CAB data set

Note that hub circles can overlap, and there may be "small nodes" in the overlapping region that would appear as potential hubs in more than one constraint. An example of this would be in the figure above if there was a "small node" city between Los Angeles and San Francisco.

For IRCBS algorithm, the nodes eligible for set $H_p$ (i.e., those not in $S_{17}$, $S_4$, or $H_S$), in decreasing order of importance, are {7, 8, 10, 1, 15, 19, 24, 16, 11, 13}. From these, we form set $H_p$ by adding $p=5$ nodes in order and ignoring any nodes within distance $PM^A = 241.96$ of any more important node, and the result is $H_p = \{7, 8, 1, 15, 19\}$. We use set $H_p$ for the IRCBS algorithm and add to the original formulation, constraints (10)–(12) and the following:

$$\sum_{k \in \{N \setminus \{10,11,13,16,24\}\}} X_{kk} = 5 \qquad (16)$$

For IARCBS, the set $H_p = \{7, 8, 1, 15, 19\}$ is same as for this example, so we add to the original formulation, constraints (10), (11), (14) and the following:

$$\sum_{k \in \{N \setminus \{10,11,13,16\}\}} X_{kk} = 5 \qquad (17)$$

For this illustration, Table 9 lists the optimal hub locations and the resulting hub locations with the six methodologies using four different values of $\alpha$. In all of these instances, CBS, IRCBS, ACBS, and IARCBS found optimal hub locations. On the other hand, RCBS and ARCBS did not find all the optimal hub locations in any of the instances, because neither node 7 (Dallas, which ranks 15th in importance) nor node 1 (Atlanta, which ranks 11th in importance) is among the subsets $S_{17}$, $S_4$ or $H_S$ ($H'_S$). Thus, RCBS and ARCBS select the node 21 (St. Louis) instead of node 7 (Dallas), and node 14 (Miami) or node 24 (Tampa) instead of node 1 (Atlanta)

**Table 9** Hub locations with different methodologies for $p=5$ with CAB data set

| $p$ | $\alpha$ | Optimal hub locations | CBS | RCBS | IRCBS | ACBS | ARCBS | IARCBS |
|---|---|---|---|---|---|---|---|---|
| 5 | 0.2 | 4,7,12,14,17 | * | 4,12,14,17,**21** | * | * | 4,12,14,17,**21** | * |
| | 0.4 | 4,7,12,14,17 | * | 4,12,14,17,**21** | * | * | 4,12,14,17,**21** | * |
| | 0.6 | 4,7,12,14,17 | * | 4,12,14,17,**21** | * | * | 4,12,14,17,**21** | * |
| | 0.8 | 1,4,7,12,18 | * | 4,12,**14**,18,**21** | * | * | 4,12,18,**21,24** | * |
| Avg. CPU time (sec)** | | 29.08 | 32.99 | 2.81 | 5.90 | 31.40 | 3.39 | 11.50 |

*: same as the optimum solution, **: average of the four instances

# Appendix 2

**Table 10** Optimal objective function values, solution time of original formulation (seconds), % differences (gap) in objective function values and CPU time improvement (%) with the CAB data set

| p | α | Optimum objective function value | Solution time of original form (sec) | CBS Gap (%) | CBS Imp (%) | RCBS Gap (%) | RCBS Imp (%) | IRCBS Gap (%) | IRCBS Imp (%) | RCBS-2 Gap (%) | RCBS-2 Imp (%) | ACBS Gap (%) | ACBS Imp (%) | ARCBS Gap (%) | ARCBS Imp (%) | IARCBS Gap (%) | IARCBS Imp (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 0.2 | 767.4 | 3.1 | 0.0 | 54.4 | 0.0 | 64.5 | 0 | 52.4 | 0.0 | 54.4 | 0.0 | 32.9 | 0.0 | 40.1 | 0.0 | 34.5 |
|   | 0.4 | 901.7 | 3.6 | 0.0 | 51.4 | 0.0 | 63.8 | 0 | 57.7 | 0.0 | 50.8 | 0.0 | −3.0 | 0.0 | 17.7 | 0.0 | 7.7 |
|   | 0.6 | 1033.6 | 8.1 | 0.0 | 70.8 | 0.0 | 81.3 | 0 | 78.8 | 0.0 | 71.6 | 0.0 | 4.0 | 0.0 | 25.3 | 0.0 | 27.9 |
|   | 0.8 | 1158.8 | 18.2 | 0.0 | 81.6 | 0.0 | 89.0 | 0 | 87.6 | 0.0 | 81.8 | 0.0 | 8.7 | 0.0 | 53.4 | 0.0 | 28.0 |
| 4 | 0.2 | 629.6 | 1.8 | 0.0 | 1.7 | 1.1 | 37.5 | 1 | 28.4 | 0.0 | −2.8 | 0.0 | −2.8 | 0.0 | 23.9 | 0.0 | 10.8 |
|   | 0.4 | 787.5 | 4.6 | 0.0 | 11.6 | 0.9 | 74.8 | 0 | 66.3 | 0.0 | 11.6 | 0.0 | 14.7 | 0.9 | 47.1 | 0.0 | 33.7 |
|   | 0.6 | 939.2 | 8.4 | 0.0 | −8.6 | 1.4 | 82.4 | 0 | 68.5 | 0.0 | 0.5 | 0.0 | −21.7 | 1.4 | 19.6 | 0.0 | 19.1 |
|   | 0.8 | 1087.66 | 32.2 | 0.0 | 29.7 | 2.0 | 89.1 | 0 | 84.3 | 0.0 | 27.2 | 0.0 | 20.8 | 1.6 | −9.8 | 0.0 | 37.2 |
| 5 | 0.2 | 538.4 | 1.5 | 0.0 | 0.7 | 5.1 | 29.1 | 0 | 20.5 | 0.0 | 0.7 | 0.0 | −4.0 | 5.1 | 21.9 | 0.0 | 14.6 |
|   | 0.4 | 707.7 | 3.3 | 0.0 | −5.1 | 2.9 | 51.1 | 0 | 33.5 | 0.0 | 4.8 | 0.0 | −10.0 | 2.9 | 36.9 | 0.0 | 32.0 |
|   | 0.6 | 876.6 | 14.3 | 0.0 | −1.8 | 1.6 | 80.1 | 0 | 62.6 | 0.0 | 13.0 | 0.0 | −3.0 | 1.6 | 75.6 | 0.0 | 60.4 |
|   | 0.8 | 1034.1 | 97.2 | 0.0 | −15.7 | 1.6 | 94.1 | 0 | 84.7 | 0.9 | −7.3 | 0.0 | −8.7 | 1.5 | 93.0 | 0.0 | 62.1 |
| 6 | 0.2 | 491.0 | 1.8 | 0.0 | 0.0 | 0.0 | 21.8 | 0 | 6.7 | 0.0 | 0.0 | 0.0 | −7.3 | 0.0 | 9.5 | 0.0 | −1.1 |
|   | 0.4 | 659.8 | 4.0 | 0.0 | −21.5 | 0.0 | 49.1 | 0 | 32.7 | 0.0 | −11.7 | 0.0 | −29.9 | 0.0 | 36.7 | 0.0 | 4.7 |
|   | 0.6 | 828.1 | 17.2 | 0.0 | 23.0 | 0.0 | 73.6 | 0 | 61.4 | 0.0 | 18.2 | 0.0 | 19.6 | 0.0 | 49.4 | 0.0 | 43.3 |
|   | 0.8 | 991.0 | 100.5 | 0.0 | −7.4 | 0.5 | 73.4 | 0 | 53.8 | 0.0 | 8.1 | 0.0 | 16.3 | 0.5 | 60.1 | 0.0 | 27.6 |
| 7 | 0.2 | 448.2 | 1.7 | 0.0 | 3.6 | 0.0 | 21.6 | 0 | 4.2 | 0.0 | 1.2 | 0.0 | −2.4 | 0.0 | 14.4 | 0.0 | 4.2 |

**Table 10** (continued)

| p | α | Optimum objective function value | Solution time of original form (sec) | CBS Gap (%) | CBS Imp (%) | RCBS Gap (%) | RCBS Imp (%) | IRCBS Gap (%) | IRCBS Imp (%) | RCBS-2 Gap (%) | RCBS-2 Imp (%) | ACBS Gap (%) | ACBS Imp (%) | ARCBS Gap (%) | ARCBS Imp (%) | IARCBS Gap (%) | IARCBS Imp (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.4 | 621.9 | 4.9 | 0.0 | 7.4 | 0.0 | 63.6 | 0 | 48.7 | 0.0 | 12.5 | 0.0 | 17.6 | 0.0 | 55.4 | 0.0 | 44.8 |
| | 0.6 | 795.1 | 45.6 | 0.0 | 23.7 | 0.0 | 91.9 | 0 | 83.0 | 0.0 | 65.8 | 0.0 | 19.4 | 0.0 | 89.0 | 0.0 | 79.7 |
| | 0.8 | 959.9 | 122.5 | 0.0 | 17.2 | 0.6 | 87.4 | 0 | 64.3 | 0.0 | 28.7 | 0.0 | 23.9 | 0.6 | 85.1 | 0.0 | 60.3 |
| 8 | 0.2 | 414.6 | 1.8 | 0.0 | 6.6 | 0.0 | 12.1 | 0 | 12.6 | 0.0 | 0.0 | 0.0 | 2.2 | 0.0 | 14.3 | 0.0 | 11.0 |
| | 0.4 | 589.0 | 4.1 | 0.0 | 13.3 | 0.0 | 31.0 | 0 | 29.7 | 0.0 | −8.1 | 0.0 | 19.2 | 0.0 | 28.8 | 0.0 | 29.0 |
| | 0.6 | 763.5 | 40.9 | 0.0 | 20.7 | 0.0 | 84.1 | 0 | 71.5 | 0.0 | 0.1 | 0.0 | 22.4 | 0.0 | 64.0 | 0.0 | 71.4 |
| | 0.8 | 929.0 | 106.0 | 0.0 | 7.1 | 0.0 | 81.5 | 0 | 55.3 | 0.0 | 33.5 | 0.0 | 10.2 | 0.0 | 73.8 | 0.0 | 55.3 |
| 9 | 0.2 | 382.8 | 2.1 | 0.0 | 6.6 | 0.0 | 14.6 | 0 | 16.0 | 0.0 | 2.8 | 0.0 | 8.0 | 0.0 | 19.8 | 0.0 | 4.3 |
| | 0.4 | 557.7 | 3.1 | 0.0 | 0.0 | 0.0 | 30.7 | 0 | 16.1 | 0.0 | −9.4 | 0.0 | 1.0 | 0.0 | 29.7 | 0.0 | 15.5 |
| | 0.6 | 732.6 | 27.5 | 0.0 | 5.7 | 0.0 | 80.0 | 0 | 73.2 | 0.0 | −4.6 | 0.0 | 14.7 | 0.0 | 76.2 | 0.0 | 68.5 |
| | 0.8 | 901.8 | 106.1 | 0.0 | 4.5 | 0.0 | 78.9 | 0 | 73.2 | 0.0 | 39.9 | 0.0 | 39.5 | 0.0 | 71.4 | 0.0 | 62.1 |
| 10 | 0.2 | 353.6 | 1.5 | 0.0 | 0.0 | 0.0 | 3.4 | 0 | −20.3 | 0.0 | −8.1 | 0.0 | −9.5 | 0.0 | 2.0 | 0.0 | −20.3 |
| | 0.4 | 528.8 | 2.5 | 0.0 | 6.3 | 0.0 | 33.6 | 0 | 15.0 | 0.0 | −1.2 | 0.0 | 9.9 | 0.0 | 24.1 | 0.0 | 15.0 |
| | 0.6 | 703.4 | 7.4 | 0.0 | 19.6 | 0.0 | 56.4 | 0 | 31.7 | 0.0 | 5.6 | 0.0 | 17.7 | 0.0 | 41.9 | 0.0 | 31.7 |
| | 0.8 | 875.1 | 38.8 | 0.0 | 28.3 | 0.0 | 82.5 | 0 | 43.0 | 0.0 | −6.9 | 0.0 | 16.3 | 0.0 | 55.1 | 0.0 | 43.0 |
| Average | | | | 0.00 | 13.60 | 0.60 | 59.60 | 0.00 | 46.80 | 0.00 | 14.80 | 0.00 | 7.40 | 0.50 | 42.00 | 0.00 | 31.80 |

# References

Alumur S, Kara BY (2008) Network hub location problems: the state of the art. Eur J Oper Res 190:1–21

Batta R, Lejeune M, Prasad S (2014) Public facility location using dispersion, population, and equity criteria. Eur J Oper Res 234:819–829

Bowen JT (2012) A spatial analysis of FedEx and UPS: hubs, spokes and network structure. J Transp Geogr 24:419–431

Bryan D (1998) Extensions to the hub location problem: formulations and numerical examples. Geogr Anal 30(4):315–330

Campbell JF (2009) Hub location for time definite transportation. Comput Oper Res 36:3107–3116

Campbell JF (2013) Modeling economies of scale in transportation hub networks. Proceedings of the 46th Annual Hawaii International Conference on System Sciences, IEEE Computer Society, 1154–1163

Campbell JF, O'Kelly ME (2012) Twenty-five years of hub location research. Transp Sci 46(2):153–169

Campbell JF, Ernst AT, Krishnamoorthy M (2002) Hub location problems. In: Drezner Z, Hamacher H (eds) Facility location: applications and theory. Springer, Berlin, pp 373–407

Chen SH (2010) A heuristic algorithm for hierarchical hub-and-spoke network of time-definite common carrier operation planning problem. Netw Spat Econ 10:509–523

Contreras I, Díaz JA, Fernández E (2011) Branch and price for large scale capacitated hub location problems with single assignment. INFORMS J Comput 23:41–55

Correia I, Nickel S, Saldanha da Gama F (2010) The capacitated single-allocation hub location problem revisited: a note on a classical formulation. Eur J Oper Res 207(1):92–96

Correia I, Nickel S, Saldanha da Gama F (2014) Multi-product capacitated single-allocation hub location problems: formulations and inequalities. Netw Spat Econ 14:1–25

Ernst AT, Krishnamoorthy M (1996) Efficient algorithms for the uncapacitated single allocation p-hub median problem. Locat Sci 4:139–154

Ernst AT, Krishnamoorthy M (1998) An exact solution approach based on shortest-paths for p-hub median problems. INFORMS J Comput 10(2):149–162

Fleming DK, Hayuth Y (1994) Spatial characteristics of transportation hubs: centrality and intermediacy. J Transp Geogr 2(1):3–18

Freeman LC (1978) Centrality in social networks conceptual clarification. Soc Networks 1:215–239

Garrison W (1960) Connectivity of the interstate highway system. Pap Reg Sci 6(1):121–137

Geoffrion AM (1976) The purpose of mathematical programming is insight, not numbers. Interfaces 7:81–92

Hillsman E (1980) Heuristic solutions to location-allocation problems: a user's guide to alloc IV, V, and VI. Monograph No. 7. Department of Geography, The University of Iowa, Iowa City

Horner MW, O'Kelly ME (2005) A combined cluster and interaction model: the hierarchical assignment problem. Geogr Anal 37:315–335

Ilic A, Urošević D, Brimberg J, Mladenovic N (2010) A general variable neighborhood search for solving the uncapacitated single allocation p-hub median problem. Eur J Oper Res 206:289–300

Kim H, O'Kelly ME (2009) Reliable p-hub location problems in telecommunication networks. Geogr Anal 41(3):283–306

Kissling CC (1969) Linkage importance in a regional highway network. Can Geogr 13:113–127

Kratica J, Stanimirović Z, Tošić D, Filipović V (2007) Two genetic algorithms for solving the uncapacitated single allocation p-hub median problem. Eur J Oper Res 182(1):15–28

Kuby M (1987) Programming models for facility dispersion: the p-dispersion and p-defense problems. Geogr Anal 19:315–329

Lordan O, Sallan JM, Simo P (2014) Study of topology and robustness of airline route networks from the complex network approach: a survey and research agenda. J Transp Geogr 37:112–120

Maertens S, Grimme W, Jung M (2014) An economic-geographic assessment of the potential for a new air transport hub in post-Gaddafi Libya. J Transp Geogr 38:1–12

Maliszewski P, Kuby M, Horner M (2012) A comparison of multi-objective spatial dispersion models for managing critical assets in urban areas. Comput Environ Urban Syst 36:331–341

Martin JC, Voltes-Dorta A (2008) Theoretical evidence of existing pitfalls in measuring hubbing practices in airline networks. Netw Spat Econ 8:161–181

Nitsch V (2005) Zipf zipped. J Urban Econ 57:86–100

O'Kelly ME (1992) Hub facility location with fixed costs. Pap Reg Sci 71(3):293–306

O'Kelly ME (2010) Routing traffic at hub facilities. Netw Spat Econ 10:173–191

O'Kelly ME (2014) Network hub structure and resilience. Netw Spat Econ. doi:10.1007/s11067-014-9267-1

O'Kelly ME, Miller HJ (1994) The hub network design problem: a review and synthesis. J Transp Geogr 2(1): 31–40

O'Kelly ME, Luna HP, Camargo RS, Miranda G (2014) Hub location problems with price sensitive demands. Netw Spat Econ. doi:10.1007/s11067-014-9276-0

O'Kelly ME (1986a) Activity levels at hub facilities in interacting networks. Geogr Anal 18(4):343–356

O'Kelly ME (1986b) The location of interacting hub facilities. Transp Sci 20:92–105

O'Kelly ME (1987) A quadratic integer program for the location of interacting hub facilities. Eur J Oper Res 302:393–404

O'Kelly ME, Lao Y (1991) Mode choice in a hub-and-spoke network: a zero–one linear programming approach. Geogr Anal 23(4):283–297

Parvaresh F, Golpayegany H, Husseini S, Karimi B (2013) Solving the p-hub median problem under intentional disruptions using simulated annealing. Netw Spat Econ 13:445–470

Rodríguez-Déniz H, Suau-Sanchez P, Voltes-Dorta A (2013) Classifying airports according to their hub dimensions: an application to the US domestic network. J Transp Geogr 33:188–195

Rosing KE, ReVelle CS (1997) Heuristic concentration: Two stage solution construction. Eur J Oper Res 97(1):75–86

Sa EM, Camargo RS, Miranda G (2013) An improved benders decomposition algorithm for the tree of hubs location problem. Eur J Oper Res 226(4):185–202

Sorensen P, Church R (1995) A comparison of strategies for data storage reduction in location-Allocation problems. National Center for Geographic Information and Analysis Technical Report

Tan P, Kara BY (2007) A hub covering model for cargo delivery systems. Networks 49:28–39

Vidovic M, Zecevic S, Kilibarda M, Vlajic J, Bjelic N, Tadic S (2011) The p-hub model with hub-catchment areas, existing hubs, and simulation: a case study of Serbian intermodal terminals. Netw Spat Econ 11: 295–314

Yu B, Zhu H, Cai W, Ma N, Kuang Q, Yao B (2013) Two-phase optimization approach to transit hub location – the case of Dalian. J Transp Geogr 33:62–71