Chapter 4
Optimal Hub-Spoke Network Design with Hub Reroute and Point-Point Connection: A Physical Perspective with Social Relevance

Guoqiang Shen

4.1 Introduction

A network, consisting of nodes and links, is a structural representation of physical or social relationships. Depending on application domains, nodes can be specifically referred to as agents, actors, markets, cities, etc. or generally as origins, destinations, points, or vertices. Likewise, links are viewed as edges, arcs, connections, or interactions. Network performance is described by nodal or link attributes such as size, length, path, direction, flow, capacity, etc. and measured by congestion, degree, centrality, clustering, density, etc. While graph theory is the branch mathematics behind the network science, it is combinatorial optimization that provides practical tools or algorithms for network applications possible in physical, biological, and social worlds. For example, transportation networks physically move people and goods, communication and social networks digitally transmit bits of information between electronic machines or individuals (organizations).

The hub-spoke (H-S) network is a special yet widely used network typology in both physical and social worlds. Hubs are those nodes that can transship or switch goods, people, or bits between nodes. Spokes are links that connect to hubs. Hubs in a H-S network often connect many more links than non-hub nodes. For examples, of the six representative network configurations in Fig. 4.1, (a) is a point-point (P-P) or fully connected network, allowing direct interaction between any pair of nodes. However, transshipment or flow switching is also possible at any node. For instance, node ① can only reach node ③ by node ②. (b)–(f) are partially connected networks, allowing some disconnections, and hence, hubs. Also, (a)–(b) and (e) may have one hub and (c)–(d) may have two hubs, depending on certain nodal interactions (physical or social).

G. Shen (✉)
The University of Texas, Arlington, TX, USA
e-mail: Guoqiang.shen@uta.edu

(a)   **(b)**   **(c)**   **(d)**   **(e)**
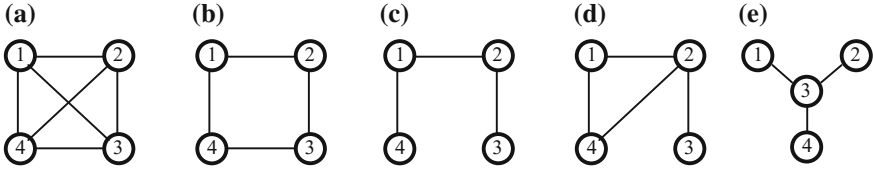
Fig. 4.1 Sample configurations for a hypothetical 4-node networks

The hub location and hub-spoke (H-S) network design problem has been drawing attention from many researchers over the past three decades. A H-S network uses a relatively small number of links (thus, paths) and hubs to serve many interacting origin-destination (O-D) nodes. Express mail delivery and passenger airline in transportation, computer and telecom systems in communication, or social networks in social media (i.e., Facebook, Twitter, Wechat, QQ) are well-known examples with certain H-S configurations. The H-S network design problem is to locate or identify a certain number of hubs out of a known set of potential hub locations and allocate non-hub origins and destinations to the hubs. All routes that visit at least one hub include an origin-to-hub link and a hub-to-destination link, both of which can be regarded as hub-to-nonhub (H-N) links. A route that visits more than one hub also includes at least a hub-to-hub (H-H) link. A link or route from a nonhub origin to a nonhub destination (N-N) is usually not allowed in a pure H-S network. However, in a mixed H-S network, N-N, H-N, and H-H links, or in other words, O-D point-to-point (P-P) connections, are all possible. Almost all social networks today are mixed H-S networks.

The pros of H-S networks can be attributed to: (1) cost discount for flow concentration on a smaller number of H-H links; hence (2) very likely smaller overall investment cost for network implementation and operation; and (3) powerful flow (traffic, data, etc.) control and management functions at hub facilities. The cons of H-S networks lie in: (1) larger investments (mainly fixed and operating costs) at hubs; (2) longer distances or time required for O-D movements routed via hubs than otherwise via P-P connections.

Generally, the following information must be specified for any H-S network design problem: (1) a set of origins, destinations, and potential hub locations; (2) O-D matrices on flows, distances, costs, or time, etc.; and (3) one or more rational network design strategies (i.e., cost minimization, utility and/or reliability maximization).

Although the concept of H-S network has been increasingly utilized in transportation and communication networks and beyond, due to the complexity of the H-S network design problem, operational models with efficient algorithms for determining optimal H-S networks with large O-Ds are yet to be found. On the one hand, most current studies focus on H-S networks with single-hub or multiple-hub without hierarchy, that is, even though multiple hubs are considered, they are treated as at the same level in terms of hub size, function, transshipping capability, so are the H-H links in terms of flow concentration, cost discount rate, and carrying

capacity. On the other hand, the existing studies are based on the cost minimization rationale solely from the hub network investment perspective. O-D flow (traffic, data, etc.) delays generated by hub(s) in a H-S network with respect to its P-P counterpart have been largely ignored. Moreover, many important common network modeling issues, such as network reliability, flow patterns, hub queuing, system optimal and equilibrium, and P-P versus H-S, efficient algorithms, etc. have just been addressed recently and yet to be investigated further.

This paper thus focuses on some of the important issues mentioned previously. Specifically, section two reviews the literature on discrete hub location and network design initialized by O'Kelly (1987). Section three presents several discrete quadratic models which consider flow delay cost generated by routing through hubs, P-P versus H-S network, and mixed hub network design. Section four gives linearized programs to the quadratic models developed in Section three. Conclusions and remarks for future efforts completes this paper.

4.2 Literature Review

4.2.1 Hub Location and Network Design Literature

Since transportation and communication are the major domains for hub location and network design research and applications, the literature review here focuses on these two fields with research from transportation, geography, operations research, management science, and regional science. For comprehensive reviews on the hub location and network design problem, the readers are referred to Campbell (1994a), Campbell et al. (2002) and Campbell and O'Kelly (2012).

In communication, Miehle (1958, p. 232) perhaps is the earliest paper on optimally locating hubs or "communication centers, road junctions, or distribution centers" in a network. Hakimi (1964, 1965) modeled the location of a single switching center in a communication network, showing that its optimal location is always at a network node, and then extended this work to the case of multiple centers. Goldman (1969), analyzing multi-center location and multi-stage (origin-to-center, center-to-center, and center-to-destination) problems in a communication network, recognized the likely lower unit cost of hub-hub (H-H) links and the importance of scale economies. He developed a model to locate n centers in a network while minimizing the total multi-stage transportation cost.

In transportation, Marsten and Muller (1980) developed a mixed-integer program for hub-spoke (H-S) network design and fleet deployment. Their study was probably the first to recognize the nature and advantage of a H-S structure, discussing pure and mixed H-S networks, single and multiple hub allocations, interactions between hubs, and airplane assignments. The deregulation of transportation in America, such as the Air Cargo Deregulation Act in 1977, the Airline

Deregulation Act in 1978, and the Motor Carrier Act in 1980 stimulated the interests in and adoptions of hub networks in the airline, trucking, railroad, and express delivery businesses, whose applied efforts provided cases and motivations for hub location and network design research (Chan and Ponder 1979; Mason et al. 1997; Fisch 2005).

In social networks and social network analysis (Moreno 1956; Granovetter 1976; Granovetter 1983; Wasserman and Faust 1994; Freeman 2004; Scott and Carrington 2011), hubs and hub-spoke networks are discussed more through network centrality, degree, influence, or connection (Freeman 1977, 1979; Friedkin 1991; Kadushin 2012). Research on markets, facilities, and cities as hubs in geographical or urban contexts can be found in (Easley and Kleinberg 2010; Berlingerioa et al. 2011). Selected studies on hierarchical urban system based on the central place theory and related to social networks include Christaller (1966) Brian and Parr (Brian et al. 1988), Batten (1995) and Fujita et al. (1999).

Concerted academic research on hub location and network design formally started with O'Kelly (1986). His seminal piece (O'Kelly 1987) developed the first integer quadratic programming hub model. For given O-D flow and unit transportation cost matrices, this model minimizes the total transportation cost from origin-to-hub, hub-to-hub, and hub-to-destination. One distinct feature of this model is a discount rate associated to H-H links to reflect scale economies due to flow concentration. The integer quadratic program is NP-hard, and has not been solved exactly for large problems. A related quadratic programming model was proposed by Helme and Magnanti (1989) to design satellite communication networks. Campbell (1994) presented the hub location and network design problem as an un-capacitated mixed integer linear program hub location problem. Aykin and Brown (1992) and Aykin (1994) developed a capacitated hub-spoke model allowing for non-hub to non-hub links. Ernst and Krishnamoorthy (1996) made some smaller hub location formulations with linear integer programming and solved for larger problems.

The hub location and network design models reviewed above are variants in one way or another of O'Kelly (1987) and Campbell (1994b) which spurred the development of many alternative formulations and solution heuristics or algorithms, such as Klincewicz (1991), Aykin (1994) Skorin-Kapov et al. (1996), Jaillet et al. (1996) and more recently Bolan et al. (2004), Wagner (2007) Alumur et al. (2009) Contreras et al. (2011) Also, recent reviews on linearization of quadratic integer programming can be found in Eliane et al. (2007) and Fischetti et al. (2012). Although diverse hub location and network design models have been formulated, "the considerably more difficult equilibrium adjustment of interactions to network structure is largely unexamined" (Campbell and O'Kelly 2012, p. 156). Also, "more complex and less idealized hub location models provide strong challenges" (Campbell and O'Kelly 2012, p. 165).

4.2.2   O'Kelly's Quadratic Integer H-S Model

O'Kelly first formulated a quadratic integer program for discrete p-hub facilities location and network design problem. The objective function is to minimize the total transportation investment costs incurred on all links in H-N and H-H sub-networks, including origin-to-hub, hub-to-hub, and hub-to-destination links.

The original notation and problem formulation in O'Kelly's model can be rewritten as:

$$\text{Min}\quad Z = \sum_i \sum_j W_{ij}\left(\sum_k X_{ik}C_{ik} + \sum_m X_{jm}C_{jm} + a\sum_k \sum_m X_{ik}X_{jm}C_{km}\right) \quad (4.1)$$

$$\text{S.t.}\quad (n-p+1)X_{jj} - \sum_i X_{ij} \geq 0 \quad \text{for all } j, \quad (4.2)$$

$$\sum_j X_{ij} = 1 \quad \text{for all } i, \quad (4.3)$$

$$\sum_j X_{jj} = p \quad \text{for all } j, \quad (4.4)$$

$$X_{ij} \in \{0,1\} \quad \text{for all } i,j, \quad (4.5)$$

$$0 \leq a \leq 1 \quad (4.6)$$

where
$$X_{ik} = \begin{cases} 1 & \text{if node } i \text{ is linked to a hub at } k \\ 0 & \text{other} \end{cases}$$
$$X_{jj} = \begin{cases} 1 & \text{if node } j \text{ is a hub} \\ 0 & \text{other} \end{cases}$$

W_{ij} = unit flow from node i and j, exogenously given with $W_{ii} = 0$

C_{ij} = unit transportation cost from node i to j, exogenously given with $C_{ii} = 0$
p, n = the total number of hubs and demand points, respectively.

The objective function contains three terms. The first two terms evaluate the costs of allocating a node to its hub for incoming and outgoing flows respectively. The third term evaluates the costs of H-H link flows. The a is a parameter reflecting the discount policy for interhub connections and flows. Constraint (4.2) ensures that no node is assigned to a location unless a hub is opened at that node. This constraint

requires no N-N links. Constraint (4.3) specifies that each node can only be assigned to one hub. Constraint (4.4) ensures that p hubs be located. Constraint (4.5) specifies that $X_{ik} = 1$ if node i is linked to a hub at k, $X_{ik} = 0$ otherwise.

Of the three terms in the objective function, the first two are linear and represents the interaction between O-D nodes and hubs. The third one, representing the interactions between hubs, contains an integer quadratic part $X_{ik}X_{km}$, which makes the model NP-hard, and thus, very difficult to solve exactly. With given n origin and destination points and p hubs in the model, the exact p-hub location is considered. The O'Kelly's model has n^4 variables, $2n^2 + n$ constraints, and $p(2n - p + 1)/2$ H-N and H-H links, on which flows are considered. However, since the flows on H-H links are concentrated and encouraged, their costs are associated with a discount rate a ($0 \le a \le 1$).

Most models stemmed from O'Kelly's (1987) model stick to the convention of minimizing total distance- or flow-based transportation investment cost. This design strategy can shed some lights on the H-S network design on the one hand, but one the other hand, can also lead to biased "optimal" hub locations, and thus, biased H-S network configurations. The biased design strategy lies in the fact that it only considers minimizing the total transportation investment cost from the investment perspective, rather than minimizing the total system-wide cost from a system optimality consideration, in which the flow delay costs caused by flows' rerouting through hubs should also be included.

Also, most of these studies focus on pure hub networks in which only flows routed by one or more hubs are allowed. Little hub research has been on such fundamental issues as under what condition(s) a H-S network rather than a P-P configuration should be used and vice versa according to the cost minimization rationale, nor has been on mixed H-S networks, in which some parts may have H-S configurations while other parts may take P-P configurations. In reality, a pure H-S network without any N-N or P-P links between some O-D nodes is indeed rarely seen.

As such, in the next section several quadratic optimization models are formulated and corresponding hub network configurations are presented. Specifically, a less biased total cost minimization program considering the system-wide optimality is formulated first. In this program, both the total investment cost and the total flow delay cost (often regarded as social externality cost) are considered. Then, the decisions on whether to use a P-P network or a H-S network are analyzed. Finally, a linearization-based solution strategy is developed to solve the programs formulated. The notations and definitions used in this research are based on the discrete integer quadratic hub facilities location and network design model developed by O'Kelly (1987).

4.3   New Hub Location and Network Design Models

### 4.3.1   System-Wide Optimal Design

Previous research on hub facilities location and network design uses a common objective—to minimize total transportation investment cost, either flow-based or distance-based. This strategy is valid from investment point of view. However, it is biased if assessed from the standing of system-wide optimization, in which the network investment costs and flow delay costs must be considered at the same time and their total be minimized. The typical routing scheme in O'Kelly's model can be shown in Fig. 4.2.

From this diagram we can see that the flow W_{ij} goes from origin i along the bold route containing hub k and m to the destination j. Considering a discount scale for the flow on H-H link L_{km} and the special utilities of hubs (k and m), O'Kelly's model investigates the locations of hubs and the allocations of O-D nodes to the hubs such that the total transportation investment cost is minimized. The resultant H-S network is thus considered to be superior to its direct P-P counterpart.

However, the diagram also shows that $L_{ij} < L_{ik} + L_{km} + L_{jm}$. This suggests that the flow W_{ij} may take more time by the bold route than by the P-P direct path from i to j. This flow delay, of course, generates a cost, which is incurred to the flow (users, cargoes, data, etc.) rerouted via hubs and generally ignored in previous H-S network studies. It appears intuitively that to minimize this flow delay cost, we should use a P-P network rather than a H-S configuration.

However, a closer look at the problem leads us to a dilemma. On the one hand, while a P-P configuration minimizes the transportation time or distance between each O-D pair, it requires more links, and thus, may lead to a higher total investment cost. Moreover, we must ensure that the functions that a H-S network provides

Fig. 4.2 A typical path in O'Kelly's model with a P-P route
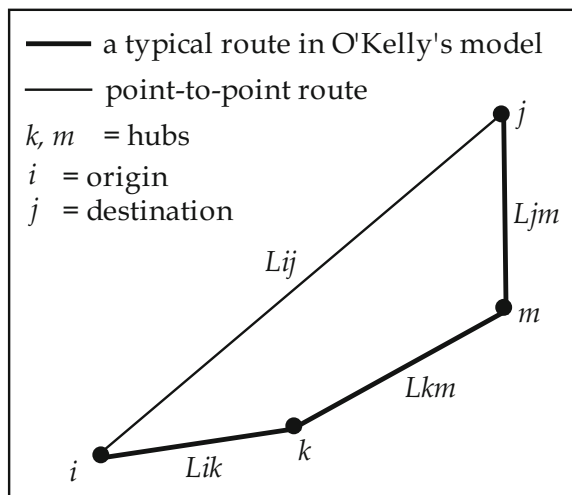
(i.e., sorting, switching, etc.) be equally furnished by the P-P network. To be so, it is necessary that each node in the P-P system be regarded and designed as a hub with certain size and functions on the one hand. These hubs are likely smaller in size and simpler in functions (due to lower levels of traffic concentration at these hubs) as compared with hubs in the H-S network. On the other hand, the H-S network needs much fewer links, and thus fewer paths, to serve the flows of all O-D pairs, but it requires relatively large, functionally sophisticated, and financially expensive hubs.

Intuitively, this dilemma is also reflected in the relationships between the number of hubs (N), the transportation investment cost (TTIC), and the flow delay cost (TFDC) in a H-S network. These two relationships can be graphically represented in Fig. 4.3.

This chart tells us that generally the larger the number of hubs, the higher the total transportation investment cost, but the lower the total flow delay cost. This is particularly true when fixed hub costs are not considered in the optimization process. When fixed hub costs are considered, however, there are trade-offs between the fixed hub costs, transportation cost, flow delay cost, and the optimal number of hubs. This dilemma can only be solved by utilizing some network design strategies that are to compromise the advantages and disadvantages of the two network configurations. One such a strategy is to construct and minimize an objective function which includes the total transportation investment cost and the total flow delay cost.

The following discussion will focus on two design strategies that are based on system-wide optimality considering both the transportation investment cost and flow delay cost. One idea is to minimize the total system-wide cost: the sum of the total transportation investment cost and the total flow delay cost. The other idea is to balance or "equilibrate" TTIC and TFDC in a way that neither the TTIC nor the TFDC is emphasized one over the other. This strategy can be changed slightly into one that minimizes the absolute difference between the TTIC and TFDC. Mathematical programs are formulated using both strategies with consideration of fixed hub costs.
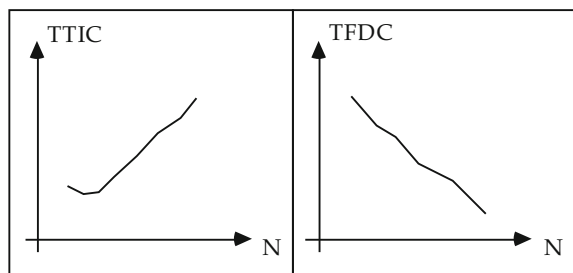
Define:

L_{ij} = Euclidean distance between point i and j,
C_{ij} = unit cost moving a unit flow over a unit distance from i to j
F_{h-s} = fixed cost for a hub in the H-S network

Fig. 4.3 Relationships between TTIC, TFDC, and number of hubs (N)

$L_{ij} = L_{ji}, C_{ij} = C_{ji}, L_{jj} = 0, C_{jj} = 0$ by definition
$i, j, k, m = 1, 2,\ldots, n$

Then, the objective function of O'Kelly's hub facilities location and network design model (4.1) becomes:

$$Z_{h-s} = \sum_i \sum_j W_{ij} (\sum_k C_{ik} X_{ik} L_{ik} + \sum_m C_{jm} X_{jm} L_{jm} + a \sum_k \sum_m C_{km} X_{ik} X_{jm} L_{km})$$
$$+ \sum_i \sum_j X_{ij} F_{h-s}$$

$$(4.7)$$

The user delay in terms of distance for a route from i to j through hub k and m is:

$$D_{ij}^{km} = L_{ik} X_{ik} + L_{km} X_{ik} X_{jm} + L_{jm} X_{jm} - L_{ij} \qquad (4.8)$$

The total flow delay cost is:

$$Z_d = \sum_i \sum_j \sum_k \sum_m W_{ij} C_{ij} D_{ij}^{km}$$
$$= \sum_i \sum_j W_{ij} C_{ij} (\sum_k X_{ik} L_{ik} + \sum_k \sum_m X_{ik} X_{jm} L_{km} + \sum_m X_{jm} L_{jm} - L_{ij}) \qquad (4.9)$$

Therefore, the complete program using the strategy one can be expressed as:

$$\text{Min} \quad Z_{s1} = Z_{h-s} + Z_d = \sum_i \sum_j \sum_k W_{ij} X_{ik} L_{ik} (C_{ij} + C_{ik}) + \sum_i \sum_j \sum_m W_{ij} X_{jm} L_{jm} (C_{ij} + C_{jm})$$
$$+ \sum_i \sum_j \sum_k \sum_m W_{ij} X_{ik} X_{jm} L_{km} (C_{ij} + a C_{km}) + p F_{h-s}$$

$$(4.10)$$

$$\text{S.t.} \quad (n - p + 1) X_{jj} - \sum_i X_{ij} \geq 0 \quad \text{for all } j, \qquad (4.11)$$

$$\sum_j X_{ij} = 1 \quad \text{for all } i, \qquad (4.12)$$

$$\sum_j X_{jj} = p \quad \text{for all } j, \qquad (4.13)$$

$$X_{ij} \in \{0, 1\} \quad \text{for all } i, j, \qquad (4.14)$$

$$0 \leq a \leq 1 \tag{4.15}$$

$$L_{ij} \leq L_{ik} + L_{km} + L_{jm} \quad \text{for all } i, j, \tag{4.16}$$

For strategy two, the absolute difference of the total transportation investment cost and the total flow delay cost can be written as:

$$Z_{s2} = |Z_{h-s} - Z_d| = |\sum_i \sum_j \sum_k W_{ij} X_{ik} L_{ik} (C_{ij} - C_{ik}) + \sum_i \sum_j \sum_m W_{ij} X_{jm} L_{jm} (C_{ij} - C_{jm})$$
$$+ \sum_i \sum_j \sum_k \sum_m W_{ij} X_{ik} X_{jm} L_{km} (C_{ij} - aC_{km}) + pF_{h-s}| \tag{4.17}$$

Suppose Z_s is the smallest, the complete program can be expressed as:

$$\text{Min:} \quad Z_s \tag{4.18}$$

$$\text{S.t.} \quad Z_s < \sum_i \sum_j \sum_k W_{ij} X_{ik} L_{ik} (C_{ij} - C_{ik}) + \sum_i \sum_j \sum_m W_{ij} X_{jm} L_{jm} (C_{ij} - C_{jm})$$
$$+ \sum_i \sum_j \sum_k \sum_m W_{ij} X_{ik} X_{jm} L_{km} (C_{ij} - aC_{km}) + pF_{h-s} \tag{4.19}$$

$$Z_s > -\sum_i \sum_j \sum_k W_{ij} X_{ik} L_{ik} (C_{ij} - C_{ik}) - \sum_i \sum_j \sum_m W_{ij} X_{jm} L_{jm} (C_{ij} - C_{jm})$$
$$- \sum_i \sum_j \sum_k \sum_m W_{ij} X_{ik} X_{jm} L_{km} (C_{ij} - aC_{km}) - pF_{h-s} \tag{4.20}$$

plus: constraints (4.11)–(4.16) as shown above.

The underlying notions of these two models are of total system-wide cost minimization and equilibrium. Since a H-S network and a P-P network are simultaneously considered in both optimization formulations, some parts of the resulting optimal hub networks may be H-S configuration while other parts may be P-P configuration. In other words, the resultant optimal networks from the above two programs are mixed, which allow not only H-N and H-H links but also N-N links, and therefore, are more realistic in reflecting the real world situations.

4.3.2 P-P Network Versus H-S Network

On the one hand, if only link cost is considered, it is clear that a P-P configuration is inferior to its H-S counterpart, judged from the investment cost minimization point

of view. This is simply because that the P-P network requires more links. But on the other hand, if fixed hub cost and transportation cost are also taken into consideration, it is not clear whether the H-S network has advantages (in terms of cost-efficiency) over the P-P network since the total fixed hub cost plus the total transportation cost in the H-S network may be well over its total link cost and not easily be compared with that in the P-P network. This means that when the number of hubs is over a certain threshold, the H-S configuration may be more costly, and thus, inferior to its corresponding P-P network. This suggests an interesting problem: under what condition(s) should the P-P network and/or the H-S network be used according to the total cost minimization rationale? The solutions to this problem will undoubtedly affect hub location and network configuration. The following mathematical model is formulated specifically for this problem.

Define:

A_{ij} = the unit cost for a link from i to j.
$A_{ij} = A_{ji}$ by convention.
F_{p-p} = the fixed cost for a hub at each node in the P-P network.

Assume that the fixed hub cost for each node in the P-P network is smaller than the fixed hub cost in the H-S network, that is, $F_{p-p} < F_{h-s}$. Specifically, assume that the fixed cost for each hub in the P-P network is $1/c$ $(c > 1)$ of the fixed hub cost in the H-S network, or $F_{p-p} = \frac{1}{c} F_{h-s}$. Then, the total fixed hub cost in the P-P network can be written as $TF_{p-p} = \frac{n}{c} F_{h-s}$.

Thus, the total cost including transportation flow cost, fixed hub cost, and link cost in the P-P network is:

$$Z_{p-p} = \sum_i \sum_j W_{ij} C_{ij} L_{ij} + \sum_i \sum_j A_{ij} L_{ij} + \frac{n}{c} F_{h-s} \qquad (4.21)$$

and the total cost including flow transportation cost, fixed hub cost, and link cost in the H-S network is:

$$Z_{h-s} = \sum_i \sum_j \sum_k W_{ij} X_{ik} L_{ik} (C_{ik} + A_{ik}) + \sum_i \sum_j \sum_m W_{ij} X_{jm} L_{jm} (C_{jm} + A_{jm})$$
$$+ \sum_i \sum_j \sum_k \sum_m W_{ij} X_{ik} X_{jm} L_{km} (a C_{km} + b A_{km}) + p F_{h-s}$$

$$(4.22)$$

Here, $b \geq 1$ is assumed for up scaling the link costs for the H-H links. This assumption is reasonable since intuitively the costs for implementing and operating H-H links are higher than other links. $0 \leq a \leq 1$ is for down scaling the transportation costs for H-H links, on which flows are concentrated and encouraged.

Obviously, we need to compare Z_{p-p} and Z_{h-s}. If $Z_{p-p} > Z_{h-s}$, we choose the H-S configuration, otherwise if $Z_{p-p} < Z_{h-s}$, we select the P-P network, and if $Z_{p-p} = Z_{h-s}$, we are indifferent of the H-S and P-P networks. However, this type of

comparison is not trivial since we have parameters p, a, b, c, and F_{h-s}, which need to be exogeously specified or derived in a certain manner. Two ideas may be used to carry out the comparison.

On the one hand, for a specified p, we can always experiment with various sets of reasonable a, b, c, and F_{h-s} values, calculate the corresponding Z_{p-p}'s and Z_{h-s}'s, and identify the relationships between the Z_{p-p}'s and Z_{h-s}'s for appropriate network configurations (P-P, H-S, or a mix of the two). On the other hand, for a given set of values a, b, c, and F_{h-s} the number of hub p can be derived and used for network selection. To do so, we can increment the number of hubs by one each time and minimize (4.22) subject to the set of constraints (4.11)–(4.16). Then, compare each of $Z_{h-s}^1, Z_{h-s}^2, \ldots, Z_{h-s}^{n^*}, \ldots, Z_{h-s}^n$ with the Z_{p-p}. If $Z_{h-s}^{n^*}$ is the closest to Z_{p-p}, then the corresponding p^* specifies the number of hubs that can be used as the threshold for determining the network configuration. Generally, for a large n, if $p > p^*$, it is good to design a P-P network, otherwise, a H-S network is the right network configuration.

To find out p^* using the second idea, we can build an objective function $Z = |Z_{h-s} - Z_{p-p}|$, the absolute difference of Z_{h-s} and Z_{p-p}. Suppose Z_1 is the smallest of all Z, then, mathematically the second idea can be expressed as:

$$\text{Min}: \quad Z_1 \tag{4.23}$$

$$\begin{aligned}
\text{S.t.} \quad Z_1 < &\sum_i \sum_j \sum_k W_{ij} X_{ik} L_{ik} (C_{ik} + A_{ik}) + \sum_i \sum_j \sum_m W_{ij} X_{jm} L_{jm} (C_{jm} + A_{jm}) \\
&+ \sum_i \sum_j \sum_k \sum_m W_{ij} X_{ik} X_{jm} L_{km} (aC_{km} + bA_{km}) + pF_{h-s} \\
&- \left(\sum_i \sum_j W_{ij} C_{ij} L_{ij} + \sum_i \sum_j A_{ij} L_{ij} + \frac{n}{c} F_{h-s} \right)
\end{aligned} \tag{4.24}$$

$$\begin{aligned}
Z_1 > &-\sum_i \sum_j \sum_k W_{ij} X_{ik} L_{ik} (C_{ik} + A_{ik}) - \sum_i \sum_j \sum_m W_{ij} X_{jm} L_{jm} (C_{jm} + A_{jm}) \\
&- \sum_i \sum_j \sum_k \sum_m W_{ij} X_{ik} X_{jm} L_{km} (aC_{km} + bA_{km}) + pF_{h-s} \\
&+ \left(\sum_i \sum_j W_{ij} C_{ij} L_{ij} + \sum_i \sum_j A_{ij} L_{ij} + \frac{n}{c} F_{h-s} \right)
\end{aligned} \tag{4.25}$$

$$(n - p + 1) X_{jj} - \sum_i X_{ij} \geq 0 \quad \text{for all } j, \tag{4.26}$$

$$\sum_j X_{ij} = 1 \quad \text{for all } i, \tag{4.27}$$

$$\sum_j X_{jj} = p \quad \text{for all } j, \tag{4.28}$$

$$X_{ij} \in \{0,1\} \quad \text{for all } i,j, \tag{4.29}$$

$$0 \le a \le 1, \, b \ge 1 \tag{4.30}$$

$$L_{ij} \le L_{ik} + L_{km} + L_{jm} \quad \text{for all } i,j, \tag{4.31}$$

For a given set of a, b, c, and F_{h-s} values, we can solve the program (4.23)–(4.31) up to p times by increasing the number of hubs by one each time from one until p. The corresponding solutions $Z_1^1, Z_1^2, \ldots, Z_1^{P^*}, \ldots, Z_1^P$ can then be compared and the minimal be selected. The number of hubs p^* corresponding to the minimal solution $Z_1^{P^*}$ can then be regarded as the threshold. For any $p > p^*$, the P-P network is the right choice. For any $1 \le p < p^*$, the H-S network is the rational choice. For $p = p^*$, the mixed H-S and P-P should be used.

Notice that for simplicity the above discussion on P-P and H-S networks only focuses on the comparison of a pure H-S network and a P-P network in terms of the total cost (transportation cost, fixed hub cost, and link cost), without being linked to the flow delay cost. In fact, the total flow delay cost can be added to (4.22) as part of the total system-wide cost of the H-S network, which, can then be compared with the total cost (4.21) of the P-P network to determine the appropriate conditions for using the pure P-P network or the H-S network or a mixed configuration of the two.

4.4 Solution Strategies

All the mathematical models developed in Sects. 4.3.1 and 4.3.2 contain one or more zero-one integer quadratic terms, which make these models belong to a family of non-polynomial programs that are very hard to solve exactly when the number of O-D pairs become large. However, linearization heuristics and algorithms are available to solve these programs. Specifically, the two modified algorithms based on Christofides et al. (1980) and Kaufman and Broeckx (1978). For the sake of simplicity, the following solution strategies are based on the second approach discussed and formulated in Shen (1996).

For the discussions on system-wide optimality in Sect. 4.3.1 and P-P versus H-S network in Sect. 4.3.2, we define one binary decision variable for a H-H link as:

$$X_{km} = \begin{cases} 1 & \text{if hub } k \text{ is linked to hub } m \\ 0 & \text{otherwise} \end{cases}$$

Then, Eq. (4.7) in Sect. 4.3.1 can be rewritten as:

$$\text{Min} \quad Z'_{h-s} = \sum_i \sum_j W_{ij} \left(\sum_k X_{ik} C_{ik} L_{ik} + \sum_m X_{jm} C_{jm} L_{jm} + a \sum_k \sum_m X_{km} C_{km} L_{km} \right)$$
$$+ \sum_i \sum_j X_{ij} F_{h-s} \tag{4.41}$$

Similarly, Eq. (4.9) can be rewritten as:

$$Z'_d = \sum_i \sum_j W_{ij} C_{ij} \left(\sum_k X_{ik} L_{ik} + \sum_k \sum_m X_{km} L_{km} + \sum_m X_{jm} L_{jm} - L_{ij} \right). \tag{4.42}$$

Notice that (4.41) and (4.42) are different from (4.7) and (4.9) in that they do not contain any integer quadratic terms. In fact, all terms in (4.41) and (4.42) are in linear form. Therefore, the quadratic objective function (4.10) can be transformed into the following linear expression:

$$Z'_{s1} = Z'_{h-s} + Z'_d \tag{4.43}$$

For constraints, however, not only (4.11) and (4.16) should be kept, but also the following constraint:

$$X_{km} = X_{ik} X_{jm} \tag{4.44}$$

must be added to ensure that at least one of the k and m be allocated on each O-D path. However, (4.44) is still in integer quadratic format, and thus, has to be linearized.

To do so, the integer quadratic term on the right-hand side of Eq. (4.44) should be linearized. The linearization procedure is shown below.

Since:

$$X_{ik}^2 = X_{ik}, \; X_{jm}^2 = X_{jm},$$

$$(X_{ik} - X_{jm})^2 \geq 0, \tag{4.45}$$

we have:

$$(X_{ik}^2 + X_{jm}^2)/2 \geq X_{ik} X_{jm}, \tag{4.46}$$

that is:

$$(X_{ik} + X_{jm})/2 \geq X_{km}. \tag{4.47}$$

Then, after replacing constraint (4.44) by (4.47) and replacing objective function (4.10) by (4.43), we get the following linearized program (A) for the integer quadratic program (4.10)–(4.16).

$$\text{Min} \quad Z'_{s1} = Z'_{h-s} + Z'_d$$
$$\text{S.t.} \quad (11)-(16) \tag{A}$$
$$(47)$$

Similarly, for strategy two, Eq. (4.17) can be written as:

$$Z'_{s2} = |Z'_{h-s} - Z'_d| = |\sum_i \sum_j \sum_k W_{ij} X_{ik} L_{ik}(C_{ij} - C_{ik}) + \sum_i \sum_j \sum_m W_{ij} X_{jm} L_{jm}(C_{ij} - C_{jm})$$
$$+ \sum_i \sum_j \sum_k \sum_m W_{ij} X_{km} L_{km}(C_{ij} - aC_{km}) + pF_{h-s}| \tag{4.48}$$

Suppose Z'_s is the smallest of Z'_{s2}, then, the linearized program (B) using strategy two discussed in Sect. 4.3.1 for the system-wide optimality can be written as:

$$\text{Min} \quad Z'_s$$
$$\text{S.t.} \quad Z'_s \geq Z'_{h-s} - Z'_d$$
$$\qquad Z'_s \leq -Z'_{h-s} + Z'_d \tag{B}$$
$$\qquad (11)-(16)$$
$$\qquad (47)$$

Similarly for P-P versus H-S analysis in Sect. 4.3.2, the Eq. (4.22) can be written as:

$$Z'_{h-s} = \sum_i \sum_j \sum_k W_{ij} X_{ik} L_{ik}(C_{ik} + A_{ik}) + \sum_i \sum_j \sum_m W_{ij} X_{jm} L_{jm}(C_{jm} + A_{jm})$$
$$+ \sum_i \sum_j \sum_k \sum_m W_{ij} X_{km} L_{km}(aC_{km} + bA_{km}) + pF_{h-s} \tag{4.49}$$

To find out p^*, we build an objective function using Eqs. (4.49) and (4.21):

$$Z' = |Z'_{h-s} - Z_{p-p}| \tag{4.50}$$

Assuming that Z'_1 is the smallest of all Z', then, mathematically the program (4.23)–(4.31) can be reformulated into the following linearized program (C):

$$\text{Min} \quad Z'_1$$
$$\text{S.t.} \quad Z'_1 \leq Z'_{h-s} - Z_{p-p}$$
$$\qquad Z'_1 \geq -Z'_{h-s} + Z_{p-p} \tag{C}$$
$$\qquad (26)-(31)$$
$$\qquad (47)$$

4.5   Conclusions and Remarks

Hubs and Hub-spoke networks can be found in virtually all physical or social networks. Hubs connect many spokes, transship people, goods, or bits at a large volume, and incur a reroute delay or cost. P-P connections allow for direct interactions, which may generate a large number of links with small O-D interactions. This paper extends O'Kelly (1986, 1987) models for optimal hub-spoke network designs with hub reroutes and P-P connections and provides an algebraic linearization.

This paper, instead of using the biased transportation investment cost minimization convention for the H-S network design problem, proposed and mathematically constructed the system-wide optimality concept, which considers both transportation investment cost and flow delay cost. Since the flow delay phenomena is common to all H-S networks due largely to flow consolidation on H-N/N-H or H-H links, it is indeed worth of investigating. Solving the H-S network design model formulated for the system-wide optimality will result in a mixed H-S network in which some parts may have H-S configuration while other parts may take P-P configuration.

A H-S network may not be superior to its P-P counterpart if the total cost including transportation cost, fixed hub cost, and link cost is minimized. Thus, a program was developed to derive the favorable conditions for a P-P, H-S, or a mix of the two. Generally, most social networks are mixed H-S networks, with the degree of mixture varies according to the link/hub cost, flow, and capacity.

The programs developed in this paper are integer quadratic programs, which are NP-hard and very difficult to solve exactly for a large number of O-D pairs. Therefore, a linearization strategy based on the binary nature of the decision variables was developed to solve these programs. However, the validity of these programs and the linearization strategy should be verified by future studies, particularly those application oriented studies for real world problems, such as in transportation and social networks.

Immediate future studies should also be on the time impact on the flow delay cost, and thus, on the hub network modeling. This is because that on the one hand the flow delay cost is actually affected by the life cycle of the hub network. In general, within the life cycle the total fixed hub and link costs increase occasionally and slightly verse the total flow delay cost which is accumulated steadily through O-D distances and flows. The longer the system's life cycle, the larger the total flow delay cost. On the other hand, the total flow delay cost is also affected by the flow concentrations on individual links. Some H-H links generate more flow delay costs than others, therefore, the higher frequencies of flow using the H-H links with higher levels of flow concentrations, the larger the total flow delay cost generated. Finally, the behavioral variations and optimal configuration patterns of a H-S network when applied to physical networks (i.e., highway freight logistics or passenger transportation) or social networks (i.e., web chat messages and direct sales) need to be captured for better domain applications.

References

Alumur, S., Kara, B. Y., & Karasan, O. E. (2009). The design of single allocation incomplete hub networks. *Transportation Research Part B, 43*(10), 936–951.

Aykin, T., & Brown, G. F. (1992). Interacting new facilities and location allocation problems. *Transportation Science, 26*(3), 212–222.

Aykin, T. (1994). Lagrangian relaxation based approaches to capacitated hub-spoke network design problem. *European Journal of Operational Research, 79,* 501–523.

Batten, D. F. (1995). Network cities: Creative urban agglomerations for the 21st century. *Urban Studies, 32*(2), 313–327.

Berlingerio, M., Coscia, M., Giannotti, F., Monreale, A., & Pedreschi, D. (2011). The pursuit of hubbiness: Analysis of hubs in large multidimensional networks. *Journal of Computational Science, 2*(3), 223–237.

Boland, N., Krishnamoorthy, M., Ernst, A. T., & Ebery, J. (2004). Preprocessing and cutting for multiple allocation hub location problems. *European Journal of Operational Research, 155*(3), 638–653.

Brian, J., Berry, L., & Parr, J. B. (1988). *Market centers and retail location-theory and application*.

Campbell, J. F. (1994a). A survey of network hub location. *Studies in Locational Analysis, 6,* 31–49.

Campbell, J. F. (1994b). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research, 72*(2), 387–405.

Campbell, J. F., Ernst, A. T., & Krishnamoorthy, M. (2002). Hub location problems. In Z. Drezner & H. Hammacher (Eds.), *Facility location: Applications and theory*. Berlin: Springer.

Campbell, J. F., & O'Kelly, M. E. (2012). Twenty-five years of hub location research. *Transportation Research Part B, 46*(2), 153–169.

Chan, Y., & Ponder, R. (1979). The small package air freight industry in the United States: A review of the Federal Express experience. *Transportation Research Part A, 13*(4), 221–229.

Christaller, W. (1966). *Central places in southern Germany*. Upper Saddle River: Prentice-Hall.

Christofides, N., Mingozzi, A., & Toth, P. (1980). Contributions to the quadratic assignment problem. *European Journal of Operational Research, 4,* 243–247.

Contreras, I., Cordeau, J.-F., & Laporte, G. (2011). Benders decomposition for large-scale un-capacitated hub location. *Operations Research, 59*(6), 1477–1490.

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge: Cambridge University Press.

Eliane, M. M., Nair, A., Paulo, O. B.-N., & Tania, Q. (2007). A survey of the quadratic assignment problem. *European Journal of Operational Research, 176*(2), 657–690.

Ernst, A. T., & Krishnamoorthy, M. (1996). Efficient algorithms for the uncapacitated single allocation *p*-hub median problem. *Location Science, 4*(3), 139–154.

Freeman, L. C. (2004). *The development of social network analysis. A study in the sociology of science*.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40*(1), 35–41.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks, 1*(3), 215–239.

Friedkin, N. E. (1991). Theoretical foundations for centrality measures. *American Journal of Sociology, 96*(6), 1478–1504.

Fisch, J. E. (2005). How do corporations play politics? The FedEx story. *Vanderbilt Law Review, 58*(5), 1495–1570.

Fischetti, M., Monaci, M., & Salvagnin, D. (2012). Three ideas for the quadratic assignment problem. *Journal of Operations Research, 60*(4), 954–964.

Fujita, M., Krugman, P., & Mori, T. (1999). On the evolution of hierarchical urban systems. *European Economic Review, 43*(2), 209–251.

Goldman, A, J. (1969). Optimal location for centers in a network. *Transportation Science*, *3*(4), 352–360.

Granovetter, M. (1976). Network sampling: Some first steps. *American Journal of Sociology, 81,* 1287–1303.

Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory, 1*(1), 201–233.

Hakimi, S. L. (1964). Optimal locations of switching centers and the absolute centers and medians of a graph. *Operations Research, 12*(3), 450–459.

Hakimi, S. L. (1965). Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations Research, 13*(3), 462–475.

Helme, M. P., & Magnanti, T. L. (1989). Designing satellite communication networks by zero-one quadratic programming. *Networks, 19,* 427–450.

Jaillet, P., Song, G., & Yu, G. (1996). Airline network design and hub location problems. *Location Science, 4* (3), 195–211.

Kadushin, C. (2012). *Understanding social networks: Theories, concepts, and findings*. USA: Oxford University Press.

Kaufman, L., & Broeckx, F. (1978). An algorithm for the quadratic assignment problem using bender's decomposition. *European Journal of Operational Research, 2,* 207–211.

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

Marsten, R. E., & Muller, M. R. (1980). A mixed-integer programming approach to air cargo fleet planning. *Management Science, 26*(11), 1096–1107.

Miehle, W. (1958). Link-length minimization in networks. *Operations Research, 6*(2), 232–243.

Moreno, J. L. (Ed.). (1956). *Sociometry and the science of man*. Beacon House.

O'Kelly, M. E. (1986). The location of interacting hub facilities. *Transportation Science, 20*(2), 92–106.

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

Scott, J., & Carrington, P. J. (2011). *The SAGE handbook of social network analysis*. SAGE publications.

Skorin-Kapov, D., Skorin-Kapov, J., & O'Kelly, M. E. (1996). Tight linear programming relaxations of uncapacitated *p*-hub median problems. *European Journal of Operational Research, 94*(3), 582–593.

Shen, G. (1996). A path-based hub-and-spoke network design model and its linearization. Conference Proceedings of the 1996 IEEE 15th Annual International Phoenix Conference on Computers and Communications, Scottsdale, AZ, USA. https://ieeexplore.ieee.org/document/493655/

Wagner, B. (2007). An exact solution procedure for a cluster hub location problem. *European Journal of Operational Research, 178,* 391–401.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge University Press: Cambridge.

Guoqiang Shen is a tenured full professor in College of Architecture, Planning, and Public Affairs, the University of Texas at Arlington (UTA). He once taught in University of Oklahoma, University of North Carolina, and Beijing Institute of Civil Engineering and Architecture. His interests include architecture design, physical planning, urban economics and real state, transportation and logistics, GIS and applied OR, organizational behaviors, risk and resiliency, and international comparative urban issues. Shen received his BARCH from Tsinghua University, MLA from Beijing University, and Ph.D. in Urban Planning from the Ohio State University. He is a balanced planning scholar, practitioner, and educator.