CLUSTER-BASED OPTIMIZATION OF URBAN TRANSIT HUB LOCATIONS: METHODOLOGY AND CASE STUDY IN CHINA

Jie Yu
Department of Civil Engineering
The University of Maryland, College Park, MD 20742
yujie@umd.edu

Yue Liu
Department of Civil Engineering
The University of Maryland, College Park, MD 20742
troybest@umd.edu

Gang-Len Chang
Department of Civil Engineering
The University of Maryland, College Park, MD 20742
gang@umd.edu

Xiao-guang Yang
Department of Traffc Engineering
Tongji University, Shanghai, PRC 200092
yangxg@mail.tongji.edu.cn

Word Count: 3941+250*11=6691

Cluster-based Optimization of Urban Transit Hub Locations: Methodology and Case Study in China

Jie Yu, Yue Liu, Gang-Len Chang and Xiao-guang Yang

Abstract. Choosing proper locations of urban transit hubs has always been one of the critical concerns facing urban transportation planning agencies in China. This study proposes a mixed integer optimal location model for urban transit hubs, with the objective to minimize the demand-weighted total travel time, when explicitly taking into account Traffic Analysis Zones (TAZs) as demand origins or destinations in a target urban area. An Integer Non-linear Programming (INLP) reformulation was developed to significantly reduce the number of variables. Bilinear constraints in the proposed INLP formulation were then re-modeled into linear functions to ensure global optimal solutions obtained. The model was successfully applied to optimize the hub locations in Suzhou Industrial Park, China, with the result of significantly improved system performance. The impacts of several critical factors, such as the number of hubs and the travel time discount coefficient on the system performance are also investigated.

1. INTRODUCTION

In a hub network, centrally located service facilities serve as the hubs. Traffic flows from various origins to destinations are consolidated to hubs, and after regrouping, leave the hub facilities bound either to other hubs or to their ultimate destinations *(1)*. Compared with a completely-interconnected network, the characteristic of such a hub network is the bundling of flows on the inter-hub links, resulting in more efficient utilization of the corresponding transportation resources. Transit hub location planning is a branch of the hub location problems. Since O'Kelly *(2)* first formulated a quadratic single assignment model for interacting hub facilities from an operations research point of view, this work has attracted the attention of researchers from such a variety of fields as telecommunications, airline passenger management and logistics etc..

In review of the literature, studies of hub location problems are concentrated mainly on two basic models depending on how non-hub nodes are connected to the hubs *(3)*. In the single assignment model each node is connected to a single hub *(2)*, and there is no sorting at the origin because all flow must travel to the same hub. However, the multiple-assignment model allows each node to be connected to more than one hub, and sorting must occur at each origin that interacts with more than one hub *(4)*. With the objective of minimizing the total travel cost, these two basic models require all services between the non-hub nodes to be hub connected, which is known to be strict hubbing policy. To deal with more realistic characteristics of hub networks, researchers discussed different extensions including a fixed cost added into the objective function so that the tradeoffs between travel costs and fixed costs are considered *(5)*, a capacity constraint incorporated in the model by limiting the flows entering a hub under its capacity *(6)*, and the non-restrictive hubbing policy which allows every pair of nodes to interact directly with each other *(1)*. Sung et al. *(7)* proposed a cluster-based hub location model with non-restrictive policy. In their model, exactly one hub location is assigned to a cluster to be opened, and traffic flows between nodes can either be routed directly or via hubs. Besides, a very recent study by Kim et al. *(8)* developed a hub network design model considering economies of scale stemming from the consolidation of traffic flows. The model is designed to determine the cost-decreasing effect due to the consolidation of flows endogenously and to include several cost components and capacity constraints relative to the uniqueness of hub networks.

Due to the nondeterministic polynomial time complete (NP-complete) property of the hub location problems, many researchers have devoted great attention to the development of efficient algorithms. O'Kelly *(2)* was the first to develop two heuristics that were able to determine optimal locations given a possibly suboptimal allocation pattern for the single assignment model. Moreover, lots of heuristics such as simulated annealing *(9)*, genetic algorithms *(10)*, tabu search *(11)*, and neural networks *(12)* have been used for near optimal solutions. Campbell *(4)* initially accomplished the linearization of the quadratic hub location model, which makes global optimal solutions possible. In order to avoid fractional solutions of Campbell's model, some researchers *(13-15)* further modified this linear model by eliminating redundant and impractical routes and by exploiting the symmetry of the available test data. These modifications reduced computation time and the number of variables without sacrificing integrality. Recently, Sung et al. *(7)* proposed a dual-based approach for a hub network design problem under non-restrictive policy, which consists of a dual ascent procedure and a dual adjustment procedure. Numerical experiments from multiple data sets have proved the efficiency of their algorithm in obtaining good solutions.

In the developing China, traffic congestion has emerged as a critical during the process of urbanization. More and more researchers have realized that development of transit-oriented urban transport systems is one of the most effective strategies to relieve the traffic congestion in China. In recent years, many big cities in China are dedicated to proposing policies and measures for developing efficient public transport systems from both planning and operation perspectives. Transit hubs are the fundamental facilities in the urban transit system, which are designed to provide switching points for inter-modal flows and to feature seamless pedestrian connections. Properly located transit hubs will significantly improve the operational efficiency of limited transportation resources and the quality of transit services. Therefore, transit hub locating problem usually serves as the basis and the first step during the urban transit planning process. However, despite the increasing needs for development of transit-oriented urban transport systems in China, the critical issue of determining optimal transit hub locations has not been adequately addressed yet in the literature. Thus, along the line of previous location researches, this study will focus on the following critical issues:

- *Formulate a cluster-based urban transit hub location model, based on the formulation by Sung et al. (7), with the objective to minimize the demand-weighted total travel time, when explicitly taking into account Traffic Analysis Zones (TAZs) as demand origins or destinations;*
- *Design an efficient solution approach to the proposed model to make it tractable for large-scale real-world applications; and*
- *Test the applicability of the proposed model through an illustrative case to assist planners in best understanding and applying the proposed model during the planning process;*

This paper is organized as follows. Next section will list the assumptions for formulating the hub location optimization model. With these preparation efforts, Section 3 presents the detailed formulation of the transit hub location model, including the objective function as well as operational constraints. A new Integer Non-linear Programming (INLP) formulation of the proposed model and its bilinear constraint processing are presented in Section 4 to make it tractable for general Mixed Integer Programming (MIP) solvers. Section 5 shows results of a case study in Suzhou Industrial Park, China that demonstrate the applicability of the proposed model. Concluding comments along with future extensions are reported in the last section.

2. MODEL ASSUMPTION

In this paper, we study the discrete location problem for urban transit hubs of a given study area, which can be divided into different TAZs. The demand origins and destinations, as well as the hubs are assumed to occur only at the centroids of those TAZs, denoted as nodes. To ensure that the proposed formulations for optimization of hub locations can be tackled and also realistically reflect the real-world constraints, this study has employed the following five assumptions in the model formulation.

Assumptions 1: Single hub allocation policy

More specifically, all nodes are partitioned into clusters in advance based on the geographic relations between neighboring TAZs, land use restrictions or other political and administrative reasons, as shown in Figure 1 (This situation often exists in the real world, since control and management on local networks are heavily influenced by the associated local communities). For any cluster, only one node has to be selected as a hub and all other nodes are assigned to this hub. Moreover, all the hubs are assumed to be fully inter-connected.

Assumptions 2: Non-restrictive policy, which means flows between origins and destinations may be sent either directly or through hub(s), and the number of hub-stops is no more than two.

Under the non-restrictive policy, if a node is assigned to a hub, any flows to or from this node have to go via the hub, or do not involve hubs at all (non-stop service). Therefore, the paths from an origin node i to a destination node j could have three possible choices, as shown in Figure 2: (1) nonstop: transit flows are transported directly from i to j; (2) one-hub stop: transit flows are transported from i to j with the hub k as the transfer point; (3) two-hub stop: transit flows are transported from node i to j via both hub k and m along the route $i \to k \to m \to j$.

Assumption 3: The transit network is composed of two types of links (see Figure 1): arterial links (connection from hubs to hubs) and branch links (connection from hubs to nodes or from nodes to nodes directly), and There exists a traffic time discount coefficient for the arterial link.

Assumption 4: The number of hubs and clustering rules are pre-determined.

Assumption 5: The transfer time (including walking time and waiting time) at hubs is assumed to be constant.

3. MODEL FORMULATION

Previous studies by Sung et al. *(7)* proposed a hub network design problem, which is called cluster-based hub location problem. Along the line of their research, this study applies the cluster-based concept in designing the urban transit network.

Notations

To facilitate the following illustration, all definitions and notations used hereafter are summarized below.

Parameters and Sets:

G : Target transit network;
N : Set of TAZs (origin or destination nodes);
i : Index of each TAZ (node);
(i, j) : Link (route) between node i and j ;
p : The number of clusters (hubs) in the target network;
C_r : Cluster r in the target network ($r = 1,\cdots, p$);
C_i : The cluster to which TAZ i is assigned to ($i = 1,\cdots, N$);
w_{ij} : Flows from node i to j (unit: trips);
α : The travel time discount coefficient between hubs;
t_{ij} : Non-stop average travel time from node i to j (unit: min);
t_k : Transfer time at hub k (unit: min);
t_{ijkm} : Hub-stop average travel time from node i to j via hubs k and m (unit: min),

$t_{ijkm} = t_{ik} + t_k + \alpha t_{km} + t_m + t_{mj}$ ($t_{km} = 0$, if $m = k$);

Model Variables:

Three sets of binary decision variables x_{ij} , x_{ijkm} and y_k are defined:

$$x_{ij} = \begin{cases} 1 & \text{if flows between nodes } i \text{ and } j \text{ are transported with non-stop} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{ijkm} = \begin{cases} 1 & \text{if flows between node } i \text{ and } j \text{ are transported with one or} \\ & \text{two hub stops } (k = m, \text{one hub-stop}; k \neq m, \text{two hub-stops}) \\ 0 & \text{otherwise} \end{cases}$$

$$y_k = \begin{cases} 1 & \text{if a hub is located at node } k \\ 0 & \text{otherwise} \end{cases}$$

Formulation

The location model for transit hubs is formulated as follows:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} \left(t_{ij} x_{ij} + \sum_{k \in C_i} \sum_{m \in C_j} t_{ijkm} x_{ijkm} \right) w_{ij} \tag{1}$$

s.t.

$$\sum_{k \in C_r} y_k = 1 \qquad r = 1,\cdots, p \tag{2}$$

$$x_{ij} + \sum_{k \in C_i} \sum_{m \in C_j} x_{ijkm} = 1 \qquad \forall i, j \tag{3}$$

$$\sum_{m \in C_j} x_{ijkm} \le y_k \qquad \forall j,k; i \in C_k \tag{4}$$

$$\sum_{m \in C_i} x_{ijkm} \le y_m \qquad \forall i,m; j \in C_m \tag{5}$$

$$x_{ij}, x_{ijkm}, y_k \in \{0,1\} \qquad \forall i, j, k, m \tag{6}$$

The objective function (1) aims to minimize the demand-weighted total travel time in the system. Constraint (2) limits that only one of the nodes in each cluster should be selected as a hub. Constraint (3) means that flows between node i and j are transported either via the non-stop service or the hub-stop service. Constraints (4) and (5) prevent any hub services unless the node is selected as a hub. Constraint (6) is a standard integrality constraint.

4. SOLUTION APPROACH

In this section, this study presents a new INLP formulation to significantly reduce the number of variables in the original model by introducing a set of new variables and bilinear constraints. Also, the bilinear constraints in the proposed INLP formulation were re-modeled by linear functions to ensure global optimal solutions obtained.

The New INLP Formulation

Note that, the original model formulation (*Eq.* (1) – (6)) could result in a huge number of variables when the number of nodes is large, due to the high dimensionality of x_{ijkm}. For example, for a regularly normal instance with 50 nodes and 8 clusters, the number of binary variables and constraints is more than 100 million, which will make the solving process inefficient and time consuming. However, if we observe the model formulation in an alternative way, we can find that all routes between two clusters or within a cluster can be easily determined if the hub is determined for those clusters. As shown in Figure 3, if node k and m ($k < m$ and $C_k \neq C_m$) are selected as hubs, demand-weighted total travel time to transport flows from i to j is:

$$T_{ij} = \min\{w_{ij}t_{ij}, w_{ij}(t_{ik} + t_k + \alpha t_{km} + t_m + t_{mj})\} + \min\{w_{ji}t_{ji}, w_{ji}(t_{jm} + t_m + \alpha t_{mk} + t_k + t_{ki})\} \tag{7}$$

Then, the total demand-weighted total travel time between clusters C_k and C_m will be:

$$T^{km} = \sum_{i \in C_k} \sum_{j \in C_m \cap k<m, C_k \neq C_m} T_{ij} \tag{8}$$

For the special case with $k = m$, T^{km} becomes T^k :

$$T^k = \sum_{i \in C_k} \sum_{j \in C_k} T_{ij} \tag{9}$$

Variables δ_{km} ($k < m$ and $C_k \neq C_m$) are introduced to replace the variables x_{ij} and x_{ijkm}. They take a value of 1 if and only if hubs are set in both nodes k and m. Then, the original model can be transformed into the following INLP:

$$\min \sum_{k<m \cap C_k \neq C_m} T^{km}\delta_{km} + \sum_{k=1}^{n} T^k y_k \tag{10}$$

s.t.

$$\sum_{k\in C_r} y_k = 1 \qquad r = 1,\cdots,p \tag{11}$$

$$\delta_{km} = y_k \cdot y_m \qquad k < m, C_k \neq C_m \tag{12}$$

$$y_k, y_m \in \{0,1\} \qquad k < m \tag{13}$$

The objective function *Eq.* (10) did the same job as the original one *Eq.* (1) with consideration of all the demand-weighted travel time between different clusters and within any cluster. Constraints *Eq.* (11) are the same as constraints *Eq.* (2). Constraints *Eq.* (12) ensure that δ_{km} is 1 if and only if y_k and y_m are both equal to 1. After the transformation, the number of decision variables can be significantly reduced. For the same instance with 50 nodes and 8 clusters, only thousands of variables are involved. However, the original problem was transformed into a INLP with bilinear constraints, as shown in *Eq.* (12). Therefore, it is hard to get the global optimal solutions because they combine all the difficulties of both of their subclasses: the combinatorial nature of MIP and the difficulty in solving non-convex (and even convex) nonlinear programs (NLP).

Bilinear Constraint Re-modeling

To address the difficulty in obtaining global optimal solutions for the above INLP formulation, this section re-modeled its bilinear constraints Eq. (12) into the following linear form functions:

$$\delta_{km} \geq 0 \tag{14}$$

$$\delta_{km} \leq y_m \tag{15}$$

$$\delta_{km} \geq y_k + y_m - 1 \tag{16}$$

$$\delta_{km} \leq y_k \tag{17}$$

Note that, *Eq.* (14) – (17) can always assure *Eq.* (12) hold. The INLP is transformed to the following MIP:

$$\min \sum_{k<m \cap C_k \neq C_m} T^{km}\delta_{km} + \sum_{k=1}^{n} T^k y_k \tag{18}$$

s.t.

$$\sum_{k\in C_r} y_k = 1 \qquad r = 1,\cdots,p \tag{19}$$

$$\delta_{km} \leq y_m \qquad k < m, C_k \neq C_m \tag{20}$$

$$\delta_{km} \geq y_k + y_m - 1 \qquad k < m, C_k \neq C_m \tag{21}$$

$$\delta_{km} \leq y_k \qquad k < m, C_k \neq C_m \tag{22}$$

$$y_k, y_m \in \{0,1\} \qquad k < m \tag{23}$$

The new MIP has significant few variables and constraints that could be solved with the existing MIP solvers.

5. CASE STUDY

This section presents the application of the proposed location model and solution approach for Suzhou Industrial Park, China. This case study intends to assist transit system planners in best understanding and applying the proposed model. The presentation hereafter will include the following parts:

- *Detailed description of the study network, including TAZ information, clustering rules, demand distribution, and average travel time between zones;*
- *Model inputs, optimization results, and performance analyses;*
- *Assessment of the impact of the travel time discount coefficient and hub numbers on the optimization results for the given network;*

Study Network

The proposed model and solution approach was applied to determine the optimal location of transit hubs in Suzhou Industrial Park, China. The network is divided into 58 TAZs, as shown in Figure 4.

Considering political as well as administrative factors, the entire Industrial Park consists of one downtown area, four suburban areas, and one college park. Moreover, from a geographic point of view, the downtown area is further divided by the Jinji Lake and the Xinhua Expressway into three parts, denoted as clusters 1, 3, and 4 in Figure 5. The four suburban areas are represented by clusters 2, 6, 7, and 8, and the college park falls into the cluster 5. Each of the clusters will contain exactly one hub. The clustering rules of the study network and the distribution of TAZs among clusters are shown in Table 1 and Figure 5.

Model Inputs, Outputs and Analysis

In order to implement the proposed model for optimal transit hub locations in the study network, the following information should be available as inputs:

- *The O-D matrix of the study network;*
- *The average travel time matrix between any pair of TAZs;*
- *The travel time discount coefficient for arterial links, denoted by α, here we use $\alpha = 0.5$;*
- *The number of hubs and clustering rules discussed in the above section;*
- *Constant transfer time at hubs, here we use 3 mins;*

Based on the input information, responsible agencies can then use our proposed model to obtain the optimal hubbing policy, which includes the following three types of information:

- *System MOE, i.e. the total demand-weighted travel time of the study network;*
- *The list of opened hubs as well as their locations and scales (flows in-and-out)); and*
- *The route assignment of transit flows between TAZs;*

The proposed model was implemented in the LINGO MIP Solver, and the optimal locations (TAZ IDs) of opened hubs are shown in Figure 6. The total demand-weighted travel time under the optimal hubbing policy is 74,050 *hrs*, and it has been reduced about 13% compared with the system performance of 84,731 *hrs* without the hubbing policy. Furthermore, based on the optimal locations of hubs, the route assignment of each transit OD pair can be easily determined from the model, which serves as the basis for generating the scales of the opened hubs, as shown in Table 2. This information will assist planners to determine the resources allocated to each hub to ensure it operate under its flow capacity.

As indicated in Table 2, hubs located at clusters 7, 17, and 21 in the downtown area accommodate much larger transit flows than other hubs in the study network. Therefore, more resources should be allocated to design these hubs at a higher level.

Impact of the Number of Clusters and α Values

Based on the geographic and political interrelations between neighboring TAZs, this case study has partitioned the study network into 8 clusters, which require 8 hubs opened. However, during the real-world planning process, budget constraints are always among the most critical factors affecting the planning results. The cost for opening and operating 8 hubs may already exceed budget limit in this case. Thus, planners need to find a new clustering rule with less opened hubs and its corresponding optimal hubbing policy to best fit into the budget requirements. However, with the decrease of the number of opened hubs, the economies of scale from consolidation of transit flows will be reduced, so a trade-off exists between the above two aspects. Also revealed as a critical factor, the travel time discount coefficient α can significantly affect the willingness of people to make transfers at hubs, and further influence the performance of the hubbing policy.

Considering the potential sensitivity of the system performance to the above two critical factors, this section tested the system performance under different scenarios with various combinations of hub numbers and α values. Table 3 shows the clustering rules of the transit network under different number of hub numbers. Table 4 shows the objective function values under hub numbers p from 2 to 8, and α values from 0.5 to 0.9. The varying of objective function values with different p and α values and a comparison with no hubbing policy are shown in Figure 7.

As indicated in Table 4 and Figure 7, one can reach the following findings:

- *For a given value of α, the objective function will decrease when the number of opened hubs increases. Notably, there exists a threshold of p, below which setting hubs has no advantage. For example, given $\alpha = 0.5$, when $p \leq 4$, the objective function values of setting hubs are even higher than setting no hubs. Only after $p \geq 5$, the hubbing policy begins to outperform. This information will help planners find the best number of hubs and clustering rules to both satisfy the budget constraints and maximize the economies of scale from consolidation of transit flows;*
- *For a given value of p, the objective function will increase when the value of α increases. Also, there exists a threshold of α, above which setting hubs has no advantage. For example, given $p = 6$, when $\alpha \geq 0.8$, the objective function values of setting hubs are even higher than setting no hubs. Only after $\alpha < 0.8$, the hubbing policy begins to outperform. This information will provide guidelines to planners in designing speeds for transit routes.*

6. CONCLUSIONS

This study has presented a cluster-based transit hub location optimization model that explicitly takes into account TAZs as demand origins or destinations in the network. With the objective of minimizing the network total demand-weighted travel time, the proposed model generates not only the optimal locations of opened hubs in the target network, but also determines the scales of those hubs to assist planners for proper design and resource allocations.

To solve the proposed model in cost-effective way, a new INLP reformulation was designed to efficiently reduce the number of variables in the original model formulation, and the bilinear constraints in the INLP were then re-modeled by linear functions to make the problem easily handled by existing MIP solvers.

The model was successfully applied to designing the hub network for Suzhou Industrial Park, China, with the result of significantly improved system performance. Furthermore, the impacts of various hub numbers and travel time discount coefficients on the system performance were evaluated. The results from the impact analyses can provide critical information for planners to choose the best number of hubs and clustering rules, as well as to properly design the transit routes.

To deal with more realistic situations, further research along this line will be focused on the following critical issues:

- *The costs for building and operating hubs can be integrated into the objective function so that the budget constraint is considered;*
- *The limitation of hub capacities can also be incorporated in the model by adding a capacity constraint;*
- *A hierarchical hub location model can be formulated with detailed consideration of different hub types and their corresponding interactions; and*
- *Actual traffic conditions and their impacts on the route choices of transit flows in the network should be taken into account in the model.*

REFERENCES

1. Aykin, T. Networking Policies for Hub-and-Spoke Systems with Applications to the Air Transportation System. *Transportation Science*, Vol. 29, 1995, pp. 201-221
2. O'Kelly M.E. A Quadratic Interger Program for the Location of Interacting Hub Facilities. *European Journal of Operational Research*. Vol. 32, 1987, pp. 393-404
3. Deborah B. and O'Kelly M.E. Hub-and-spoke Networks in Air Transportation: An Analytical Review. *Journal of Regional Science*. Vol. 39, 1999, pp. 275-295
4. Compbell J.F. Interger Programming Formulations of Discrete Hub Locaton Problems. *European Journal of Operational Research*. Vol. 72, 1994, pp. 387-405
5. O'Kelly M.E. Hub Facility Location with Fixed Costs. *Papers in Regional Science: The Journal of the Regional Science Association International*. Vol. 71, 1992, pp. 293-306
6. Rodriguez V., Alvarez M.J. and Barcos L. Hub Location under Capacity Constraints. *Transportation Research Part E*. Vol. 43, 2007, pp. 495-505.
7. Sung C.S. and Jin H.W. Dual-based approach for a hub network design problem under non-restrictive policy. *European Journal of Operational Research*. Vol.132, 2001, pp. 88~105
8. Kim D.K., Chang H.P. and Kim T.J. Single Allocation Hub Network Design Model with Consolidated Traffic Flows. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007
9. Abdinnour-Helm S. and Munirpallam A. V. Using Simulated Annealing to Solve the p-Hub Location Problems. No.9304, IRMIS
10. Topcuoglu H., Corut F., Ermis M. and Yilmaz G. Solving the Uncapacitated Hub Location Problem Using Genetic Algorithms. *Computers & Operations Research*. Vol. 32, 2005, pp. 967-984
11. Skorin-Kapov D. and Skorin-Kapov J. On Tabu Search for the Location of Interacting Hub Facilities. *European Journal of Operational Research*. Vol. 73, 1994, pp. 502-509
12. Kate S., Krishnamoorthy M. and Palaniswami M. Neural Versus Traditional Approaches to the Location of Interacting Hub Facilities. *Location Science*. Vol. 4, 1996, pp. 155~171
13. Ernst A. and Mohan K. Efficient Algorithms for the Uncapacitated Single Allocation p-hub Median Problem. *Location Science*. Vol.4, 1996, pp.139-154
14. O'Kelly M.E., Deborah B., Skorin-Kapov D. and Skorin-Kapov J. Hub Network Design with Single and Multiple Allocation: A Computational Study. *Location Science*. Vol.4, 1996, pp.125-138
15. Skorin-Kapov D., Skorin-Kapov J. and O'Kelly M.E. Tight Linear Programming Relaxation of Uncapacitated p-Hub Median Problems. *European Journal of Operational Research.* Vol. 94, 1996, pp. 582-593

LIST OF TABLES

LIST OF FIGURES

TABLE 1 Clustering rules of the study network

Cluster ID	TAZ ID
1	{1-9, 12}
2	{37-39}
3	{10,11,13-19}
4	{20-28}
5	{40-46}
6	{32-36}
7	{47-50}
8	{51-58, 29-31}

TABLE 2 Hub scales generated from the model

Hub Location (TAZ ID)	Hub Scale (in trips/hr)	Hub Location (TAZ ID)	Hub Scale (in trips/hr)
7	9862	37	4916
17	14850	42	5939
21	8979	47	4087
33	5303	58	5667

TABLE 3 Clustering rules under different numbers of hubs

Number of Hubs (p)	Clustering Rules (TAZ IDs)
2	{1-28, 37-46}, {29-36, 47-58}
3	{1-9,12,37-39}, {10-11, 13-28, 40-46}, {29-36, 47-58}
4	{1-9,12,37-39}, {10-11, 13-28, 40-46}, {32-36}, {29-31, 47-58}
5	{1-9,12,37-39}, {10-11, 13-28}, {40-46}, {32-36}, {29-31, 47-58}
6	{1-9,12,37-39}, {10-11, 13-28}, {40-46}, {32-36}, {47-50}, {29-31, 51-58}
7	{1-9,12}, {37-39}, {10-11, 13-28}, {40-46}, {32-36}, {47-50}, {29-31, 51-58}
8	{1-9,12}, {37-39}, {10-11, 13-19}, {20-28},{40-46}, {32-36}, {47-50}, {29-31, 51-58}

TABLE 4 Objective function values under different scenarios (in *hrs*)

p \ α	0.5	0.6	0.7	0.8	0.9
2	114298	114628	114905	115176	115289
3	95922	97045	97768	98256	98475
4	90735	92194	93281	94052	94533
5	81370	83436	84914	86023	86764
6	79857	82099	83742	84944	85730
7	78456	80827	82657	84029	84931
8	74050	76967	79150	80823	81958

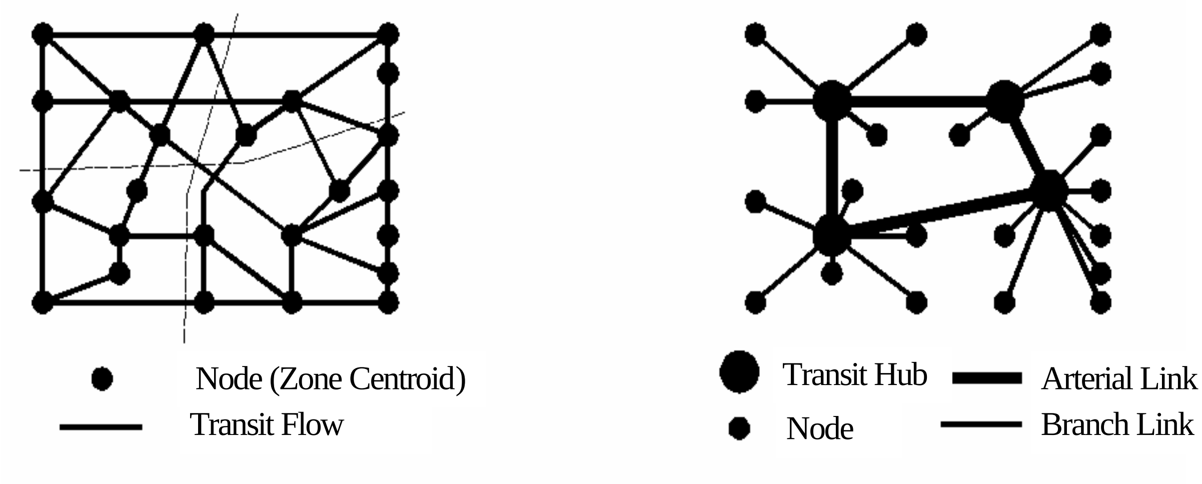
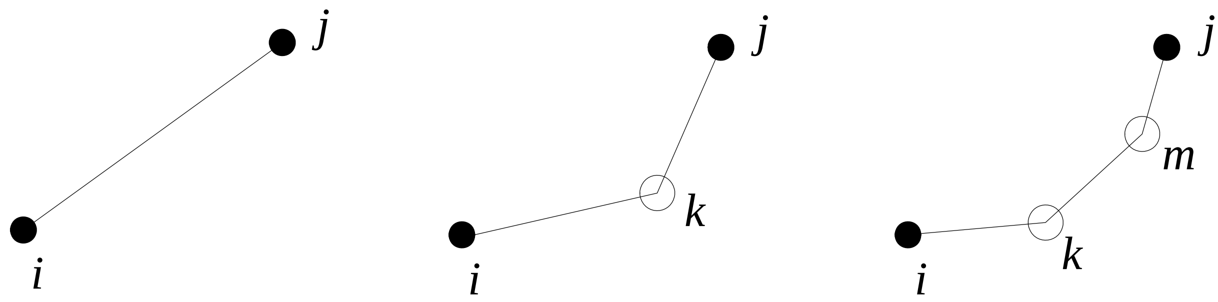


FIGURE 1 Illustration of the cluster partition and single hub allocation



(a) Non-stop （b）One-hub Stop （c）Two-hub Stop

FIGURE 2 Paths for transit flows between origins and destinations

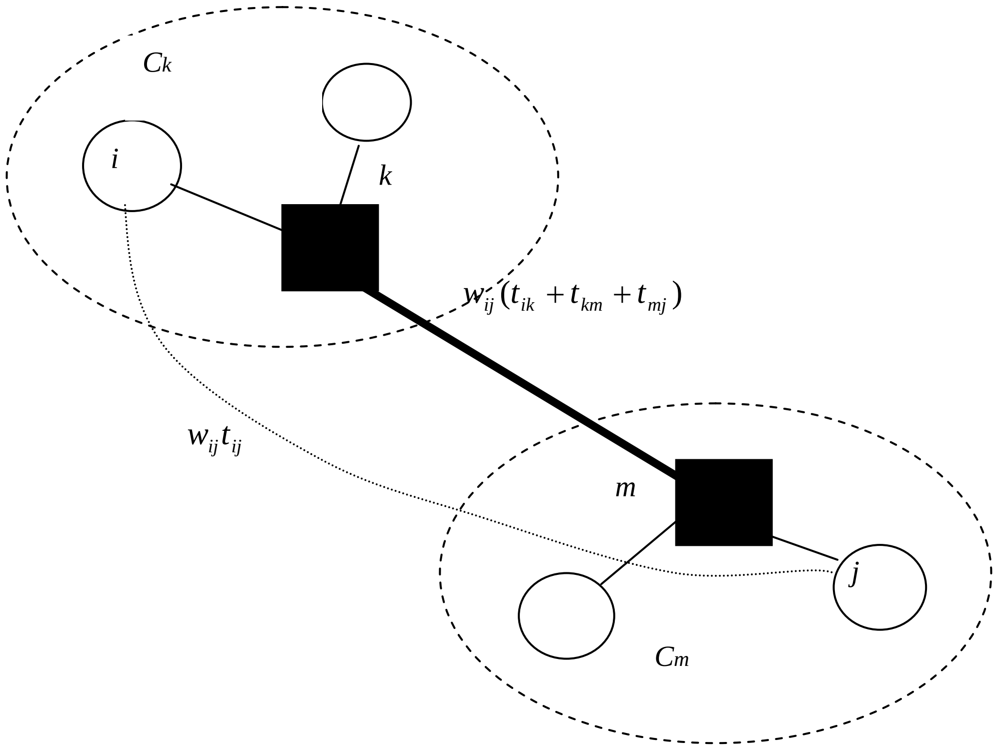


FIGURE 3 The minimum demand-weighted total travel time between clusters

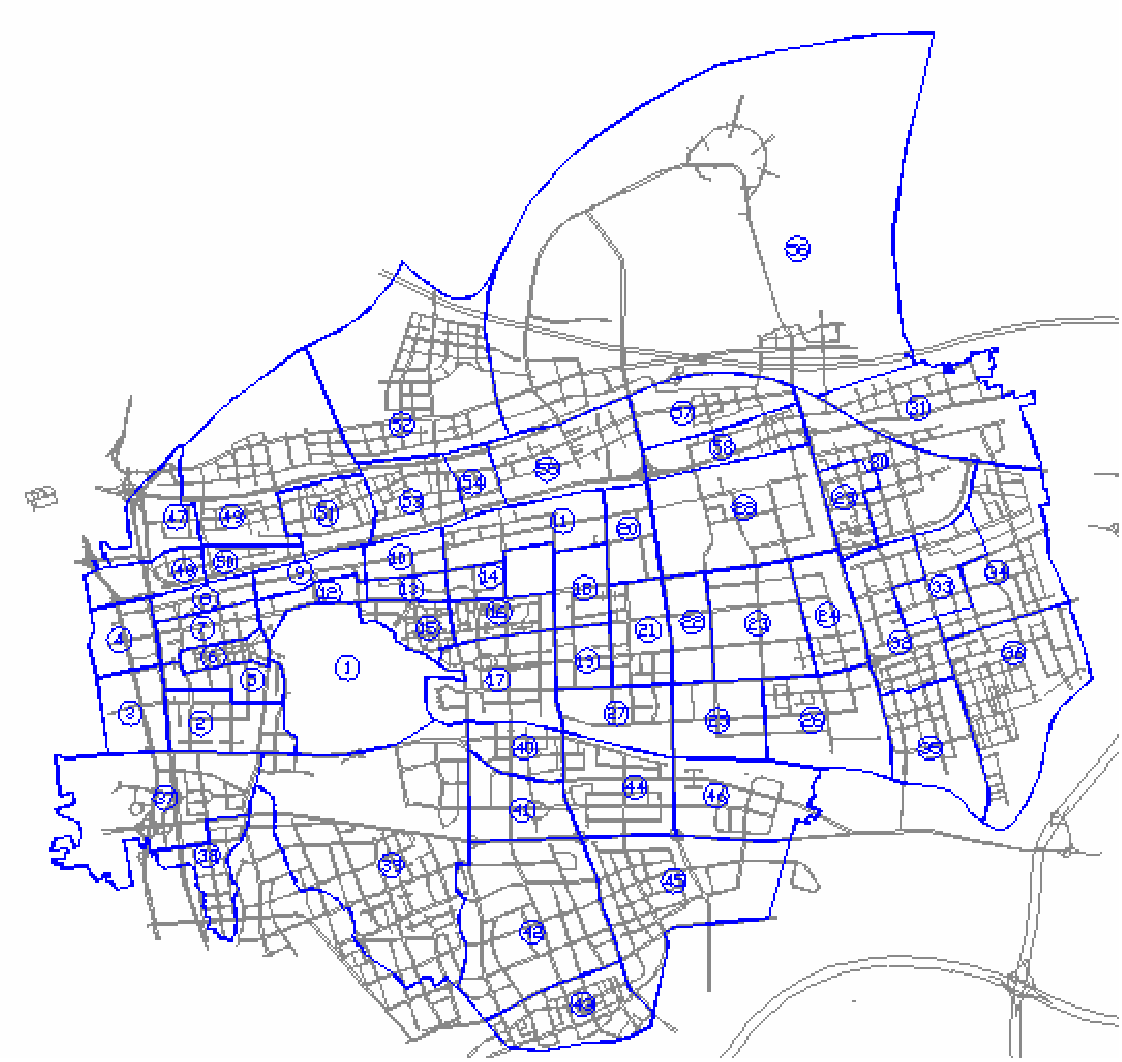


FIGURE 4 Distribution of TAZs on the study network

FIGURE 5 Clustering rules for the study network

FIGURE 6 Optimal locations of the transit hubs from the model

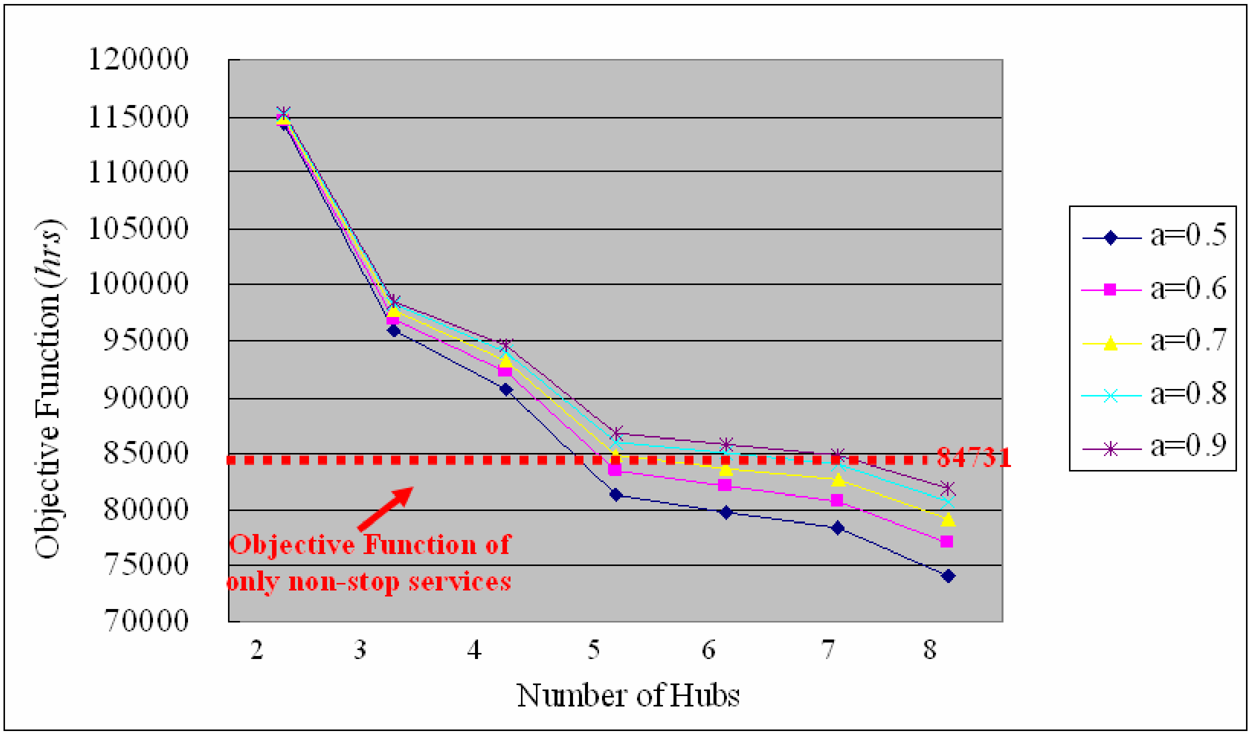


FIGURE 7 Objective function values under different scenarios