

Research Article

Design of a Support System for Complicated Logistics Location Integrating Big Data

Yusi Cheng ¹ and Xinwei Pan ²

¹School of Civil Engineering, Southeast University, 2# Sipailou, Nanjing, Jiangsu 210096, China

²School of Defense Engineering, PLA University of Army Engineering, 88# Houbiaoying Road, Nanjing, Jiangsu 210014, China

Correspondence should be addressed to Yusi Cheng; 101012017@seu.edu.cn

Received 7 December 2020; Revised 23 December 2020; Accepted 3 January 2021; Published 15 January 2021

Academic Editor: Wen Yi

Copyright © 2021 Yusi Cheng and Xinwei Pan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Logistics location is an important component of logistics planning that affects traffic pressure and vehicle emissions. To date, there has not been an adequate study of the integration of big data into the location for a complicated logistics system. This study developed a decision support system that can address location problems for complicated logistics systems, e.g., a multilevel urban underground logistics system (ULS), using logistics big data. First, information needed in the logistics location, such as the traffic performance index (TPI) and the origin/destination (OD) matrix, was collected and calculated using a big data platform, and this information was digitized and represented based on a geographic information system (GIS) tool. Second, a two-stage location model for a ULS was designed to balance the construction costs and traffic congestion. The first stage is establishing a set-covering model to identify optimum locations for secondary hubs based on the ant colony optimization algorithm, and the second stage is clustering of the secondary hubs to determine locations for primary hubs using the iterative self-organizing data analysis technique algorithm (ISODATA). Finally, the Xianlin district of Nanjing, China, was chosen as a case study to validate the effectiveness of the proposed system. The system can be used to facilitate logistics network planning and to promote the application of big data in logistics.

1. Introduction

With the rise of e-commerce and the growth of urbanization in various countries, urban logistics has become critical in ensuring the quality of people's lives and the sustainability of city development [1]. On the one hand, the demand for logistics services is growing rapidly, and the delivery efficiency requirement has become higher [2]. For example, the annual express delivery business growth rate in China exceeded 50% in 2016 [3]. On the other hand, the increasing freight volumes in limited urban areas exacerbate traffic congestion, which has an inevitable impact on energy consumption and environmental pollution. In the Guidelines for National Greenhouse Gas Inventory, it was indicated that petrol consumption in traffic jams is almost twice that during normal driving [4]. Moreover, as traffic congestion increases, CO₂ emissions [5] and PM_{2.5} [6] increase.

To address the above problems, new logistics techniques have been proposed, such as the underground logistics system (ULS) [7, 8]. Although these new techniques are innovative and appealing, they present challenges in logistics engineering, including logistics planning, construction, and maintenance, because of the potentially high costs and the technical complexities and difficulties.

A ULS is an infrastructure that utilizes underground urban space to build a dedicated freight network [8]. A ULS network mainly consists of logistics parks, logistics hub nodes, underground transportation tunnels, and dedicated vehicles. The location-allocation of hubs is key to network construction because it guides and controls the logistics activities of the entire system. Moreover, the location of the hubs affects the number of underground nodes and the total length of the tunnels, and underground engineering is extremely expensive. Determining the logistics location for a

ULS is a complicated and systematic task for which optimization algorithms have been designed considering various parameters, such as costs, distances, regional congestion, and freight volumes [9–12]. However, the current location approaches assume that important parameters, such as regional congestion and freight volumes, are known parameters. These parameters, which are obtained using traditional statistical methods, may not reflect the real circumstances. As a result, the accuracy of the selected logistics location is undermined by the errors in the parameter values. It is challenging to accurately and promptly collect and process massive amounts of data that are related to these parameters and are produced in the process of daily transportation and logistics.

Big data has been implemented in logistics, and various applications have shown tremendous value [13]. It is characterized by a high data volume, a rapid data flow (velocity), and diverse data (variety) [14]. With the maturity of information technologies, the collection, storage, computation, and visualization of big data can basically meet the requirements [13], while its core values include integrating data, extracting the features of data, and recognizing patterns from the features to support decision-making. In logistics systems, data from a variety of sources, such as the Global Positioning System (GPS), the Internet of Things (IoT), or traffic conditions, are time-variant and large. In logistics, big data has been utilized to optimize crew and vehicle routing, predict consumer demands, and optimize warehousing layout networks [13, 15]. However, studies regarding the utilization of logistics big data for complicated logistics locations are limited. Since logistics big data is needed to make decisions about complicated logistics systems, integrating it into a decision support system is challenging.

The aim of this study is to fill the literature gap, i.e., the fact that there is no integrated decision support system for complicated logistics location problems in which big data and geographic information system (GIS) data are efficiently integrated. The system is composed of logistics data collection, data processing architectures, data storing services, location optimization algorithms, and data visualization. The location optimization algorithms combine the set-covering problem (SCP) and clustering to solve the multi-stage hub location problem in a limited-capacity network based on the operational mechanism design of a hierarchical ULS network. The main contributions of this study are the following:

- (i) A big data platform is designed that integrates data from various sources, e.g., GPS, applications (APPs) in smartphones, GIS, and third-party data providers, to support decisions concerning ULS hub locations.
- (ii) A two-stage location optimization model is developed for a complicated ULS network that produces a more scientific layout. The model takes traffic congestion into consideration to estimate the actual demand of the ULS network. Moreover, the model is

applied to a city in China in order to validate the effectiveness of the model.

This study integrates location selection into a big data platform for large e-commerce or logistics companies for actual logistics planning. It realizes the application of big data to the optimization of complicated logistics systems. Moreover, due to the ability of the ULS network to speed up delivery time and alleviate traffic congestion, the study provides a basis for the sustainable development of the urban transportation system.

2. Literature Review

2.1. Big Data in Logistics Systems. In recent years, there have been many attempts made to improve logistics performances using the advancements in information technologies. Logistics ontologies were designed to systematically formalize domain knowledge of logistics [16], especially optimization-related knowledge [17]. IoT technology has been proposed to improve the productivity and efficiency of logistics management. For example, radio frequency identification (RFID) sensors and barcode sensors were attached to cargo to assist the operation flow of cargo inventory [18]. In addition, RFID technology and wireless sensors were combined to track and trace parts, semifinished goods, and finished goods to provide highly flexible information updating for order changes and picking problems [19]. Using the positioning capability of GPS and the tag identification capability of RFID, information about vehicle lines, logistics distribution, and the origin/destination can be directly managed to support space decisions in logistics [20]. Another useful tool in logistics is GIS, which captures, manages, analyses, and displays all forms of geographically referenced information. As spatial location information is useful in logistics, GIS is applied to assist with the positioning of facilities and vehicle route planning [20]. The application of these information technologies produces large-scale logistics data for various scenarios and operations. The significance of big data technology lies in mining the hidden values of the big data, rather than mastering huge amounts of information [13, 21].

In transportation and logistics, big data has received more attention in recent years, and research on its practical implementation has been conducted. Cell phone location data and license plate recognition data were collected and fused to predict traffic flow using a zero-shot transfer learning model [22]. A highly skewed speed-density dataset was processed using reproducible sample generation and the least squares method to obtain accurate traffic flow fundamental diagrams for various traffic flow conditions [23]. Inspection records for port state control have been used to predict the number of deficiencies each inspector can identify for each ship [24] and the ship detention probability [25]. Zhong et al. [26] introduced RFID-Cuboids to represent the logistics information and mined the frequent trajectory from the cuboids. The frequent trajectory

knowledge can then be used to determine the logistics plans and the layout of distribution facilities. Hopkins and Hawking [27] documented the integration of big data analytics and the IoT to improve driver safety, lower operating costs, and reduce the environmental impact of vehicles for a large logistics firm. A combination of truck telematics and geospatial information was used to monitor dangerous roads, driving conditions, and driver behaviour and to send alerts to drivers [27]. Furthermore, congestion maps developed from truck data records (GPS locations and dates) and road data records (usage of roads and dates) were used to optimize the schedules and routes of trucks in order to reduce congestion. Zhao et al. [28] analysed the management model, influencing factors, and development status of the B2C e-commerce logistics distribution, using big data, and optimized the resource allocation taking into consideration production sales and logistics distribution based on a big data platform. Shahparvari et al. [29] identified the optimal location for a logistics hub at a larger geographical scale using GIS. Thirty-four criteria related to the spatial-structural and functional attributes of the logistics hub location were input into the GIS, and the values associated with the performance of alternatives based on the criteria were automatically output [29]. In summary, different types of logistics big data models have been investigated for different scenarios and logistics processes. However, few studies have developed an integrated platform that includes logistics data collection, storage, optimization, and visualization. Similar to Shahparvari et al.'s study [29], this study focuses on the location selection of logistics hubs using GIS; nevertheless, ULS networks are more complicated than the logistics centres in Shahparvari et al.'s study [29].

2.2. Logistics Location for ULS Networks. ULS is an effective approach for alleviating the negative effects of urban freight traffic, improving the efficiency and safety of urban logistics, and saving the ground space [30]. In recent years, Switzerland, Italy [8], and China [31] have actively prepared for the construction of underground logistics networks. Given the uniqueness of the transport mechanism and because the service capacity of hubs and channels is restricted by the underground space, the traditional location-allocation method used for ground logistics networks is not suitable for ULS networks. In the following, we briefly review the model construction and method selection in order to define the boundary of the urban underground logistics hub location problem (ULHLP). As a multilayer network, the ULHLP has been exploited via referable methods by various researchers. The optimal ULS nodes, tunnel layout, and transport route network flows were formulated via a biobjective mixed-integer linear programming model considering minimal costs and maximal system utilization [9]. Liang et al. [10] established a multiobject ULS network planning model, including hub location and tunnel linking, and used agglomerative hierarchical clustering to determine the location of the first-level hubs and a greedy algorithm (GA) to determine the locations of the second-level hubs covered by each first-level hub. Ren et al. [11] constructed a set-covering

problem (SCP) model to determine the locations of the first-level hubs, and their results were optimized according to freight volume and the cargo handling capacity of the hubs. Tunnel length, regional congestion, and weighted distance were considered to determine whether to set second-level hubs [11]. He et al. [12] used the SCP to determine the number of underground logistics centres under a redefined underground logistics structure; then, a 0-1 planning model was used to connect the logistics centre with a distribution service centre. An improved bat algorithm was used to solve the layout model of the centres [12]. In previous studies, most two-layer hub-locating studies have been conducted from top to bottom, with calculations based on road freight data or standardized data. This means the calculated network carrying capacity may be much larger than the actual demand. The solution to congestion has been to add hubs or establish tunnel connections in the congestion area. While this can effectively solve local congestion, it lacks overall optimization for the entire area.

The SCP and clustering approaches are often used to determine the location of facilities to solve the problems of facility locations and customer allocation [32]. The SCP model can effectively deal with the strategy of selecting the minimum and optimal locations from candidate nodes and can provide service for the maximum demand in the region [33, 34]. The exact method of the SCP is inefficient due to the significant time needed for large problems [35]. As a result, heuristic algorithms, such as the GA [36, 37] and the ant colony optimization (ACO) [38–40], were proposed to solve the SCP. Compared to the GA, which is prone to being trapped in local optimization in the SCP, the ACO can obtain better results [41]. Clustering approaches have been widely used in the division of spatial regions, and they are useful for classifying demand points according to certain characteristics. O'Kelly [42] used clustering to minimize the sum of the squared deviation from the clustering mean based on a set of interacting spatial points. A probabilistic clustering method was proposed to solve the multifacility location problem, and the probability at each iteration depends on the travel costs based on hub locations [43]. The iterative self-organizing data analysis technique algorithm (ISODATA) can effectively eliminate transmission errors through the dynamic correction of the cluster centre [44, 45]. This gives it an advantage over k -means clustering without a given cluster quantity, which meets the requirement of basic ULS hub division.

3. Materials and Methods

The transportation process of a ULS-embedded urban logistics system can be organized into a three-layer network (Figure 1). In the first layer, goods are transported via underground tunnels (pipeline) from logistics parks (LPs) outside of the city to the hubs in the ULS network. Such hubs are defined as primary hubs (PHs). In the second layer, depending on the flow direction, the goods are transferred from the PHs to hubs serving the logistics demand point (DP). These hubs, which serve the DP, are defined as secondary hubs (SHs). The SHs are not connected to the LPs. In

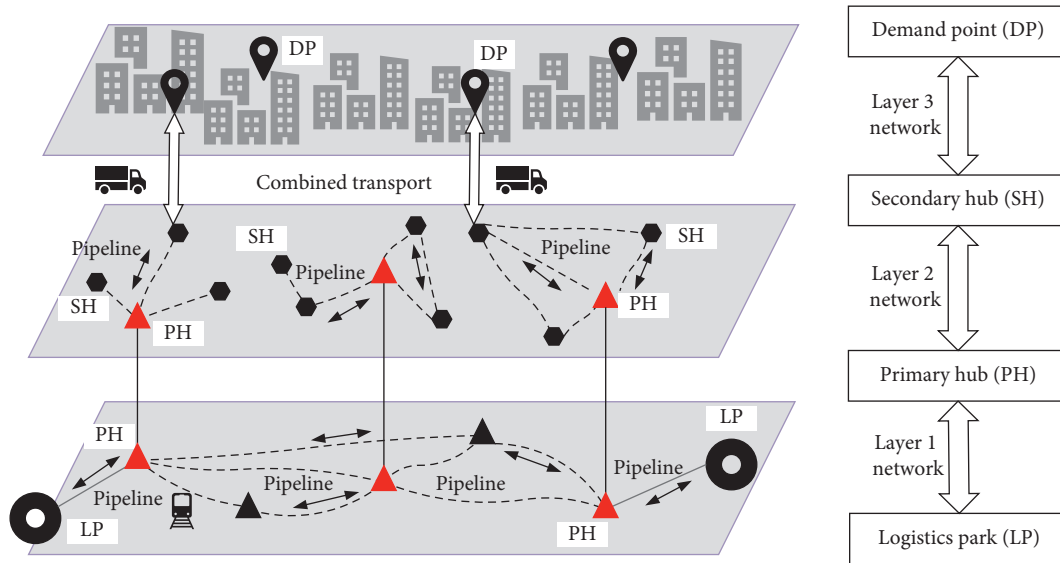


FIGURE 1: ULS network structure.

the third layer, the goods are transported from the SHs to the ground and finally to the customer via ground transportation (e.g., by hand or minivan), in such a way that the ground traffic load does not increase. The reverse process is also feasible, that is, goods flow from the SHs to the LPs.

3.1. Data Preparation. Since the ULS networks cooperate with traditional freight transportation systems to deliver goods from their origins to their destinations with the aim of alleviating traffic congestion problems, the optimization of the hub locations begins with determining two factors, i.e., the daily traffic congestion and the one-day freight volume with the origin/destination (OD). This step consists of data collection and analysis, and the values of the two factors are input into the ULS hub location model as influencing constraint conditions.

Traffic congestion is measured using the traffic performance index (TPI) in China [46]. The required data to calculate the TPI include the set of roads in a certain region A ; the road grade j ; the length of a road l ; the average speed on a road v ; and the traffic flow on a road x . The details of the calculation method can be found in the literature [47]. The TPI is calculated every 15 min during the morning peak (from 7:00 to 9:00) and the evening peak (from 17:00 to 19:00) on workdays, and thus, the values of v and x of each road in A must be updated every 15 min. The remaining variables, i.e., A , j , and l , can be extracted from the Amap API (<https://lbs.amap.com/>) that is used as GIS software. Amap also provides v in real time, while x can be collected from DiDi (<https://github.com/didi>).

The freight OD matrix in which each element represents the quantity of the goods delivered from an origin region (row) to a destination region (column) is estimated using data collected by a logistics tracking system. The logistics tracking system is developed to collect and manage the information of the present location of each delivery item and is supported by a combination of technologies, including the quick response (QR) code,

GPS, scanning equipment, and GIS. The QR code is attached to every item and stores the information about the item, such as the item id, origin address, destination address, and weight. When goods are loaded for transportation, a scanner is used to scan the QR code and to connect the good id to the vehicle id. Given that each vehicle is equipped with a GPS receiver and the vehicle id is connected to the good id, real-time location data for goods can be recorded and stored. Moreover, GPS coordinates are converted into Amap coordinates using the Amap API, and then, regional information can be inferred from the Amap coordinates. The OD information stored in the QR codes and inferred from the GPS is fused to ensure the accuracy. The freight OD matrix is obtained by grouping the delivered goods according to their origin and destination regions and by summing the weights of the goods in each group.

3.2. Big Data Platform Architecture. This section proposes a big data architecture and an analytic framework to support ULS hub location selection. Because large e-commerce enterprises (e.g., Amazon.com and JD.com) and logistics companies (e.g., FedEx and SF Holding) are prone to plan, construct, and maintain ULS networks due to the high cost and the complexity of ULS networks, the support system for ULS hub location selection based on big data and the existing logistics information systems of large enterprises can be integrated into a big data platform associated with big data technologies in order to support flexible logistics processes.

Figure 2 presents the architecture framework. The overall architecture is organized into six parts: the data source, data pipeline, data processing, data storage, data application, and data service layers. The data source acknowledges the data produced using different techniques, such as the QR code, GPS, text input to APPs in smartphones, sensors, and data from third parties, during logistics processes. Kafka is a distributed event-streaming platform for high-performance data pipelines, and it

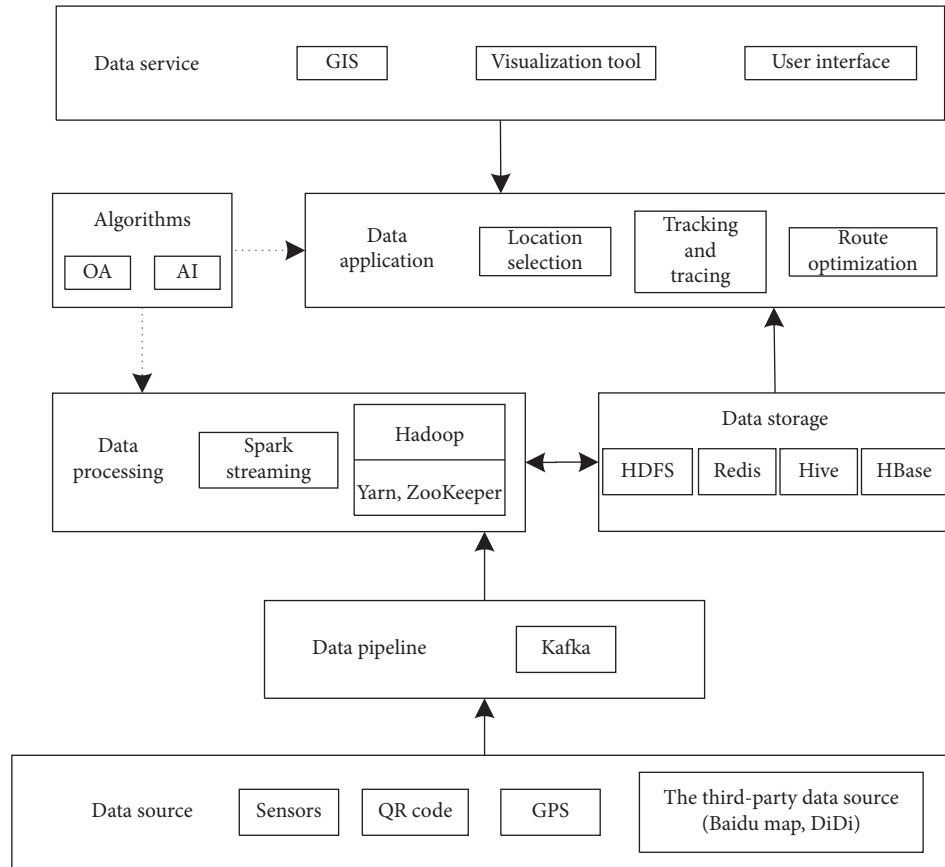


FIGURE 2: Architecture of the proposed big data platform.

has been adopted in logistics big data platforms [48, 49]. Kafka distinguishes between producers and consumers to improve scalability. The data sources in the data generation layer are producers that publish events that are organized and stored by topics on Kafka. These topics include average speed, location, and traffic flow. Spark and Hadoop receive the records on the topics in Kafka separately for front-end applications, e.g., ULS hub location selection, logistics tracking and tracing, and logistics route optimization. Optimization algorithms (OAs) and artificial intelligence (AI) technologies for these applications are implemented in Spark or Hadoop depending on the real-time requirement. Hadoop processes large amounts of static data collected over time [48]. The two core components of Hadoop are the HDFS (Hadoop distributed file system), which is responsible for storing data, and the MapReduce paradigm, which distributes the data across many servers. In contrast, Spark gathers and processes the data dynamically as they appear [48]. In the case of ULS hub location selection, Hadoop was selected based on historical data, and the algorithms implemented in Hadoop for the ULS hub location model are presented in Section 3.3. However, regarding route optimization for driving vehicles, Spark is more appropriate since the most optimal route should be calculated as quickly as possible. Finally, the results of the calculations in Spark or Hadoop are stored and sent to

corresponding applications. Moreover, a GIS provides geographic information, such as road length, roads in a region, and the GPS in a region, to the ULS hub location model and visualizes the results of the hub locations.

3.3. *Two-Stage ULS Hub Location Model.* To simplify the model, the following assumptions were made:

- (1) Since the freight demand per region is too small in some regions, it is deemed to cover the entire demand region if the service range (SR) of the hub covers the DP coordinate.
- (2) The flow of goods within each demand region is not calculated.
- (3) The length of the underground channel between hubs is assumed to be a straight line.
- (4) A DP is served by only one SH, and one SH is served by only one PH.
- (5) The maximum SR and cargo handling capacity (CHC) of the SH are capped. The SR is the maximum area covered by the service provided by the hub, and the CHC refers to the maximum daily capacity of a

hub in regard to the delivery or transport of goods, including sorting, handling, and other operations.

(6) The PH has an unlimited CHC.

A two-stage hybrid set-covering and clustering model was proposed for ULS hub cluster classification and location. First, based on the regional ground freight OD data and the TPI discussed in Section 3.1, the possible underground freight OD data undertaken by the ULS network were calculated and generated. Second, a set-covering model was established to obtain the minimum SH number that can cover all of the areas under a given SR. Third, the number and location of PHs were determined by clustering the freight volume of the SHs and the distance from the PH to the SHs it covers.

3.3.1. OD Matrix for ULS. Based on the readily available freight OD matrix, this study proposes a calculation method using the regional TPI to generate an OD matrix for underground freight, which is the basis for the ULS network calculation. By weighting the TPI and the underground construction cost, the TPI values of all the demand regions can

be reduced to $\eta_0 = 4$, which indicates that the traffic congestion state is clear. In addition, it is approximately considered that the TPI has a linear relationship with the regional freight volume [30]. In this way, according to the reduction rate of the TPI, a regional freight volume of the same proportion can be transferred underground. Thus, the freight volume to be transferred underground in region i is

$$d_i = \frac{\eta_i - \eta_0}{\eta_i} Q_i = \frac{\eta_i - \eta_0}{\eta_i} \left(\sum_{k \neq i} \text{OD}(i, k) + \sum_{k \neq i} \text{OD}(k, i) \right), \quad i = 1, 2, \dots, n, \quad (1)$$

where Q_i is the freight volume of region i ; $\text{OD}(i, k)$ is the element in the i th row and the k th column of the OD matrix; and η_i is the TPI of region i . Considering the adverse effect of the large number of decision variables [50] and the influence of $\text{OD}(i, k)$ on the TPI of origin region i and the TPI of destination region k , the OD matrix of the underground freight can be simplified to

$$\text{OD}_u(i, k) = \begin{cases} \max\left(\frac{\eta_i - \eta_0}{\eta_i} \text{OD}(i, k), \frac{\eta_k - \eta_0}{\eta_k} \text{OD}(i, k)\right), & \text{if } \eta_i > \eta_0 \text{ or } \eta_k > \eta_0, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where $\text{OD}_u(i, k)$ is the freight volume transported from region i to region k that needs to be transferred through the ULS. When the TPI of origin region i or destination region k is larger than η_0 , the larger TPI of i and k is used to calculate the value of $\text{OD}_u(i, k)$; otherwise, $\text{OD}_u(i, k) = 0$ is used.

3.3.2. SH Location Based on the ACO Algorithm. According to the assumptions, all of the SHs have a maximum service radius, which is the maximum coverage of one hub as a subset. Since the distance from the DP to the SH does not exceed the service radius of the SH and hub construction costs far exceed the transportation costs, the SH construction cost is taken as the total cost. Since the total construction cost of hubs is proportional to the quantity, the objective function can be established to minimize the total cost. The SCP model is as follows:

$$\min Z = \sum_{j \in M} c_j x_j, \quad (3)$$

subject to

$$\sum_{j \in M} a_{ij} x_j \geq 1, \quad \forall i \in N, \quad (4)$$

$$\sum_{j \in B(i)} y_{ij} = 1, \quad \forall i \in N, \quad (5)$$

$$\sum_{j \in A(j)} d_i y_{ij} \leq D_j x_j, \quad \forall j \in M, \quad (6)$$

$$\sum_{j \in A(j)} d_i y_{ij} = \sum_{i' \in A(j)} \sum_{i \in N} V_{ii'} + \sum_{i \in A(j)} \sum_{i' \in N} V_{ii'} - \sum_{i, i' \in A(j)} V_{ii'}, \quad \forall i, i' \in N, j \in M, \quad (7)$$

$$x_j \in \{0, 1\}, \quad \forall j \in M, \quad (8)$$

$$y_{ij} \in \{0, 1\}, \quad \forall i \in N, j \in M. \quad (9)$$

In the model, c_j is the construction cost of SH j , which can be set to 1 since the construction of the SH costs the same. If DP i is covered by SH j , then $a_{ij} = 1$; otherwise, $a_{ij} = 0$. N is the set of DPs in the district, $N = \{1, 2, \dots, n\}$. M is the set of candidate nodes for the SH, $M = \{1, 2, \dots, m\}$. d_i is the freight demand volume of DP i , and D_j is the CHC of SH j . $V_{ii'}$ is the freight volume transported from DP i to DP i' . $A(j)$ is the set of DP i covered by SH j . $B(i)$ is the set of SH j that can cover DP i .

The objective function (equation (3)) describes the SCP model with the goal of achieving the lowest total cost for the hubs. The constraint described by equation (4) ensures that each secondary hub covers at least one DP. Equation (5) means that each DP can only be serviced by one SH. The

constraint described by equation (6) indicates that the total freight volume of the DPs covered by the SH cannot exceed its capacity. The model is an SCP with nonlinear constraints rather than a simple SCP. It is used to cover all of the demand regions with the least number of SHs, which reduces the large construction investment and the difficulties in the early stage of ULS construction.

The principle of the ACO is to mimic the foraging behaviour of ant colonies based on information exchange.

The principle of the ACO is to mimic each candidate node regarded as a node on the path. The ACO in this study is described as follows:

- (i) Initialization: m ants are randomly placed on n candidate nodes, that is, each ant randomly selects a candidate node as the starting node of the path. The initial pheromone of each candidate node is set to be the same.
- (ii) Probabilistic choice: using a random probability selection mechanism, ant k selects a subset S_i in S with probability P . P is calculated using equation (10), where $\eta_i(t) = |C_i|$. $|C_i|$ is the number of nodes in subset S_i .

$$P(t) = \begin{cases} \frac{\tau_i(t)^\alpha \eta_i(t)^\beta}{\sum_{s \in \text{Allow}_k} \tau_s(t)^\alpha \eta_s(t)^\beta}, & i \in \text{Allow}_k, \\ 0, & \text{else.} \end{cases} \quad (10)$$

- (iii) Update the pheromone: after each iteration, the pheromone of the path through which the ant passed is increased. At $t + 1$, the pheromone adjustment rule for each subset is

$$\tau_j(t+1) = \rho \cdot \tau_j(t) + \Delta\tau_j(t, t+1), \quad (11)$$

where $\Delta\tau_j(t, t+1)$ is the number of pheromones left in solution set i by ant k at $t + 1$, which is calculated using the following equation:

$$\Delta\tau_j(t, t+1) = \begin{cases} \frac{Q}{|C_j|}, & \text{if ant } k \text{ chooses set } C_j, \\ 0, & \text{else.} \end{cases} \quad (12)$$

- (iv) Termination condition for the ACO: the SCP has difficulty obtaining the optimal solution in practical applications, and the given iteration number is usually taken as the termination condition for the algorithm. For example, the algorithm terminates when the quality of the solution is no longer improved after N consecutive iterations.

3.3.3. PH Location Based on ISODATA Clustering. SHs can be grouped based on certain rules to locate the PH. A combination of the clustering method and the improved ISODATA was used to solve the PH localization problem.

The regional centre formed by the clustering and grouping of the SHs is the PH's position.

The positioning of the PH should minimize the overall transportation costs of the ULS network, which is related to the freight volume and the transportation distance. The PH should be set in an SH area with a large and concentrated freight volume. Therefore, the clustering model considers the freight demand of the SH and the distance between the SH and the PH. The distance between hubs is the Euclidean distance [39]. The objective function of the clustering model is

$$\min F = \sum_{j \in I} \sum_{k \in J} w_j s_{jk}, \quad (13)$$

where w_j is the freight demand of SH j and s_{jk} represents the distance between SH j and PH k . I is the set of SHs in the region, and J is the set of PHs.

The ISODATA has advantages in terms of automatic calculation and the acquisition of a reasonable number of clusters [40]. The distance in the ISODATA is calculated by weighting $h_{jk} = w_j d_{jk}$, where h_{ij} is the clustering distance from SH j to the cluster centre k ; w_j is the freight volume of SH j ; and d_{jk} is the Euclidean distance between j and k .

Step 1. Identify and determine some of the initial values that can be artificially modified during the iteration. N samples $\{x_\gamma, \gamma = 1, 2, \dots, N\}$ are allocated to the clusters according to the initial values.

- (1) Preselect N_c initial cluster centres $\{z_\gamma, \gamma = 1, 2, \dots, N_c\}$, which can be selected from all of the samples.

N_c : the number of cluster centres

K : the expectant number of cluster centres

θ_N : the minimum number of samples in the per cluster domain

θ_S : the standard deviation of the sample distance distribution in a cluster domain

θ_c : the minimum distance between two cluster centres

L : the maximum pairs of cluster centres that can be merged in an iterative operation

I : the number of iterations, $I \leq I_0$

- (2) Distribute N samples to the nearest cluster S_λ . If $\|x - z_\lambda\| = \min\{\|x - z_\gamma\|, \gamma = 1, 2, \dots, N_c\}$, then $x \in S_\lambda$.

If the number of samples in S_λ is less than θ_N , S_λ should be deleted, and $N_c = N_c - 1$.

Step 2. Calculate the distance index function of all of the samples.

- (1) Modify the cluster centre:

$$z_\lambda = \frac{1}{N_\lambda} \sum_{x \in S_\lambda} x, \quad \lambda = 1, 2, \dots, N_c. \quad (14)$$

- (2) Calculate the average distance between each sample and the cluster centre in each cluster domain S_λ :

$$\overline{D}_\lambda = \frac{1}{N_\lambda} \sum_{x \in S_\lambda} \|x - z_\lambda\|, \quad \lambda = 1, 2, \dots, N. \quad (15)$$

Step 3. Perform the split operation.

- (1) Calculate the standard deviation vector of the sample distance of each cluster:

$$\delta_\lambda = (\delta_{1\lambda}, \delta_{2\lambda}, \dots, \delta_{n\lambda})^t,$$

$$\delta_{\gamma\lambda} = \sqrt{\frac{1}{N_\lambda} \sum_{\mu=1}^{N_\lambda} (x_{\gamma\mu} - z_{\gamma\lambda})^2}, \quad \gamma = 1, 2, \dots, n; \quad (16)$$

$$\lambda = 1, 2, \dots, N_c,$$

where $x_{\gamma\lambda}$ is the γ component of S_λ and $z_{\gamma\lambda}$ is the γ component of z_λ . $\delta_{\gamma\lambda}$ is the standard deviation of the γ component of x , and x is an n -dimensional vector.

- (2) Calculate the maximum component of each standard deviation component. The maximum component of δ_λ is $\delta_{\lambda \max}$.
- (3) In the maximum component set $\{\delta_{\lambda \max}, \lambda = 1, 2, \dots, N_c\}$, if $\delta_{\lambda \max} > \theta_s$ and one of the following conditions is met, then split z_λ into two centres, z_λ^+ and z_λ^- , and use N_c plus one. Conditions: (i) $\overline{D}_\lambda > \overline{D}$ and $N_\lambda > 2(\theta_N + 1)$; (ii) $N_c \leq (K/2)$.

Step 4. Perform the merge operation.

- (1) Calculate the distance between each cluster centre:

$$D_{\gamma\lambda} = \|z_\gamma - z_\lambda\|, \quad \gamma = 1, 2, \dots, N_c - 1; \quad (17)$$

$$\lambda = \gamma + 1, 2, \dots, N_c.$$

- (2) Compare $D_{\gamma\lambda}$ with θ_c , and sort the values of $D_{\gamma\lambda}$ ($D_{\gamma\lambda} < \theta_c$) from small to large; that is, $\{D_{\gamma_1\lambda_1}, D_{\gamma_2\lambda_2}, \dots, D_{\gamma_L\lambda_L}\}$, where $D_{\gamma_1\lambda_1} < D_{\gamma_2\lambda_2} < \dots < D_{\gamma_L\lambda_L}$.
- (3) Merge z_{ik} and z_{jk} . The distance between them is D_{ikjk} . The new centre is calculated as follows:

$$z_\mu^* = \frac{1}{N_{\gamma\mu} + N_{\lambda\mu}} [N_{\gamma\mu} z_{\gamma\mu} + N_{\lambda\mu} z_{\lambda\mu}], \quad \mu = 1, 2, \dots, L. \quad (18)$$

Step 5. Iterate again and judge whether the clustering results meet the requirements. After several iterations, if the result converges, the operation exits, and the result is preserved.

4. Results and Discussion

The proposed methodology was validated through a case study of the selection of ULS hub locations in the Xianlin district of Nanjing, China. Figure 3 shows the freight transportation map of the Xianlin district in Amap, which can be used as a GIS tool. The black squares in the figure not only represent the geographical centre of the 110 regions officially included in the urban planning but also correspond to freight DPs in these regions. The number represents the code of the region and its DP. Four LPs are located in different directions outside of the district. The TPI of each region is displayed in Figure 4. Using different colours corresponding to different TPI levels, Figure 4 shows that a majority of the regions are congested. A 114×114 OD matrix represents the volumes of the goods transported between the 114 regions, including 110 DPs and 4 LPs. The sum over each column corresponds to the goods leaving in the region and is illustrated in Figure 5(a), and the sum over each row corresponds to the goods entering the region and is illustrated in Figure 5(b).

Based on equation (2), when $\eta_0 = 4$, the total underground freight volume is 67,112 tons, accounting for 41% of the original freight volume. The CHC and SR of the underground logistics hub were temporarily set as 3000 tons and 3 km, respectively, to calculate the location-allocation. In Section 4.3, the calculation results for different combinations of these two parameters are further compared and discussed.

4.1. Optimization Results for SH Locations. When SR = 3 km and CHC = 3000 t, 40 SHs were screened out by the ACO (Figure 6, in which red circles represent SHs selected from DPs). Figure 6 provides an example of 791 DPs that were selected as the SHs, in which the circle indicates the service coverage centered on the SH and the covered regions are represented by the blue squares inside the circle. The actual freight volume (AFV) of an SH is defined as the sum of the freight turnover volume of the DPs covered by the SH. Moreover, another important index, i.e., the average saturation of the cargo handling capacity (ASCHC), was used to measure the actual freight handling performance of the SH, which can be calculated using the following equation:

$$\text{ASCHC} = \frac{\sum_n (\text{actual freight turnover}/\text{cargo handling capacity})}{n} \times 100\%, \quad \text{ASCHC} \in (0, 1), \quad (19)$$

where n is the number of SHs. The larger the ASCHC, the higher the hub utilization and the better the ULS network performance. In contrast, a low ASCHC indicates that the designed capacity of the hub is far larger than the actual demand, which will cause huge waste. Table 1 shows the SH sets and their coverage regions calculated by the ACO; the summary is shown in Table 2. The average single runtime for the ACO was 414.5400 s.

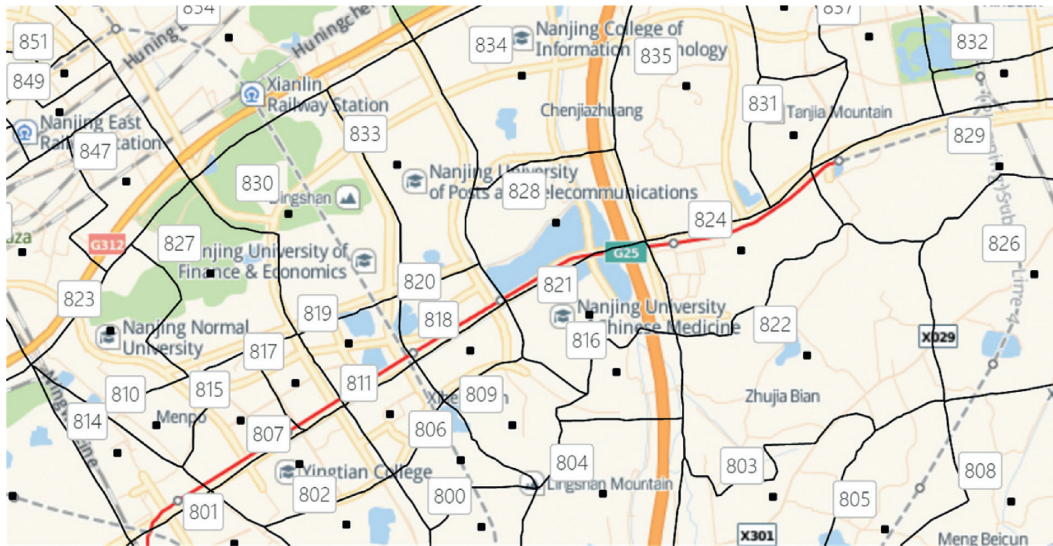


FIGURE 3: Distribution map of the logistics DPs in the Xianlin district.

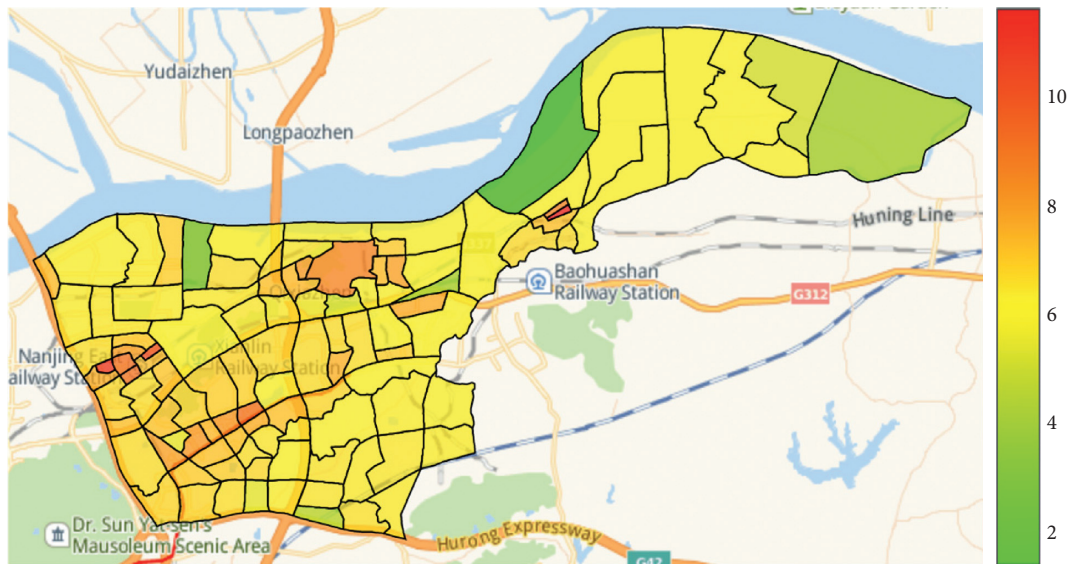


FIGURE 4: TPIs of the regions in the Xianlin district.

There are three reasons for the quantitative difference in the hubs serving only one DP in the results. First, the distance from the other nearest DPs to the hub is more than its service range. There is only one such hub in this case (number 896). Second, the capacity of a hub can only meet the demand of the region where it is located. For instance, DP 886 has a freight volume of 3088.07 t (over 3000 t) in the ACO result, and thus, it requires the help of the SHs nearby to meet this volume, including SHs 885, 890, and 891, which are dedicated to it. Third, following the logic and search order of the chosen algorithm, some of the hubs were generated through local optimization. Three SHs (nos. 808, 825, and 892) in the ACO result remained after the points nearby were searched and covered. The superiority of the ACO's pheromones enables individuals to communicate indirectly through the environment and use the probabilistic search method to effectively avoid local optimal solutions.

This study aimed to propose a two-stage method suitable for the selection of hub locations in a ULS. Thus, the next stage was calculated based on the set of results obtained using the ACO.

4.2. Optimization Results for PH Location. The PH locations were obtained by reclustering the optimized SH group and by considering the freight volume and distance. The goal of clustering is to minimize the total distance between the PHs obtained and the subordinate SHs and to maximize the freight volume. Given that the PH needs to be linked with the LPs outside of the district, in order to reduce the difficulty of cargo scheduling and improve the clustering effect, the expectant number of clustering centres K was set to 4. In this way, the PH can be regarded as a controlling intermediate node connecting the logistics park and the SHs.

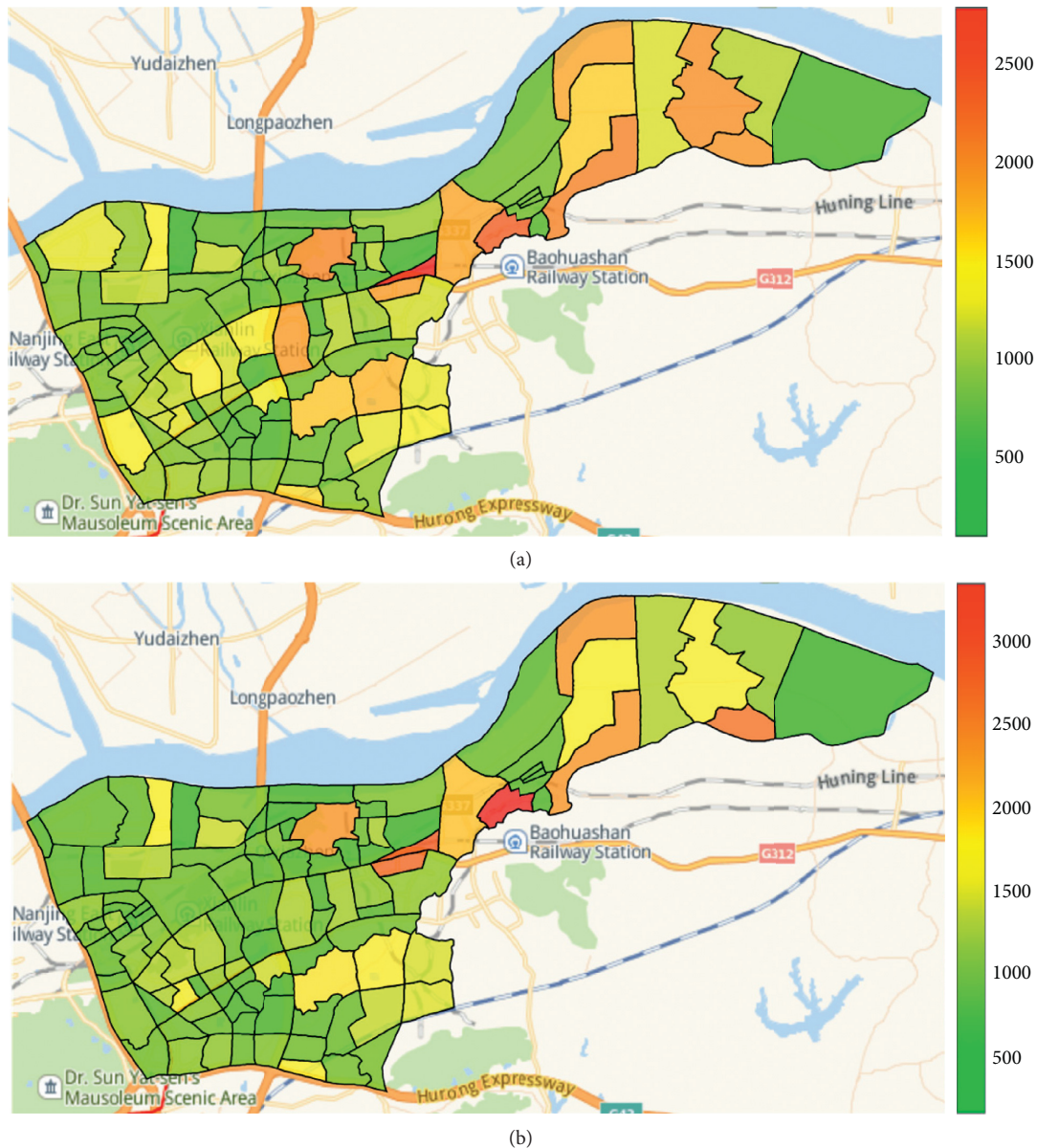


FIGURE 5: OD matrix of the regions in the Xianlin district. Heat maps of the volumes of goods (a) leaving and (b) entering the region.

From the perspective of management convenience, a one-to-one correspondence between parks and PHs is reasonable.

Table 3 lists the four PHs and their subordinate SHs, which were calculated using the ISODATA. Their distributions are marked separately using different colours in Figure 7. The locations of the four generated PHs do not coincide with the existing DPs. However, it turns out that the positions of some of the SHs are very close to the PHs. In actual planning or further calculations, these SHs could be replaced by PHs. For example, the distance between SH 800 and PH A is only 0.087 km, so it is economical to let PH A assume the function of SH 800. When the PH location obtained through the clustering is not suitable under actual conditions, it is acceptable to adjust its position within the clustering area.

4.3. Parameter Effects. CHC and SR are worth discussing in terms of the modelling and computational analysis. The hub is usually equipped with an automated sorting and transmission system, and some of the goods that are transported from the SH to the customers still need to be delivered on the ground. Therefore, in addition to the constraint imposed by the underground space, these two parameters are largely subject to the automated logistics technology and the carrier distribution capacity. Given its similarity to the current express delivery process, the common data for general express delivery (Table 4) are used for the discussion.

Table 5 summarizes the calculation results for the parameter combinations. Several valuable observations are as follows:

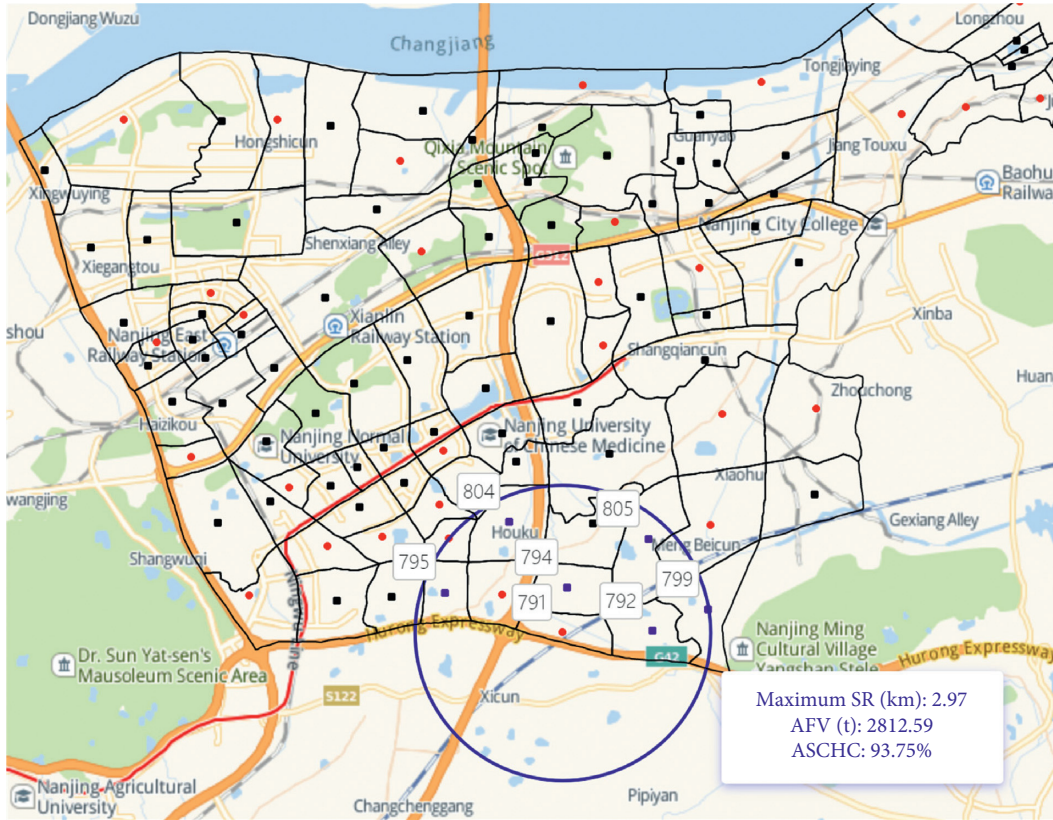


FIGURE 6: Distribution of SHs obtained using the ACO.

TABLE 1: Solutions to the SCP obtained using the ACO.

SH	Covered region	AFV (t)	ASCHC
791	804, 805, 795, 794, 799, 792	2812.59	93.75
796	803, 809, 811, 816, 797	2736.97	91.23
798	813, 814	2961.4	98.71
800	820, 821	2937.7	97.92
801	823, 827	2991.21	99.71
802	807, 793	2198.47	73.28
806	815, 819	2934.17	97.81
808	—	1069.46	35.65
810	817	1374.02	45.80
818	833, 834, 824	2894.47	96.48
825	—	1440.68	48.02
826	812, 832	2895.84	96.53
831	822, 828, 829	2996.37	99.88
836	835, 837	2907.65	96.92
838	840	1618.46	53.95
841	848, 849, 845	2833.22	94.44
846	852, 858, 844	2951.12	98.37
851	830, 850	2602.31	86.74
853	847, 842, 843	2738.23	91.27
855	854, 856, 860, 861, 864, 870	2794.79	93.16
863	875, 878, 881, 865	2847.29	94.91
867	859, 866	2759.53	91.98
868	862, 869, 857	2559.67	85.32
871	839, 873	2911.62	97.05
883	880	2921.35	97.38
884	872, 874, 876, 877, 879, 882	2799.14	93.30
885	888, 889	1527.48	50.92

TABLE 1: Continued.

SH	Covered region	AFV (t)	ASCHC
886	—	3088.07	102.94
890	887	2202.37	73.41
891	893, 895	2269.43	75.65
892	—	2165.61	72.19
894	898	2723.1	90.77
896	—	80.04	2.67
899	897, 900	2626.28	87.54

TABLE 2: Summary of SH results.

Algorithm	Number of SHs	Number of single-service SHs	ASCHC	Runtime (s)
ACO	34	5	85.52	414.5400

TABLE 3: PHs and their covered SHs calculated using the ISODATA.

PH	Covered SHs	AFV (t)
A	791, 796, 798, 800, 801, 802, 806, 810, 818	23,841
B	808, 825, 826, 831, 836, 838, 871, 883, 884, 890	23,763
C	885, 886, 891, 892, 894, 896, 899	14,420
D	841, 846, 851, 853, 855, 863, 867, 868	22,086

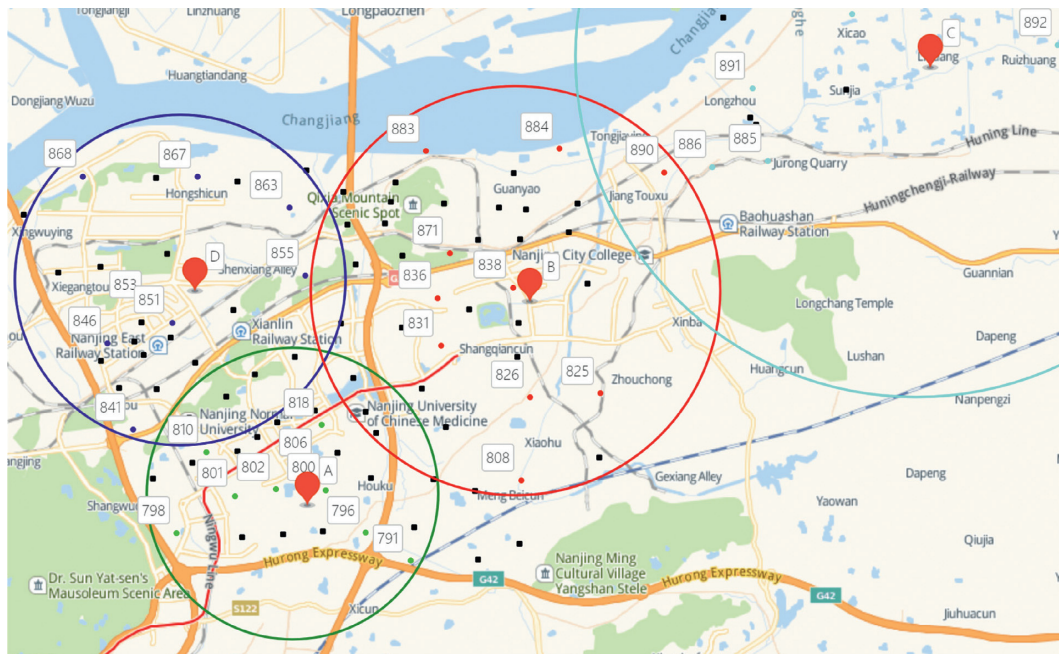


FIGURE 7: Distribution of the four PHs obtained using the ISODATA.

TABLE 4: Parameter settings for CHC and SR.

CHC (t)	3000	4000	5000
SR (km)	3	4	5

TABLE 5: Summary of the SH results under different parameters.

CHC (t)	SR (km)	Number of SHs	ASCHC (%)	Number of isolated hubs	PH label	Number of covered SHs	AFV (t)	SD of PH AFV	
3000	3	34	82.52	5	A	9	23,841	3853	
					B	10	23,763		
					C	7	14,480		
					D	8	22,086		
	4	35	80.59	5	A	9	23,537	5552	
					B	10	22,329		
					C	6	11,966		
					D	10	26,793		
	5	33	85.46	5	A	9	23,693	3702	
					B	9	24,267		
					C	7	15,062		
					D	8	19,530		
4000	3	29	72.89	6	A	6	19,248	15,726	
					B	15	50,288		
					C	8	15,012		
	4	24	88.08	3	A	10	38,646	9808	
					B	8	30,845		
					C	6	15,066		
	5	25	84.65	3	A	9	31,313	9683	
					B	5	15,115		
					C	11	38,219		
	5000	3	24	70.24	5	A	7	26,093	5796
						B	5	20,246	
						C	6	11,915	
D						6	26,036		
4		21	80.48	1	A	7	31,228	7683	
					B	6	25,660		
					C	4	12,025		
					D	4	15,589		
5		20	84.50	2	A	10	38,503	8822	
					B	6	29,044		
					C	4	16,948		

- (1) An obvious conclusion is that the amount of SH decreases significantly with increasing CHC, and the relationship is almost linear. However, the amount of SH is not inversely proportional to the SR of the hub. CHC is more sensitive to the number of SHs than SR. Therefore, effort should be made to improve the freight capacity of the hubs to reduce the number of SHs and thus to reduce construction costs.
- (2) The number of hubs serving only one DP varies considerably under the same CHC but a different SR, especially when the SH has sufficient capacity to handle the AFVs of all of the DPs. For instance, when CHC = 4000 t and SR = 3 km, there are six such hubs, which is three more than when SR = 4 km. Similarly, when CHC = 5000 t and SR = 3 km, there are five such hubs, which is three to four more than under other SR conditions. The lower the number of these hubs, the more efficient the SH group. Therefore, the CHC and SR of a hub should be matched; otherwise, they will seriously affect the network construction costs and the operational efficiency.
- (3) It is necessary to reserve some of the spare capacity of the node. Global search results always make full use

of the CHC under computational logic, aiming at the minimum number of nodes. In this case, the AFVs of all of the hubs account for over 80% of the designed capacity on average, and the utilization rate of a few hubs even exceeds 90%. Although increasing facility utilization is a common goal, future risks of rapid growth in urban freight demand cannot be ignored when operating at near-full capacity. A surge in freight volume that exceeds the capacity limits can lead to the collapse of the hub operations, as can equipment failures. Only a small number of non-adjacent hubs collapse. The other hubs can still maintain the network operation when the capacity is exceeded, but the collapse of multiple or adjacent hubs will lead to the inefficient operation of the entire ULS network, or even paralysis of the network.

- (4) In addition, the standard deviations (SDs) of the PH AFVs presented in Table 5 reflect the balance degree of the freight volume of each PH. As can be seen, the SD of the AFV varies irregularly. When CHC = 4000 t and SR = 3 km, the SD reaches the maximum value, which is 15,726 t. An imbalance of the AFV would limit the overall operating efficiency of the network. The operating efficiency of the PH

with the largest AFV would become the limiting component, which could negatively affect the operation of the entire network. Thus, it is necessary to choose a hub group set with a balanced AFV distribution. Furthermore, the PH capacity is not set as a fixed value in this study. Rather, each PH should be independently designed according to the practical engineering requirements, and the balance of the AFV distribution of the PH should be maintained as much as possible.

5. Conclusions

In this paper, an integrated support system for complicated logistics location is proposed, i.e., ULS hub location determination, based on big data technologies. First, to calculate the TPI and OD matrix, the types of data generated across a typical data-driven logistics system and methods of collecting these data were investigated. Second, a platform for big data analytics in logistics systems, including hub location selection, was developed. The data for the TPI and OD matrix were produced, collected, stored, and analysed using different modules in the platform. Third, a bottom-up two-stage hub location method and a visualization tool based on the GIS were incorporated into the platform. In this method, the underground cargo flow data used for the calculation were obtained by proportionally processing the OD matrix according to the TPIs of different demand regions. An SCP model under nonlinear constraints was used for the SH selection, which was solved using the ACO algorithm. Compared with the genetic algorithm (GA), the ACO algorithm gives better results, but it takes more time. The PH locations were determined using the ISODATA to cluster the SHs, and the ISODATA allows for different numbers of clusters.

In addition, this study highlighted the significance of developing a big data analysis platform for large e-commerce enterprises and logistics companies to improve the efficiency of logistics management. Complicated logistics location is one of the important applications supported by the platform. Traditional logistics planning focuses on designing optimization algorithms to obtain improved solutions or choosing evaluation criteria for decision-making, but it neglects some basic variables, such as the TSCPPI and OD matrix, the values of which are assumed to be known. The complete decision-supporting process for ULS hub locations, including data producing, collecting, storing, processing, and visualization, was integrated into the platform, which enables the smooth flow of information between different stages of the process. Moreover, considering the growing number of applications in logistics management, the platform was built using the Kafka system, which has the ability to plug in different data sources and applications. In this way, the platform can address the interoperability across the logistics data, the logistics process, and the logistics services. Different users of the platform, such as delivery vehicles, managers, and planners, can cooperate with each other.

This study provides an initial step in the investigation of the use of a big data platform for complicated logistics location determination. In future studies, the performance analysis of the big data platform for logistics management will be considered since determining the main factors that affect the quality of the platform is important for optimizing the design of the platform. More efforts are also needed to optimize the matching of the CHC and SR in the two-stage ULS hub location method before the hub group calculation. Other factors influencing location selection should also be investigated, such as the distance from a PH to its corresponding LP, the actual cost, and urban land use planning.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (grant nos. 71601047, 71601095, and 71631007).

References

- [1] W. J. Rose, J. E. Bell, C. W. Autry, and Cherry, "Urban logistics: establishing key concepts and building a conceptual framework for future research," *Transportation Journal*, vol. 56, no. 4, pp. 357–394, 2017.
- [2] K. Zheng, Z. Zhang, and B. Song, "E-commerce logistics distribution mode in big-data context: a case analysis of JD.COM," *Industrial Marketing Management*, vol. 86, pp. 154–162, 2020.
- [3] Z. Chen, J. Dong, and R. Ren, "Urban underground logistics system in China: opportunities or challenges?" *Underground Space*, vol. 2, no. 3, pp. 195–208, 2017.
- [4] Intergovernmental Panel on Climate Change, *IPCC Guidelines for national Greenhouse Gas Inventories*, Intergovernmental Panel on Climate Change, Geneva, Switzerland, 2006, <https://www.ipcc.ch/report/2006-ipcc-guidelines-for-national-greenhouse-gas-inventories/>.
- [5] M. Barth and K. Boriboonsomsin, "Real-world carbon dioxide impacts of traffic congestion," *Journal of the Transportation Research Board*, vol. 2058, no. 1, pp. 163–172, 2008.
- [6] M. Li and C. Mao, "Spatial effect of industrial energy consumption structure and transportation on haze pollution in Beijing-Tianjin-Hebei region," *International Journal of Environmental Research and Public Health*, vol. 17, no. 15, Article ID 5610, 2020.
- [7] I. Zandi, J. A. Warner, and R. Mersky, "Freight pipeline, rail, and truck cost comparison," *Journal of Transportation Engineering*, vol. 105, no. 4, pp. 411–425, 1979.
- [8] J. G. S. N. Visser, "The development of underground freight transport: an overview," *Tunnelling and Underground Space Technology*, vol. 80, pp. 123–127, 2018.

- [9] W. Hu, J. Dong, B. Hwang, R. Ren, and Z. Chen, "Hybrid optimization procedures applying for two-echelon urban underground logistics network planning: a case study of Beijing," *Computers & Industrial Engineering*, vol. 144, Article ID 106452, 2020.
- [10] H. Liang, G. Yuan, J. Han, and L. Sun, "A multi-objective location and channel model for ULS network," *Neural Computing and Applications*, vol. 31, no. S1, pp. 35–46, 2019.
- [11] M. Ren, Z. Fan, J. Wu, L. Zhou, and Z. Du, "Design and optimization of underground logistics transportation networks," *IEEE Access*, vol. 7, pp. 83384–83395, 2019.
- [12] M. He, L. Sun, X. Zeng, W. Liu, and S. Tao, "Node layout plans for urban underground logistics systems based on heuristic bat algorithm," *Computer Communications*, vol. 154, pp. 465–480, 2020.
- [13] Z. Yan, H. Ismail, L. Chen, X. Zhao, and L. Wang, "The application of big data analytics in optimizing logistics: a developmental perspective review," *Journal of Data, Information and Management*, vol. 1, no. 1-2, pp. 33–43, 2019.
- [14] R. Y. Zhong, S. T. Newman, G. Q. Huang, and S. Lan, "Big data for supply chain management in the service and manufacturing sectors: challenges, opportunities, and future perspectives," *Computers & Industrial Engineering*, vol. 101, pp. 572–591, 2016.
- [15] G. Wang, A. Gunasekaran, E. W. T. Ngai, and T. Papadopoulos, "Big data analytics in logistics and supply chain management: certain investigations for research and applications," *International Journal of Production Economics*, vol. 176, pp. 98–110, 2016.
- [16] N. Anand, M. Yang, J. H. R. van Duin, and L. Tavasszy, "GenCLOn: an ontology for city logistics," *Expert Systems with Applications*, vol. 39, no. 15, pp. 11944–11960, 2012.
- [17] H. I. Hendi, A. Ahmad, M. Bouneffa, and C. Fonlupt, "Logistics optimization using ontologies," in *Proceedings of ICCSA 2014*, Le Havre, France, June 2014.
- [18] S.-L. Chen, Y.-Y. Chen, and C. Hsu, "A new approach to integrate internet-of-things and software-as-a-service model for logistic systems: a case study," *Sensors*, vol. 14, no. 4, pp. 6144–6164, 2014.
- [19] C. K. M. Lee, Y. Lv, K. K. H. Ng, W. Ho, and K. L. Choy, "Design and application of internet of things-based warehouse management system for smart logistics," *International Journal of Production Research*, vol. 56, no. 8, pp. 2753–2768, 2018.
- [20] F. Shen and M. Sui, "Improved modern logistics system based on the integration of GPS, GIS, and RFID technology," in *Proceedings of the International Conference of Logistics Engineering and Management (ICLEM) 2010*, pp. 2353–2359, Chengdu, China, October 2010.
- [21] H. Gholizadeh, H. Fazlollahab, and M. Khalilzadeh, "A robust fuzzy stochastic programming for sustainable procurement and logistics under hybrid uncertainty using big data," *Journal of Cleaner Production*, vol. 258, Article ID 120640, 2020.
- [22] Z. Liu, Y. Liu, Q. Meng, and Q. Cheng, "A tailored machine learning approach for urban transport network flow estimation," *Transportation Research Part C: Emerging Technologies*, vol. 108, pp. 130–150, 2019.
- [23] J. Zhang, X. Qu, and S. Wang, "Reproducible generation of experimental data sample for calibrating traffic flow fundamental diagram," *Transportation Research Part A: Policy and Practice*, vol. 111, pp. 41–52, 2018.
- [24] R. Yan, S. Wang, and K. Fagerholt, "A semi-"smart predict then optimize" (semi-SPO) method for efficient ship inspection," *Transportation Research Part B: Methodological*, vol. 142, pp. 100–125, 2020.
- [25] R. Yan, S. Wang, and C. Peng, "An artificial intelligence model considering data imbalance for ship selection in port state control based on detention probabilities," *Journal of Computational Science*, vol. 48, Article ID 101257, 2020, In press.
- [26] R. Y. Zhong, G. Q. Huang, S. Lan, Q. Y. Dai, X. Chen, and T. Zhang, "A big data approach for logistics trajectory discovery from RFID-enabled production data," *International Journal of Production Economics*, vol. 165, pp. 260–272, 2015.
- [27] J. Hopkins and P. Hawking, "Big data analytics and IoT in logistics: a case study," *The International Journal of Logistics Management*, vol. 29, no. 2, pp. 575–591, 2018.
- [28] Y. Zhao, Y. Zhou, and W. Deng, "Innovation mode and optimization strategy of B2C e-commerce logistics distribution under big data," *Sustainability*, vol. 12, no. 8, Article ID 3381, 2020.
- [29] S. Shahparvari, A. Nasirian, A. Mohammadi, S. Noori, and P. Chhetri, "A GIS-LP integrated approach for the logistics hub location problem," *Computer & Industrial Engineering*, vol. 146, Article ID 106488, 2020.
- [30] B. Vernimmen, W. Dullaert, E. Geens et al., "Underground logistics systems: a way to cope with growing internal container traffic in the port of Antwerp?" *Transportation Planning and Technology*, vol. 30, no. 4, pp. 391–416, 2007.
- [31] J. Dong, W. Hu, S. Yan, R. Ren, and X. Zhao, "Network planning method for capacitated metro-based underground logistics system," *Advances in Civil Engineering*, vol. 2018, Article ID 6958086, 14 pages, 2018.
- [32] R. Sahraeian and M. S. Kazemi, "A fuzzy set covering-clustering algorithm for facility location problem," in *Proceedings of the 2011 IEEE Industrial Engineering and Engineering Management*, pp. 1098–1102, Singapore, December 2011.
- [33] Y. Park, P. Nielsen, and I. Moon, "Unmanned aerial vehicle set covering problem considering fixed-radius coverage constraint," *Computers & Operations Research*, vol. 119, Article ID 104936, 2020.
- [34] J.-F. Cordeau, F. Furini, and I. Ljubić, "Benders decomposition for very large scale partial set covering and maximal covering location problems," *European Journal of Operational Research*, vol. 275, no. 3, pp. 882–896, 2019.
- [35] M. Rahoual, R. Hadji, and V. Bachelet, "Parallel ant system for the set covering problem," in *Proceedings of the Third International Workshop, ANTS 2002*, pp. 262–267, Brussels, Belgium, September 2002.
- [36] L. A. Wolsey, "An analysis of the greedy algorithm for the submodular set covering problem," *Combinatorica*, vol. 2, no. 4, pp. 385–393, 1982.
- [37] M. I. Sviridenko, "Worst-case analysis of the greedy algorithm for a generalization of the maximum p-facility location problem," *Operations Research Letters*, vol. 26, no. 4, pp. 193–197, 2000.
- [38] Z.-G. Ren, Z.-R. Feng, L.-J. Ke, and Z.-J. Zhang, "New ideas for applying ant colony optimization to the set covering problem," *Computers & Industrial Engineering*, vol. 58, no. 4, pp. 774–784, 2010.
- [39] L. Lessing, I. Dumitrescu, and T. Stützle, "A comparison between ACO algorithms for the set covering problem," in *Proceedings of the 4th International Workshop, ANTS 2004*, pp. 1–12, Brussels, Belgium, September 2004.
- [40] S. Sundar and A. Singh, "A hybrid heuristic for the set covering problem," *Operational Research*, vol. 12, no. 3, pp. 345–365, 2012.

- [41] Y.-C. Chang, V. C. Li, and C.-J. Chiang, "An ant colony optimization heuristic for an integrated production and distribution scheduling problem," *Engineering Optimization*, vol. 46, no. 4, pp. 503–520, 2014.
- [42] M. E. O'Kelly, "A clustering approach to the planar hub location problem," *Annals of Operations Research*, vol. 40, no. 1, pp. 339–353, 1992.
- [43] C. Iyigun, "The planar hub location problem: a probabilistic clustering approach," *Annals of Operations Research*, vol. 211, no. 1, pp. 193–207, 2013.
- [44] G. H. Ball and D. I. Hall, *ISODATA, a Novel Method of Data Analysis and Classification*, Stanford Research Institute, Menlo Park, CA, USA, 1965.
- [45] S. Wang and H. Jiang, "Determine level of thief zone using fuzzy ISODATA clustering method," *Transport in Porous Media*, vol. 86, no. 2, pp. 483–490, 2011.
- [46] BTMB, *Urban Road Traffic Performance Index*, Beijing Municipal Commission of Transport, Beijing, China, 2011, in Chinese.
- [47] X. Li, Y. Lv, W. Sun, and L. Zhou, "Cordon- or link-based pricing: environment-oriented toll design models development and application," *Sustainability*, vol. 11, no. 1, Article ID 258, 2019.
- [48] K. Redding, B. Dikunow, and K. Krzykowska, "Proposal of big data route selection methods for autonomous vehicles," *Internet Technology Letters*, vol. 1, no. 5, Article ID e36, 2018.
- [49] M. A. M'hand, A. Boulmakoul, H. Badir, and A. Lbath, "A scalable real-time tracking and monitoring architecture for logistics and transport in RoRo terminals," *Procedia Computer Science*, vol. 151, pp. 218–225, 2019.
- [50] S. Mahdavi, M. E. Shiri, and S. Rahnamayan, "Metaheuristics in large-scale global continues optimization: a survey," *Information Sciences*, vol. 295, pp. 407–428, 2015.