This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Ranking Station Importance With Human Mobility Patterns Using Subway Network Datasets

Feng Xia, *Senior Member, IEEE*, Jinzhong Wang, Xiangjie Kong, *Senior Member, IEEE*, Da Zhang, and Zhibo Wang *Senior Member, IEEE*

***Abstract*— Complex networks have become an active interdisciplinary field of research inspired by the empirical study of various networks. A subway network is a real-world example of complex networks in the transportation domain, which has attracted growing attention in network analysis recently. Analyzing human mobility patterns, specifically in ranking subway stations closely bounded by urban subway planning and individuals' travel experience, is still an open issue. In this paper, we propose a novel ranking method of station importance (SIRank) by utilizing human mobility patterns and improved PageRank algorithm. Specifically, by analyzing human mobility patterns of the subway system in Shanghai, we demonstrate both static and dynamic characteristics using two network models (Shanghai subway static network and Shanghai subway passenger network). In particular, the SIRank focuses on bi-directional passenger flow analysis between origins and destinations to iteratively generate the importance value for each station. We implement a range of the experiments to illustrate the effectiveness of SIRank using the real-world subway transaction datasets. The results demonstrate that the hit ratio in SIRank reaches 60% in the top five stations, which is much higher than that of ranking by a weighted mixed index (WMIRank) and ranking by node degree (NDRank) approaches.**

***Index Terms*— Human mobility patterns, complex networks, PageRank, subway networks.**

I. Introduction

COMPLEX networks have become an active interdisciplinary field of research inspired by the empirical study of real-world networks, e.g. transportation networks, cellular networks, brain networks, computer networks, and social networks [1]. Complex network analysis has drawn more attention and become a hot spot of research [2]. Subway networks are a real-life example of complex networks, in which human mobility patterns can be analyzed and ranked by their station importance (SI).

Manuscript received March 31, 2017; revised November 17, 2017, April 8, 2018, August 17, 2018 and April 6, 2019; accepted May 31, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61572106 and Grant 61872274, in part by the Dalian Science and Technology Innovation Fund under Grant 2018J12GX048, and in part by the Fundamental Research Funds for the Central Universities under Grant DUT18JC09. The Associate Editor for this paper was B. F. Ciuffo. *(Corresponding author: Xiangjie Kong.)*

F. Xia and X. Kong are with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian 116620, China (e-mail: xjkong@ieee.org).

J. Wang is with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian 116620, China, and also with the School of Management, Journalism & Communication, Shenyang Sport University, Shenyang 110102, China.

D. Zhang is with the Electrical and Computer Engineering Department, University of Miami, Coral Gables, FL 33146 USA.

Z. Wang is with the School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China.

Digital Object Identifier 10.1109/TITS.2019.2920962

Commuting by subway in urban areas is a popular choice because of its economic and practical efficiency [3]. Increasingly more subway lines and stations are being constructed to improve transportation efficiency as well as relieve traffic congestions. This trend gradually forms a complex subway system. Therefore, it is vital to understand network topology, station importance, and human mobility patterns of a subway network. In turn, this knowledge subsequently aids a wide range of functions including city planning [4], [5], functional region identification [6], emergency response strategies [7], and public transportation management [8], [9]. On the other hand, it can also shed valuable light in the field of network science.

Ranking station importance is closely related to a range of applications, i.e., station investment [10], subway construction planning [11], and route planning [12]. First, SI ranking is an essential criterion for investing in the maintenance of subway stations. Stations of substantial importance should acquire more funding to improve their service quality. Second, SI ranking is helpful in subway construction planning. For example, the locations of transfer stations must be carefully considered when constructing a new subway line. Third, SI ranking score is quite useful when planning infrastructure construction and shuttle bus routes around subway stations.

Additionally, understanding the importance of stations also contributes to network robustness analysis [13], city planning [6], and advertising placement [14]. In recent years, the expansion of subway networks makes transit security face severe challenges. Natural disasters, human errors, and malicious attacks cause very serious consequences which threaten the safety of people's lives and property. To evaluate the robustness of subway networks, we usually remove critical and transit stations and compare the routing planning performance before and after removing these stations. Furthermore, researchers also propose evacuation plans and detour routing strategies in the event of emergency. Moreover, SI ranking can help city planners solve the contradiction between urban regional functions and people's living needs. In addition, city planners can analyze shortcomings in local development and prioritize the construction of the corresponding service venues. It is common practice to design the smart city plan today. Additionally, advertising is critical to business development.

However, how to select a good place is still an open issue in subway networks. From this perspective, our SI ranking method provides useful information for advertising agencies, which not only saves the marketing expenses for the company but also enhances the product promotion and influence of the company.

Based on passenger mobility patterns, centrality, and connectivity, we leverage the SI metric to identify the most important station. Schematically, a subway station is defined as a node and a journey is regarded as a link. The spatio-temporal distribution of passenger flow and network topological characteristics determine the strength of ties between nodes.

In the past few years, several types of research methods have been conducted to analyze subway networks and passenger flow using complex network theory. Such work mainly focuses on network topology [15], network robustness [16], [17], network evolution [18], network efficiency [19], [20], network node importance [17], [21], [22], passenger flow patterns [23]–[25], and urban event detection [26]. The research provides a valuable reference and lays a solid foundation for this paper.

The current findings also exhibit limitations in exploring topology and human mobility patterns. To be specific, primary methods typically focus on investigating static network topology without considering passenger movement trajectories, and usually ignore the dynamic nature of subway networks. Additionally, although passenger flow patterns have been examined in previous research [23]–[25], it is limited to adjacent stations and passenger flow with a single state. The interrelationship between origins and destinations are typically ignored [24]. Furthermore, the trip direction is another critical factor for evaluating station importance, which has always been omitted in the current research.

In this paper, we use subway smart card transaction datasets consisting of 14 subway lines and 288 subway stations from Shanghai to analyze human mobility patterns and station importance. First, we construct two topological network models: (1) a Shanghai Subway Static Network (SSSN) based on physical infrastructure configurations, and (2) a Shanghai Subway Passenger Network (SSPN) based on passengers' origins and destinations. According to [27], we assume that passenger flow follows the geodesic paths in SSSN and SSPN. Next, we leverage three macroscopic metrics and three microscopic metrics to compare and analyze these two network topologies and acquire human mobility patterns. Then, we develop a novel Ranking Method of Station Importance (SIRank) to evaluate station importance during operation.

Our main contributions can be generalized as follows:

- We propose a definition of SI and a new Ranking Method of SI (SIRank) to rank subway stations by utilizing human mobility patterns and improved PageRank algorithm. Moreover, we verify the effectiveness of SIRank by comparing it to Ranking by a Weighted Mixed Index (WMI-Rank) and Ranking by Node Degree (NDRank) methods.
- We leverage microscopic and macroscopic indicators to analyze static and dynamic characteristics of Shanghai subway network.
- We propose a novel network model referred to as SSPN, which is a directed weighted graph based on the geodesic paths between origins and destinations. SSPN breaks geographical spatial limits in the network.
- We discover a heavy-tailed phenomenon for the degree distribution in SSSN and a relatively high clustering coefficient in SSPN through analyzing the characteristics of the two models.

The remainder of the paper is structured as follows. First, we briefly review the related work and outline their limitations in Section II. Subsequently, we provide a detailed description of the two network models and six indicators in section III. In Section IV, we define our SI metric and describe the proposed SIRank. Section V illustrates the datasets and the experiments we conduct to compare SIRank with other two methods. Finally, we conclude in Section VI.

II. Related Work

Ranking subway stations is an interesting research area primarily motivated by the empirical study of real-world networks such as social networks [1], the world economic networks [28], and transportation networks [17], [18], [29], [30] in the last decade. In this section, we review relevant literature and highlight related techniques from the following two aspects.

A. Subway Network Analysis

Subway network analysis has experienced a rapid development due to three significant network models, namely, random networks [31], small-world networks [32], and scale-free networks [33]. Specifically, random networks show that the number of vertices and edges are distributed randomly in a graph. Small-world networks propose a power-law degree distribution feature, and scale-free networks discover a characteristic of short path length and high clustering coefficient in a network. In general, these discoveries are landmarks in the field. Meanwhile, several researchers concentrate on analyzing the topology, connectivity, reliability, robustness, and node importance of subway networks.

The authors in [16] examine betweenness centrality of 28 subway networks and analyze the regularities of the evolution of centrality when network size varies. Louf *et al.* [15] evaluate the scaling properties of subway systems by utilizing three indicators: the number of stations, the total length, and the ridership.

The authors in [18] investigate the evolutionary patterns of Beijing subway network and propose a growth model utilizing an expanding and intensifying mode. Furthermore, Vragović *et al.* [19] introduce the concept of network efficiency for information exchange in the Boston subway network. Yang *et al.* [17] assess the robustness of Beijing subway network under random failures as well as malicious attacks. They verify their model's high reliability and robustness and also propose a new weighted mixed index (degree and betweenness centrality) to evaluate node importance. Wei *et al.* [21] focus on a grading method of subway stations by utilizing static factors (location, surrounding facilities, etc.)

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

and dynamic features (such as passenger flow), but they only perform rough classifications of stations instead of ranking them.

The research discussed above mainly focuses on the subway network topology, while neglecting the spatio-temporal changes in passenger flow and the interrelationships of subway stations. In a subway system, travel characteristics of people and their social activities usually play an important role in analyzing network topology and evaluating detour planning strategies.

B. Human Mobility Patterns

Human mobility patterns have recently attracted researchers partially because traffic trajectory data can be easily accessed. Exploring patterns of human mobility improves urban traffic conditions [20], identifies urban functional areas [34], predicts human mobility [35], and increases the efficiency of our lives [26], [36].

Lévy flight model lays a solid foundation for human travel regularities, which demonstrates trip distance following a power law distribution [37]. Furthermore, Song *et al.* [38] find that human mobility exhibits spatio-temporal characteristics, and they introduce the concept of mobile phone users' entropy which can predict mobility patterns with an accuracy rate of 93%. In addition, Simini *et al.* [39] prove that the radiation model usually has a higher forecast precision than the gravity law in human mobility patterns.

Zhao *et al.* [40] examine four transportation modes and find out a single mode following a log-normal distribution and a mixed mode following a power law distribution. These findings shed light on the understanding of human mobility patterns. Later, Calabrese *et al.* [41] prove that the trip displacement of people follows a power law with an exponential cut-off by analyzing traffic trajectory data. Wang *et al.* [42] analyze taxi trajectory data from five cities and propose an exponential distribution that fits trip displacement and a log-normal distribution that fits trip duration.

Veloso *et al.* [43] find that Gamma distribution fits trip distance, and an exponential distribution fits trip interval by using a taxi dataset in Lisbon. Meanwhile, Csáji *et al.* [44] illustrate that commuting distances follow a log-normal distribution based on mobile phone call records in Portugal. Additionally, Chen *et al.* [26] propose a tensor co-factorization based data fusion framework, for urban event detection by using crowd mobility and social activity data. Lenormand *et al.* [36] utilize credit-card records to analyze socio-demographic phenomena in human mobility patterns based on three human attributes (gender, age, and occupation).

The authors in [29] construct a complex weighted network to analyze passenger flow in the rail transit system of Singapore and notice a heavy volume of passengers on hub nodes. However, it was limited by the adoption of a coarse-grained method with a time slot of a day. Hasan *et al.* [45] propose a human mobility model that predicts the visiting locations of people through the analysis of smart subway card transactions. Lee *et al.* [46] develop a master equation approach to analyze passenger flow distributions in the subway system of Seoul.

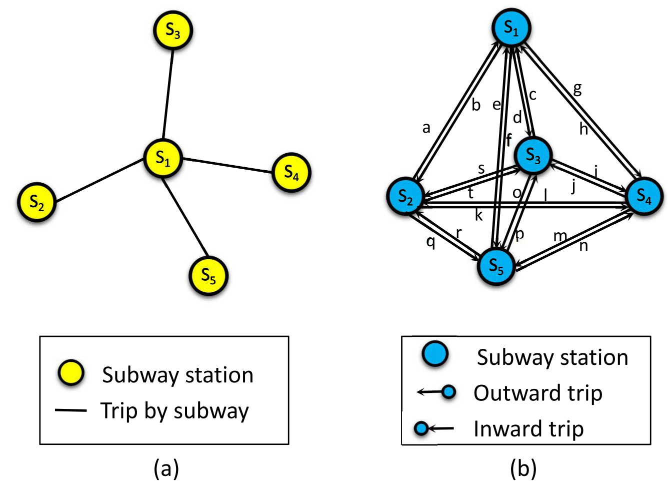

Fig. 1. Two different network topologies of Shanghai subway. (a) SSSN topology. (b) SSPN topology.

Roth *et al.* [47] study the structure and formation of the city based on trip data from the subway system of London. Xu *et al.* [24] utilize human mobility data to explore individual passenger movement patterns in the Beijing subway system and identify ten cluster structures in the network.

The research mentioned above has not considered the underlying physical topology of subway systems in great detail and lacks a refined validation method to evaluate network node importance. In this paper, we leverage travel regularities to thoroughly analyze the subway system of Shanghai by reevaluating the topology and node importance in the network.

III. Preliminary

A. Network Model

The first subway line built in Shanghai was in 1993. Today, Shanghai subway system has grown to a complex system with 14 subway lines and 288 subway stations. The total length of Shanghai subway is 567 km and ranks the first in the world as of 2014. To illustrate the topological characteristics of Shanghai subway network, we represent it using two models (SSSN and SSPN) leveraging L-space [48] and spatio-temporal patterns of human mobility. As shown in Fig. 1, SSSN displays topological and connectivity features, while SSPN provides further details including the direction and volume of passenger flow.

Definition 1: (SSSN) $G_s = (V_s, E_s)$ is an undirected graph, where V_s denotes the set of subway stations ($V_s \neq \emptyset$) and E_s represents the set of edges ($\| E_s \| \geq 1$). According to the theory of L-space, there exists an edge $e = (v_{si}, v_{sj}) \in E_s$ linked directly between an origin v_{si} and a destination v_{sj} in a real subway network, where $v_{si}, v_{sj} \in V_s$. As shown in Fig. 1(a), SSSN describes the original topology of Shanghai subway.

Definition 2: (SSPN) $G_p = (V_p, E_p)$ is a directed weighted graph, where V_p denotes the set of subway stations ($V_p \neq \emptyset$) and E_p represents the set of edges ($\| E_p \| \geq 1$). According to the distribution of subway passenger flow, there exists a directed edge $e_p = (v_{pi} \rightarrow v_{pj}) \in E_p$, if there exists an origin-destination trip from an origin υ_{pi} to a destination υ_{pj} in a real subway network, where $\upsilon_{pi}, \upsilon_{pj} \in V_p$. Meanwhile, the weight of e_p is equal to the volume of passenger flow from υ_{pi} to υ_{pj}. As shown in Fig. 1(b), SSPN analyzes Shanghai subway from the perspective of passenger flow.

L-space represents the original network topology of real-world transport networks, in which stations are vertices and edges link any two consecutive stations on a specific route. In contrast, P-space represents the spatial topology of real-life transport networks [49]. Two nodes are connected if they are linked with at least a subway line. The path weight between two nodes denotes the transfer frequency in the trip. However, P-space only considers the distribution of passenger flow using a single physical subway line. In this paper, we investigate a different network topology concerning the relationship between a pair of stations from different subway lines. In addition, Borgatti [27] mainly focus on the pattern of traffic flow within a network and categorize traffic into four types, i.e. geodesics, paths, trails, or walks. In an urban subway scenario, a passenger who travels for his/her social activities has fixed destinations and origins. In addition, a passenger normally knows and selects the shortest route, so we can assume that the travel trajectory follows the shortest path (with the utilization of transfer stations). To be specific, we acquire the coordinates of a trip's origin and destination and calculate the geodesic path D as follows:

$$D = r \times \arccos(\sin(\arg(\beta_2))\sin(\arg(\alpha_2)) + \cos(\arg(\beta_2)) \times \cos(\arg(\alpha_2))\cos(\arg(\beta_1 - \alpha_1))), \quad (1)$$

where r denotes the Earth radius with the value of 6370 km, α_1 and α_2 represent the origin's longitude and latitude respectively, and the destination's longitude and latitude are expressed by β_1 and β_2 respectively.

For SSSN, a passenger's travel distance is equal to the sum of the shortest geodesic path between any two stations. For SSPN, we directly calculate the geodesic path between an origin and a destination. Our proposed SSPN mainly focuses on how to analyze network architecture and rank station importance by utilizing human mobility patterns and network indicators in Shanghai subway system. In SSPN, we use a bi-directional weighted graph to illustrate both incoming and outgoing passenger flow at subway stations.

B. Human Mobility Macroscopic Indicators

1) Time: Time is one of the macroscopic indicators of human mobility patterns. In general, the daily activities of most commuters exhibit temporal regularities on weekdays. However, people tend to possess different movement patterns on weekends. Extensive experiments are conducted on weekdays and weekends at different time slots. We set δ hour(s) as an interval for each day as shown in Equation (2):

$$t_h = [h\delta, (h+1)\delta), h = 0, 1, \ldots, (24/\delta) - 1, \quad (2)$$

where t_h is the hth time slot. We define one hour to be the time interval and acquire transaction records spanning 20 time slots from 4:00 to 23:00.

2) Location: Location is another macroscopic indicator of human mobility patterns and reflects the spatial law regarding the movement of people. In this paper, this metric is closely related to the geographic coordinates of subway stations. Moreover, it is well known that different locations possess different social functions.

3) Volume: The subway passenger volume is the third important macroscopic indicator of human mobility patterns. In particular, we evaluate station importance concerning the passenger flow of incoming links. In other words, the frequency of passenger flow is measured against node importance.

C. Network Microscopic Indicators

Network indicators are often used in complex network analysis [2]. In this paper, we introduce three microscopic network indicators motivated in the following three aspects. First, subway networks are real-world examples of complex networks. A subway station is deemed as a node and a trip as an edge in complex networks. Therefore, we can take the problem of ranking station importance in subway networks as an issue of evaluating node centrality in complex networks. When analyzing node centrality, several network indicators (degree centrality, closeness centrality, and clustering coefficient) are crucial factors [27]. Second, some scholars also utilize network indicators to analyze the topology of subway networks and evaluate the node centrality [15]. Furthermore, some approaches are also proposed to rank station importance [17], [21], [22]. Third, we combine network indicators with human mobility patterns and conduct extensive experiments to verify the effectiveness of our proposed algorithm SIRank.

1) Node Degree and Degree Centrality: Node degree depicts a significant level of nodes' transferability. In SSSN, the degree of station i (D_i) represents the number of the stations linked with it. In SSPN, we introduce the in-degree ($InD(i)$) and out-degree ($OutD(i)$) values to measure the corresponding importance level of station i. The equations of node degree D_i, in-degree InD_i, and out-degree $OutD_i$ are defined as follows:

$$D_i = \sum_{j=1, j\neq i}^{N} E_{ij}, \quad (3)$$

$$InD_i = \sum_{j=1, j\neq i}^{N} E_{i\leftarrow j}, \quad (4)$$

$$OutD_i = \sum_{j=1, j\neq i}^{N} E_{i\rightarrow j}, \quad (5)$$

where N denotes the number of subway stations. If there exists an undirected link between stations i and j, $E_{ij} = 1$; Otherwise, $E_{ij} = 0$. InD_i denotes the total number of connections which terminate at station i, and $OutD_i$ denotes the total number of connections originated from station i.

To analyze the distribution of node degree thoroughly, we introduce degree centrality (DC_i) to normalize the links

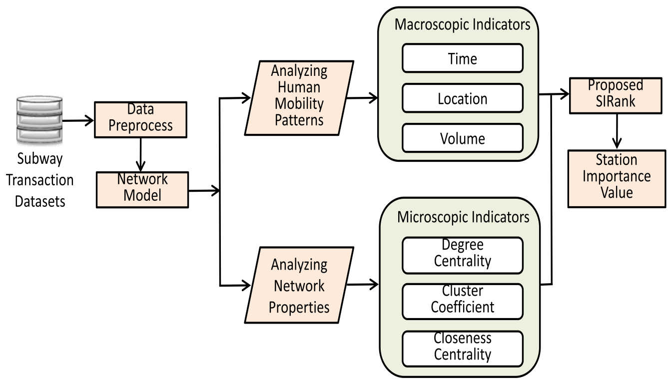

Fig. 2. Framework for ranking station importance.

of nodes for SSSN and SSPN. DC_i is defined as follows:

$$DC_i = \frac{D_i}{N-1}. \tag{6}$$

2) Clustering Coefficient: In this paper, we utilize this metric to identify groups with a relatively high density of ties within the subway network. In addition, this metric can be used as the characteristic of a scale-free network or small-world phenomenon with high connectivity. C_i represents the clustering coefficient of node i. C denotes the average clustering coefficient and reflects the density relationships between all subway stations in SSSN and SSPN.

$$C_i = \frac{2D_i}{N(N-1)} \quad 0 < C_i \leq 1, \tag{7}$$

$$C = \frac{\sum_{i \in G} C_i}{N}. \tag{8}$$

3) Closeness Centrality: Closeness centrality (CC) reflects a station's closeness to other stations in SSSN and SSPN. A larger value of CC represents a shorter total distance to all the other stations and thus indicates a higher importance level. The metric is defined as the reciprocal of the sum of the shortest path length from a station to all the other stations. The formula is defined as follows:

$$CC_i = \frac{1}{\sum_{i,j \in G, i \neq j} d_{ij}}, \tag{9}$$

where d_{ij} represents the shortest path length between stations i and j. In addition, the average of CC indicates the overall operation efficiency of the network.

IV. Framework of the Ranking Method

A. Overview

As shown in Fig. 2, we first filter out erroneous and irrelevant records and utilize a statistical analysis method to acquire real passenger transaction datasets from subway smart cards. Next, we construct SSSN and SSPN to analyze static and dynamic characteristics of the subway network, extract human mobility patterns, and acquire degree centrality, clustering coefficient, and closeness centrality of all stations. Finally, we define the SI metric and propose a ranking method SIRank to evaluate the importance level of each subway station, which is valuable for city planning and public transport management.

Nowadays, subway networks contribute to the acceleration of urbanization process in China (Shanghai being no exception). To our knowledge as the authors, these indicators have been validated as being practically effective [22]. Based on the analogy between subway passenger networks and complex networks, we use three microscopic metrics to rank the SI of each station. These are degree centrality, clustering coefficient, and closeness centrality.

Considering the spatio-temporal distribution of passengers, there also exist dynamic characteristics in subway networks. These dynamic features are essential for evaluating which subway stations are more important. People usually travel for their social activities such as working, shopping, and entertainment. Furthermore, we find that trajectory data has some spatio-temporal regularities. For example, in the morning rush hour, the majority of passenger flow is from residential regions to commercial districts. Therefore, we introduce three macroscopic indicators (time, location, and volume) to analyze human mobility patterns and identify the most important stations. We elaborate the ranking mechanism for the SI in the following subsections.

As shown in Fig. 3, the 14 Shanghai subway lines are denoted in different colors. Transfer stations, where multiple subway lines or branches meet or cross, are shown in large black circles (as opposed to small color circles for non-transfer stations). It should be noted that physical transfers take place in connected subway lines, such as in Xinzhuang Station which connects subway Lines 1 and 5. Hence, stations with connecting subway lines are also considered transfer stations.

B. PageRank

It is well known that PageRank algorithm [50] has been developed by Google to rank websites based on their search engine results. PageRank is an algorithm for quantifying the importance of web pages by calculating the number and quality of links to that page. Its basic assumption is that websites of greater importance are likely to obtain higher rankings by obtaining a larger quantity of links from influential websites. This algorithm can not only be applied to rank the importance of websites on the World Wide Web but also calculate the importance of stations in traffic networks. The ranking score of each node is generated iteratively by considering the number and importance level of neighboring nodes.

Considering all the nodes in the network, the model develops an iterative process shown in Equation (10).

$$PR' = \alpha \mathbf{M} PR + (1-\alpha)q, \tag{10}$$

where PR' is the ranking score vector for the next step, and q is a row vector (0, …, 1, …, 0). $\mathbf{M}$ is the transfer matrix that denotes the probability for each node to go to the neighboring node. The process will not stop until each node is assigned a constant ranking score.

For PageRank algorithm, there exists a hypothesis that web pages with more important information are likely to obtain higher ranking scores. In other words, PageRank assumes

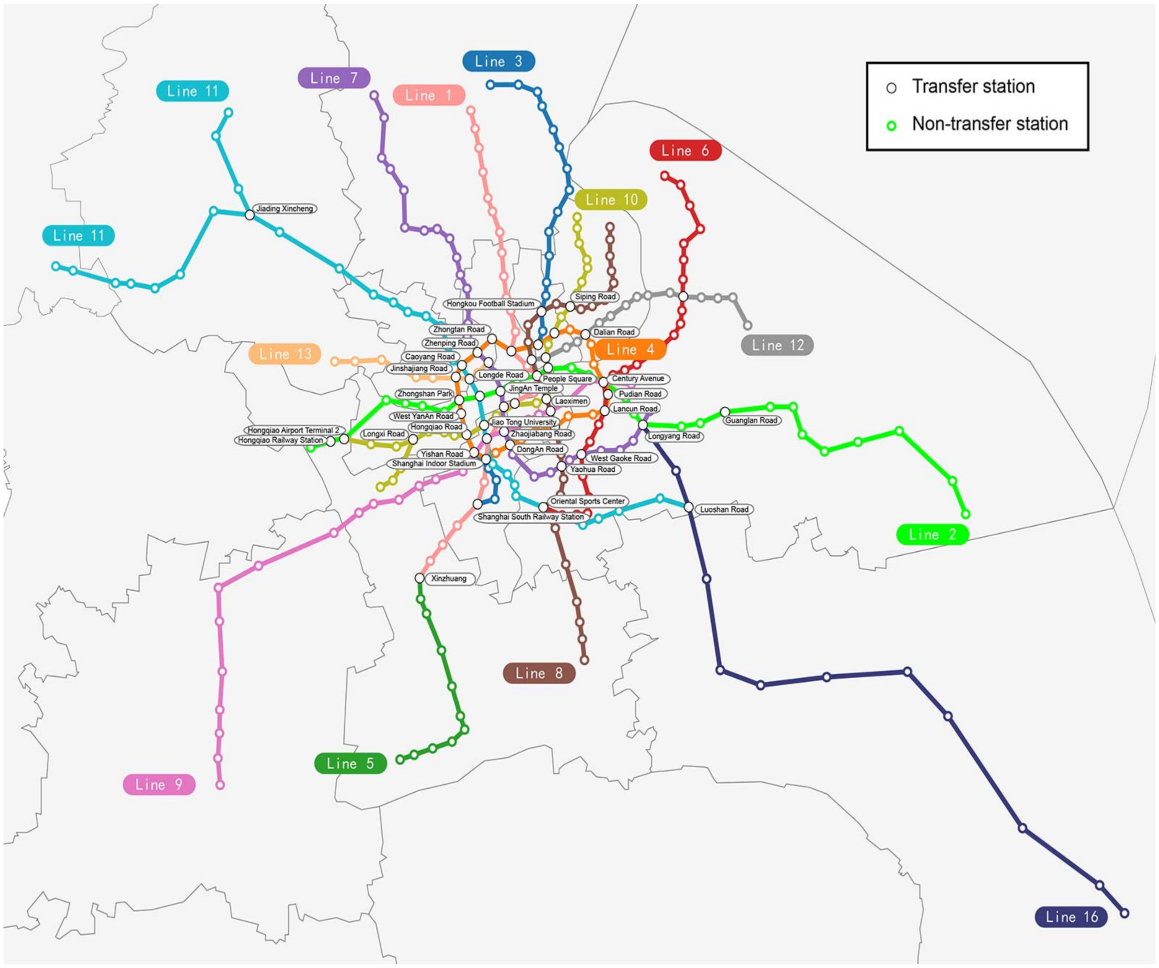

Fig. 3. Shanghai subway network map.

that the web pages receive more links for its degree of correlation with a given web page. However, passenger flow in urban traffic scenarios is quite different from web page traffic. In daily life, people tend to travel in regards to their social activities along with their predetermined path and station sequence. Therefore, in addition to static properties of stations, the dynamic distribution of passenger flow also significantly impacts SI. Thus, we introduce human mobility patterns to reevaluate the SI metric for each station. In SIRank, we allocate various weights to edges in attempt to reflect the importance of each node.

C. Proposed SIRank Method

In this paper, SIRank mainly assesses the importance level of stations based on their topology and human mobility patterns. SIRank is inspired by two factors: one, passengers usually travel from one station to another for their social activities, and the other, passenger flow indicates the strength of ties between origin and destination stations. Moreover, SIRank stems from PageRank algorithm, which has been proved to be suitable for ranking the importance of nodes in complex networks [50]. According to the above-mentioned metrics, we rank subway stations by utilizing static network properties and dynamic laws of human mobility and assess the topology of the subway network with greater accuracy.

Definition 3: (Station Importance (SI)) In subway networks, SI is defined as follows:

$$SI(i) = \alpha SI\_H(i) + \beta SI\_N(i), \tag{11}$$

$$SI\_H(i) = d \sum_{t \in T} \sum_{j \in P(i)} \frac{SI(j) \overleftarrow{W_{i,j}}^{t}}{20 \sum\limits_{k \in P(j)} \overleftarrow{W_{k,j}^{t}}} + \frac{(1-d)}{N}, \tag{12}$$

$$SI\_N(i) = \gamma DC_i^t + \varphi C_i^t + \lambda CC_i^t. \tag{13}$$

In Equation (11), $SI(i)$ denotes the ranking score of station i. $SI\_H(i)$ denotes the ranking score of station i considering human mobility patterns. $SI\_N(i)$ denotes the ranking score of station i considering the characteristics of the subway network topology. In Equation (12), $P(i)$ is a set containing all of the neighbors of station i. In addition, N is the total number of stations in the networks and d is a dampening factor that is usually set as 0.85 [51]. For Pagerank, the transfer matrix is comprised of the inverse of out-degree of nodes. However, the transfer matrix for SIRank includes dynamic factors (distribution of spatio-temporal passenger flow) that denote the probability of passengers to go from a given station to other stations with an average value of total slots T. In Equation (13), we define $\overleftarrow{W_{i,j}}^{t}$ as the volume of passenger flow from station j to i at the tth time slot, which ranges from 1 to 27,530. Meanwhile, DC_i^t, C_i^t, and

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

CC_i^t are degree centrality, clustering coefficient, and closeness centrality of station i at the tth time slot, respectively. The value of each of the three indicators is between 0 and 1.

For five parameters α, β, γ, φ and λ, $\alpha + \beta = 1$ and $\gamma + \varphi + \lambda = 1$. And their values range from 0 to 1. To obtain good experimental results, we set the probability of random jump as 0.1. At present, the commonly used parameter estimation methods are linear regression, multiple linear regression, and vector machine regression [52]. According to our experimental characteristic, we apply multivariate linear regression to estimate the parameters of SIRank. We discover that α obtains a relatively high value, and β gets a relatively low value. We also estimate the optimal parameters to compare the ranking results with WMIRank and NDRank. Specifically, stations with higher node degree acquire higher ranking scores for $\beta \geq 0.5$. When β is 0.9, the performance of SIRank is almost equal to that of NodeRank. If $\beta \leq 0.3$, SIRank has a skewed ranking result and stations in residential areas get higher ranking scores. For γ, φ and λ, degree centrality has a greater impact on ranking scores of stations than the other two parameters.

Algorithm 1 Pseudocode of SIRank

Input: Subway Transaction Datasets
Output: The Importance of Stations

procedure :Generate the SI value for each Station
  Assume all stations as nodes
  $\epsilon \leftarrow 0.00001$
  $d \leftarrow 0.85$
  for each node $i \in V_p$ **do**
    $SI(i) \leftarrow 1/N$
  end for
  for each node $i \in V_p$ **do**
    $SI(i)' \leftarrow 0$
    for each node $j \in node\ i's\ neighbors$ **do**
      $S = \frac{1}{20} \sum_{t \in T} \sum_{j \in P(i)} \frac{\overleftarrow{W_{i,j}}^t}{\sum_{k \in P(j)} \overleftarrow{W_{k,j}}^t}$
      $SI\_H(i) \leftarrow dSI(j)S + (1-d)/N$
      $SI\_M(i) \leftarrow \gamma DC_i^t + \varphi C_i^t + \lambda CC_i^t$
      $SI(i)' \leftarrow SI(i)' + \alpha SI\_H(i) + \beta SI\_M(i)$
      if $|SI(i)' - SI(i)| \leq \epsilon$ **then**
        $Exit$
      end if
      $SI(i) \leftarrow SI(i)'$
    end for
  end for
  return $SI(i)', i \in V_p$
end procedure

The proposed SIRank is illustrated in Algorithm 1. We analyze the stability and time complexity of calculating the SI index. We select two stations (People Square and Shanghai Railway Station) as test objects which are denoted by S_1 and S_2 respectively. We utilize different datasets to verify whether the relative positions of S_1 and S_2 change in the ranking list. The results show that S_1 is always in front of S_2, which confirms the stability of SIRank. In addition, there is a problem of artificially cheating in PageRank to improve the ranking of websites, and this problem does not exist in SIRank in public transit scenarios. Meanwhile, we assume that $|S|$ denotes the number of non-zero elements in transfer matrix S. The importance propagation of SIRank utilizing multiplication of transfer matrix needs $O(n|S|)$ time with n iterations. Actually, SIRank converges only after a limited number of iterations. Therefore, the efficiency is mainly affected by the passenger flow and network metric values. SIRank is able to effectively rank the station importance in the subway network from a data-driven perspective.

TABLE I
FORMAT OF THE DATASETS

Name	Type	Example
CardID	Varchar	2602022534
Date	Date	2015-04-01
Time	Time	07:20:49
Line	Char	No.9
Station	Varchar	Middle Yanggao Road
Longitude	Float	121.555042
Latitude	Float	31.233833

V. EXPERIMENTS AND ANALYSIS

A. Datasets Description and Preprocessing

In this paper, we collect transaction records from the subway smart cards in Shanghai subway system during April 2015. Shanghai subway included 14 subway lines and 288 stations at the time of data collection. As listed in Table I, the dataset consists of seven fields with more than 451 million transaction records, containing details regarding spatio-temporal patterns of human mobility.

To improve the statistical precision, we first cleaned the dataset and excluded error data generated by faulty devices and human failure. The percentage of these error logs is about 0.95%. Based on statistical analysis, we extracted the origins and destinations of each person on a daily basis by analyzing passenger flow between pairs of subway stations, with consideration for human behavioral regularities. Furthermore, we conducted an in-depth analysis of three time slots (morning rush hour, noon, and evening rush hour) on weekdays and weekends. Particularly, we introduced indicators to assess SI using static network metrics and dynamic passenger flow respectively.

B. Exploring Human Mobility Patterns

According to three-dimensional properties (time, location, and volume), we quantitatively analyze passenger movement patterns in Shanghai subway network on weekends and weekdays. We then construct a chord diagram to illustrate the characteristics of passenger flow. Each circle consists of arcs of various colors, with each color representing a different subway line. The links denote passenger flow between each pair of subway lines. The thickness of each link denotes

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

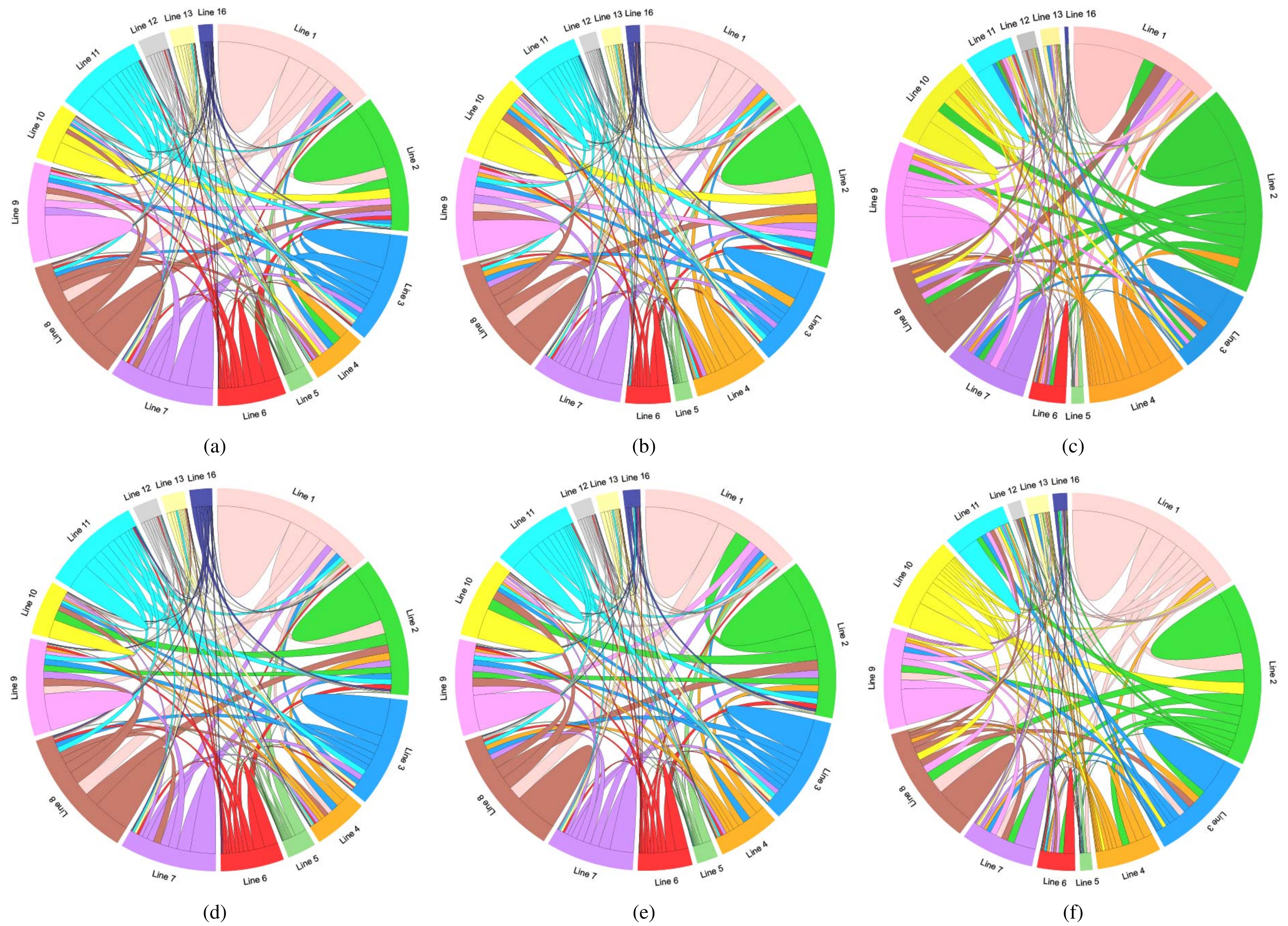

Fig. 4. Spatio-temporal patterns of human mobility from the perspective of subway lines. (a) Morning rush hours on weekdays. (b) Noon break hours on weekdays. (c) Evening rush hours on weekdays. (d) Morning rush hours on weekends. (e) Noon break hours on weekends. (f) Evening rush hours on weekends.

the volume of passenger flow between two subway lines; thicker links represent larger passenger flow. The color of links denotes the direction of trips. Links with the same color in a circle represent an outgoing passenger flow. Simultaneously the inverse corresponds to incoming passenger flow.

As shown in Fig. 4, we observe that Lines 1 and 2 have relatively more passenger flow than that of Line 16. Line 2 has a close tie with Line 12 as compared with other lines as seen in Figs. 4(a) – 4(c). On weekends, most people engage in non-work related activities, which results in a skewed human mobility pattern. For example, the trip direction between Lines 2 and 10 on weekdays is opposite to that on weekends according to Figs. 4(d) – 4(f).

We explore the patterns of human mobility from the perspective of an association between stations and shed light on the knowledge concerning Shanghai subway network. As shown in Fig. 5, larger passenger flow exists in the morning and evening rush hours on weekdays than that on weekends. As shown in Figs. 5(a) – 5(c), we notice that several hot spot subway routes exist, namely, from Tonghe Xincun to Shanghai Railway Station, from Xinzhuang to People's Square, and from Jiuting to Caohejing Hi-Tech. Interestingly, round trips on the same routes exist during the evening rush hours. This implies that the three originating stations are located in residential areas and therefore possess large travel demands towards the city, during the morning rush hours.

As shown in Figs. 5(d)–5(f), it can be seen that two popular subway trips exist, namely, from Xinzhuang to People's Square and from East Xujin to Zhongshan Park. People's Square resides in the central business district, an area to which more people travel for leisure and entertainment.

C. Analyzing Network Indicators

1) Degree and Degree Centrality Distribution: As shown in Fig. 6(a), we observe that degree distribution does not follow the power or Poisson laws, which are characteristics of scale-free and small-world networks, respectively. As the size of network increases, subway stations cannot be connected preferentially due to the limitation of geographical structure of the city. This factor contributes to the characteristics of degree distribution. In addition, other factors such as popular station considerations, balanced coverage, and costs, may also impact the distribution of degree. In SSSN, nodes with a degree of 2 have a high proportion of 77% in all nodes. Furthermore, we calculate the Cumulative Probability Distribution (CPD) of degree and find out a heavy-tailed phenomenon exists. The average degree of the network is also computed as 2.3125.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

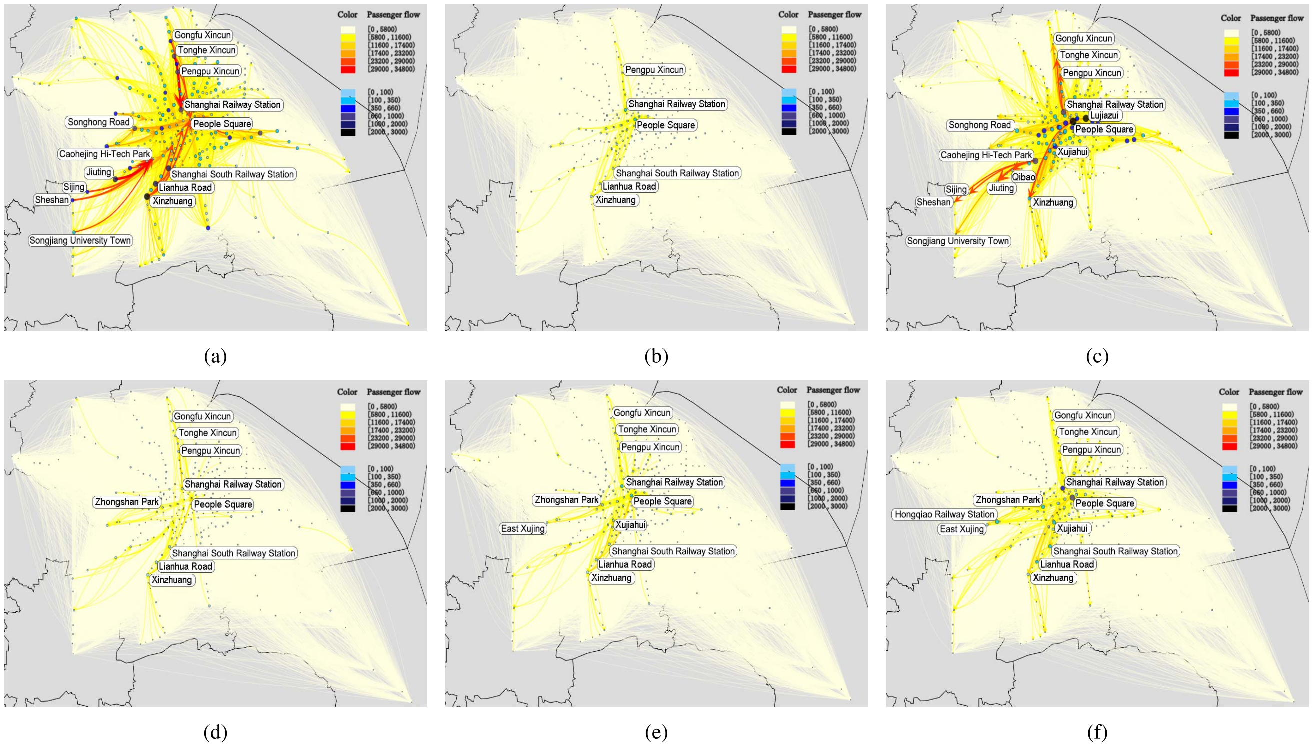

Fig. 5. Spatio-temporal patterns of human mobility with respect to subway stations. (a) Morning rush hours on weekdays. (b) Noon break hours on weekdays. (c) Evening rush hours on weekdays. (d) Morning rush hours on weekends. (e) Noon break hours on weekends. (f) Evening rush hours on weekends.

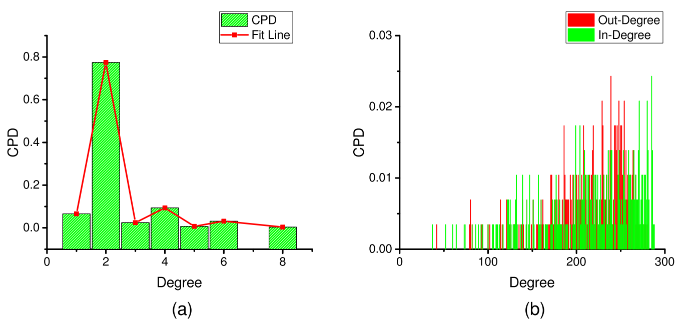

Fig. 6. Distribution of node degree. (a) SSSN. (b) SSPN.

Different CPD for in-degree and out-degree exists for SSPN according to Fig. 6(b). The average values of in-degree and out-degree are 203 and 197 respectively, which are much larger than the node degree in SSSN. Through comparative analysis, we find that popular stations are mainly located near People's Square, Shanghai Railway Station, and Lujiazui.

We normalize the degree centrality and compare the difference between weekdays and weekends. As shown in Fig. 7(a), without considering passenger flow factor in SSSN, the degree centrality remains low. In SSPN, we notice that the degree centrality in the evening rush hours is obviously larger than that in other time slots. Furthermore, the degree centrality fluctuates more frequently on weekends, as shown in Fig. 7(b). This implies that travel patterns are more regular on weekdays than that on weekends.

2) Clustering Coefficient Distribution: As shown in Fig. 8, we acquire the distribution of clustering coefficients in SSSN.

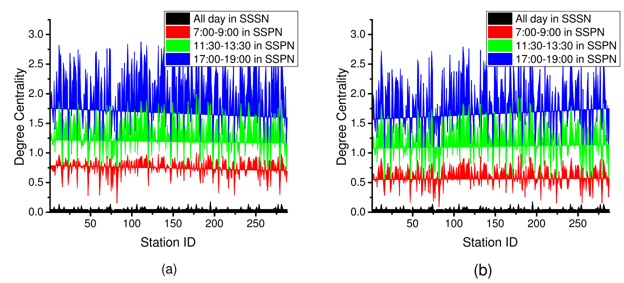

Fig. 7. Distribution of degree centrality. (a) On weekdays. (b) On weekends.

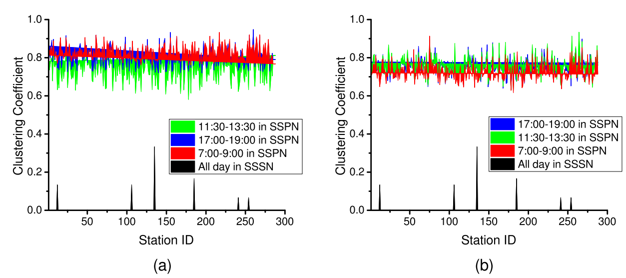

Fig. 8. Distribution of clustering coefficient. (a) On weekdays. (b) On weekends.

To our surprise, only six nodes exist with a non-zero value. These six nodes form two triangular routines, i.e., Shanghai Indoor Stadium, Yishan Road, and Xujia Hui. Furthermore, the average value is only 0.003125 which is much less than that of small-world networks. It also reflects the loose connectivity of the network.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

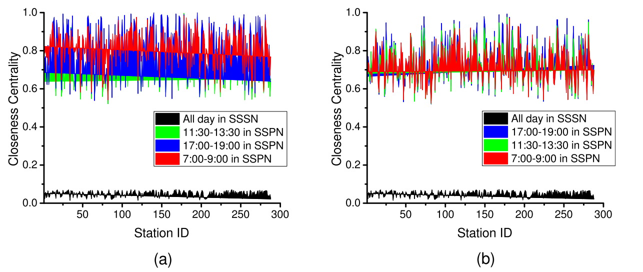

Fig. 9. Distribution of closeness centrality. (a) On weekdays. (b) On weekends.

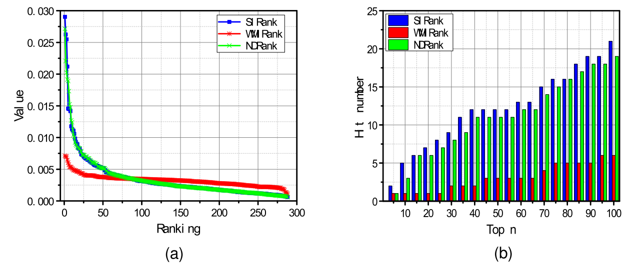

Fig. 10. Experimental results of three ranking algorithms. (a) Rank value of all stations. (b) Number of hit stations.

The clustering coefficient in SSPN is 0.828249 on weekdays and 0.804645 on weekends, which demonstrates the close connectivity of the network. We also note that a different distribution law in SSPN exists, where the value during morning rush hours is higher than that at noon on weekdays and is the opposite on weekends. This indicates that the travel patterns of people are similar in the mornings during weekdays due to their commuting routines. On the other hand, it has a higher clustering coefficient during the noon time on weekends, which demonstrates travel preferences and social activities of people.

3) Closeness Centrality Distribution: This metric reflects the efficiency of the network as it correlates to the inverse of the length of the shortest path. As shown in Fig. 9, the value of SSPN falls within the interval [0.6, 1.0] whereas the value of SSSN drops to the interval [0.0, 0.5]. Furthermore, we find that the value of closeness centrality is the largest during rush hours on weekday mornings and rush hours on weekend evenings. It partially shows traffic congestion condition during these period. Specifically, Shanghai Railway Station is the most popular station with a value of 0.9829 between 7:00 and 9:00 on Monday, while Century Avenue Station reaches the highest value of 1 during the evening rush hours, as it is essentially the traffic hub of four subway lines.

D. SIRank Performance Analysis

Ranking station importance is crucial when trying to comprehend network topology and the integration of a new subway line. In this paper, we propose a novel ranking method SIRank by introducing human mobility patterns and network properties to evaluate the influence of subway stations in the network. To verify the effectiveness of SIRank, we conduct a series of experiments to analyze the performance of SIRank as compared to WMIRank [17] and NDRank [22]. WMIRank utilizes a weighted compound method to evaluate SI by using two metrics (node degree and node betweenness). NDRank leverages node degree to rank SI. However, SIRank considers static network topological characteristics as well as dynamic human mobility patterns.

We propose an evaluation indicator (hit ratio) to measure the performance of SIRank. The definition of hit ratio is shown in Equation (14).

Definition 4 (Hit Ratio): This metric is defined as the ratio of the number of selected existing stations as transfer stations for new subway lines to the total number of stations considered.

$$Hit\ Ratio = \frac{N_{tr}}{N_{se}}, \tag{14}$$

where N_{tr} is the number of subway stations which are selected as transfer stations linking new subway lines. N_{se} is the total number of stations considered in our experiment.

To the best of our knowledge, there is no existing standard for ranking the importance of subway stations. However, the SI for each station is obviously different and is quite helpful to manage and analyze the subway networks. The decision of how to construct a new line must take into account many factors including geographical structure, construction cost, and operation efficiency. In this paper, we assume that the improvement of operation efficiency of subways is the primary element for the selection of transfer stations. The selected transfer stations should intuitively acquire a higher SI ranking. Therefore, we identify the ranking of transfer stations as a critical evaluation factor. According to Shanghai subway transit development plan,[1] 12 subway lines and 26 transfer stations will be completed by 2020. Traffic management departments have announced the locations of the transfer stations. Based on the information, we calculate the hit ratio of the top n stations ranked by SIRank. The higher the hit ratio, the better the performance of the ranking algorithm.

As shown in Fig. 10, we acquire the experimental results of SIRank, WMIRank, and NDRank. Fig. 10(a) mainly reflects the degree of difference in rank value. X-axis represents the ranking of 288 stations. Y-axis indicates the corresponding rank value for each station. We discover that SIRank and NDRank have a higher discrimination score than WMIRank. The value of SIRank ranges from 0.02903 to 0.00061, and the value of NDRank is between 0.02715 and 0.00059. However, the value of WMIRank is only between 0.00713 and 0.00112. Fig. 10(b) shows the number of hit stations with different top n such as 5, 10, ..., 100. As seen from this figure, the performance of SIRank outperforms other two methods regarding hit ratio. For example, if n is equal to 10, the number of transfer stations for three algorithms are 5, 1, and 3 respectively. SIRank has the most selected transfer stations in the top 10 stations.

We list the SI ranking of the top 10 stations in Table II. Hit stations are underlined. Furthermore, we also calculate the hit ratio in different top numbers as shown in Fig. 11(a). Through statistical analysis, we find that the hit ratio of the top 10 stations in SIRank is about 60% as opposed to 10% in WMIRank and 30% in NDRank. Based on the improved PageRank and bi-directional passenger flow, we evaluate the SI

[1] http://www.envir.gov.cn/docs/2016/20160418709.htm

TABLE II
THE TOP 10 RANKING OF STATION IMPORTANCE

Ranking	SIRank	WMIRank	NDRank
1	People Square	Century Avenue	People Square
2	Shanghai Railway Station	*Longyang Road*	Shanghai Railway Station
3	*Hongqiao Railway Station*	Oriental Sports Center	Xujia Hui
4	*Longyang Road*	People Square	*Hongqiao Railway Station*
5	*South Shanxi Road*	Jufeng Road	East Nanjing Road
6	Xujia Hui	Xujia Hui	*Jingan Temple*
7	*Jingan Temple*	Siping Road	*South Shanxi Road*
8	Zhongshan Park	West Gaoke Road	Zhongshan Park
9	*Lujia Zui*	Yaohua Road	Yishan Road
10	*Shanghai South Station*	Dalian Road	Shanghai Indoor Stadium

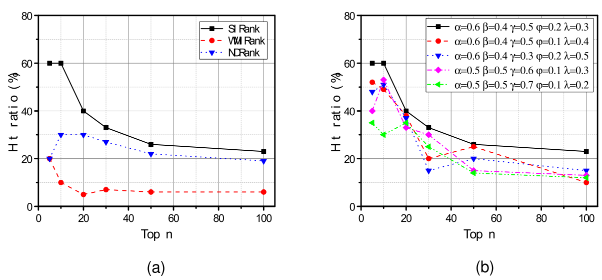

Fig. 11. Analysis of experimental results. (a) Performance analysis of algorithms. (b) Sensitivity analysis of parameters.

metric in terms of hit ratio, and the results of top 100 stations show that SIRank (at 23%) outperforms WMIRank (at 6%) and NDRank (at 19%). In general, SIRank has a better performance when compared with other two approaches.

We choose some representative weight combinations and analyze the sensitivity of the five parameters on the SI ranking as shown in Fig. 11(b). The different weights of five parameters cause the SI ranking to fluctuate between 10% and 60%. Extensive experiments prove that when $\alpha = 0.6, \beta = 0.4, \gamma = 0.5, \varphi = 0.2$ and $\lambda = 0.3$, SI ranking method achieves the best performance. Specifically, human mobility patterns make greater contributions than other indicators. Among these three network metrics, degree centrality outperforms the other two on impacting the hit ratio. On the contrary, the clustering coefficient does not influence the SI ranking as closeness centrality.

SIRank utilizes macroscopic and microscopic metrics to evaluate SI of each station. The microscopic metrics primarily focus on static characteristics of network topology but cannot reflect the relationship between origins and destinations. The macroscopic metrics determine the strength of ties between nodes through analyzing the spatio-temporal distributions of passenger flow at different time intervals. In other words, we focus on the importance level of stations from two perspectives, travel behavior and network science, to acquire an in-depth multidimensional understanding. It is entirely different from traditional methods. In particular, this paper mainly focuses on real-world subway transaction data. SIRank can also be used to evaluate node importance of other public transportation networks.

VI. CONCLUSION AND FUTURE WORK

By incorporating network science and human mobility patterns, we propose a new ranking method SIRank to analyze the subway network in Shanghai. The network topology is presented and network node importance is evaluated. With respect to passenger flow patterns, we constructed two network models denoted by SSSN and SSPN respectively. SSSN mainly reflects the native connectivity structure of the subway network, while SSPN mainly focuses on commuter's travel characteristics including time, location, and volume. Then, we comprehensively analyzed the network properties of SSSN and SSPN by utilizing various evaluation metrics.

We discover that the values of in-degree and out-degree in SSPN are much higher and do not fit the power and normal distribution, respectively. For clustering coefficients, the value of SSPN follows small-world characteristics of subway networks. Moreover, we notice that the value of closeness centrality in SSSN is also far less than that in SSPN. We conduct extensive experiments to verify the performance of SIRank using real transaction datasets. We implement a hit ratio metric to measure the effectiveness of SIRank by varying the number of stations. The results show that SIRank outperforms other two algorithms in terms of hit ratio.

Although we only apply our method to subway networks, other transportation systems such as buses, railway, and taxi networks can also benefit practically from our research results. By comparison, subway networks have fewer restrictions on variables such as congestions and traffic lights.

In the future, we plan to analyze other transportation networks by using our proposed SIRank method. Furthermore, we will consider human mobility patterns within different networks along with exterior factors such as weather conditions, regional functions, and traffic congestions to understand transport systems comprehensively.

REFERENCES

[1] F. Pasqualetti, S. Zampieri, and F. Bullo, "Controllability metrics, limitations and algorithms for complex networks," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 1, pp. 40–52, Mar. 2014.

[2] A. Háznagy, I. Fi, A. London, and T. Nemeth, "Complex network analysis of public transportation networks: A comprehensive study," in *Proc. IEEE Int. Conf. Models Technol. Intell. Transp. Syst.*, Jun. 2015, pp. 371–378.

[3] J. Han, S.-B. Kwon, and C. Chun, "Indoor environment and passengers' comfort in subway stations in Seoul," *Building Environ.*, vol. 104, pp. 221–231, Aug. 2016.

[4] X. J. Kong, X. Song, F. Xia, H. Guo, J. Wang, and A. Tolba, "LoTAD: Long-term traffic anomaly detection based on crowdsourced bus trajectory data," *World Wide Web*, vol. 21, no. 3, pp. 825–847, 2017. doi: 10.1007/s11280-017-0487-4.

[5] F. Xia, A. Rahim, X. Kong, M. Wang, Y. Cai, and J. Wang, "Modeling and analysis of large-scale urban mobility for green transportation," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1469–1481, Apr. 2018.

[6] J. Wang, X. Kong, A. Rahim, F. Xia, A. Tolba, and Z. Al-Makhadmeh, "IS2Fun: Identification of subway station functions using massive urban data," *IEEE Access*, vol. 5, pp. 27103–27113, 2017.

[7] Z. Ning, F. Xia, N. Ullah, X. J. Kong, and X. P. Hu, "Vehicular social networks: Enabling smart mobility," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 16–55, May 2017.

[8] X. Kong, F. Xia, J. Wang, A. Rahim, and S. K. Das, "Time-location-relationship combined service recommendation based on taxi trajectory data," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1202–1212, Jun. 2017.

[9] X. J. Kong *et al.*, "Mobility dataset generation for vehicular social networks based on floating car data," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3874–3886, May 2018. doi: 10.1109/TVT.2017.2788441.

[10] X. Wang, "Comparative analysis on investment and financing models of urban rail transportation," in *Recent Developments in Data Science and Business Analytics*. Cham, Switzerland: Springer, 2018, pp. 217–224.

[11] W. Xu, J. Zhou, and G. Qiu, "China's high-speed rail network construction and planning over time: A network analysis," *J. Transp. Geography*, vol. 70, pp. 40–54, Jun. 2018.

[12] X. Yang, A. Chen, B. Ning, and T. Tang, "Measuring route diversity for urban rail transit networks: A case study of the Beijing metro network," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 2, pp. 259–268, Feb. 2017.

[13] L. Sun, Y. Huang, Y. Chen, and L. Yao, "Vulnerability assessment of urban rail transit based on multi-static weighted method in Beijing, China," *Transp. Res. Part A, Policy Pract.*, vol. 108, pp. 12–24, Feb. 2018.

[14] R. Jiang *et al.*, "Data-driven digital advertising with uncertain demand model in metro networks," *IEEE Trans. Big Data*, to be published. doi: 10.1109/TBDATA.2017.2725913.

[15] R. Louf, C. Roth, and M. Barthelemy, "Scaling in transportation networks," *PloS One*, vol. 9, no. 7, Jul. 2014, Art. no. e102007.

[16] S. Derrible and C. Kennedy, "The complexity and robustness of metro networks," *Phys. A, Stat. Mech. Appl.*, vol. 389, no. 17, pp. 3678–3691, Sep. 2010.

[17] Y. Yang, Y. Liu, M. Zhou, F. Li, and C. Sun, "Robustness assessment of urban rail transit based on complex network theory: A case study of the Beijing subway," *Saf. Sci.*, vol. 79, pp. 149–162, Nov. 2015.

[18] B. Leng, X. Zhao, and Z. Xiong, "Evaluating the evolution of subway networks: Evidence from Beijing subway network," *Europhys. Lett.*, vol. 105, no. 5, Mar. 2014, Art. no. 58004.

[19] I. Vragović, E. Louis, and A. Díaz-Guilera, "Efficiency of informational transfer in regular and complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 71, no. 3, Mar. 2005, Art. no. 036122.

[20] X. Kong, Z. Xu, G. Shen, J. Wang, Q. Yang, and B. Zhang, "Urban traffic congestion estimation and prediction based on floating car trajectory data," *Future Generat. Comput. Syst.*, vol. 61, pp. 97–107, Aug. 2016.

[21] Y. Wei, S. Lin, R. Chu, Q. Tian, and W. Fei, "A method of grading subway stations," *Procedia Eng.*, vol. 137, pp. 806–810, Jul. 2016.

[22] H. Wang, J. Huang, X. Xu, and Y. Xiao, "Damage attack on complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 408, pp. 134–148, Aug. 2014.

[23] F. Xia, J. Wang, X. Kong, Z. Wang, J. Li, and C. Liu, "Exploring human mobility patterns in urban scenarios: A trajectory data perspective," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 142–149, Mar. 2018.

[24] Q. Xu, B. H. Mao, and Y. Bai, "Network structure of subway passenger flows," *J. Stat. Mech.: Theory Exp.*, vol. 16, no. 3, Mar. 2016, Art. no. 033404.

[25] J. K. K. Yuen, E. W. M. Lee, S. M. Lo, and R. K. K. Yuen, "An intelligence-based optimization model of passenger flow in a transportation station," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1290–1300, Sep. 2013.

[26] L. Chen, J. Jakubowicz, D. Yang, D. Zhang, and G. Pan, "Fine-grained urban event detection and characterization based on tensor cofactorization," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 3, pp. 380–391, Jun. 2017.

[27] S. P. Borgatti, "Centrality and network flow," *Soc. Netw.*, vol. 27, no. 1, pp. 55–71, 2005.

[28] O. Oullier, A. P. Kirman, and J. A. S. Kelso, "The coordination dynamics of economic decision making: A multilevel approach to social neuroeconomics," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 6, pp. 557–571, Dec. 2008.

[29] H. Soh *et al.*, "Weighted complex network analysis of travel routes on the Singapore public transportation system," *Phys. A, Stat. Mech. its Appl.*, vol. 389, no. 24, pp. 5852–5863, Dec. 2010.

[30] L. Li, H. Zhang, X. Wang, W. Lu, and Z. Mu, "Urban transit coordination using an artificial transportation system," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 374–383, Jun. 2011.

[31] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.

[32] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[33] A.-L. Barabási and E. Bonabeau, "Scale-free networks," *Sci. Amer.*, vol. 288, no. 5, pp. 60–69, May 2003.

[34] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 712–725, Mar. 2015.

[35] Q. Lv, Y. Qiao, N. Ansari, J. Liu, and J. Yang, "Big data driven hidden Markov model based individual mobility prediction at points of interest," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5204–5216, Jun. 2017.

[36] M. Lenormand *et al.*, "Influence of sociodemographic characteristics on human mobility," *Sci. Rep.*, vol. 5, May 2015, Art. no. 10075.

[37] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, "On the Levy-walk nature of human mobility," *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 630–643, Jun. 2011.

[38] C. Song, Z. Qu, N. Blumm, and A. L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[39] F. Simini, M. C. González, A. Maritan, and A. L. Barabási, "A universal model for mobility and migration patterns," *Nature*, vol. 484, no. 7392, pp. 96–100, Apr. 2012.

[40] K. Zhao, M. Musolesi, P. Hui, W. Rao, and S. Tarkoma, "Explaining the power-law distribution of human mobility through transportation modality decomposition," *Sci. Rep.*, vol. 5, no. 9136, pp. 1–7, Mar. 2015.

[41] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira, and C. Ratti, "Understanding individual mobility patterns from urban sensing data: A mobile phone trace example," *Transp. Res. C, Emerg. Technol.*, vol. 26, pp. 301–313, Jan. 2013.

[42] W. Wang, L. Pan, N. Yuan, S. Zhang, and D. Liu, "A comparative analysis of intra-city human mobility by taxi," *Phys. A, Stat. Mech. Appl.*, vol. 420, pp. 134–147, Feb. 2015.

[43] M. Veloso, S. Phithakkitnukoon, C. Bento, N. Fonseca, and P. Olivier, "Exploratory study of urban flow using taxi traces," in *Proc. 1st Workshop Pervasive Urban Appl.*, Jun. 2011, pp. 1–8.

[44] B. C. Csáji, A. Browet, V. A. Traag, J. C. Delvenne, E. Huens, P. Van Dooren, Z. Smoreda, and V. D. Blondel, "Exploring the mobility of mobile phone users," *Phys. A, Stat. Mech. Appl.*, vol. 392, no. 6, pp. 1459–1473, Mar. 2013.

[45] S. Hasan, C. M. Schneider, S. V. Ukkusuri, and M. C. González, "Spatiotemporal patterns of urban human mobility," *J. Stat. Phys.*, vol. 151, nos. 1–2, pp. 304–318, Apr. 2013.

[46] K. Lee, S. Goh, J. S. Park, W.-S. Jung, and M. Y. Choi, "Master equation approach to the intra-urban passenger flow and application to the metropolitan Seoul subway system," *Phys. A, Math. Theor.*, vol. 44, no. 11, 2011, Art. no. 115007.

[47] C. Roth, S. M. Kang, M. Batty, and M. Barthélemy, "Structure of urban movements: Polycentric activity and entangled hierarchical flows," *PloS One*, vol. 6, no. 1, 2011, Art. no. e15923.

[48] J. Lin and Y. Ban, "Complex network topology of transportation systems," *Transp. Rev.*, vol. 33, no. 6, pp. 658–685, Oct. 2013.

[49] M. Barthélemy, "Spatial networks," *Phys. Rep.*, vol. 499, nos. 1–3, pp. 1–101, Feb. 2011.

[50] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the Web," *Social Netw.*, vol. 9, no. 1, Nov. 1999.

[51] J. Lei and H.-F. Chen, "Distributed randomized PageRank algorithm based on stochastic approximation," *IEEE Trans. Autom. Control*, vol. 60, no. 6, pp. 1641–1646, Jun. 2015.

[52] H. Purwins *et al.*, "Regression methods for virtual metrology of layer thickness in chemical vapor deposition," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 1, pp. 1–8, Feb. 2014.

Feng Xia (M'07–SM'12) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China. He was a Research Fellow with the Queensland University of Technology, Australia. He is currently a Full Professor with the School of Software, Dalian University of Technology, China. He is also the (Guest) Editor of several international journals. He serves as the General Chair, the PC Chair, the Workshop Chair, or the Publicity Chair for number of conferences. He has published two books and over 200 scientific papers in international journals and conferences. His research interests include data science, big data, knowledge management, network science, and systems engineering. He is a Senior Member of ACM.

Jinzhong Wang received the B.S. degree in computer education from Anshan Normal University, Anshan, China, in 2002, and the M.S. degree in computer application technology from Liaoning University, Shenyang, China, in 2005. He is currently pursuing the Ph.D. degree with the School of Software, Dalian University of Technology, Dalian, China. He is currently an Associate Professor with Shenyang Sport University, Shenyang. His research interests include traffic data analytics and mobile social networks.

Xiangjie Kong (M'13–SM'17) received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2009. He is currently an Associate Professor with the School of Software, Dalian University of Technology, China. He has published over 100 scientific papers in international journals and conferences (with over 70 indexed by ISI SCIE). His research interests include big traffic data, social computing, and cyber-physical systems. He is a Member of ACM.

Da Zhang received the B.E. degree (Hons.) in software engineering from the Dalian University of Technology, in 2010, and the M.S. degree in computer science and engineering from Ohio State University, USA, in 2012. She is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Miami (UM), Coral Gables, Florida. She is currently a Teaching Assistant with UM. Her research interests include knowledge graph management, big data, Web semantics, and deep learning.

Zhibo Wang (M'10–SM'18) received the B.E. degree in automation from Zhejiang University, China, in 2007, and the Ph.D. degree in electrical engineering and computer science from the University of Tennessee, Knoxville, in 2014. He is currently an Associate Professor with the School of Cyber Science and Engineering, Wuhan University, China. His current research interests include mobile crowdsensing systems, the Internet of Things, network security, and privacy protection. He is a Member of ACM.