Social Network Analysis and Connection Strength Evaluation of Urban Tourist Attractions Using Car-Hailing Data: A Case Study of Beijing

Shixia Ma, Xuedong Yan, Xiaobing Liu and Deqi Chen

Abstract For large cities with huge tourism market, tourism travels can have significant impact on the urban traffic, which, cannot be ignored in both urban transportation and tourism planning. This paper first proposes a simple and effective spatial matching method to identify tourist travel patterns based on massive online car-hailing data. Then we construct the tourist attractions network based on the tourist movement by car. From the perspective of social network analysis, the development status of the holistic tourist attractions network and the influence of the attractions are evaluated. Finally, connection strength based Jenks natural breaks classification method is employed to divide the attractions into four levels: unconnected, weak, moderate and strong connection. Taking Beijing as case study, the main factors that affect the connection strength among attractions are the popularity of the attraction and the spatial proximity of the attraction. These findings in tourist movement can facilitate authorities and planners to develop tourism destinations and manage tourism traffic better.

Keywords Tourist attractions · Social network analysis · Connection strength · Tourist trip patterns · Car-hailing

1 Introduction

With the increase of people tourist traveling, massive tourist data can play a more important role for both urban transport planning and tourism spots management. Cities with huge tourism demand like Beijing, tourist travel patterns and travel destinations are important components of urban travel [1], but little attention is paid to this aspect. There is a difference between tourism transportation demand and urban commuting demand. When tourism transportation demand is neglected, it will lead

S. Ma · X. Yan (✉) · X. Liu · D. Chen
MOT Key Laboratory of Transport Industry of Big Data Application
Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044,
People's Republic of China
e-mail: xdyan@bjtu.edu.cn

to the dislocation of tourism transportation demand and urban transportation supply space, and the utilization of tourism resources is not balanced. Traditional travel surveys focus on urban residents. Tourist surveys are also concentrated on the destination choice [2], as well as tourist behavior analysis in particular tourist attraction. Less research focuses on the tourist movement patterns among attractions.

With the rapid development of information technology, massive data is recorded and stored in a structured or unstructured form, making tourism research enter the era of big data. Far more existing research study tourist behavior based on diverse data, including spatial behavior [3], temporal behavior [4] and spatial-temporal behavior [5]. Travel recommendation is another hot topic. Yoon et al. [6] presented an effective recommendation based on GPS trajectories. Zheng et al. [7] predicted the next destination of individual tourists using the GPS tracking data. Tourism destination has five dimensions-spatial, temporal, compositional, social and dynamic [8]. In the destination management, tourist movements and flows act as one of the significant issues to form attraction networks in a destination [9]. Raun et al. [8] applied roaming data to analyze tourist flows, and proposed a model to forecast the number of hotel nonresident registrations. As the potential importance of understanding attraction networks in a destination, some studied attempted to look into tourist attractions network as informed by tourist movement [10] identified the spatial structure of the tourist attraction system in Seoul, South Korea based on social network analysis techniques with spatial statistics. Similarly [11] applies the Quadratic Assignment Procedure of SNA to test the relationships between region proximity, grade proximity, and tenure proximity, and the attraction network Juan and Hernández [12] reveals attractions cluster by segmenting this network using network analysis tools. In addition, limited by the data, on the mining of connection in tourist attraction visits, existing research mainly considers association rules [13]. Compared with the existing literature, this paper evaluates the connection strength among attraction and analyzes the factors affecting the relationship among attractions from the perspective of tourist movement.

The aim of this study is to explore tourist travel behavior from massive car-hailing data. Firstly, a spatial matching method is presented to identify and collect tourist trip, which shows the possibility of using the floating car data to explore the behavior of tourists. Then, the attractions network is constructed by tourist movement. From the perspective of the overall network and network nodes of social network analysis, we evaluate the development of the attraction network and the status of the attractions in the network. Finally, a simple method is proposed to evaluate the strength of the connection among the attractions.

2 Study Area and Data Preprocess

2.1 Beijing Tourist Attractions

Beijing is the capital of China, the national political center, cultural center, international exchange center, and science and technology innovation center. It is one of the first batch of national historical and cultural cities with rich tourism resources, attracting a large number of domestic and overseas tourists. According to the latest statistics of the Beijing Tourism Public Information and Consultation Platform, there are 242 A-level tourist attractions in Beijing. In this paper, taking into account the activity of the attraction, select three levels of attractions 5A, 4A, part of the 3A, a total of 98 attractions for research. The overview of the study area and attractions is shown in Fig. 1. This paper studies the behavior of tourists visit attractions during the June 2017 and National Day holiday. This period includes working days, weekends and holidays, and the source of tourists is more diverse, including local residents and foreign tourists, more comprehensively reflecting the spatial distribution structure of attractions.
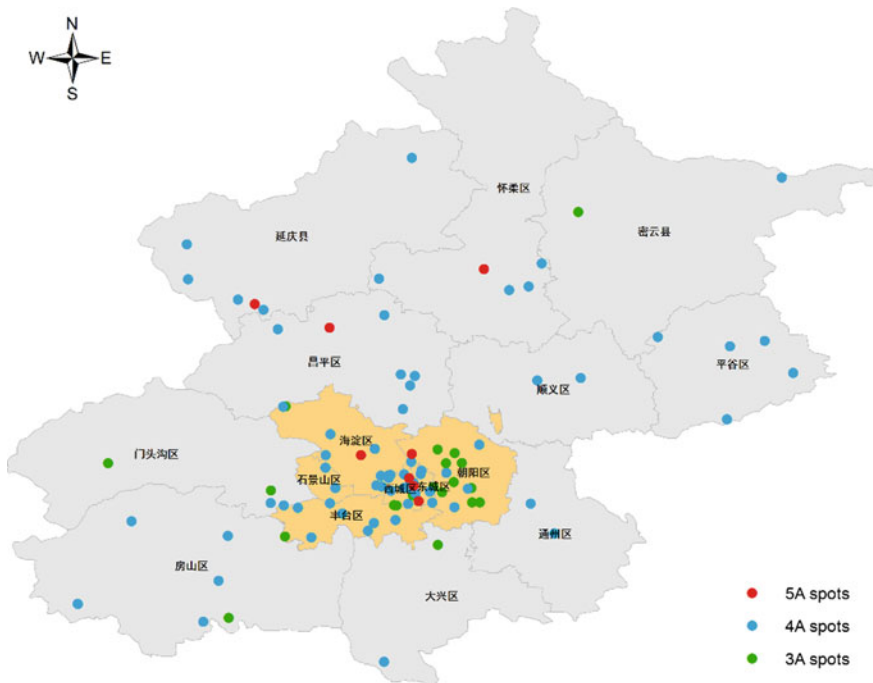

Fig. 1 The overview of the study area and attractions

2.2 Data Description

In this paper, two datasets are established. The first one is the tourist attractions infor-
mation in Beijing, which includes the coordinates of the attraction and its entrance.
The coordinates are obtained from the Baidu API open platform. The other is express
ride orders (from 2017/06/01 to 2017/06/30 and from 2017/10/01 to 2017/10/08) in
Beijing. Each express ride order includes eleven variables such as: the order ID,
passenger ID, departure time, arrival time, departure latitude and longitude, arrival
latitude and longitude, travel time, travel distance, and average travel speed. The
passenger ID has character of uniqueness, which would be applied to distinguish
different passengers and it has been data desensitization processing.

In the process of data collection and storage, the errors are caused by signal
blocking or equipment failure and so on. To ensure the quality of the research data,
the following express ride orders are viewed as outliers and removed from datasets:

(1) Orders with Departure time before arrival time;
(2) Orders with Longitude or latitude is 0;
(3) Orders with travel distance larger than 100 km or smaller than 0.01 km;
(4) Orders where with travel time longer than 2 h or shorter than 1 min;
(5) Orders with an average speed over 100 km/h or below 1 km/h;
(6) Orders with route circuity less than 1. (The route circuity means the ratio between
    the actual travel distance and the Euclidian distance, and the Euclidean distance
    is calculated by the coordinates of the origin and destination).

2.3 Tourist Trips Identification

In the navigation map, the tourist attraction is a point. When we set the attraction as
the destination, the navigation system may choose its entrance or the interior as the
destination. Therefore, destinations of the order to the attraction are concentrated in
the attraction or its entrance. Meanwhile, there is an intersect tool in the Arcgis10.2
overlay toolset. The intersect tool calculates the geometric intersection of any num-
ber of feature classes and feature layers, extracting the parts of the two layers that
intersect the spatial relationships. The inputs can be any combination of geometry
types (point, multipoint, line, or polygon). The output geometry type can only be of
the same geometry or a geometry of lower dimension as the input feature class with
the lowest dimension geometry (point = 0 dimension, line = 1 dimension, and poly
= 2 dimension). As shown in Fig. 2, the inputs are a shadow polygon and red points,
and the outputs is points within the shadow polygon.

According to the above analysis, a spatial matching method to identify tourist
trips is proposed, the specific steps are as follows:

Step 1: Generate a minimum geometry polygon from the attraction and the
entrance to represent the attraction area. As shown in Fig. 2, The Tiantan Park is
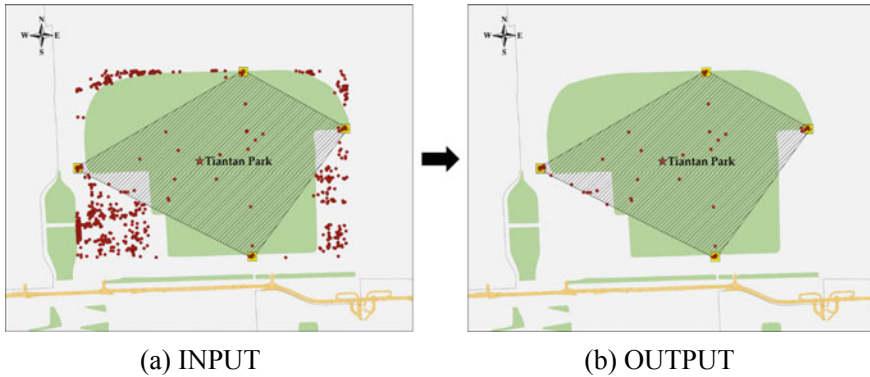
(a) INPUT   (b) OUTPUT

Fig. 2 Diagram of intersect

taken as an example. The five-pointed star is the location of attraction. The yellow point is the four main entrances of the Tiantan Park. The shadow is the minimum geometry polygon that represents the Tiantan Park. Although the polygon and the attraction area are not completely identical, it can be seen that the orders for the Tiantan Park as the origin or destination is in this polygon.

Step 2: The order destinations constitute a point layer, and intersect with the attraction polygon layer to get tourist orders. The output orders are labeled with attraction ID. Attraction ID represents which attraction belongs to.

Step 3: In the short-term, tourists will not visit the same attraction multiple times, and the trip may be generated by the tourist attraction staff or nearby residents. It is not a tourist trip, so it is excluded from the dataset.

The trip extraction step for leaving the attraction is the same except that the point feature in step 2 is the origin. Finally, we get two trip datasets, one is the trip to the attraction, and the other is the trip from the attraction.

3 Method

3.1 Social Network Analysis

A tourist attraction system is defined as an empirical connection of tourist, nucleus, and marker [14]. The movement of tourists among different attractions represents the connectivity among attractions [12]. In this study, if a passenger has both an attraction A order and an attraction B order, or an order from attraction A to attraction B, it can believe that the passenger visited both attraction A and attraction B. Accordingly, a tourist attraction connection matrix was constructed by the amount of tourists among attractions. As the visit sequence is not considered in the study, our matrix is symmetric.

Social Network Analysis (SNA) is the process of investigating social structures through the use of networks and graph theory [15]. If the tourist attraction is regarded as a node, the connection among the attractions is the tie, SNA can be used to analyze the characteristics of the tourist attraction network, reveal the role and status of the attractions in the network, and compare the influence of different tourist attractions. This process can be carried out by UCINET6.0 and visualized in Netdraw. Since in the social network analysis, the matrix is required to be a binary matrix, we transformed the connection matrix to a binary matrix using the mean as the cut-off value. The evaluation is carried out from two perspectives: the overall network and the network node.

3.1.1   Overall Network

Network density refers to the proportion of direct ties in a network relative to the total number possible [16], which measures the overall tightness of the network. The higher the density in a network, the more connections, the closer it is to 1, indicating that the more frequent the flow of tourists among the various attractions. The network centralizations characterize the difference in various nodes in the network and describe the consistency of the overall network. Examples of network centralizations commonly include degree centralization and betweeness centralization. The degree centralization is gathering trend of the nodes in the network. If the degree centralization is larger, the greater the possibility that the flow of tourists will gather or spread around the attraction with high degree. The betweeness centralization is the quantification of the betweeness centrality of the node. The higher the value, the more nodes in the network may be divided into multiple small groups and rely on one node to transfer the relationship. In our network, it means that there may be clustering in the tourist attraction system. The purpose of Core-Periphery structural analysis is to study which nodes in the social network are at the core and which are at the edge. Through the Core-Periphery, we can judge the hierarchical structure of the attraction and analyze the core attractions that mainly attract tourists.

3.1.2   Network Node Analysis

Centrality is one of the focuses of SNA, referring to a group of metrics that aim to quantify the "importance" or "influence" of a particular node within a network [17]. According to different calculation methods, there are three common types of centrality: degree centrality, betweeness centrality, and closeness centrality. The degree centrality is the number of immediate connections an attraction has in a network, which can be used to measure an attraction's level of involvement, prestige, or dependence in the network [10]. The betweeness centrality refers to the times of a node acts as the shortest bridge between the other two nodes. The betweeness centrality measures the ability of a node as an intermediary. An attraction with high betweeness centrality have strong control over other attractions. The closeness centrality

is meaningful when all nodes in the network are connected. In this study, there are isolated nodes in the network. Thus, closeness centrality was not computed. The degree centrality can be formulated as follows:

Degree centrality, for attraction A_i:

$$C_D(A_i) = \sum_{j=1}^{n} x_{ij}(i \neq j) \tag{1}$$

where, x_{ij} is the value of the tie from attraction i to attraction j (the value being either 0 or 1). n is the number of attractions in the network.

3.2 Connection Strength Evaluation

In social network analysis, the connection among attractions are binary. The value is 1 when two attractions have common tourists and there is a tie between them, otherwise, the value is 0. We know which attractions have great influence, but we don't know the close connection pairs, and we don't know the strength of the connection among the attractions. Therefore, to further reveal the connection strength among attractions, the Jenks natural breaks classification method (JNBC) is employed. It is one of the most common methods for data classification. JNBC seeks to reduce the variance within classes and maximize the variance between classes [18]. Using the Jenkspy to calculate the natural breaks of the connection value among the attractions, according to which the connection between the two attractions can be divided into four levels: unconnected, weak, moderate and strong connection, labeled by 0, 1, 2 And 3 respectively. Thus, the connection strength can be formulated as follows:

We define a variable $S(C_{ij})$ to represent connection strength index between attraction A_i and attraction A_j, as shown in Eq. (2), Where b_0, b_1, b_2, b_3 are natural breaks of connection value, sorted in ascending. c_{ij} is connection value.

$$S(C_{ij}) = \begin{cases} 0, c_{ij} = 0 \\ 1, c_{ij} \in [b_0, b_1) \\ 2, c_{ij} \in [b_1, b_2) \\ 3, c_{ij} \in [b_2, b_3] \end{cases} \tag{2}$$

As illustrated in Fig. 3, for example, a simple case is shown in four tourist attractions (labeled A1, A2, A3 and A4). In the connection matrix (see Fig. 3a), the rows and columns index are attractions and the cell value refer to the number of tourists among attractions. Then, we transform the connection matrix into a connection strength matrix (see Fig. 3b) by Jenks. After obtaining the connection strength matrix, the Arcgis10.2 is applied for visual display to reflect the connection strength distribution among the attractions.

	A1	A2	A3	A4
A1	-	0	25	5
A2	0	-	172	53
A3	25	172	-	33
A4	5	53	33	-

	A1	A2	A3	A4
A1	-	0	2	1
A2	0	-	3	2
A3	2	3	-	2
A4	1	2	2	-

(a) Connection matrix   (b) Connection strength matrix

Fig. 3   Data matrix

4   Results

4.1   Tourist Attractions Network Profiles

Applying the proposed method of tourist trip identification, orders related to 63 attractions was extracted, accounting for 64.3% of the given tourist attractions. Some of the remaining attractions are missing information, while others are far from the city center where there are almost no tourists taking taxis. The travel data of attractions include seven 5A attractions, forty-two 4A attractions, thirteen 3A attractions and Lama Temple which is not an A-level attraction, but it is very active. These travel datasets cover almost all the popular attractions in Beijing, which are representative and can be used to study the tourist movement characteristics among tourist attractions.

The density of Beijing attractions network is 21.44%, indicating that the tourist movement network is relatively loose, and the connection among the attractions are not tight enough. The degree centralization index is 25.15%, which is relatively high. The network has a concentrated trend, and the central attractions have obvious influence. The betweeness centralization is 3.07%, which is relatively small, indicating that the interdependence among the attractions is relatively low and most of attractions are directly connected. Through the analysis of the overall network, it is noticed that the overall connection among Beijing tourist attractions is weak, the influence of the attractions varies greatly, and tourists may focus on several important attractions.

After calculating the core degree of 98 attractions, Core-Periphery module divides them into core and edge levels, of which 22 attractions are core nodes: Summer Palace, Tiantan Park, Forbidden city, Chaoyang Park, Ditan Park, Jingshan Park, Beijing Zoo, etc., and other attractions are the edges. The density of the core attraction is 66.2%, which is much higher than the overall network density by 21.44%, indicating that the core attractions are closely connected. The network density of the edge attractions is only 0.4%, which is far lower than the overall network density. The relationship among the edge attractions is extremely weak, the connections among

the attractions are not tight and even some attractions are not connected to each other. The polarization of network density in the core and edge attractions reveals that the network structure of Beijing tourist attractions is extremely uneven. The density among the core attractions and the edge attraction is 5.1%, showing that the core attraction and the edge attraction are not closely connected, and the core attractions hardly promote the development of the edge attractions. Therefore, while maintaining the good development of the core attractions, on the one hand, it is necessary to enhance the competitiveness of the edge attractions, on the other hand, it is vital to strengthen the cooperation between the core attraction and the edge attraction. The core attractions drive the development of the edge attractions, thus promoting the healthy development of the city tourism system.

4.2 Tourist Attractions Profile

The centrality of tourist attractions were calculated using UCINET6.0 and top 10 are shown in Table 1. Employing the Netdraw, we provide a visual network diagram in Fig. 4 based on the results of degree centrality and attraction grade. In this figure, the isolated node is removed. The average value of degree centrality is 5.102, which implied that each attraction is might be directly connected to five attractions. The top 10 attractions in degree centrality are at central in the attraction system, with strong agglomeration and radiation function, and are concentrated destinations for tourists. As shown in Table 1 and Fig. 4, attractions with high degree centrality are a popular attractions that we are familiar with, and high-level attractions may have high degree centrality. In addition to the Forbidden City, the Summer Palace and the Tiantan Park, the Jingshan Park and the Beihai Park are also important nodes because they are close to the Forbidden City, attracting a large number of tourists. An interesting

Table 1 Centrality of tourist attractions

Rank	Attractions	Degree	Rank	Attractions	Betweeness
1	Tiantan Park	29	1	Tiantan Park	151.10
2	Beijing Zoo	28	2	Chaoyang Park	142.56
3	Chaoyang Park	27	3	Beijing Zoo	65.72
4	Beihai Park	26	4	Xinglong Park	65.15
5	Summer Palace	25	5	Summer Palace	57.34
6	Forbidden City	25	6	Jingshan Park	55.17
7	Jingshan Park	21	7	Beihai Park	50.24
8	Lama Temple	21	8	Forbidden City	38.55
9	Taoranting Park	17	9	Ritan Park	31.98
10	Ritan Park	17	10	Lama Temple	24.99

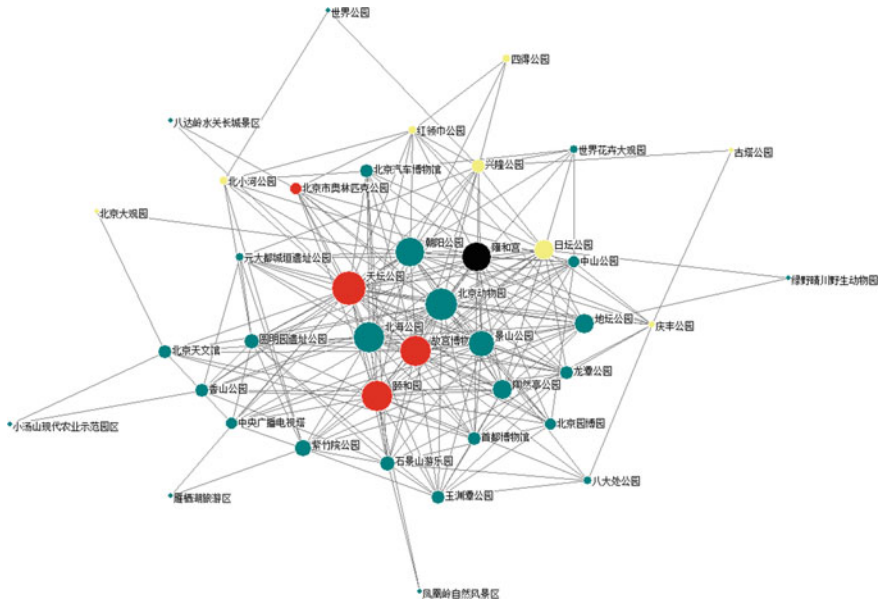*Note* degree: degree centrality; betweeness: betweeness centrality

Fig. 4 Tourist network of Beijing

finding is that the Lama Temple is a non-A-level attraction, but an important node. The Lama Temple is the highest-ranking Buddhist temple in the middle and late Qing Dynasty, where attracts lots of people to pray.

The betweeness centrality refers to the ability of the attraction to act as an intermediary in the tourist movement network, which can reflect the control ability of the node to tourists flow in the network to some extent. The Tiantan Park and the Chaoyang Park are not only at the central position, but also control the connection among other attractions and play an important role in the attractions system. However, Xinglong Park is a node with low centrality but high centrality, indicating that it controls the connection with some relative edge nodes, plays an important role in the local, and improves the overall connection of the network.

It should be noted that the size of nodes depends on the value of degree centrality. Nodes in red refer to 5A attractions; nodes in green refer to 4A attractions; nodes in yellow refer to 3A attractions; node in black refers to The Lama Temple.

4.3 Connection Strength Distribution

The connection strength of the attractions are shown in Fig. 5. There are 5 pairs of strong connections, 11 pairs of moderate connections, and 196 pairs of weak connections. From Table 2, it shows 16 attraction pairs with strong or moderate connection. The Forbidden City, the Summer Palace and the Tiantan Park is strongly connected
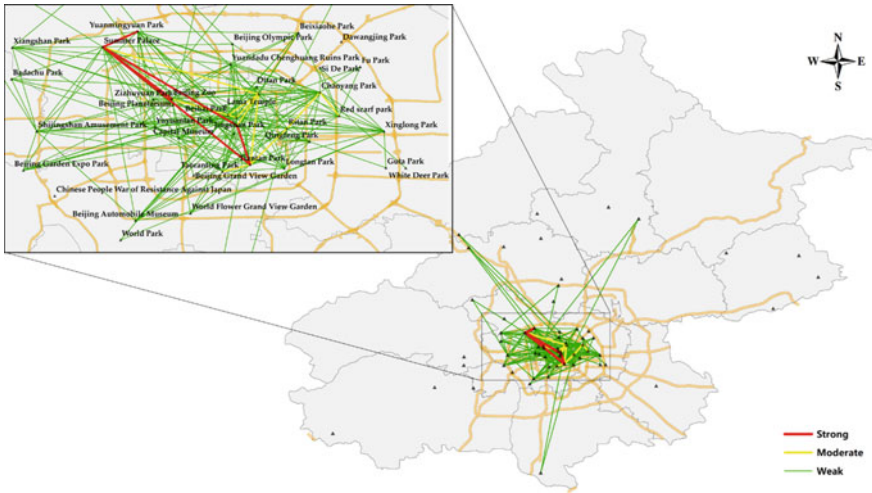
Fig. 5 Distribution of connection strength

Table 2 Connection strength of tourist attractions

Connection strength	Attraction pairs	Connection strength	Attraction pairs
Strong	Summer Palace—Forbidden City	Moderate	Lama Temple—Forbidden City
	Summer Palace—Tiantan Park		Lama Temple—Summer Palace
	Summer Palace—Yuanmingyuan Park		Lama Temple—Beihai
	Summer Palace—Beijing Zoo		Lama Temple—Jingshan
	Forbidden City—Tiantan Park		Beihai Park—Jingshan Park
Moderate	Tiantan Park—Taoranting		Beihai Park—Beijing Zoo
	Tiantan Park—Ritan		Chaoyang Park—Ritan Park
	Tiantan Park—Lama Temple		Chaoyang Park—Red Scarf Park

to each other. The Lama Temple is strongly connected with the Forbidden City, the Summer Palace, and the Tiantan Park. In addition to the strong connection among the popular attractions, the connection among attractions with close geographical location is also strong, forming a tour route of the group. For example, the Summer

Palace and Yuanmingyuan Park; Tiantan Park, Taoranting and Ritan Park; Beihai, Jingshan Park and Beijing Zoo; Chaoyang Park, Ritan and Red Scarf Park.

5   Discussion and Conclusion

The contribution of this study can be summarized in three aspects. First, although taxi trip data is widely used for resident travel research, especially commuting, there is little attention on other travel purposes, such as leisure. This study proposes a method for spatial matching using Arcgis10.2 to identify tourists travel from taxi travel data, making up for the lack of data in traditional tourism research. The travel data obtained from the GPS of the vehicle has a long period and a large spatial scale, which can provide data support for more travel research. The method proposed in this study provides new insight for taxi travel data mining, especially for tourism travel.

Second, the tourist attractions network is very important in the destination management, and the increasing research begins to focus on demand-driven network relationships among attractions in a destination [11]. This study builds a tourist attractions network based on tourists movement, and evaluates the network from the perspectives of overall network and network nodes. New knowledge was contributed to the field of tourism network. In the analysis of the overall structure of the network, four indicators of density, degree centralization, betweeness centralization and Core-Periphery are applied. The results show that the overall structure of the tourist movement network in Beijing is loose and unevenly developed, and there is great potential for developing tourist routes. The core attractions are closely connected to each other, but the edge attractions are not only weakly attractive to tourists, but also not closely connected to the core attractions. In the analysis of network nodes, the influence of the attractions are compared and analyzed with two indicators: the degree centrality and the betweeness centrality. There are many tourist attractions in Beijing, but tourists are only interested in some attractions. These attractions are traditional attractions with high reputation and brand awareness. They are at the center of the network. These attractions also have strong intermediary capabilities and control the flow of tourists in other attractions in the network.

Third, the analysis of the connection among the attractions can provide support for traffic management among the attractions to meet the tourist traffic demand during the holiday. Based on the amount of tourists among attractions, this study proposes a method to divide the connection strength among the attractions, identify the strong connection pairs of attractions, and try to analyze the factors affecting the strength of the relationship among the attractions. According to the results, it is found that an attraction with strong or moderate connection strength is almost a central attraction in the network. In addition, when the central attraction has an enough influence on the nearby attraction, there will be a strong connection between them, such as the Yuanmingyuan Park and Red Scarf Park, which are closely connected with the Summer Palace and the Chaoyang Park respectively. Based on this, we speculate on

two main factors that cause the strength of the connection among the attractions: one is the attraction of the tourist attractions to tourists and its influence in the network; the second is the proximity of geospatial location.

6 Suggestion and Future Work

Based on the analysis of the tourist attraction network and relationship among attractions, the problems of tourism existing in tourism management and traffic planning are explored, and some suggestions with practical application value can be provided for managers. Beijing has abundant tourism resources, but its development is uneven. It is necessary to strengthen the overall network connection and establish a good competition and cooperation system to enable the core attractions to drive the development of the edge attractions. There are many tourists in the core attraction, which leads to the pressure of reception in the attraction. However, there may be almost no tourists in the surrounding attractions. Thus, it is necessary to enhance the publicity and popularity of edge attractions. At the same time, form some regional tourist routes, evacuate the tourists gathered in the popular attractions, increase the tourists in the surrounding attractions, and enhance the tourist experience.

Traffic is the basic guarantee for the tourist movement among attractions. For traffic managers, on the one hand, it is imperative to strengthen the transportation infrastructure among the attractions with high tourist flow, to meet the demand for tourism travel, and to do the traffic diversion work during the holidays to ensure the smooth flow of the urban transportation system. On the other hand, it is useful to depend on traffic to guide the tourist flow, to provide transportation services for isolated nodes or edge attractions, and to enhance their centrality in the tourism network.

Although this study explored the attraction network and puts forward some insights, it is not deep enough and comprehensive. Future research should consider more factors affecting the attraction, such as basic service facilities, traffic accessibility, etc.

Conflicts of Interest The authors declare that there is no conflict of interest in any aspect of the data collection, analysis, or the funding received regarding the publication of this paper.

References

1. Hasnat MM, Hasan S (2018) Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. Transp Res Part C 96:38–54

2. Yu P, Xiao L, Yan-Yang MA et al (2016) The study on the factors influencing the citizen elderly' choice for tourism destination—based on a market survey of middle-aged and elderly people in Beijing Chaoyang. Sci Technol Ind

3. East D, Osborne P, Kemp S, Woodfine T (2017) Combining GPS & survey data improves understanding of visitor behaviour. Tour Manag 61:307–320

4. Edwards D, Griffin T (2013) Understanding tourists' spatial behaviour: GPS tracking as an aid to sustainable destination management. J Sustain Tourism 21(4):580–595

5. Shoval N, McKercher B, Birenboim A, Ng E (2015) The application of a sequence alignment method to the creation of typologies of tourist activity in time and space. Environ Plan 42(1):76–94

6. Yoon H, Zheng Y, Xie X, Woo W (2010) Smart itinerary recommendation based on user-generated GPS trajectories. In: Proceedings of the 7th international conference on ubiquitous intelligence and computing, Xi'an, China

7. Zheng W, Huang X, Li Y (2017) Understanding the tourist mobility using GPS: Where is the next place? Tour Manag 59:267–280

8. Raun J, Ahas R, Tiru M (2016) Measuring tourism destinations using mobile tracking data. Tour Manag 57:202–212

9. Lew A, McKercher B (2006) Modeling tourist movements: a local destination analysis. Ann Tourism Res 33(2):403–423

10. Kang S, Lee G, Kim J, Park D (2018) Identifying the spatial structure of the tourist attraction system in South Korea using GIS and network analysis: An application of anchor-point theory. J Destination Mark Manag 9:358–370

11. Liu B, Huang S, Fu H (2017) An application of network analysis on tourist attractions: the case of Xinjiang, China. Tourism Manag 58:132–141

12. Hernández JM, Kirilenko AP, Stepchenkova S (2018) Network approach to tourist segmentation via user generated content. Ann Tourism Res 73:35–47

13. Versichele M, de Groote L, Bouuaert MC, Neutens T, Moerman I, Van de Weghe N (2014) Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: a case study of Ghent. Belgium Tourism Manag 44:67–81

14. Leiper N (1990) Annals of tourism research. Tourist Attraction Syst 17(3):367–384

15. Otte E, Rousseau R (2002) Social network analysis: a powerful strategy, also for the information sciences. J Inf Sci 28(6):441–453

16. Xu G et al (2010) Web mining and social networking: techniques and applications. Springer, Berlin, p 25. ISBN 978-1-4419-7734-2

17. Liu B (2011) Web data mining: exploring hyperlinks, contents, and usage data. Springer, Berlin, p 271. ISBN 978-3-642-19459-7

18. Jenks GF (1967) The data model concept in statistical mapping. Int Yearbook Cartogr 7:186–190