Received November 17, 2020, accepted December 4, 2020, date of publication December 14, 2020,
date of current version December 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3044613

Tourism Flow Between Major Cities During China's National Day Holiday: A Social Network Analysis Using Weibo Check-in Data

ANJUN LI[1], NAIXIA MOU[1], LINGXIAN ZHANG[1], TENGFEI YANG[2], WENBAO LIU[1], AND FENG LIU[1]

[1]College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China
[2]Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

Corresponding authors: Naixia Mou (mounaixia@163.com) and Feng Liu (luf3286@126.com)

This work was supported by the National Natural Science Foundation of China under Grant 41771476.

ABSTRACT Holiday tourism flow is a significant indicator to evaluate the development of tourism. The exploration of the rule of tourism flow between cities can not only provide reasonable suggestions for stimulating demand, promoting consumption and economic development, but also make crucial significance for the management of tourism destinations and the optimization of spatial structure of tourism flow. Based on Weibo check-in data, this paper, by using social network analysis, studies the spatial distribution and network structure characteristics of tourism flow in 50 major cities in China during the National Day holiday in 2018. The results show that: 1) the tourism flow connection of the main cities in China represents a diamond shaped spatial structure with ''Beijing-Shanghai-Guangzhou-Chengdu'' as the core. There exists spatial heterogeneity in different levels of tourism flow intensity and proximity and selectivity in tourism links between cities; 2) the intensity of tourism connection between cities in China is clearly divided into different levels. On the basis of index of degree of centrality, Beijing and Shanghai are far higher than other cities;3) there are obvious differences between core nodes and edge nodes, with the core nodes often composed of cities with high economic development or rich tourism resources. Although the number is small, it plays a significant role in driving the edge cities; 4) the urban tourism flow network is relatively stable, but most cities have relatively weak tourism links and more small-scale tourism flows. In the division of cohesive subgroups, the fifth and sixth subgroups are not only the main tourist sources but also the main destinations. Whether it is the internal connection of subgroups or the connection with other subgroups, tourism flow has a very high density of connection.

INDEX TERMS Tourism flow, network structure, social network analysis, city, Weibo check-in.

I. INTRODUCTION

Tourism flow [1] refers to the phenomenon of collective migration of tourists in the spatial area. Generally speaking, tourism flow is divided into two kinds, namely, narrow sense and broad sense. In a narrow sense, tourism flow means the number of tourist groups between the source and destination; in a broad sense, tourism flow includes not only tourism flow, but also information flow, material flow between the source and destination as well as various capital flows undertaken by tourists and tourism operators. As the main body of tourism, in the process of tourism, tourists also carry a variety of flows, such as information flow and material flow. Therefore, the research on tourism flow has always been the core topic of tourism flow research [2]. As an important node of tourism flow, city, with the collection of many spatial tourism elements, is the source, transit and destination of tourism, as well as the spatial carrier of collecting or diffusion of tourism flow [3]. With the improvement of the tertiary industry, the inter-city transportation network and social consumption, people are more apt to go out and play between cities, gradually forming a closer relationship of inter-city tourism flows and a stable and complex inter-city tourism flow network. Therefore, inter-city tourism is more valuable

The associate editor coordinating the review of this manuscript and approving it for publication was Massimo Cafaro.

in the domestic market [4], and tourism flow has become one of the important indicators to evaluate the improvement of urban tourism. Tourism, a sensitive industry, is very vulnerable to external factors. Holidays, climate, major events obviously disturb the local tourism flow in a short period of time, especially the short-term high-intensity holidays with more obvious differences in time and space characteristics. As the longest holiday in China, the National Day golden week has always been the peak of domestic tourism. Therefore, the study of the spatial and temporal rule, flow, flow direction, spatial level and spatial structure characteristics of tourism flow during the National Day holiday is of guiding significance to the economic benefits, cultural integration and environmental governance of tourist destinations [5].

With the advent of the ''Internet plus'' era, all kinds of social software sprung up like mushrooms. Based on the report released by China Internet Information Center (CNNIC) in September 2018, the number of Weibo users in China is 337 million in 2018, accounting for 42.3% of the total number of Internet users, indicating that Weibo has become one of the most representative and influential social communication software in China [6]. Weibo, as a kind of social software that can release messages or express the feelings in real time, by virtue of its convenience, makes tourists more willing to share the beauty of the journey on Weibo, which has become an essential social tool in the process of travel. With its own geographical location attribute, Weibo check-in data is becoming an important carrier of tourists' behavior [7]. There is no doubt that the exploration of Weibo check-in data of visitors provides a new research method and perspective for the network structure characteristics of tourism flow [8].

With the development of Internet, big data including tourism flow has been applied to various fields for the research of the tourism flow [9]. The arrival of big data era has brought opportunities for ''space of flows'' and urban network research. The check-in data on the basis of location mobile social network provides a new data collection method for tourism flow researchers. Compared with the traditional questionnaire data collection method, the digital footprints obtained in social media address the shortcomings of insufficient sample data. This paper, by using Weibo check-in data, constructs the network spatial structure of Chinese holiday city tourism flow [10], and provides reasonable suggestions for the management of short-term inter-city tourism flow through the study of the interrelation of inter-city tourism flow [11].

II. LITERATURE REVIEW

As a special social activity group flow, tourism flow, connecting the destination and the source of tourists, is the nerve center and link of the tourism system. Therefore, the research on tourism flow has always been one of the core issues in tourism research. Scholars have studied the tourism flow for a long time. In the early stage, the relevant researches mainly used the traditional methods such as questionnaire and statistical yearbook to obtain data. Shih *et al.* [12] collect data of 16 self- driving tourism destinations in Nantou County, Taiwan Province by questionnaire survey, and provide feasible suggestions for local tourism development through the analysis and evaluation of the degree, centrality and structural hole indicators of network nodes; Mckercher *et al.* [13], collecting the data of Japanese inbound tourism flow from November 2004 to December 2005 through questionnaire survey, analyze the spatial and temporal characteristics, and divide the Japanese inbound tourism flow into 11 types of flow patterns; Connell *et al.* [14], collecting data of automobile travel from March to October 2003 in Loch Lomond and Trossachs National Park in Scotland through questionnaire survey, study the spatial characteristics of tourism flow and the travel mode of tourists; Zhang *et al.* [15], collecting the statistical yearbook data of Chinese urban inbound tourism flow and domestic tourism flow from 1997 to 2007, analyze the spatial auto correlation of the two kinds of tourism flow, and divide the Chinese tourism hot cities into three urban clusters through Moran's I. Based on the traditional questionnaire survey and statistical yearbook data collection method, it is difficult to accurately reflect the spatial distribution of tourists and the characteristics of tourism flow network structure, especially the questionnaire survey method. Although long-term and large-scale tourism data are difficult to obtain, these studies lay a theoretical foundation for current tourism research.

In order to accurately reflect the time and space trajectory information of tourists, scholars begin to study the ''digital footprint of tourism'' information generated on the network during the travel of tourists [16]. Specifically, ''digital footprint of tourism'' refers to the electronic traces with geographical location or geographical labels left on the Internet by tourists in the process of playing [17]. Fabien *et al.* [18], collecting geotagged photos published by users on the photo sharing website Flickr from April 2005 to April 2007 and combining mobile phone signaling, conduct a survey of the tourism flows in Florence and Italy in time and space respectively. It is found that tourists are more inclined to visit Florence in July and August in terms of time; while in terms of space, the most tourists visit the Santa Maria in Florence Province, and Rome and Florence are most closely connected; Mou *et al.* [19], obtaining the online travel notes from Qunar.com, analyze the spatial mode of tourism flow in Qingdao by using the gravity center model and constructing the tourism flow network, and find that the tourism resources in coastal and inland areas of Qingdao are obviously different and the coastal tourism core area has been gradually formed;Chung *et al.* [20] analyze the tourism mode of Korean backpackers to Europe from the central indicators, finding that tourists are more apt to travel to London and Paris. Önder et al. [21], collecting all Flickr tourism photo data from 2007 to 2011 in Austria, establish polynomial regression prediction model, and predict the number of tourists in different cities in Austria through Flickr data;Su et al. [22], using Weibo check-in data, compare the temporal and spatial characteristics of day trippers and general tourists, tourists from Shenzhen and other tourists form Chinese mainland visiting Hong Kong.

With the increasing research results on the spatial structure of tourism flows, some scholars begin to focus on the study of the factors affecting tourism flow.Wang et al. [23], based on China's important tourist attractions data, using the nearest neighbor distance method and multi-distance analysis method, study the factors affecting the tourist flow of scenic spots, finding that the tourist flow of scenic spots is related not only to the level of urban economic development, infrastructure and other social factors but also to the spatial structure of scenic spots and natural factors such as hierarchy. Yang et al. [24], after analyzing the driving mechanism of tourism flow, construct the conceptual model of driving mechanism of tourism flow, which divides the driving force into four factors: the driving force of tourist demand, the attraction of tourist destination, the resistance between tourist destination and destination, and the spatial structure force between tourist source and destination; Papateodorou *et al.* [25] analyze the reasons for tourists' choice of travel destination diversity from the aspects of transportation convenience, tourism infrastructure, travel cost, tourists' income and tourism popularity; Marrocu *et al.* [26] count the economic development of 17 European countries and the tourism flows received by each country, concluding that the production of tourism destination countries is positively related to the tourist reception. Then, they further analyze the impact of national population, social capital, scientific and technological development and public infrastructure on tourism flows. Migu é ns et al. [27] confirm the rationality of the application of complex network in tourism flow research through the global tourism flow data in 2004, concluding that the number of outbound tourism in each country is related to the number of outbound tourism market, and the global tourism network is characterized by assortativity. Su *et al.* [28], collecting Flickr photos taken by Chinese inbound tourists from 2008 to 2013, analyze the factors affecting the urban tourism flow through the user sharing density. It is found that the mechanism of the urban economic development, traffic accessibility and infrastructure are the important factors that affect the inbound tourism flows. Liu *et al.* [29], collecting the check-in data left by Chinese tourists in LBSNS, analyze the spatial interaction of tourism flows between cities by using the gravity model, finding that the spatial interaction intensity between tourism cities is dominated by the power law attenuation effect.

With regard to the studies of tourism flows, relying on the theory of geography, integrating physics, statistics and other disciplines, scholars establish a relatively systematic behavior pattern of space-time tourism research [30], creating a theoretical framework of tourism flow. Under the framework of this theory,the spatial-temporal characteristics of tourism flows [31], [32], driving mechanism [33] and spatial structure evolution [34] are analyzed by means of gravity model [35], rank size [36] and econometric analysis. China is rich in tourism resources; urban tourism has obvious cultural differences and the demand of domestic tourism is increasing, showing that the research of domestic tourism flows in China is an important direction in tourism flow researches in the future. From the perspective of the research area, most of the researches on tourism flow are mainly confined to the cities or scenic spots, with more research results on small-scale tourism flow [37], and less research results on large-scale tourism flow nationwide [38]; from the perspective of time, most scholars focus on the long-term scale, with few on the traditional Chinese holiday scale, such as National Day and the Spring Festival, etc. From the perspective of the researches, most of the achievements mainly focus on the analysis of time and space characteristics, but few on the network characteristics. The existing researches mainly are as follows: Leung *et al.* [39], collecting the data of foreign tourists' travel notes during the Beijing Olympic Games, analyze the network structure characteristics of Beijing tourism flow, finding that foreign tourists prefer to play in traditional famous tourist attractions and the center of Beijing; Wu *et al.* [40], taking Beijing as the research area, analyze the characteristics of Beijing inbound tourism flow network through complex network analysis, providing some suggestions for tourists' tourism route planning and bus route planning. However, these studies only analyze the overall characteristics of the network, ignoring the characteristics of nodes and the relationship between nodes.

In view of this, this paper, on the basis of Weibo check-in data, selecting 50 typical Chinese cities as the research objects, uses social network analysis method to study Chinese tourism flow from the perspective of urban network structure, so as to play a guiding role in urban tourism planning, passenger flow prediction and regulation [41]. In addition, it studies the structure of domestic urban tourism flow space, trying to establish a spatial structure system of tourism flows between urban tourism destinations.

III. DATA

In this paper, a new research framework is proposed to study the network characteristics of the National Day Holiday flows, based on Weibo check-in data. The process framework consists of three steps. First, by using the API interface provided by Sina Weibo, we crawl to the research data source and work out detailed processing rules for the original data set. Second, We calculate the evaluation index of tourist flow network from two dimensions of space and network respectively. Third, as the core part of the research, we analyze the tourism flow network, which mainly includes the analysis of tourist flow spatial network and network structure. In order to understand the flow pattern of the National Day Holiday tourism flow network more clearly, when analyzing the spatial characteristics of the tourism flow network, we make a comprehensive analysis of the hierarchy and the spatial structure of the tourism flow in the network; and when analyzing the network structure of the tourism flow, we analyze the node structure and the whole structure in the network, as is shown in Figure 1.

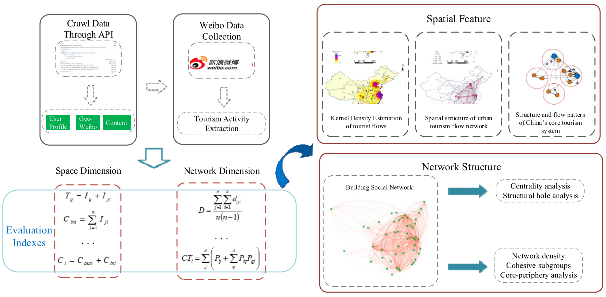

FIGURE 1. Research framework for analyzing the network characteristics of tourist flows.

TABLE 1. Properties of the data.

Data fields	Examples
Weibo ID	5448356917
Check-in time	2018/10/4 14:27
Longitude	118.081947
Latitude	24.454321
Check-in cities	Xiamen City
Weibo content	With my beloved, enjoy lovely sea breeze
Registration location	Shijiazhuang, Hebei Prov.
User gender	Female

A. RESEARCH AREA AND DATA SOURCE

The number of Weibo check-in is closely related to the city's development of network information. The higher the development of urban network, the more reasonable the sample data will be. Taking 50 cities as network node units,considering the accuracy of the experimental results and the representativeness of the sample cities, this paper selects the cities with high Internet level as the research cities.This paper, based on top 100 cities for Internet development selected by China's "Internet plus" index report (2018), combined with the existing researches on the level of China's urban tourism destinations, considering the representativeness of cities in the province and the balance of inter provincial cities, finally selects 50 cities that may become the key cities of the tour registration (excluding the Hongkong Special Administrative Region, the Macao Special Administrative Region and Taiwan province) as the research samples. On this basis, Weibo check-in data is collected from Sina Weibo platform from October 1, 2018 to October 7, 2018, obtaining a toal of 734926 raw data. Data collection content includes user ID, check-in time, longitude and latitude, check-in cities, Weibo content and user registration location information, among which user ID is used for eliminating duplication, check-in time for determining the time of check-in, Weibo check-in content for filtrating check-in information, check-in city and user registration location for forming tourist flow direction with check-in city as tourism destinations and user registration location as tourist source, and longitude and latitude for verifying the accuracy of check-in city. The properties of the data are shown in Table 1.

B. DATA CLEANING

Because the data captured by Sina Weibo API contains a large number of invalid data, it is necessary to filter and process the collected original data. The standards are as follows: 1) according to the information of the user's registration places and check-in places, the check-in information in the cities are excluded; 2) For the Weibo check-in information exclusive of local check-in, traveling during the holidays can be divided into traveling, visiting relatives and friends, going to school, traveling on business. On the basis of the content of Weibo, non-tourism check-in data including "overtime", "school", "home", "business", "visiting relatives" and other words, are further removed; 3) incomplete and non-standard check-in data is deleted; 4) Weibo check-in data of tourists outside the sample city is excluded; 5) multiple check-in information of tourists in the same city is screened out; 6) municipal districts are merged, such as Haidian District and Chaoyang District of Beijing, which should be marked as Beijing at the same time. Through the statistics, 194,231 Weibo check-in data are finally obtained, among which 21824, 28028, 39882, 37375, 29526, 22193 and 15403 effective data are respectively collected form Oct. 1 to Oct. 7. In order to verify the accuracy of the collected data, one of every 200 pieces of data is extracted from the cleaned data, thus obtaining a total of 1000 samples for verification, among which only 76 are non-tourism data with an accuracy rate of about 92.4%. The cleaning rules are shown in Table 2.

TABLE 2. Detailed cleaning rules for weibo data.

Cleaning rules	Operation fields	Remaining amount
Data of local check-in	Registration location, check-in city	434782
Non-tourism data	Check-in content	327843
Data in wrong format	All fields	315964
Tourist data beyond sample city	Registration location	212391
Repeated data	Weibo ID, check-in city	194231

Kernel Density Estimation [42] is carried out according to the number of Weibo check-in in the sample city, and according to the number of city check-in, the city is divided into five grades [43] using the Nature Breaks method, and Figure 2 is obtained, where the larger the purple dots, the more the number of city check-in, and the higher the tourist grade, and then Figure 1 is drawn, which shows that Weibo check-in data is concentrated in the east of Hu Line, especially densely scattered in areas of Sichuan-Chongqing, Pearl River Delta, Yangtze River Delta and Beijing-Tianjin-Jebei. Although the western region is rich in tourism resources, due to traffic, climate and other reasons, the number of check-in the western region is sparse.

C. METHODS

Social network analysis is to describe actors and their relationships. As an important method to quantify the network, it has been widely applied to academia [44]. According to social network analysis, actors are nodes in the whole network, and society is a network composed of mutual relations

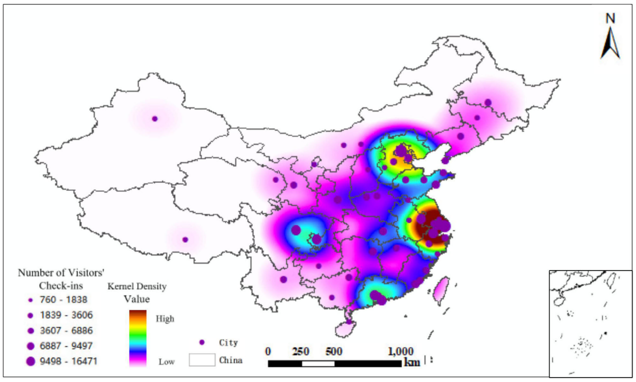

FIGURE 2. **Study on the spatial distribution of cities and the number of tourist check-in.**

between actors. Through the analysis of actors and their relations, the characteristics of the whole network structure are revealed [4]. In this paper, by using the social network analysis method, based on the Weibo check-in data during National Day data, analyzes the node structures and overall structures of the above-mentioned 50 urban tourism flow networks, in which the node structure is evaluated by node centrality and structural hole; the overall network structure is evaluated from the network density, core-periphery model and cohesive subgroups.

1) TOURISM FLOW

Tourism flow refers to the total amount of tourist links between two cities. The larger the value is, the closer the tourist links between the two cities are. The formula is as follows:

$$T_{ij} = I_{ij} + I_{ji} \tag{1}$$

where: T_{ij} refers to the tourism flow between cities i and j; I_{ij} is the tourism flow from city i to city j; and I_{ji} is the tourism flow from city$j$ to city i.

2) INFLOW DEGREE

$$C_{ini} = \sum_{j=1}^{n} I_{ji} \tag{2}$$

where: C_{ini} is the inflow of city i; I_{ji} is the tourism flow from city $j$to city i; and n is the number of network nodes excluding i.

3) OUTFLOW DEGREE

Outflow degree refers to the total flows of a node in the network to other nodes. The larger the value is, the higher the inflow degree of the node in the network is, the stronger the diffusion ability of the city is. The formula is as follows:

$$C_{outi} = \sum_{j=1}^{n} I_{ij} \tag{3}$$

where: C_{outi} is the outflow degree of city i; and I_{ij} is the tourism flows from city i to each city j;

4) DEGREE OF CENTRALITY

Degree of centrality [45] is measured by the sum of inflow and outflow of a node in the network. The larger the value, the higher the degree of centrality of the node in the network. The formula is as follows:

$$C_i = C_{outi} + C_{ini} \tag{4}$$

5) NETWORK DENSITY

Network density, reflecting the closeness of the whole network, is an effective guide to the overall structure of the network. This data is obtained by dividing the actual number of connections between network nodes by the number of theoretical connections. The result is between [0,1]. The closer the value is to 1, the greater the network density is, the greater the impact of the network on each node is; the closer the value is to 0, the smaller the network density is. The formula is as follows:

$$D = \frac{\sum_{j=1}^{n}\sum_{i=1}^{n} d_{ji}}{n\,(n-1)} \tag{5}$$

where: D represents the network density; and d_{ji} refers to the connection path between the actual network nodes i and j.

6) CENTRALITY ANALYSIS

Network centrality measures the degree to which a node is in the center of the network. In the tourism flow networks, each city is regarded as each actor. Through the analysis of its core, the power and status of different cities in the network can be gained. At present, the common central indicators include: degree centrality, proximity centrality and betweenness centrality.

Degree centrality, reflecting the ability of a tourism node to communicate with other nodes in the networks, is the simplest and most intuitive index to evaluate the centrality. In directed network, degree centrality can be divided into inward degree centrality and outward degree centrality, which means that tourists flow from other tourism nodes to a tourism node and tourists flow from a tourism node to other tourism nodes. The higher inward degree centrality indicates the node is agglomeration center, while the more outward degree centrality indicates the node is radiation center. The formula is as follows:

$$C_{RD(i)} = \frac{C_{AD(i)}}{n-1} \tag{6}$$

where: $C_{RD(i)}$ is the relative degree centrality of node i; and $C_{AD(i)}$ is the absolute degree centrality of node i, that is, the number of nodes connected to the node.

Closeness centrality means the sum of the distance between one node and other tourism nodes in the network, indicating its position in the network. The smaller the value is, the higher the closeness centrality is, the closer the connection is, and the more the node is in the center of the network. In the directed network, closeness centrality can be divided into inward approach centrality and outward approach centrality. The formula is as follows:

$$C_{APi}^{-1} = \sum_{j=1}^{n} d_{ij} \tag{7}$$

where: C_{APi}^{-1} is closeness centrality; and d_{ij} is the shortest distance between nodes i and j.

Betweenness centrality refers to the degree of control over other nodes when one node in the network contacts with other tourism nodes. The higher betweenness centrality of nodes, the stronger the control for other nodes. The formula is as follows:

$$C_{RBi} = \frac{2C_{ABi}}{(n-1)(n-2)} = \frac{2\sum_{j=1}^{n}\sum_{k=1}^{n} g_{jk}(i)/g_{jk}}{(n^2-3n+2)} \tag{8}$$

where, C_{RBi} is the relative betweenness centrality of node i; C_{ABi} is the absolute betweenness centrality; g_{jk} is the shortest path number between node k and node j; and $g_{jk}(i)$ is the shortest path number between node j and node k passing through node i.

7) CORE-PERIPHERY ANALYSIS

The analysis of core-periphery model can judge whether the tourism node is in the core area or the periphery area of the whole network, so as to realize the quantification of the location of the tourism node. The significance of core edge analysis can not only clarify the core role in the whole network, but also quantify the driving effect of core area on the periphery area.

8) COHESIVE SUBGROUPS

The analysis of cohesive subgroups is a study of the structure and characteristics of the whole network. It divides the nodes in the network into several discrete subgroups according to certain standards, and the nodes in each subgroup have relatively strong, direct, and close correlation. These subgroups are called "blocks". The analysis of cohesive subgroups can reveal the structural characteristics of inner subgroups. In this paper, the CONCOR algorithm is applied to analyze China's urban tourism network, and 50 cities in the tourism network are divided into subgroups.

9) STRUCTRUAL HOLE

Structural hole index can distinguish the advantage and disadvantage nodes in the network. It describes the fracture phenomenon between tourism nodes. For the study of structural holes, it can not only clearly point out the urban tourism nodes with irreplaceable location advantages and competitive advantages, but also guide the disadvantaged nodes to strengthen communication and cooperation with the surrounding advantageous nodes. At present, the common structural hole evaluation indexes include: effective size, efficiency and constraint.

Effective size refers to the non-redundant part of the network where the target node connects with all other tourism nodes. It reflects the importance of the node's position in the network to some extent. The greater the effective size, the stronger the overall influence of the target node in the network. The formula is as follows:

$$ES = \sum_{j}^{n}\left(1-\sum_{q}^{n} P_{iq}m_{jq}\right) = \sum_{j}^{n}\left[1-\sum_{q}^{n}\frac{(z_{iq}+z_{qi})}{\sum_{j}^{n}(z_{ij}+z_{ji})}\frac{(z_{jq}+z_{qj})}{\max(z_{jk}+z_{kj})}\right] \tag{9}$$

where: p_{iq} is the proportional relationship between tourism node i and q, that is, the number of connections between nodes i and q divided by the sum of all connections between nodes i and j; m_{jq} is the marginal strength between nodes j and q, and the number of connections between node j and q divided by the maximum number of connections between node j and other nodes; and n is the number of nodes in the tourism flow network.

Efficiency refers to the ratio of effective scale and actual scale in the tourism network. The higher the efficiency of the tourism nodes, the more obvious the competitive advantage of the node in the network. The formula is as follows:

$$ET = \frac{ES_j}{D_j} \tag{10}$$

where: ES_j represents the effective size of node j in the tourism network; and D_j refers to the number of connections between node j and other nodes in the tourism network.

Constraint reflects the number of direct and indirect dependence of nodes on other tourism nodes in the tourism city networks. The greater the constraint is, the more easily the node is dominated by other nodes in the whole network and in a disadvantage position in the tourism competition. Otherwise, with less constraint, the node is more likely to have a core position in the whole network. The formula is as follows:

$$CT_i = \sum_{j}^{n}\left(P_{ij}+\sum_{q}^{n} P_{iq}P_{qj}\right) \tag{11}$$

where: p_{ij} is the proportion between node i and node j; p_{iq} is the proportion between node i and node q; p_{qj} is the proportion between node q and node j; and n is the number of tourism nodes in the tourism flow network.

IV. RESULTS

A. SPATIAL STRUCTURE OF URBAN TOURISM FLOW NETWORK

In order to explore the position of each node's city in the whole tourism network, this paper, by using the node's degree of centrality to quantify the size of urban power, calculates the tourism flow and the city's degree of centrality between

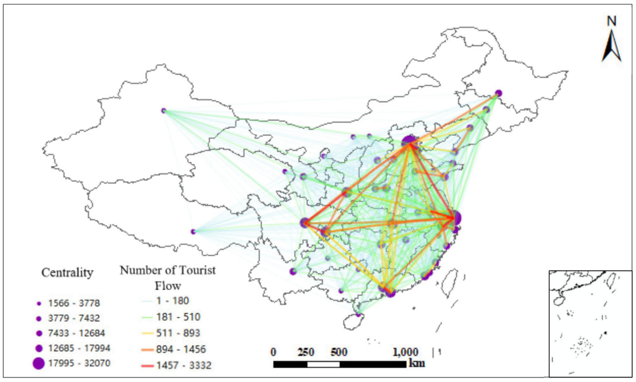

FIGURE 3. Spatial structure of urban tourism flow network.

50 urban nodes on the basis of the formulas (1) - (4). In addition, by using Jenks, it divides the inter-city tourism flow and the city's degree of centrality into five levels, and draws the spatial structure of China's urban tourism flow network (Figure 3), finding that, based on Weibo check-in data, the spatial development of urban tourism flow of data is not balanced, mainly concentrated in the eastern and central regions. The characteristics of results are as follows:

1) HIERARCHY OF ROLES IN URBAN TOURISM NODES

Different from the standard of traditional tourism city classification, this paper, on the basis of the degree of centrality index of city nodes, divides the research city into five grades (Table 3) by using the natural fracture point. Shanghai and Beijing, belonging to the first-class cities, reach 32,070 and 32,002 in the degree of centrality respectively, which is significantly higher than that of other nodes in the network. They are at the core of the whole tourism network and the core cities of national tourism links. Hangzhou, Chengdu, Chongqing, Guangzhou, Nanjing, Shenzhen, Xi'an and Wuhan, the second-class cities, with a degree of centrality between 12684 and 17994, belonging to national tourism related cities, are concentrated in the south of the Qinling Mountains-Huaihe River line. Tianjin, Suzhou, Qingdao, Changsha, Xiamen and Zhengzhou, the third-class cities, with a degree of centrality between 7432 and 12684, belonging to the core cities of regional tourism links, are concentrated in the eastern part of China. Eighteen cities, such as Jinan, Harbin and Dalian, the fourth-class cities, with a degree of centrality between 3779 and 7432, belonging to regional tourism related cities, are mainly distributed in the central and western regions. Sixteen cities, including Nanning, Guilin and Guiyang, the fifth-class cities, with a degree of centrality of between 3779 and 1566 and the weakest city tourism connection in the network, belonging to general tourism related cities, are evenly distributed in the eastern, central and western regions of China.

2) SPATIAL DISTRIBUTION CHARACTERISTICS OF URBAN TOURISM FLOW

Through the study of inter-city tourism flows, this paper concludes that there are 1,224 inter-city tourism links in the whole tourism flow network, only the tourism flow from Baotou City to Dongguan City is 0. Figure 3 shows that the main tourism flows among Chinese cities present a spatial structure with diamond as the core. Among them, Beijing (North), Chengdu (West), Shanghai (East) and Guangzhou (South) constitute four diamond apexes respectively. Wuhan is the intersection of diamond diagonal, and these cities belong to the core cities of national tourism connection or national tourism connection cities. It demonstrates that the constraint effect of distance on the spatial structure of tourism flow network between cities is not particularly obvious, and the core tourism flow mainly occurs between national tourism cities, which is not consistent with the conclusion of tourism flow research through gravity model [46]. The four top cities of the rhombus show multi-core subjects in different regional systems, and form a multi-level complex spatial network pattern together with the surrounding cities with low centrality. As is shown in Figure 4, a dual core spatial structure centered on Chengdu and Chongqing has been formed in the Sichuan-Chongqing tourism destination system; a dual core spatial structure centered on Beijing and Tianjin has been formed the tourism destination system in the Beijing-Tianjin-Hebei region; a dual core spatial structure centered on Shenzhen and Guangzhou has been formed in the Pearl River Delta tourism destination system; a dual core spatial structure centered on the Yangtze River Delta tourism destination system; and a multi-core common development spatial structure centered on Shanghai, Nanjing, Hangzhou and Suzhou has been formed in the Yangtze River Delta tourism destination system.

TABLE 3. Hierarchy of tourism cities.

Degree of centrality	Hierarchy	List of cities
>17994	Core cities of national tourism links	Shanghai, Beijing
12684~17994	Cities linked with national tourism	Hangzhou, Chengdu, Chongqing, Guangzhou, Nanjing, Shenzhen, Xi'an, Wuhan
7432~12684	The core cities of regional tourism links	Tianjin, Suzhou, Qingdao, Changsha, Xiamen, Zhengzhou
3779~7432	Cities linked with regional tourism	Jinan, Dalian, Harbin, Shenyang, Ningbo, Wuxi, Hefei, Shijiazhuang, Changchun, Fuzhou, Taiyuan, Yantai, Wenzhou, Kunming, Nanchang, Quanzhou, Lanzhou, Luoyang
<3779	Cities liked with general tourism	Jinhua, Weifang, Baoding, Dongguan, Nanning, Nantong, Guilin, Xuzhou, Guiyang, Haikou, Urumqi, Hohhot, Xining, Yinchuan, Baotou, Lhasa

In order to analyze the spatial structure of China's main tourist flows, the inter-city tourist flows are ranked to screen out the tourism links that constitute the rhombus spatial structure of China's tourism network. These tourist links account for 6 per cent of total inter-city tourism links,

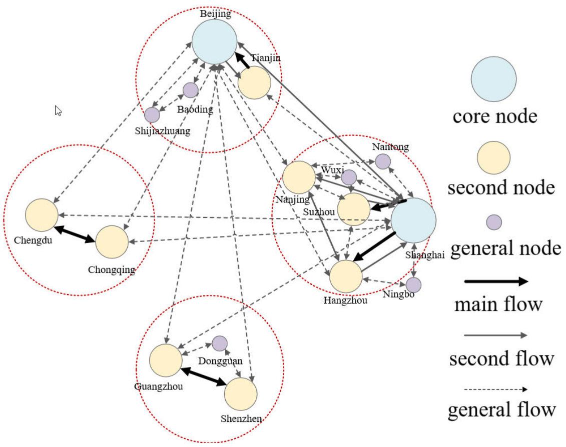

FIGURE 4. Structure and flow pattern of China's core tourism system.

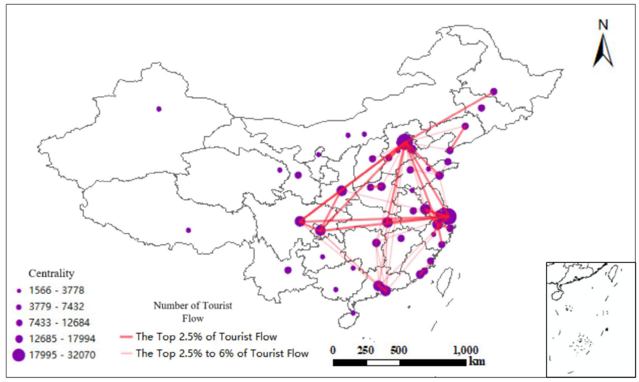

FIGURE 5. Spatial distribution of two types of tourism connection flow.

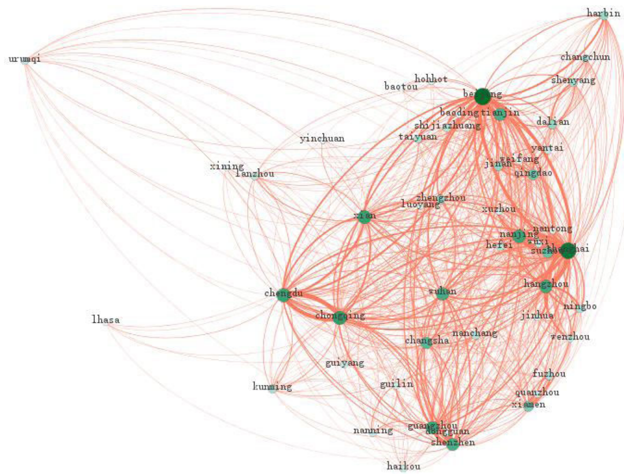

FIGURE 6. Network structure of inter-city tourism flow.

and according to the intensity of inter-city connection, they are divided into the first 2.5% and 2.5%∼6% respectively. As is shown in Figure 5, the top five flow directions include: Chongqing-Chengdu, Tianjin-Beijing, Shanghai-Suzhou, Shenzhen-Guangzhou and Shanghai-Hangzhou. The tourism flow values are all larger than 2,800, and the flow directions are all between the core cities within the adjacent urban agglomerations. It is mainly because when tourists travel, they not only consider the distance between cities, but also ponder over the level of urban tourism, showing that urban tourism flow is characterized by proximity and selectivity. There are 30 pairs of top 2.5% tourism connection flows, and their value of tourism connection flow is greater than 1,065. The main spatial direction includes the connection between the Beijing-Tianjin-Hebei region and the Sichuan-Chongqing region, the Yangtze River Delta city group and the connection between the Yangtze River Delta city group and the Sichuan-Chongqing region, forming the upper half of diamond core structure with Beijing, Shanghai and Chengdu as nodes. The top 2.5% - 6% of the tourism connection flows contain 43 pairs, and their values are larger than 670. These tourism connection flows mainly include the tourism connection between the Pearl River Delta region, the Yangtze River Delta region, the Sichuan-Chongqing region and the Beijing-Tianjin-Hebei region, forming the lower half of the diamond core structure with Guangzhou, Chengdu and Shanghai as nodes. From the perspective of spatial pattern, the intensity of tourism links between the nodes in the upper half of the spatial structure with diamond as the core formed by the main tourism links in China is significantly stronger than that in the lower half, that is to say, the size of the main tourism links in China has obvious heterogeneity in spatial distribution.

B. NETWORK STRUCTURE CHARACTERISTICS OF URBAN TOURISM FLOW

1) NETWORK STRUCTURE CONSTRUCTION OF URBAN TOURISM FLOW

According to the collected information of user's check-in and registration places, this paper constructs a directed relationship matrix for the inter-city tourism flow network. First, a threshold value of n is determined; then, if the tourism flow between city i and city j is greater than or equal to n, then it is recorded as 1, otherwise as 0; finally, a binary matrix of $50 * 50$ is obtained. If the threshold value is too large, the important tourism flow in the network structure will be lost, and the network structure features will not be fully displayed; if the threshold value is too small, the tourism paths between nodes will be too many, and it is not easy to highlight the core nodes in the network structure. When the threshold is 1, the number of tourist flow paths in the network is 2432, and the maximum number of paths composed of 50 nodes in theory is 2450. At this time, the characteristics of the tourist flow network are not very obvious. Therefore, to reflect the characteristics of the network structure more clearly, the threshold value is finally determined to be 50. With the help of UCINET software, the overall network structure diagram of inter-city tourism flow is created (Figure 6), in which the node size and color indicate the number of connections between the city and other cities; the thickness of inter-city line segment indicates the traffic of inter-city tourism.

2) EVALUATION OF NETWORK NODE STRUCTURE

In this paper, the total GDP of sample cities in 2018 is counted, the economic development of sample cities is quantified, and the inflow, outflow and net inflow of each city are counted and analyzed on the basis of the GDP ranking, and then Figure 7 is drawn. From the perspective of inflow degree and outflow degree, the size of flow and the urban economic development presents a logarithmic function distribution. The lower the level of urban economic development, the less the inflow and outflow between cities. R2 of the fitting function is 0.78 and 0.86 respectively, which has a high fitting degree. In terms of inflow, the top five cities with a high economic development and sound urban infrastructure are Beijing, Shanghai, Hangzhou, Chengdu and Nanjing. In Figure 7 (a), the inflow of Xi'an and Xiamen is much higher than the tourism inflow value of the fitting function, because Xi'an is a famous ancient capital of China, with a profound historical background and a large number of places of interest, such as the terracotta warriors of the first emperor of Qin, Lotus Garden of Tang Dynasty, etc. Xiamen, as a major coastal city in China, enjoys a proper temperature, beautiful natural scenery and strong romantic atmosphere, such as the Gulangyu Island, the Huandao Road, etc.; the top five cities in terms of outflow rate are Shanghai, Beijing, Hangzhou, Chongqing and Chengdu, which are important export destinations of tourists. These cities also have large GDP and high per capita disposable income, providing a guarantee for travel expenses. In Figure 7 (b), the outflow of Xi'an is much higher than the total outflow of the fitting function, which is mainly because Xi'an, located in the middle of China's mainland plate, is relatively close to city. In addition, this city is an important transportation hub in China, providing some convenient conditions for people's travel. In addition, another important factor is that Xi'an boasts a large population base. From Figure 7 (c), Suzhou, Xiamen, Changsha and Guilin have higher net imports, while Zhengzhou, Tianjin and Shanghai have higher net exports. Suzhou has the highest net imports of 4025 people, mainly due to its rich tourism resources, typical Jiangnan Water Town, and relatively complete urban infrastructure; Tianjin has the highest net exports of 2516 people, mainly by virtue of its location adjacent to Beijing, the capital city of China, with rich tourism resources, convenient transportation and strong cultural history, attracting the inflow of surrounding low-ranking cities to high-ranking cities.

Through the analysis of the node centrality and structural holes, Table 4 is drawn. From the degree centrality index, the mean value of degree centrality is 33%, showing that each node in the network has agglomeration and radiation connection with 33% of other nodes. Degree centrality can be divided into inward degree centrality and outward degree centrality, with variances of 668 and 494, respectively, indicating that there is a strong imbalance in the network structure. Beijing, Shanghai, Tianjin and Chongqing are highly export-oriented, with a strong radiation effect on tourism flow. Among them, Beijing, the highest inward degree centrality, up to 97.959%, is the largest tourism divergent center in the network; Lhasa, the lowest inward degree centrality, with a value of 0, has no tourism outflow under this threshold.

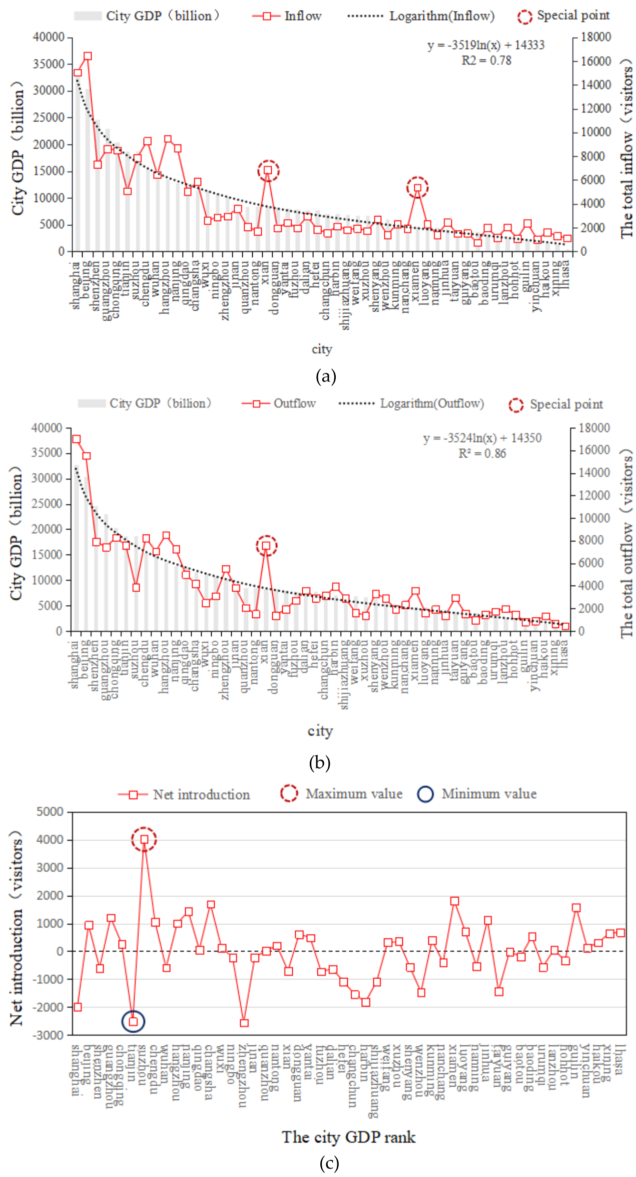

FIGURE 7. Line chart of urban inflow, outflow and net inflow. (a) Relationship between economic development and total inflow; (b) Relationship between economic development and total outflow; (c) Total number of net tourists in the city.

Beijing, Shanghai, Nanjing and Chongqing have a high degree of inward centrality. Among them, Beijing has the highest degree of export-oriented centrality, while Baotou and Baoding have the lowest degree of export-oriented centrality. From the point of view of proximity to the central indicators, Beijing, Shanghai, Chongqing and Xi'an have both high inward and outward connectivity, reflecting that these cities occupy the central position of the network, with good tourism flow liquidity and weak dependence on other nodes. From the perspective of betweenness centrality, Beijing and Shanghai are significantly higher than other node cities in the network, which has strong control over other cities. This is the key node in the network, which benefits from the establishment of rich tourism resources, perfect transportation facilities and infrastructure.

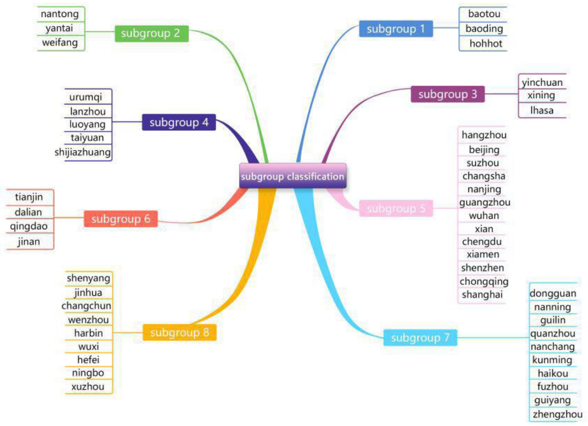

FIGURE 8. Members of eight subgroups.

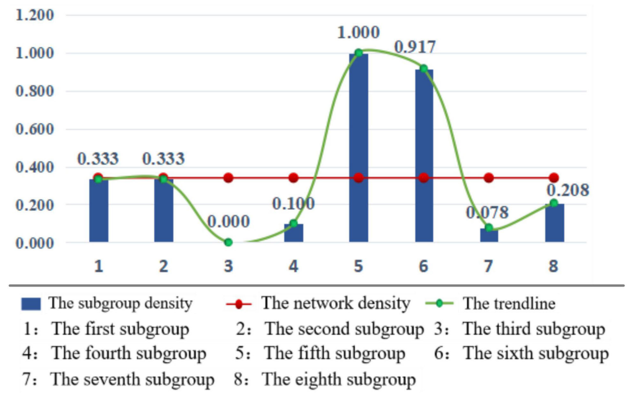

FIGURE 9. Members of eight subgroups.

Based on the structure hole index in Table 4, Beijing, Shanghai, Chongqing, Xi'an, Nanjing, Chengdu, Guangzhou, Wuhan and Tianjin have higher effective size and efficiency, and less constraints. It is shown that these cities are less limited by other cities in the network, have better competitive advantages in the tourism flow network, and play an irreplaceable role in different regions, making them in a favorable position in the network. Baotou, Baoding, Lhasa and Hohhot have lower efficiency and efficiency, with a great constraint, showing that these cities are low in efficiency, high in dependence on other cities, and in a disadvantage position in the whole network due to the greater restrictions of other cities. Therefore, these cities should strengthen the connection with surrounding core cities and promote their development, so as to avoid the negative influence of structural hole on them.

3) EVALUATION OF OVERALL NETWORK STRUCTURE

Based on Formula (5), the network density of urban tourism flow in China is 33.6%. In the 50 $*$ 50 network, only 818 travel paths in reality, reflect the low degree of the whole network structure and the lack of close tourism links between cities. In addition, most of the tourism flows are concentrated in Beijing, Shanghai and other cities with rich tourism resources, and the links between cities with poor tourism resources are relatively weak. Therefore, it is necessary to strengthen the integrity of China's urban tourism network.

Based on the CONCOR algorithm, the nodes in the network are divided into eight subgroups based on the degree of external connection of urban tourism, each of which has some cohesive characteristics, and the members of each subgroup are shown in Figure 8. There are great differences in the number of members in each subgroup. Among them, subgroup 1, subgroup 2 and subgroup 3 have three members respectively, and the number of members is the least; subgroup 5 has 13 members, and the number of members is the most.

Through the analysis of the internal connection density of the subgroups, Figure 9 is drawn. On the basis of the graph of internal connection density of each subgroup, the internal connection density of six subgroups in the tourism network is lower than that of the whole network, and only the internal connection density of two subgroups is higher than that of the whole network. Among them, the internal density of subgroup 3 is 0, without any tourism flow connection, with the lowest density; the internal connection density of subgroup 5 and subgroup 6 is higher than that of other subgroups, all of which are above 0.9, and the internal connection density of subgroup 5 is even higher than that of other subgroups, indicating that each city in the subgroup has tourism connection. These cities are responsible for the efficient operation of China's tourism flow, and their subgroups mainly include the core cities of national tourism links and the cities of national tourism links, such as Beijing, Shanghai, Chongqing and Hangzhou. With the convenient transportation, perfect infrastructure and high development of the city, these cities, mainly located in the eastern part of China, provide the necessary conditions for the development of urban tourism.

Figure 10 shows the density of tourism links among subgroups. The first subgroup is relatively weak in agglomeration and diffusion, and only connects with three subgroups, with low connection density. Further analysis indicates that the first subgroup is composed of general tourism related cities, which is limited by location conditions, economic development and other factors, resulting in low connection density with other subgroups; the second subgroup is most closely related to the sixth subgroup, and the sixth subgroup is the main source of tourists. The results show that the second and the sixth sub cluster cities belong to the geographical neighborhood relationship in space, while the third sub cluster has weak tourism relationship with other sub groups, which is mainly restricted by climate environment, economic development and the accessibility of transportation network; The fourth subgroup is most closely related to the fifth. Due to the convenient transportation and rich tourism resources, the fifth sub group becomes the first choice of the fourth sub group; the fifth sub group is not only the main tourist source sub group, but also the main tourist destination sub group. The fifth sub group is mostly a national tourism city, and its infrastructure, tourism resources, traffic network accessibility and economic development and other factors are

TABLE 4. City node centrality and structural hole index.

Cities	Degree centrality		Closeness centrality		Betweenness centrality	Structrual holes		
	Inward	Outward	Inward	Outward		Effective Size	Efficiency	Constraint
Beijing	93.878	97.959	95.92	98.98	17.255	33.803	0.69	0.091
Shanghai	87.755	93.878	92.86	96.94	12.163	31.197	0.664	0.096
Guangzhou	75.510	55.102	86.74	77.55	2.867	19.18	0.518	0.12
Shenzhen	65.306	59.184	81.63	79.59	2.937	16.754	0.493	0.126
Chengdu	73.469	65.306	85.71	82.65	4.17	20.75	0.546	0.116
Chongqing	75.510	69.388	86.74	84.69	4.388	23.676	0.577	0.112
Hangzhou	73.469	57.143	85.71	78.57	3.305	18.445	0.512	0.119
Nanjing	79.592	55.102	88.78	77.55	3.277	21.97	0.549	0.113
Xi'an	75.510	65.306	86.74	82.65	5.66	23.297	0.582	0.113
Wuhan	67.347	59.184	82.65	79.59	2.423	19.194	0.519	0.121
Tianjin	51.020	71.429	74.49	85.71	3.764	18.6	0.531	0.124
Suzhou	63.265	36.735	80.61	68.37	1.184	13.235	0.427	0.137
Qingdao	55.102	40.816	76.53	70.41	0.982	11.33	0.391	0.145
Changsha	67.347	44.898	82.65	72.45	1.759	15.391	0.466	0.133
Xiamen	57.143	30.612	77.55	65.31	0.315	9.314	0.333	0.15
Zhengzhou	30.612	38.776	64.29	69.39	0.231	4.309	0.215	0.198
Jinan	34.694	34.694	65.99	67.35	0.307	6.191	0.281	0.183
Dalian	28.571	40.816	62.93	70.41	0.264	5.382	0.245	0.181
Harbin	24.490	40.816	61.22	70.41	0.166	4.141	0.207	0.197
Shenyang	14.286	38.776	55.78	69.39	0.028	3.212	0.169	0.208
Ningbo	24.490	30.612	61.22	65.31	0.229	2.981	0.186	0.239
Wuxi	16.327	20.408	56.46	60.20	0.082	2.778	0.253	0.345
Hefei	18.367	30.612	57.82	65.31	0.031	1.958	0.122	0.246

TABLE 4. ***(Continued.)*** **City node centrality and structural hole index.**

Shijiazhuang	16.327	30.612	56.80	65.31	0.167	3.261	0.217	0.258
Changchun	14.286	32.653	55.44	66.33	0.017	2.413	0.151	0.243
Fuzhou	22.449	28.571	60.20	64.29	0.053	2.62	0.175	0.258
Taiyuan	14.286	38.776	55.78	69.39	0.021	2.885	0.152	0.21
Yantai	18.367	18.367	57.82	59.18	0.058	2.194	0.199	0.34
Wenzhou	8.163	32.653	51.36	66.33	0.054	3.075	0.192	0.245
Kunming	26.531	28.571	62.25	64.29	0.03	1.704	0.114	0.254
Nanchang	24.490	26.531	60.88	63.27	0.003	1.42	0.095	0.257
Quanzhou	16.327	18.367	56.80	59.18	0	1.088	0.109	0.37
Lanzhou	20.408	24.490	58.84	62.25	0.335	3	0.231	0.295
Luoyang	12.245	14.286	54.76	57.14	0	1	0.125	0.454
Jinhua	14.286	16.327	55.44	58.16	0.044	1.667	0.185	0.406
Weifang	12.245	12.245	54.42	56.12	0.005	1.167	0.194	0.558
Baoding	6.122	10.204	51.36	55.10	0	1	0.2	0.677
Dongguan	12.245	10.204	54.76	54.76	0	1	0.167	0.569
Nanning	18.367	18.367	58.16	59.18	0.169	1.556	0.173	0.394
Nantong	10.204	14.286	53.40	57.14	0	1.083	0.155	0.507
Guilin	24.490	8.163	61.22	47.62	0.014	1.563	0.13	0.321
Xuzhou	14.286	12.245	55.44	56.12	0.007	1.462	0.183	0.458
Guiyang	20.408	20.408	59.18	60.20	0	1	0.091	0.336
Haikou	18.367	20.408	58.16	60.20	0	1	0.1	0.364
Urumqi	16.327	24.490	57.14	62.25	0	1	0.083	0.314
Hohhot	10.204	10.204	53.74	55.10	0.069	1.8	0.36	0.624
Xining	16.327	6.122	56.80	48.30	0	1	0.125	0.464
Yinchuan	10.204	8.163	53.74	53.74	0	1	0.167	0.589
Baotou	6.122	6.122	51.36	52.72	0	1	0.25	0.814
Lhasa	12.245	0	56.12	0	0	1	0.167	0.607

TABLE 4. ***(Continued.)*** **City node centrality and structural hole index.**

Mean value	33.388	33.388	65.45	65.45	1.377	—	—	—
Maximum	93.878	97.959	95.92	98.98	17.255	—	—	—
Minimum	6.122	0	51.36	0	0	—	—	—
Variance	668	494	171.6	213.1	9.684	—	—	—

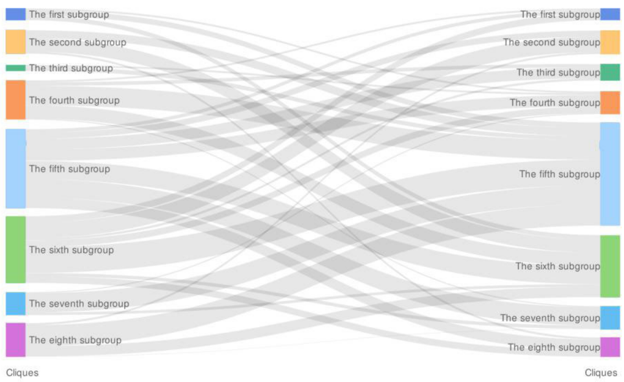

FIGURE 10. **Relationship diagram of eight subgroups.**

in the forefront of the national city ranking, so a large number of urban tourists from other subgroups are attracted to visit; the 6th, 7th and 8th subgroups are closely related to the 5th subgroups respectively, mainly for two reasons. On the one hand, the urban distribution area of these three subgroups is close to the urban spatial distribution of the fifth subgroups, mainly concentrated in the eastern part of China. The distance between cities is close, and the characteristics of spatial proximity are very obvious. In addition, the developed traffic network in the eastern region provides favorable conditions for inter-city tourism in the eastern region. On the other hand, the strengthening of inter-regional urban tourism relationship makes the national central cities in different regions rely on their own tourism agglomeration and diffusion ability, which drives the development of tourism industry in the surrounding low-ranking cities, and the tourism relationship between them is closer. For example, the Yangtze River Delta city group, Shanghai, Hangzhou, Nanjing and other high-ranking cities continue to attract tourism contacts and cooperation of surrounding low-ranking cities by virtue of their own tourism resources and location advantages, while these high-ranking cities are mainly concentrated in the fifth subgroup and the low-ranking cities in the sixth, seventh and eighth subgroups, the main source of tourists in the whole network. The fifth and sixth subgroups are the tourism destinations of the whole network.

Using the core-periphery model, this paper analyzes the inter-city tourism flow to get Table 5. In the tourism flow network, the core cities include 15 members, such as Beijing, Shanghai, Guangzhou. Most of these node cities are concentrated in the eastern coastal areas, the Beijing-Tianjin-Hebei region, the Yangtze River Delta region, the Pearl River Delta region and some cities in the southwest. The edge cities are composed of 35 cities, such as Zhengzhou, Taiyuan, Urumqi, etc., which echoes the analysis results of the above indicators. Table 6 shows that the network density among the members in the core area is 0.986, about three times of the overall network density, indicating that there are intensive tourism flows among the core members and close tourism links between cities. The network density of the edge area is only 0.053, showing that although the number of marginal cities accounts for the vast majority, there are few tourist links between marginal cities, which are almost isolated and constitute the peripheral cities outside the core tourist areas. Most of the strong tourism links in the holiday travel network occur between the core cities, but there are few tourist check-ins between the edge cities, and the tourism connection between cities needs to be strengthened. In addition, the network density between the core members and the edge members has also reached 0.596, close to twice the average network density, showing that the core area and the edge area are also closely related to tourism, and the core city has a significant impact on the tourism development ability of the edge city [47]. Therefore, it is necessary to actively enhance the tourism industry of China's core cities, actively cultivate the diffusion and linkage of the edge areas step by step, and promote the continuous upgrading of the scenic spots of the core cities, so as to improve the tourism industry of the surrounding edge cities.

V. DISCUSSION

1) The core area of the spatial pattern of tourism flow in China's typical cities is generally in the shape of rhombus. The intensity of tourism flow between network nodes is more concentrated in eastern and central China, while the intensity of tourism flow is less concentrated in western cities, indicating that tourism flow has an uneven distribution in space. Therefore, the construction of tourism transportation

TABLE 5. Core-edge analysis of city division.

Subregion	List of cities
Edge cities	Zhengzhou, Wuxi, Urumqi, Wenzhou, Weifang, Taiyuan, Shijiazhuang, Yinchuan, Shenyang, Xining, Nanchang, Luoyang, Lhasa, Lanzhou, Yantai, Quanzhou, Kunming, Hohhot, Hefei, Haikou, Harbin, Xuzhou, Ningbo, Jinhua, Guiyang, Fuzhou, Dongguan, Dalian, Xining, Nantong, Jinan, Guilin, Changchun, Baotou, Baoding
Core cities	Chongqing, Shenzhen, Shanghai, Xi'an, Xiamen, Guangzhou, Wuhan, Qingdao, Chengdu, Tianjin, Nanjing, Changsha, Suzhou, Hangzhou, Beijing

TABLE 6. Core edge density matrix.

	Core	Edge
Core	0.986	0.448
Edge	0.596	0.053

routes in the western region should be improved to enhance the accessibility of inter-city transportation, and to facilitate the tourism links with the core cities in the eastern and central regions. At the same time, the city itself should also strengthen the establishment of infrastructure to attract more tourists to visit; for the cities in the eastern and central regions, the existing intercity transportation and tourism routes and public infrastructure between Beijing, Shanghai, Guangzhou and Chengdu are to provide basic support for the tourism development of the surrounding cities and continuously promote the development of the inter-regional urban tourism.

2) China's tourism network node has an obvious hierarchical structure. The core tourism cities, led by Shanghai and Beijing, should promote the development of the relatively backward cities around. It not only promotes the upgrading of the core cities' tourism industry, but also enhances the relatively backward cities' tourism industry, forming a win-win pattern.

3) With the formation of the core area and the edge area, how to speed up the integration of the edge city into the core area has become the biggest problem facing China's urban tourism. Core regional cities play an important role in maintaining the overall stability of tourism flow network [48]. There is a weak tourism flow between the edge cities, and the strong tourism connection between the cities mainly occurs between the core cities, which controls the development direction of China's tourism flow network structure. Therefore, it is necessary to stabilize the core node, enhance the position of the core node in the network, promote the development of urban tourism, and improve the stability of the tourism flow network. Through the in-depth analysis of the generation and formation of inter-city tourism flow, it is necessary to strengthen the connection between the edge cities and the core areas, enhance the extensive cooperation in the fields of capital flow, material flow, information flow, etc., develop strengths and avoid weaknesses, and improve the strength and market competitiveness of urban tourism destinations in the overall network.

4) There are obvious agglomeration phenomena in China's urban tourism flow network, such as the Yangtze River Delta tourism group, the Pearl River Delta region, the Beijing-Tianjin-Hebei region tourism group and the Chengdu-Chongqing tourism group. High-ranking cities continue to lead the development of regional tourism industry, consolidate and strengthen the leading position of cities in the region. Finally, it is necessary to build a new tourism spatial system [49]. On the basis of the development of the existing tourism city groups and the integration of the geographical distribution of each city in the network, the degree of contact of cohesive subgroups and tourism characteristics, new tourism city groups are to be constructed and new tourism spatical inner pattern is to be formed, so as to promote city tourism of China develop in the direction of diversity.

Social network analysis relies heavily on data sources, especially detailed information on the flow of tourism flows between cities. In this paper, there are still some problems in tourism flow: 1) the distinction between tourism flow and normal flow between cities is difficult to distinguish them completely on the basis of definition of "tourism". This is also an important problem encountered by many scholars in tourism related research. 2) Some popular tourist cities are ignored, such as Dali, a famous tourist city in China. However, due to the low level of urban informatization, the number of Weibo check-in cannot accurately represent the actual number of tourists. Nevertheless, Weibo check-in data is still an important way to obtain tourism research data [50]. Although we have revealed to some extent the spatial network characteristics of China's National Day Holiday tourist flow through Sina Weibo check-in data, the lack of representativeness of Weibo check-in data has always been the focus of scholars' attention to the study of tourist flow. Because most Weibo users are post-80s and post-90s, lacking statistics on the check-in data of elderly and young children, and female Weibo users are significantly higher than male Weibo users, the conclusion of this paper may be biased by these factors. Digital footprints with geographical labels overcome the limitations of traditional surveys. However, this is not a substitute for the traditional data collection method, but a supplement to the current data acquisition method. Obviously, the next step would be to integrate more travel data sources, such as obtaining photo data with geographical location label from Flicker, or obtaining travel blog data with geographical location from Qunar.com and statistics obtained from questionnaire survey, etc., which will make the research data more representative to analyze the spatial network structure of tourism flows [51]. This paper only analyzes the network structure of China's urban tourism flow, and focuses on the spatial evolution of the network structure in the following research. Besides, this paper only discusses the influence of urban economic development level on urban tourism flow, but the factors affecting tourism flow include not only urban infrastructure, accessibility of transportation and income level of urban population, but also urban spatial structure, urban hierarchy and urban spatial distribution. Comprehensive analysis of the factors affecting tourism flow is also the key research direction in the future. And this paper only analyzes the spatial network structure of the tourist flow in the main cities of China during the National Day holiday, and does not analyze the tourist flow laws of traditional festivals such as Spring Festival, Mid-Autumn Festival. Therefore, comparing the network structure of tourism flow in different Chinese holidays is also the main problem to be explored in the future.

VI. CONCLUSION

This paper proposes a framework of spatial behavior research of tourism flow based on digital footprint of tourism. On the basis of the theory of geography, the framework combines the traditional analysis method of Geography with the social network analysis method to analyze the spatial characteristics and network structure of intercity tourism flow. Regarding Weibo check-in data of 2018 National Day holiday as the research data source, this paper studies the tourism flow between typical cities in China through this research framework, drawing the following conclusions:

1) From the point of view of the degree of centrality of urban nodes, the tourism flow rating system of typical cities in China has obvious structural classification. Cities are divided into five grades, among which high-ranking urban nodes are mainly distributed in the eastern and central regions, while low-ranking urban nodes in the western and central underdeveloped regions.
2) Based on the study of the spatial characteristics of urban tourism network, it is found that China has formed the basic framework of diamond-shaped tourism flow network with Beijing, Shanghai, Guangzhou and Chengdu as its apexes, and each vertex has formed a city tourism group with its core in different regions, which highlights the characteristics of the complex interaction of selectivity and spatial proximity of China's tourism network.
3) From the point of view of network node structure, Beijing, Shanghai, Hangzhou, Chongqing, Nanjing and other cities with developed economy or rich tourism resources play a core role in the tourism flow network, with high centrality, effective size and efficiency. These cities have obvious competitive advantages in tourism and have strong control over other cities. However, the indicators of underdeveloped economy or poor traffic accessibility nodes such as Lhasa, Yinchuan and Xining indicate that these cities are highly dependent on the core cities in the tourism flow network and need to further strengthen the tourism links with other cities.
4) From the overall structure of the network, tourism flow has a relatively stable network, but most of the tourism flow links are weak and small, so the tourism links between cities still need to be further strengthened. Based on the analysis of cohesive subgroups, it is found that the tourism density within the fifth subgroup is 1, indicating that the inner cities of the subgroup are most closely related to each other and the tourism network structure between cities is complete. From the core-edge analysis, there is a high network density between the nodes in the core area, while low network density in the edge area, showing that the division of the core area and the edge area is very significant. The core area mainly includes the eastern and southwest region. Finally, by analyzing the tourism relationship between the core-edge area, it is found that the core regional cities in China have significant ability to drive the development of the tourism industry in the edge regional cities.

ACKNOWLEDGMENT

The authors sincerely thank everyone who has provided comments and assistance to this article.

REFERENCES

[1] J. Coshall, "Spectral analysis of international tourism flows," *Ann. Tourism Res.*, vol. 27, no. 3, pp. 577–589, 2000, doi: 10.1016/S0160-7383(99)00081-X.

[2] B. Prideaux, "Factors affecting bilateral tourism flows," *Ann. Tourism Res.*, vol. 32, no. 3, pp. 780–801, 2005, doi: 10.1016/j.annals.2004.04.008.

[3] J. C. García-Palomares, J. Gutiérrez, and C. Mínguez, "Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS," *Appl. Geography*, vol. 63, pp. 408–417, Sep. 2015, doi: 10.1016/j.apgeog.2015.08.002.

[4] Y. Gao, C. Ye, X. Zhong, L. Wu, and Y. Liu, "Extracting spatial patterns of intercity tourist movements from online travel blogs," *Sustainability*, vol. 11, no. 13, p. 3526, Jun. 2019, doi: 10.3390/su11133526.

[5] H. Shao, Y. Zhang, and W. Li, "Extraction and analysis of city's tourism districts based on social media data," *Comput., Environ. Urban Syst.*, vol. 65, pp. 66–78, Sep. 2017, doi: 10.1016/j.compenvurbsys.2017.04.010.

[6] S. M. Allen, M. J. Chorley, G. B. Colombo, and R. M. Whitaker, "Opportunistic social dissemination of micro-blogs," *Ad Hoc Netw.*, vol. 10, no. 8, pp. 1570–1585, Nov. 2012, doi: 10.1016/j.adhoc.2011.04.012.

[7] D. Edwards and T. Griffin, "Understanding tourists' spatial behaviour: GPS tracking as an aid to sustainable destination management," *J. Sustain. Tourism*, vol. 21, no. 4, pp. 580–595, May 2013, doi: 10.1080/09669582.2013.776063.

[8] A. Bertone and D. Burghardt, "A survey on visual analytics for the spatio-temporal exploration of microblogging content," *J. Geovisualization Spatial Anal.*, vol. 1, nos. 1–2, pp. 1–21, Dec. 2017, doi: 10.1007/s41651-017-0002-6.

[9] P. Du, X. Bai, K. Tan, Z. Xue, A. Samat, J. Xia, E. Li, H. Su, and W. Liu, "Advances of four machine learning methods for spatial data handling: A review," *J. Geovisualization Spatial Anal.*, vol. 4, no. 1, p. 13, Jun. 2020, doi: 10.1007/s41651-020-00048-5.

[10] C. CooperR Baggio and N. Scott, *Network Analysis and Tourism: From Theory to Practice*. Clevedon, U.K.: Channel View Publications, 2008, p. 259.

[11] A. Scuttari, M. D. Lucia, and U. Martini, "Integrated planning for sustainable tourism and mobility. A tourism traffic analysis in Italy's south tyrol region," *J. Sustain. Tourism*, vol. 21, no. 4, pp. 614–637, May 2013, doi: 10.1080/09669582.2013.786083.

[12] H.-Y. Shih, "Network characteristics of drive tourism destinations: An application of network analysis in tourism," *Tourism Manage.*, vol. 27, no. 5, pp. 1029–1039, Oct. 2006, doi: 10.1016/j.tourman.2005.08.002.

[13] B. Mckercher and G. Lau, "Movement patterns of tourists within a destination," *Tourism Geographies*, vol. 10, no. 3, pp. 355–374, Jul. 2008, doi: 10.1080/14616680802236352.

[14] J. Connell and S. J. Page, "Exploring the spatial patterns of car-based tourist travel in Loch Lomond and Trossachs National Park, Scotland," *Tourism Manage.*, vol. 29, no. 3, pp. 561–580, Jun. 2008, doi: 10.1016/j.tourman.2007.03.019.

[15] Y. Zhang, J.-H. Xu, and P.-J. Zhuang, "The spatial relationship of tourist distribution in Chinese cities," *Tourism Geographies*, vol. 13, no. 1, pp. 75–90, Feb. 2011, doi: 10.1080/14616688.2010.529931.

[16] F. Girardin, F. Calabrese, F. D. Fiore, C. Ratti, and J. Blat, "Digital footprinting: Uncovering tourists with user-generated content," *IEEE Pervas. Comput.*, vol. 7, no. 4, pp. 36–43, Oct. 2008.

[17] Y. Li, Q. Sun, X. Ji, L. Xu, C. Lu, and Y. Zhao, "Defining the boundaries of urban built-up area based on taxi trajectories: A case study of Beijing," *J. Geovisualization Spatial Anal.*, vol. 4, no. 1, 2020, Art. no. 8, doi: 10.1007/s41651-020-00047-6.

[18] F. Girardin, F. D. Fiore, C. Ratti, and J. Blat, "Leveraging explicitly disclosed location information to understand tourist dynamics: A case study," *J. Location Based Services*, vol. 2, no. 1, pp. 41–56, 2008, doi: 10.1080/17489720802261138.

[19] N. Mou, Y. Zheng, T. Makkonen, T. Yang, J. J. Tang, and Y. Song, "Tourists' digital footprint: The spatial patterns of tourist flows in Qingdao, China," *Tourism Manage.*, vol. 81, Dec. 2020, Art. no. 104151, doi: 10.1016/j.tourman.2020.104151.

[20] H. ChungN Chung and Y. Nam, "A social network analysis of tourist movement patterns in blogs: Korean backpackers in Europe," *Sustainability-Basel*, vol. 9, no. 12, p. 2251, 2017, doi: 10.3390/su9122251.

[21] I. Koerbitz and A. Hubmann-Haidvogel, "Tracing tourists by their digital footprints," *J. Travel Res.*, vol. 55, no. 5, pp. 566–573, 2014, doi: 10.1177/0047287514563985.

[22] X. Su, B. Spierings, P. Hooimeijer, and S. Scheider, "Where day trippers and tourists go: Comparing the spatio-temporal distribution of Mainland Chinese visitors in Hong Kong using Weibo data," *Asia Pacific J. Tourism Res.*, vol. 25, no. 5, pp. 505–523, 2020, doi: 10.1080/10941665.2020.1741409.

[23] T. Wang, L. Wang, and Z.-Z. Ning, "Spatial pattern of tourist attractions and its influencing factors in China," *J. Spatial Sci.*, vol. 65, no. 2, pp. 327–344, May 2020, doi: 10.1080/14498596.2018.1494058.

[24] X. Yang, C. Gu, and Q. Wang, "Study on the driving force of tourist flows," *Geographical Res.*, vol. 30, no. 1, pp. 23–36, 2011.

[25] A. Papatheodorou, "Why people travel to different places," *Ann. Tourism Res.*, vol. 28, no. 1, pp. 164–179, 2001, doi: 10.1016/S0160-7383(00)00014-1.

[26] E. Marrocu and R. Paci, "They arrive with new information. Tourism flows and production efficiency in the European regions," *Tourism Manage.*, vol. 32, no. 4, pp. 750–758, Aug. 2011, doi: 10.1016/j.tourman.2010.06.010.

[27] J. I. L. Miguéns and J. F. F. Mendes, "Travel and tourism: Into a complex network," *Phys. A, Stat. Mech. Appl.*, vol. 387, no. 12, pp. 2963–2971, May 2008, doi: 10.1016/j.physa.2008.01.058.

[28] S. Su, C. Wan, Y. Hu, and Z. Cai, "Characterizing geographical preferences of international tourists and the local influential factors in China using geo-tagged photos on social media," *Appl. Geography*, vol. 73, pp. 26–37, Aug. 2016, doi: 10.1016/j.apgeog.2016.06.001.

[29] Y. Liu, Z. Sui, C. Kang, and Y. Gao, "Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data," *PLoS ONE*, vol. 9, no. 1, Jan. 2014, Art. no. e86026, doi: 10.1371/journal.pone.0086026.

[30] M. D. Sarrión-Gavilán, M. D. Benítez-Márquez, and E. O. Mora-Rangel, "Spatial distribution of tourism supply in andalusia," *Tourism Manage. Perspect.*, vol. 15, pp. 29–45, Jul. 2015, doi: 10.1016/j.tmp.2015.03.008.

[31] N. Mou, R. Yuan, T. Yang, H. Zhang, J. Tang, and T. Makkonen, "Exploring spatio-temporal changes of city inbound tourism flow: The case of Shanghai, China," *Tourism Manage.*, vol. 76, Feb. 2020, Art. no. 103955, doi: 10.1016/j.tourman.2019.103955.

[32] Y. Guo, J. Zhang, Y. Yang, and H. Zhang, "Modeling the fluctuation patterns of monthly inbound tourist flows to China: A complex network approach," *Asia Pacific J. Tourism Res.*, vol. 20, no. 8, pp. 942–953, Aug. 2015, doi: 10.1080/10941665.2014.948024.

[33] J. Bowden, "A cross-national analysis of international tourist flows in China," *Tourism Geographies*, vol. 5, no. 3, pp. 257–279, Jan. 2003, doi: 10.1080/14616680309711.

[34] K. Pavlovich, "The evolution and transformation of a tourism destination network: The Waitomo Caves, New Zealand," *Tourism Manage.*, vol. 24, no. 2, pp. 203–216, 2003, doi: 10.1016/S0261-5177(02)00056-0.

[35] M. Santana-Gallego, F. J. Ledesma-Rodríguez, and J. V. Pérez-Rodríguez, "International trade and tourism flows: An extension of the gravity model," *Econ. Model.*, vol. 52, pp. 1026–1033, Jan. 2016, doi: 10.1016/j.econmod.2015.10.043.

[36] D. R. Vining, "The rank-size rule in the absence of growth," *J. Urban Econ.*, vol. 4, no. 1, pp. 15–29, 1977, doi: 10.1016/0094-1190(77)90027-4.

[37] B. Liu, S. Huang, and H. Fu, "An application of network analysis on tourist attractions: The case of xinjiang, China," *Tourism Manage.*, vol. 58, pp. 132–141, Feb. 2017, doi: 10.1016/j.tourman.2016.10.009.

[38] X. Wang, S. Huang, T. Zou, and H. Yan, "Effects of the high speed rail network on China's regional tourism development," *Tourism Manage. Perspect.*, vol. 1, pp. 34–38, Jan. 2012, doi: 10.1016/j.tmp.2011.10.001.

[39] X. Y. Leung, F. Wang, B. Wu, B. Bai, K. A. Stahura, and Z. Xie, "A social network analysis of overseas tourist movement patterns in Beijing: The impact of the olympic games," *Int. J. Tourism Res.*, vol. 14, no. 5, pp. 469–484, Sep. 2012, doi: 10.1002/jtr.876.

[40] X. Wu, Z. Huang, X. Peng, Y. Chen, and Y. Liu, "Building a spatially-embedded network of tourism hotspots from geotagged social media data," *IEEE Access*, vol. 6, pp. 21945–21955, 2018, doi: 10.1109/ACCESS.2018.2828032.

[41] A. Lindner and C. S. Pitombo, "Sequential Gaussian simulation as a promising tool in travel demand modeling," *J. Geovisualization Spatial Anal.*, vol. 3, no. 2, Dec. 2019, Art. no. 15, doi: 10.1007/s41651-019-0038-x.

[42] X. Su, B. Spierings, M. Dijst, and Z. Tong, "Analysing trends in the spatio-temporal behaviour patterns of mainland Chinese tourists and residents in Hong Kong based on Weibo data," *Current Issues Tourism*, vol. 23, no. 12, pp. 1542–1558, 2019, doi: 10.1080/13683500.2019.1645096.

[43] J. Qin, C. Song, M. Tang, Y. Zhang, and J. Wang, "Exploring the spatial characteristics of inbound tourist flows in China using geo-tagged photos," *Sustainability*, vol. 11, no. 20, p. 5822, Oct. 2019, doi: 10.3390/su11205822.

[44] J. Wang, H. Mo, F. Wang, and F. Jin, "Exploring the network structure and nodal centrality of China's air transport network: A complex network approach," *J. Transp. Geography*, vol. 19, no. 4, pp. 712–721, Jul. 2011, doi: 10.1016/j.jtrangeo.2010.08.012.

[45] C. Jin, J. Cheng, and J. Xu, "Using user-generated content to explore the temporal heterogeneity in tourist mobility," *J. Travel Res.*, vol. 57, no. 6, pp. 779–791, Jul. 2018, doi: 10.1177/0047287517714906.

[46] J. Zhang, J. Du, Y. Zhou, S. Zhang, and B. Pan, "Spatial structure of tourist source areas for the naturally scenic sightseeing places: A case study of Jiuzhaigou," *Acta Geographica Sinica*, vol. 54, no. 4, pp. 71–78, 1999.

[47] Z. Wu, J. Liu, and J. Yuan, "Network analysis and evolutionary studies based on tourist flows of mainland residents's self-service traveling in Taiwan," *Tourism Tribune*, vol. 31, no. 10, pp. 113–121, 2016.

[48] S.-H. Lee, J.-Y. Choi, S.-H. Yoo, and Y.-G. Oh, "Evaluating spatial centrality for integrated tourism management in rural areas using GIS and network analysis," *Tourism Manage.*, vol. 34, pp. 14–24, Feb. 2013, doi: 10.1016/j.tourman.2012.03.005.

[49] Y. Yang and K. K. F. Wong, "Spatial distribution of tourist flows to China's cities," *Tourism Geographies*, vol. 15, no. 2, pp. 338–363, May 2013, doi: 10.1080/14616688.2012.675511.

[50] G. Akehurst, "User generated content: The use of blogs for tourism organisations and tourism consumers," *Service Bus.*, vol. 3, no. 1, pp. 51–61, Mar. 2009, doi: 10.1007/s11628-008-0054-2.

[51] Y. Zheng, N. Mou, L. Zhang, T. Makkonen, and T. Yang, "Chinese tourists in nordic countries: An analysis of spatio-temporal behavior using geo-located travel blog data," *Comput., Environ. Urban Syst.*, vol. 85, Jan. 2021, Art. no. 101561, doi: 10.1016/j.compenvurbsys.2020.101561.

ANJUN LI is currently pursuing the master's degree with the College of Geodesy and Geomatics, Shandong University of Science and Technology, China. His research interests focus on data mining, big data analysis, and transport geography.

NAIXIA MOU is currently a Professor with the College of Geodesy and Geomatics, Shandong University of Science and Technology, China. His main research interests are data mining, big data analysis, tourism, and transport geography.

LINGXIAN ZHANG is currently a Lab Technician with the College of Geodesy and Geomatics, Shandong University of Science and Technology, China. Her main research interests are geographical information systems and transport geography.

TENGFEI YANG is currently pursuing the Ph.D. degree with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests focus on natural language processing and spatio-temporal data mining.

WENBAO LIU is currently a Professor with the College of Geodesy and Geomatics, Shandong University of Science and Technology, China. His main research interests are data mining and big data analysis.

FENG LIU is currently a Lecturer with the College of Geodesy and Geomatics, Shandong University of Science and Technology, China. His main research interests are geographic information sharing, WebGIS, and technology service of engineering survey.