


Article

Exploring the Spatial Characteristics of Inbound Tourist Flows in China Using Geotagged Photos

Jing Qin ^{1,2}, Ci Song ^{3,4,*} , Mingdi Tang ^{1,2,*}, Youyin Zhang ⁵ and Jinwei Wang ^{1,2}

¹ School of Tourism Sciences, Beijing International Studies University, Beijing 100024, China; qinjing@bisu.edu.cn (J.Q.); wangjinwei@bisu.edu.cn (J.W.)

² Research Center for Beijing Tourism Development, Beijing 100024, China

³ State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

⁴ College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

⁵ Institute of Regional Tourism Planning and Development, China Tourism Academy, Beijing 100005, China; yyzhang@mct.gov.cn

* Correspondence: songc@reis.ac.cn (C.S.); mingditang@bisu.edu.cn (M.T.)

Received: 6 September 2019; Accepted: 16 October 2019; Published: 20 October 2019



Abstract: As important modern tourist destinations, cities play a critical role in developing agglomerated tourism elements and promoting urban life quality. An in-depth exploration of tourist flow patterns between destination cities can reflect the dynamic trends of the inbound tourist market. This is significant for the development of tourism markets and innovation in tourism products. To this end, photos with geographical and corresponding metadata covering the entire country from 2011 to 2017 are used to explore the spatial characteristics of China’s inbound tourist flow, the spatial patterns of tourist movement, and the tourist destination cities group based on data mining techniques, including the Markov chain, a frequent-pattern-mining algorithm, and a community detection algorithm. Our findings show that: (1) the strongest flow of inbound tourists is between Beijing and Shanghai. These two cities, along with Xi’an and Guilin, form a “double-triangle” framework, (2) the travel between emerging destination cities in Central and Western China have gradually become frequently selected itineraries, and, (3) based on the flow intensity, inbound tourist destination cities can be divided into nine groups. This study provides a valuable reference for the development of China’s inbound tourism market.

Keywords: Geotagged photos; inbound tourist flow; spatial characteristics; tourist transfer pattern; tourist destination cities group

1. Introduction

The core elements of tourist flow are the spatial movement of tourists and the consequent economic and social development [1]. Tourist flow connections between tourism origins and destinations, or among various tourism destinations, constitute a “nerve-center” in the tourism network [2]. Studying the movement patterns of inbound visitors between cities can provide the basis for designing international travel routes, formulating international marketing strategies, and developing regional tourism cooperation. Moreover, understanding the flow of international tourists is of great significance for improving tourism management, planning, and marketing methods of national and local agencies and administrative departments [3].

Using destination cities as the basic research unit is essential to study the spatial distribution and travel trajectories. However, because it is difficult to obtain city-level data on the spatial and temporal behavior of tourists, especially in developing countries [4], our ability to study the behavior

of tourists at the city level is significantly limited. To address the statistical data shortage, some studies have used questionnaires to analyze the tourist flow and diffusion routes among different tourist destinations [5–7]. Although this method can reflect tourist movement trajectories to some degree, it is limited when used to study the spatial and temporal characteristics of tourist behavior. Questionnaires are time-consuming and may provide only a small data sample. Using a limited random sample of responses for statistical analysis makes it difficult to accurately and comprehensively reflect the behavioral patterns of tourists [8–10]. Moreover, statistical data acquired through questionnaires show a significant scale dependency (for example, a fixed spatial scale such as a province, a city, or a scenic area is selected when designing questionnaires) [11,12]. Lastly, using questionnaire survey data to carry out city-level research may lead to biased results. Inbound tourists tend to focus on the scenic destinations themselves rather than their administrative locations. For example, visitors to Chengdu (city) who travel to Jiuzhaigou (scenic area) may not realize that this scenic area is actually located in another administrative district (the Ngawa Tibetan and Qiang Autonomous Prefecture), which might affect the analysis results. The results might show Chengdu as a major tourist destination while ignoring its function as an important tourist distribution hub. Therefore, obtaining accurate tourism behavior data and using geo-information technology for spatial expression are vital in the study of tourist flows.

The era of big data has produced large volumes of data with spatial and temporal attributes. Mobile phone data, vehicle trajectory data, and social media data provide a means to quantitatively understand our socio-economic environment [12]. Social media platforms such as Flickr and Panoramio contain millions of geotagged photographs uploaded by users. Photographs and their EXIF data, including shooting time and GPS location, can be downloaded through interfaces provided by the social media platforms. That data can be further processed using data-mining algorithms to generate tourist movement trajectories, which provides a huge dataset spanning a significant amount of time and delivering the level of precision needed to study tourist flow [11,13,14]. Many scholars have used geotagged photos to study problems related to tourism geography [15]. Some studies have focused on customized trip plans and travel recommendations [16–21], whereas others have explored popular tourist destination districts within cities [9,13,14]. However, most of these studies have emphasized small-scale tourist movement trajectories and patterns within cities or scenic areas [8,11,22–25]. Moreover, studies involving medium-scale or large-scale (urban or urban-agglomeration scale) are concentrated on the static analysis of spatial distribution of geotagged photos instead of analyzing tourist movement trajectories between cities [4,13,26,27]. Therefore, an in-depth city-level study of the spatial behavior and movement patterns of inbound tourists is necessary.

In order to get an insight into the inter-city characteristics of inbound tourist flows, a Markov chain model is used to describe the probability of transfer between cities and two data mining methods, including frequent pattern-mining and community detection, which are applied in this study. Markov chain is a “memorylessness” random process that undergoes transitions from one state to another [28]. It is widely used in many applications, such as stock predicting [29], human behavior analyzing [30], and weather forecasting [31,32]. Frequent pattern-mining is first proposed to explore the associated relationships between retail goods [33], and is extended to discover the spatial associations and spatial-temporal associations between geographical features [34,35]. Community detection provides a classical method to identify subgroups of a complex network with strong connections. It is key to understanding the structure of complex networks in virtual space [36] and, more importantly, in geographical space [37].

Based on the above methods, we carry out a city-level study of inbound tourist flows in China using prefectural level cities as the basic unit. First, we use network crawling technology to retrieve geotagged photos and their corresponding metadata from Flickr’s public Application Programming Interface (API). We extract data covering all of China from 2011 to 2017. Second, we use data-cleaning techniques to control the initial data quality and establish the basic database. Third, we explore the tourist movement trajectories between cities using the Markov chain technique and analyze the spatial

patterns of tourist movement with a frequent-pattern-mining algorithm. Lastly, we use a community detection algorithm to group cities based on the strength of their interconnections. The results of this study can provide scientific guidance for designing international tourist travel routes, formulating international marketing strategies, and developing regional tourism cooperation.

2. Data and Methods

2.1. Data and Processing

Geotagged photos with the corresponding metadata for China (latitudes 3.86°N–53.55°N and longitudes 73.66°E–135.05°E), not including Taiwan, Hong Kong, and Macao, shot between January 2011 and December 2017 were obtained from the Flickr API (provided by Flickr at www.flickr.com/services/api). The extracted metadata record includes the following fields: user ID, photo ID, user's country of origin, shoot time, upload time, shoot location (latitude and longitude), photo caption, photo description, and comments of other users. The dataset consists of 282,787 records covering a period of seven years. Users whose country of origin is China and null are excluded, leaving 80,278 records from 3680 tourists. Lastly, the geographic coordinates were converted into planar coordinates in a Universal Transverse Mercator (UTM) projection for the spatial calculation.

2.2. Methodology

2.2.1. Probability of Transfer between Cities

Using records of tourist movement between cities, we can model the tourist inflow/outflow for each city. Each record of movement is described by five fields, including <photo ID, user ID, time, longitude, and latitude>. The movement between cities can be represented by a "stationary discrete Markov chain" [28]. Let $A = \{a^1, a^2, \dots, a^m\}$ be defined as the collection of all cities and the travel route of tourist i for any given time series $t_1 < t_2 < \dots < t_k$ as $T_i = (a_i^{t_1}, a_i^{t_2}, \dots, a_i^{t_k}), a_i^t \in A$. Then, under the assumptions of the stationary discrete Markov chain theory, the probability of a tourist traveling from location $a_i^{t_{k-1}}$ to location $a_i^{t_k}$ is independent of the previous travel route and can be expressed by Equation (1):

$$P(a_i^{t_k} | a_i^{t_{k-1}}, a_i^{t_{k-2}}, \dots, a_i^{t_0}) = P(a_i^{t_k} | a_i^{t_{k-1}}) \quad (1)$$

Thus, the travel probability from city a^1 to city a^2 is simply the ratio between the number of tourists traveling from city a_1 to city a_2 and the total number of tourists departing city a_1 .

2.2.2. Mining Tourist Travel Patterns

The frequent movement patterns of tourists are analyzed using the classic Apriori method, which is a frequent-pattern-mining algorithm [38]. Given that the travel locations (cities) of all tourists have been defined as $A = \{a^1, a^2, \dots, a^m\}$, the travel trajectory data can be viewed as a transaction set, denoted as $D = \{I_1, I_2, \dots, I_n\}$. Furthermore, the movement trajectories of each tourist are recorded as $I_i = \{a_i^{t_{j_1}}, a_i^{t_{j_2}}, \dots, a_i^{t_{j_k}}\}, a_i^{t_j} \in A$, so that the location (city) $a_i^{t_j}$ of tourist i can be viewed as a single-transaction item. For a certain travel pattern $P \subset A$, if the movement trajectory of tourist i includes this pattern, then $P \subset I_i$ and the pattern has a support value of $s+1$. When the support value s exceeds a certain threshold s_0 , the travel pattern is designated as popular and is expressed using the following:

$$P = \langle \text{Shanghai} \rightarrow \text{Xi'an} \rightarrow \text{Beijing}, 3.38\% \rangle$$

Shanghai \rightarrow Xi'an \rightarrow Beijing is a popular travel pattern, found among the trajectories of 3.38% of tourists (i.e., tourist numbers choosing the travel pattern/total tourist numbers = 0.0338).

2.2.3. Detecting Tourist Destination City Groups

The grouping characteristics of destination cities can be detected based on the compactness of connectivity (number of travel flows) among cities by using a network community-partitioning algorithm. In this paper, we use the Combo algorithm [39] to identify communities among tourist destination cities in China. This community-detection algorithm provides a way of dividing a network into several communities to achieve the goal of maximizing connections within the same community and minimizing connections across different communities. The performance of this algorithm is based on a predefined modularity Q [40], which is widely used in many applications of community detection algorithms. For a network composed of n vertices and m edges, the definition of the modularity Q for an instance of community-partitioning C is represented as:

$$Q = \frac{1}{2m} \sum_{uv} [A_{uv} - \frac{k_u k_v}{2m}] \delta(C_u, C_v) \quad (2)$$

In this case, A_{uv} is used to represent whether two vertices, u and v , are connected (1 for connected, otherwise 0). $\delta(C_u, C_v)$ is a binary function used to filter the vertices belonging to the same communities. If vertices u and v belong to the same community, $\delta(C_u, C_v)$ is equal to one. Otherwise, it is zero. k_u and k_v stand for the degree of vertices u and v , respectively. Based on this concept, the details of the algorithm applied in this study to detect communities of tourist-destination cities are as follows.

The detection method [40] measures the modularity of network partitions to find the optimal partitioning scheme. Optimal partitioning maximizes connections within the same communities and minimizes connections across different communities, which results in dense community intra-connections and sparse community inter-connections. Detailed steps of Combo algorithm are described as follows:

The network data, which consists of n vertices and m edges, is the major input. Every edge has a weight, reflecting the strength of the relationship between the two vertices it connects. In this paper, tourist destination cities are the vertices, tourist flow is represented by the edges, and the number of tourists moving between two cities is represented by the edge weight. By default, all vertices are initialized to belong to the same community.

Step 1: For each tourist flow (edge), calculate the highest positive ΔQ (i.e., the highest difference in *modularity* before and after the movement) by moving each node from its initial community to a target community and search for optimal community partitioning. (That is, the outcome of each node being redistributed into a different community is iteratively tested).

Step 2: When the optimal partitioning is determined, re-distribute the vertices into each target community.

Step 3: After updating community vertices, calculate the highest positive ΔQ by moving each node to different communities to find the optimal community-partitioning scheme.

Step 4: Test whether the current community-partitioning scheme is optimal. If it meets the preset criteria (Q is large enough or the iteration number has reached a certain threshold), the loop ends. Otherwise, steps 2–3 should be repeated. A threshold can be set in this step to ensure that the modularity score of the optimal network partition is as high as possible, which guarantees the qualities of the detection results.

The final output is a group of communities.

3. Results

3.1. Spatial Distribution of Geotagged Photos

The locations of the geotagged photos were mapped using ArcGIS10.3 software based on their coordinates. Popular destination regions for inbound tourists, which were obtained using kernel density analysis, are shown in Figure 1. Cities with the highest aggregate number of geotagged photos

are distributed within three large regions, which include the Beijing-Tianjin-Hebei Region (where Beijing is the regional central city), the Yangtze River Delta Zone (where Shanghai, Suzhou, and Hangzhou are regional central cities), and the Pearl River Delta Zone (where Guangzhou and Shenzhen are regional central cities). These regions are China's political and economic centers, and constitute the core of China's cultural heritage. Secondary concentration regions are primarily centered on the cities of Xi'an and Chengdu. Moreover, small-scale concentrations are also found around Guilin (Guangxi Zhuang Autonomous Region), Lijiang (Yunnan Province), Lhasa (Tibet Autonomous Region), and Dalian (Liaoning Province). These observations show that inbound tourists seek out China's unique natural landscapes, cultural heritage, and economic development centers.

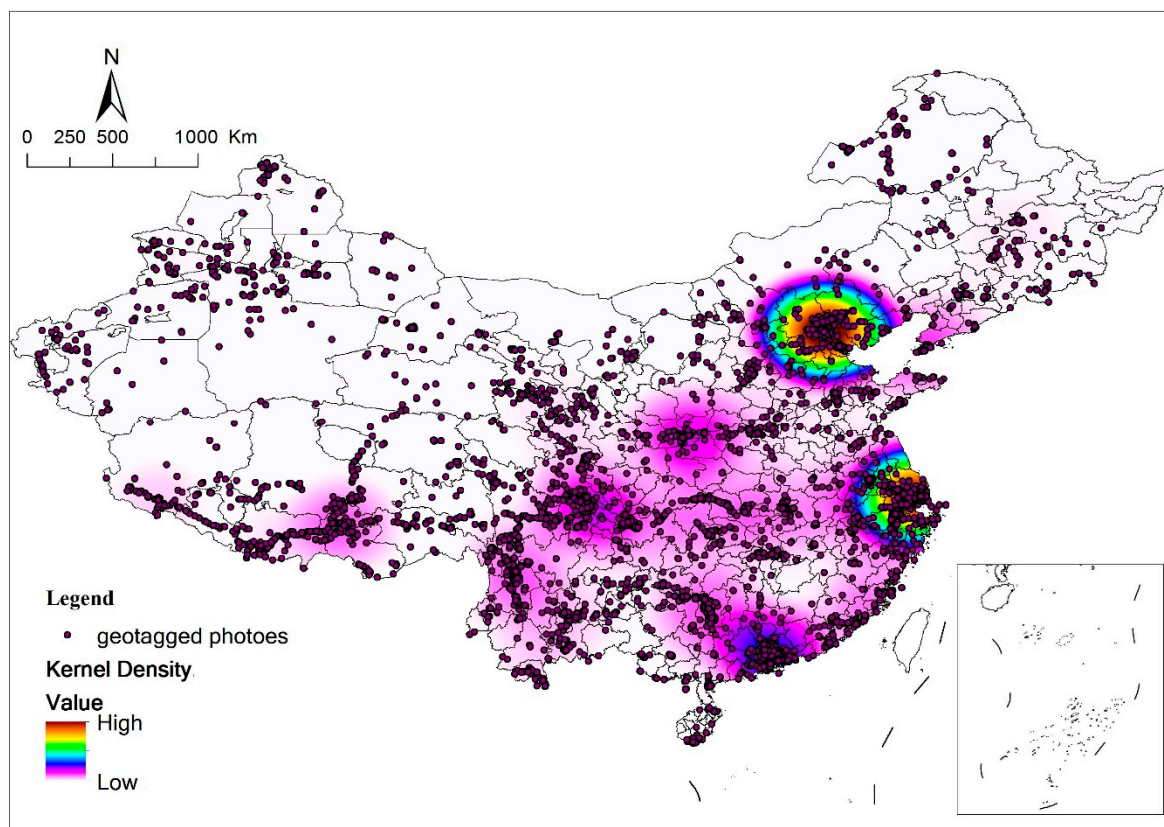


Figure 1. Spatial distribution of geotagged photos from 2011 to 2017.

To further explore the cities' popularity, we calculate the number of inbound tourists in each city. The top 20 most visited cities are shown in Table 1. We then classify those cities into three tiers using the Nature Breaks method. Natural Breaks classes are based on natural groupings inherent in the data. Class breaks are identified that best group similar values and that maximize the differences between classes [41]. Beijing and Shanghai are in the first tier since the number of inbound tourists in Beijing and Shanghai is much larger than in any other city. The second tier is constituted of traditional tourist destinations such as Xi'an, Guilin, Guangzhou, Suzhou, and Hangzhou. The third tier consists of newly emerging destination cities, which are generally in the central and western regions of China such as Lijiang, Lhasa, Kunming, Diqing, Dali, and Aba. In this case, the beautiful natural scenery and unique ethnic cultures have attracted growing attention from foreign visitors. In particular, Lijiang and Lhasa are already among the top 10 destination cities. Overall, traditional destination cities are still the most popular among inbound tourists. However, with the development of airports, high-speed railway, and other transportation services, the newly emerging destination cities have gradually become more popular.

Table 1. Popularity rankings of tourist destination cities.

Popularity Rank	City Name	Percentage (%)	Tier
1	Beijing	38.26	First
2	Shanghai	31.41	
3	Xi'an	7.39	Second
4	Guangzhou	7.04	
5	GuiLin	6.17	
6	Shenzhen	5.87	
7	Hangzhou	5.38	
8	Suzhou	5.00	
9	Chengdu	4.81	
10	Lijiang	3.02	
11	Lhasa	2.83	
12	Chongqing	2.61	
13	Nanjing	2.53	
14	Kunming	2.45	
15	Xiamen	2.15	
16	Dali bai autonomous prefecture	2.04	
17	Ngawa Tibetan and Qiang autonomous prefecture	1.93	
18	Diqing Tibetan autonomous prefecture	1.74	
19	Fuzhou	1.60	
20	Wuhan	1.60	

3.2. Spatial Structure of Inbound Tourist Flows

The various different routes chosen by tourists converge into tourist flows. The convergence of flows ultimately forms a tourist transition network. The major tourist flow connections between tourism nodes constitute the main spatial characteristics of the tourist flows network. In addition, the number of small tourist flow cities connected to the node cities with high tourist flows determine the tourist collecting and distributing ability of cities with high tourist flows. Therefore, when exploring the spatial characteristics of tourist flow, we first calculate the tourist flow connections between cities and describe the spatial framework formed by major inbound flows. Next, we discuss the characteristics of tourist movements between major tourist-node cities. Lastly, we identify urban agglomeration regions with high spatial concentrations of inbound tourists to examine the travel patterns in these regions.

3.2.1. Overall Spatial Structure of Inbound Tourist Flow

Since the travel trajectories of inbound tourists have been formed by connecting locations in succession for each tourist, we create the flow map of travels by aggregating the number of flows between prefectural cities in China. As shown in Figure 2, travel flows are divided into five categories by travel numbers, which are rendered in different colors. Beijing and Shanghai have the highest inbound tourist flows. The bi-directional flow between them constitutes the backbone of China's inbound tourism network. These two cities, together with Xi'an and Guilin, constitute the "double-triangle" inbound tourist flow framework. The remaining significant flow connections of inbound tourists are established through the above framework.

- The strong intra-regional flows form three major and two minor tourist city agglomeration regions. The three major regions are Beijing-Tianjin-Hebei Region, the Yangtze River Delta, and the Pearl River Delta, which represent China's economically developed urban agglomerations. Higher inbound tourist flows between Diqing-Lijiang-Kunming-Dali in Yunnan Province and Lhasa-Shigatse-Shannan in the Tibet Autonomous Region have created two relatively higher-flow triangles in the central and western regions of China and formed increasingly popular destination agglomeration regions.
- In the eastern region of China, Beijing and Shanghai are the main tourist flow concentration and diffusion cities. In addition, Guangzhou, Shenzhen, Suzhou, and Hangzhou are also important

nodes of the inbound tourist flow network in this region. In the central and western regions, the most important tourist flow concentration and diffusion nodes are in the traditional tourism cities like Xi'an and Guilin. Moreover, Chongqing, Chengdu, Yichang, Lijiang, Kunming, and Lhasa are gradually becoming important transfer cities for inbound tourists.

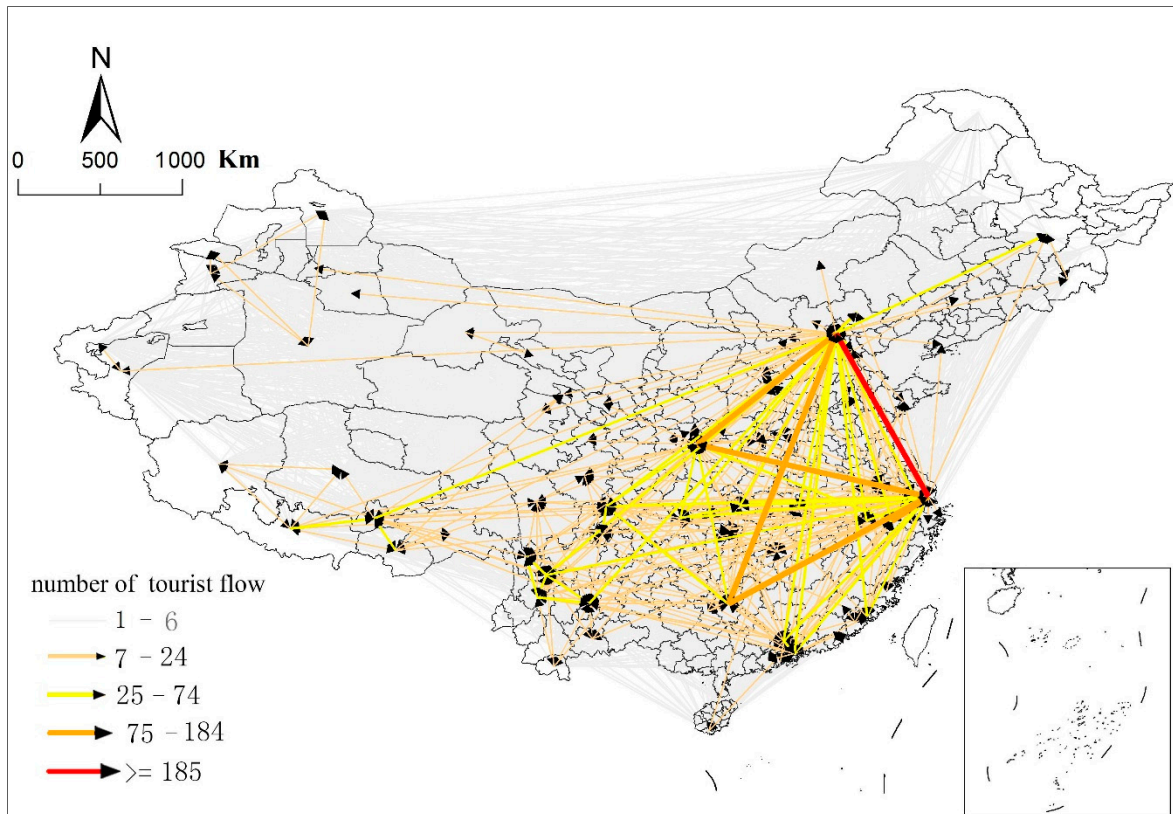


Figure 2. Overall spatial structure of inbound tourist flow.

3.2.2. Inbound Tourist Flow between Major Cities

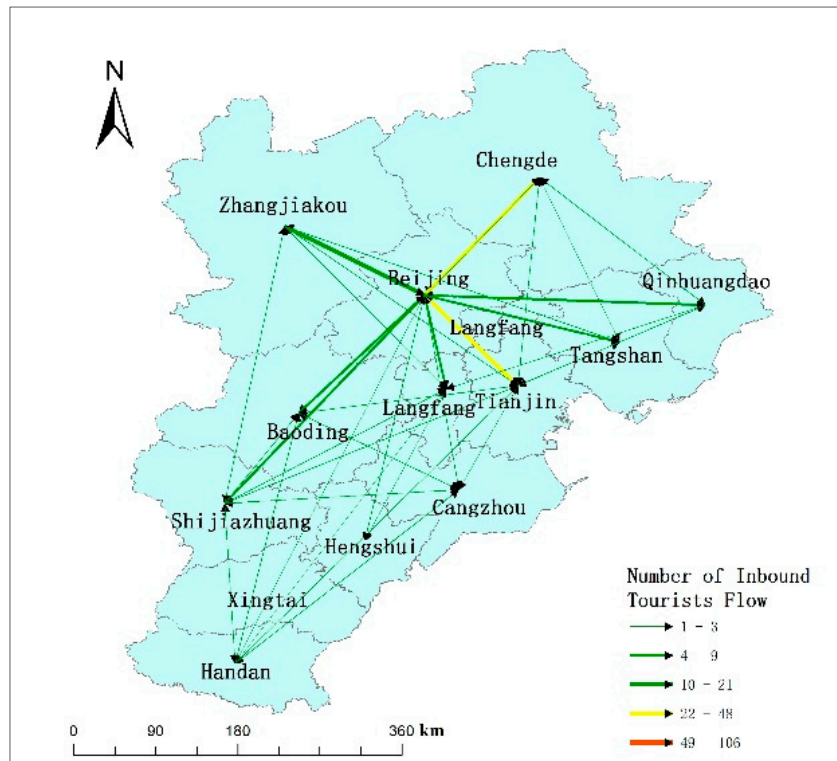
To further explore tourist flow, we calculate the number of tourist flows and the transition probability between each combination of two cities. The 20 cities with the highest number of tourist flows are shown in Table 2. Beijing and Shanghai have the highest north-south tourist flow and transfer probability, which means these two large cities are China's leading inbound tourist destinations. Therefore, cooperation and communication on inbound tourism between these two cities is essential. Other city connections with relatively high tourist flows and transfer probabilities include Beijing-Xi'an, Suzhou-Shanghai, Hangzhou-Shanghai, Shanghai-Hangzhou, and Xi'an-Beijing. The high flows indicate the close relationships between the two ancient capitals of Beijing and Xi'an, and between the central cities of the Yangtze River Delta, Shanghai, Hangzhou, and Suzhou. The high tourist flow and low transfer probability between Shanghai and Jiaxing, Nanjing, Guilin, and Shenzhen show that Shanghai is an active tourist flow distribution center with a strong ability to radiate outward. The high tourist flows from Lijiang to Dali and Diqing and from Lhasa to Shigatse indicate close tourism cooperation and development in the western regions.

Table 2. Tourist transition between destination cities.

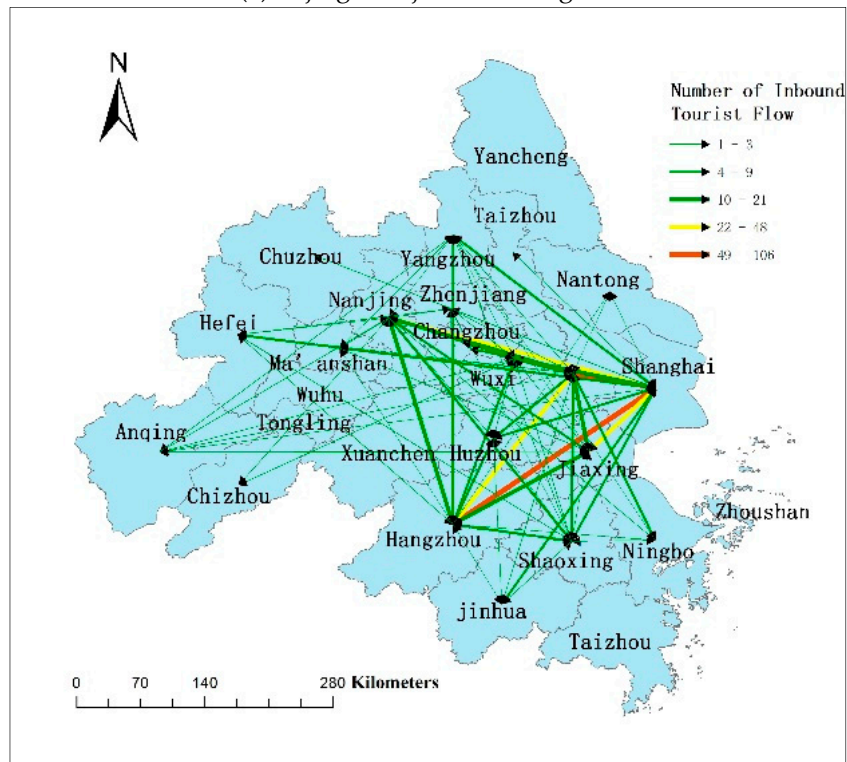
Rank	Start City	Destination City	Transition Probability	Number of Flow
1	Beijing	Shanghai	0.236	178
2	Shanghai	Beijing	0.241	163
3	Beijing	Xi'an	0.129	97
4	Suzhou	Shanghai	0.521	85
5	Hangzhou	Shanghai	0.394	69
6	Shanghai	Hangzhou	0.101	68
7	Xi'an	Beijing	0.257	67
8	Shanghai	Suzhou	0.096	65
9	Xi'an	Shanghai	0.138	36
10	Jianxing	Shanghai	0.614	35
11	Shanghai	Nanjing	0.050	34
12	Shanghai	Jiaying	0.049	33
13	Dali bai autonomous prefecture	Lijiang	0.400	32
14	Guilin	Shanghai	0.200	32
15	Lhasa	Shikatsé	0.307	31
16	Lijiang	Diqing Tibetan autonomous prefecture	0.292	31
17	Shanghai	Shenzhen	0.046	31
18	Shanghai	Guilin	0.044	30
19	Nanjing	Shanghai	0.318	28
20	Diqing Tibetan autonomous prefecture	Lijiang	0.368	25

3.2.3. Tourist Flow in Major Urban Agglomerations

Tourist flows show a degree of distance attenuation whereby cities that are further away are less likely to be visited. Hence, a number of tourist flow aggregation regions are generated based on the regional central city. Beijing-Tianjin-Hebei Region, Yangtze River Delta, and Pearl River Delta are the three biggest regions, which all have concentrations of strong intra-regional inbound tourist flow. In order to understand flows within these regions better, the tourist flows within China's three large economic zones are analyzed and the results are shown in Figure 3. In the Beijing-Tianjin-Hebei region, the dominant inbound tourist flow is bi-directional between Beijing and Tianjin and between Beijing and Chengde (a city famous for the emperor's Summer Palace of the Qing Dynasty). By contrast, Shijiazhuang, which is the provincial capital of Hebei, has a relatively low-density tourist flow with Beijing and Tianjin. In this region, the tourist flow shows a radial pattern emanating from Beijing and the tourist flow between other cities is rather sparse. In the Yangtze River Delta, the inbound tourist flow is spread between four cities including Shanghai, Hangzhou, Nanjing, and Suzhou, which shows a relatively balanced spatial distribution of the tourist flow network. Lastly, in the Pearl River Delta, the inbound tourist flow is relatively low and mainly occurs as a bi-directional flow between Guangzhou and Shenzhen. The diffusion of tourist flow from Guangzhou-Shenzhen to places outside of the region far exceeds the intra-regional tourist flow.

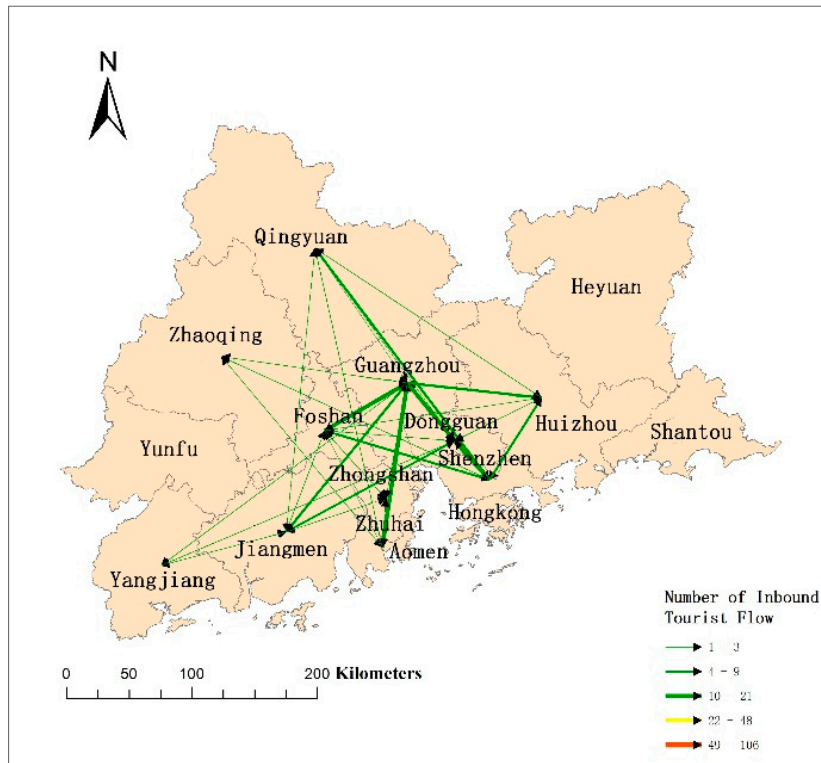


(a) Beijing-Tianjin-Hebei Region.



(b) Yangtze River Delta Region.

Figure 3. Cont.



(c) Pearl River Delta Region.

Figure 3. Distribution of tourist flow within urban agglomerations.

3.3. Spatial Pattern of Tourist Movement

Based on the tourist flows obtained from our results, the tourist’s trajectory patterns were further analyzed. Figure 4 shows the number of inbound tourists based on the number of cities they visited during a single trip. As the number of cities increases, the number of tourists decreases. Most tourists visit 1–4 cities during a single tour, and few travel to more than five cities. Because the number of visitors to five or more cities is so small, no statistical pattern could be detected. The trajectory patterns with four or less destination cities are analyzed using the frequent-pattern-mining Apriori algorithm.

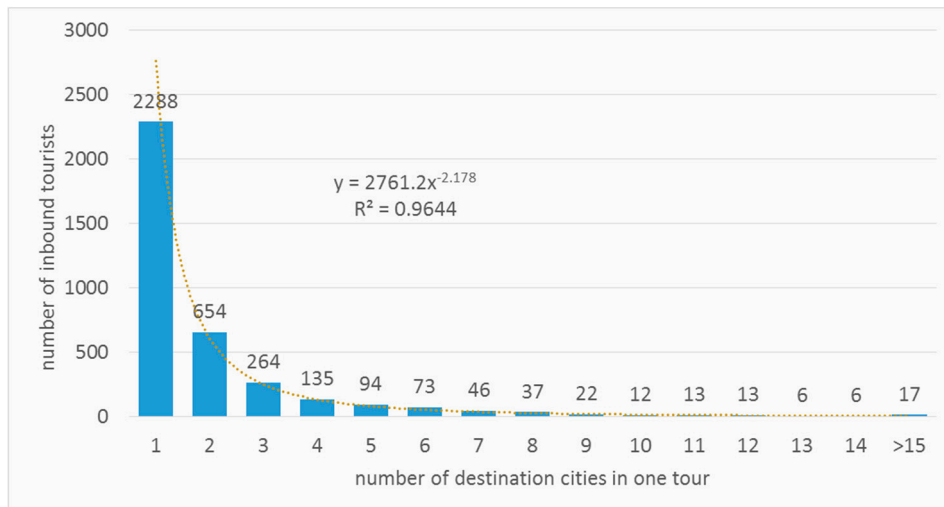


Figure 4. Distribution of tourist flow within urban agglomerations.

The movement pattern of tourists between cities reflects their itineraries and the role of each city in the inbound tourism network. The results of the frequent-pattern-mining algorithm applied to travel routes consisting of different destination cities are shown in Tables 3–5. The two destination travel patterns mostly occur in traditional destinations such as Shanghai, Beijing, Xi’an, Suzhou, Hangzhou, and Guilin, with the Beijing→Shanghai and Shanghai→Beijing route being the two most frequently occurring patterns. In terms of travel direction, the most frequent pattern is from other cities such as Chengdu, Suzhou, and Hangzhou to Beijing or Shanghai. The most frequently selected travel pattern of three destinations is Shanghai→Xi’an→Beijing, which indicates that these three cities are the nucleus cities of China’s inbound tourism. In addition, other routes favored by inbound tourists were from Suzhou, Hangzhou, and Guilin to Beijing and Shanghai. Apart from traditional destination cities, a travel route from Chengdu–Yichang to Beijing or Shanghai also occurred frequently, which demonstrates the important effect of Yangtze River cruise tourism on China’s inbound tourism market and the potential for Yichang to become a popular international tourist destination. Moreover, the Shigatse→Shannan→Lhasa route is among the top 20 high-frequency travel patterns, which indicates that Tibet’s unique natural scenery and cultural heritage are highly attractive for international visitors. The remaining frequently occurring patterns involve destination cities connected to Beijing, Shanghai, and Xi’an, which illustrates the important role of these three cities as popular destinations and distribution hubs for China’s inbound tourist flow. There are only six frequent travel patterns of four destinations. In addition to Beijing, Shanghai, and Xi’an, the cities of Suzhou and Hangzhou are the most attractive cities in Eastern China. Chongqing and Guilin are the most attractive nodes in the middle of China and in Western China.

Table 3. Travel pattern of two destination cities.

City 1	City 2	Percentage (%)
Beijing	Shanghai	18.67
Shanghai	Beijing	14.87
Xi’an	Beijing	7.33
Suzhou	Shanghai	6.75
Hangzhou	Shanghai	5.53
Shanghai	Xi’an	5.46
Guilin	Beijing	4.38
Guilin	Shanghai	4.02
Chengdu	Beijing	3.52
Chengdu	Shanghai	3.23
Suzhou	Beijing	2.87
Shenzhen	Beijing	2.80
Nanjing	Shanghai	2.80
Shenzhen	Shanghai	2.66
Hangzhou	Beijing	2.51
Chongqing	Shanghai	2.51
Guangzhou	Shanghai	2.44
Chongqing	Beijing	2.30
Guangzhou	Beijing	2.30
Beijing	Tianjin	2.16

Table 4. Tourist travel pattern of three destination cities.

City 1	City 2	City 3	Percentage (%)
Shanghai	Xi’an	Beijing	3.38
Suzhou	Shanghai	Beijing	2.23
Guilin	Shanghai	Beijing	1.80
Hangzhou	Shanghai	Beijing	1.58

Table 4. Cont.

City 1	City 2	City 3	Percentage (%)
Suzhou	Hangzhou	Shanghai	1.51
Guilin	Shanghai	Xi'an	1.51
Chongqin	Shanghai	Beijing	1.36
Chengdu	Shanghai	Beijing	1.29
Shenzhen	Shanghai	Beijing	1.29
Suzhou	Shanghai	Xi'an	1.29
Nanjing	Shanghai	Beijing	1.15
Chongqin	Yichang	Shanghai	1.15
Chengdu	Xi'an	Beijing	1.08
Chongqin	Xi'an	Beijing	1.08
Guilin	Xi'an	Beijing	1.08
Shikatse	Shannan	Lhasa	1.08
Hangzhou	Shanghai	Xi'an	1.01
Chendu	Shanghai	Xi'an	1.01
Chongqing	Shanghai	Xi'an	1.01
Chongqing	Yichang	Beijing	0.93

Table 5. Tourist travel pattern of four destination cities.

City 1	City 2	City 3	City 4	Percentage (%)
Guilin	Shanghai	Xi'an	Beijing	2.30
Chongqing	Yichang	Shanghai	Beijing	1.44
Chongqing	Shanghai	Xi'an	Beijing	1.72
Suzhou	Hangzhou	Shanghai	Beijing	1.44
Suzhou	Shanghai	Xi'an	Beijing	1.65
Hangzhou	Shanghai	Xi'an	Beijing	1.65

3.4. Identifying Groups of Inbound Tourist Destination Cities

Based on the intensity connection between cities of inbound tourist flow networks, the destination cities were divided into different tourism groups using the Combo community detection algorithm, as shown in Figure 5. The groupings are: the Pan-Bohai Sea group (including the Bohai Rim, Eastern Inner Mongolia, and the south part of Shaanxi Province), the southeast coastal group (including the Pearl River Delta, Xiamen in Fujian Province, and the southern Zhejiang Province), the Yangtze River Delta group, the Northeast China group, the Northwest China group (including Xinjiang, Qinghai, Ningxia, and Western Inner Mongolia), the Sichuan-Yunnan group, the Chongqing-Hubei-Hunan group, the Tibet group, and the Guizhou group.

The groups indicate that China's inbound tourist flow network is related not only to natural and cultural factors, but also to regional economies. In each of the groups, the tourist flow connections between cities within the groups are relatively intense, which indicates a strong correlation between tourism resources and regional economic development. Therefore, promoting regional cooperation can facilitate further development of integrated urban tourism. It is noteworthy that the cities in Hainan Island show evidence of weak connections, which might be related to the vacation travel preferences of international tourists, and need further exploration.

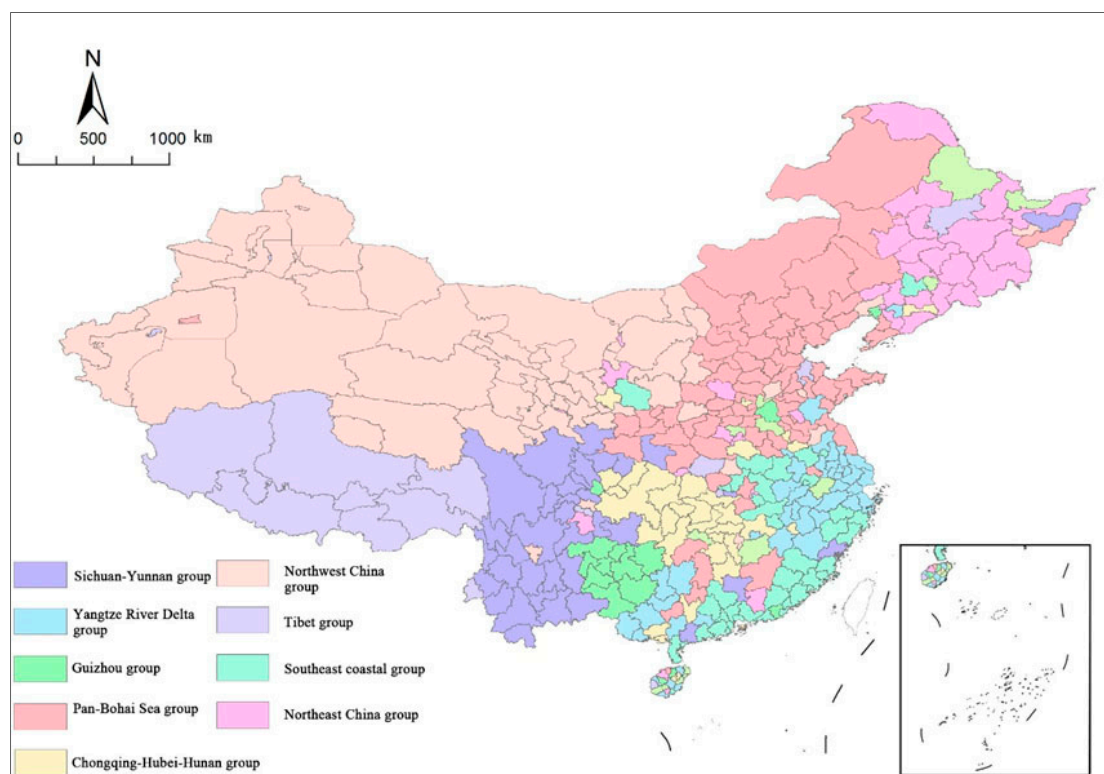


Figure 5. Tourist destination cities group detection.

4. Discussion

4.1. Distributions and Mechanisms of Tourist Flows

The results show that the new emerging hot tourist destination cities in Central and Western China, such as Lijiang, Lhasa, Dali, and aba, have gradually become popular. This may be due to the construction of high-speed railway networks and airports in Central and Western China, which are improving the accessibility of these cities to inbound tourists. Considering the intra-regional tourist flow, the Yangtze River delta urban agglomeration has more connections among cities with high flow, which forms an inbound tourism network with close connections. However, the connection between the Pearl River delta and the Beijing-Tianjin-Hebei urban agglomeration is relatively weak. The application of the frequent pattern mining algorithm showed that inbound tourists generally enter China directly through their destination city and leave from Beijing or Shanghai. The 20 most frequently selected itineraries include Chongqing-Yichang to Beijing or Chongqing-Yichang to Shanghai. This shows the significant effect of the Yangtze River cruise on the tourism market, which demonstrates that, sometimes, tourism products can promote the development of destination cities. Moreover, the Shigatse→Shannan→Lhasa pattern is also among the 20 most frequently selected itineraries, which shows that inbound tourists are attracted to Tibet's unique natural environment and cultural heritage. From the analysis of the group division of tourism cities, we can see that China's inbound tourist flow network is not only related to natural and cultural factors, but is also related to regional economic relevance. Breaking the cooperation of cross-provincial administrative divisions will benefit the further cooperation of intra-group urban tourism. The state and regional tourism departments should correctly guide cooperation among city groups, integrate regional resources, and promote the joint development of regional tourism. Different city groups should pay attention to the characteristic and differentiated development of tourism.

4.2. Scale Effects of Tourist Flows

The spatial flow characteristics of tourist flow can be summarized into three different scale flow patterns, according to the behavior characteristics of tourists. The first scale flow pattern is the cross regional scale. This is according to the strength sorting for the twin cities of Beijing - Shanghai, Beijing-Shanghai and Xi'an, Guilin, Beijing-Shanghai and Guangzhou-Shenzhen, and the Chengdu-Chongqing urban group. Corresponding to this, the inbound market, which has been stagnating, overseas marketing and tourism investment are basically concentrated in the traditional tourist hot spots, which lack endogenous development power. The second is the regional scale. The strong tourist flow within the region forms an active group of "three big and two small" tourism cities. The "three big" economically developed city clusters on the east coast form an active flow within the region by providing better public services and business travel opportunities. In Western Yunnan and Tibet, "two small" urban agglomeration with resources' advantages and integration formed the regional tourism cooperation. Through government-led transportation infrastructure improvement and differentiation of product design and marketing, a breakthrough to the overseas market was realized. The third point is the scale of city nodes. The frequent-pattern-mining results show that most tourists only choose one city as their destination in one trip. When inbound tourists arrange multiple cities in one trip, they generally choose a traditional tourist city as the entry and choose Beijing and Shanghai, which are two major cities in China, as the exit. This study also found that the design of featured tourism products can also contribute to the emergence of inbound tourism node cities. For example, the cruise products promote the development of emerging inbound tourism cities such as the Three Gorges Yangtze. In addition, the study classified different tourism groups according to the closeness of the tourist flow in the above-mentioned inbound cities. The results were generally consistent with the urban agglomeration construction vigorously promoted at the national level during the 13th five-year plan period, which could be a booster for regional synergy.

4.3. Policy Suggestion

Inbound tourism plays an important role in enhancing the soft power and international influence of a country (region). Based on the above research results, this paper proposes the following suggestions: (1) The government needs to pay attention to the formulation and implementation of overseas marketing strategies, actively expand overseas markets, and, on the basis of consolidating the existing traditional tourism cities, vigorously support emerging or potential inbound tourism destinations, especially for central and western cities, to make use of resource advantages to ensure sustainable development. (2) The government needs to give full play to the role of the center/key hub city. On the one hand, the government needs to deeply explore and improve the cultural connotation and quality of urban tourism products themselves. On the other hand, the government needs to actively guide tourists to transfer and spread to the next level of destination. (3) The government needs to carry out institutional innovation and technological innovation to further optimize the product system of policies, environment, infrastructure, and the natural and cultural resources that are conducive to the development of inbound tourism, in order to enhance China's international tourism competitiveness and conduct spatial drainage through a product portfolio. (4) The government needs to break the cross-provincial administrative division of the cooperation, integrate regional resources, guide the characteristics of tourism products and differential development, and promote further regional cooperation.

5. Conclusions

Understanding the spatial distribution and transfer trajectory characteristics of inbound tourists between cities, plays an important role in improving tourism management and upgrading the planning and marketing of the national tourism management authority and local tourism management departments. To this end, using cities as the basic research unit, this study explores the spatial

characteristics of China's inbound tourist flow, and identifies the spatial patterns and network structures of tourism movements based on comprehensive data mining methods, including the Markov random chain, frequent-pattern-mining, and community detection. The results provide scientific guidance for tourism managers to design international tourism routes, formulate international marketing strategies, and develop regional tourism cooperation.

The spatial relationship of strong tourist flows are mainly discussed to uncover the general characteristics and structures of tourism behaviors. Tourist flows can objectively reflect the prevailing trends in the international tourism market. However, based on the long tail theory, the individualized demands of weak tourist flows can help promote the innovation of tourism products, encourage friendly competition, and improve regional cooperation. These issues have great research significance and are worth further investigation. Moreover, this study takes the inbound tourist data as a whole to analyze the spatial characteristics of inbound tourist flow. However, travel behaviors vary among tourists from different countries and regions. Therefore, tourists from different countries should be analyzed in detail in future research. Lastly, transportation, tourism marketing, urban economic development, urban public support services, and other factors have a great impact on inbound tourists. The quantitative analysis of the influence of these factors should also be a focus of subsequent research.

Author Contributions: J.Q., designed and conceived this research; C.S. designed the experiments and analyzed the data; J.Q., C.S. and M.T. wrote the manuscript; Y.Z. and J.W. reviewed this research. All authors read and approved the final manuscript.

Funding: The National Natural Science Foundation of China (No. 41801168), the Philosophy and Social Science Foundation of Beijing, No.17JDGLB004, and the Key Research Program of Frontier Science, Chinese Academy of Sciences (Grant No. QYZDY-SSW-DQC007) funded this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mckercher, B.; Lew, A.A. Tourist flows and the spatial distribution of tourists. In *A Companion to Tourism*; Blackwell Publishing Ltd.: Hoboken, NJ, USA, 2004.
2. Boniface, B.; Cooper, C. *The Geography of Travel and Tourism*; Butterworth-Heinemann: Oxford, UK, 1994; pp. 1–6.
3. Li, X.; Meng, F.; Uysal, M. Spatial pattern of tourist flows among the Asia-Pacific countries: An examination over a decade. *Asia Pac. J. Tour. Res.* **2008**, *13*, 229–243. [[CrossRef](#)]
4. Su, S.; Wan, C.; Hu, Y.; Cai, Z. Characterizing geographical preferences of international tourists and the local influential factors in China using geo-tagged photos on social media. *Appl. Geogr.* **2016**, *73*, 26–37. [[CrossRef](#)]
5. Liu, B.; Huang, S.; Fu, H. An application of network analysis on tourist attractions: The case of Xinjiang, China. *Tour. Manag.* **2017**, *58*, 132–141. [[CrossRef](#)]
6. Wu, J. Distribution of inbound foreigner tourist flows and its property and structure characters in China. *J. Arid Land Res. Environ.* **2014**, *28*, 177–182.
7. Yang, Y.; Fik, T.; Zhang, J. Modeling sequential tourist flows: Where is the next destination? *Ann. Tour. Res.* **2013**, *43*, 297–320. [[CrossRef](#)]
8. Vu, H.Q.; Li, G.; Law, R.; Ye, B.H. Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tour. Manag.* **2015**, *46*, 222–232. [[CrossRef](#)]
9. Hu, Y.; Gao, S.; Janowicz, K.; Yu, B.; Li, W.; Prasad, S. Extracting and understanding urban areas of interest using geotagged photos. *Comput. Environ. Urban Syst.* **2015**, *54*, 240–254. [[CrossRef](#)]
10. Xu, Z.; Zhang, J. Antecedents and consequences of place attachment: A comparison of Chinese and western urban tourists in Hangzhou, China. *J. Destin. Mark. Manag.* **2016**, *5*, 86–96. [[CrossRef](#)]
11. Zheng, Y.T.; Zha, Z.J.; Chua, T.S. Mining travel patterns from geo-tagged photos. *ACM Trans. Intell. Syst. Technol.* **2012**, *3*, 1–18. [[CrossRef](#)]
12. Liu, Y.; Liu, X.; Gao, S.; Gong, L.; Kang, C.; Zhi, Y.; Chi, G.; Shi, L. Social sensing: A new approach to understanding our socioeconomic environments. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 512–530. [[CrossRef](#)]

13. García-Palomares, J.C.; Gutierrez, C.; Mínguez, C. Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Appl. Geogr.* **2015**, *63*, 408–417. [[CrossRef](#)]
14. Paldino, S.; Bojic, I.; Sobolevsky, S.; Ratti, C.; Gonzalez, M.C. Urban magnetism through the lens of geo-tagged photography. *EPJ Data Sci.* **2015**, *4*, 5. [[CrossRef](#)]
15. Brabyn, L.; Mark, D.M. Using viewsheds, GIS, and a landscape classification to tag landscape photographs. *Appl. Geogr.* **2011**, *31*, 1115–1122. [[CrossRef](#)]
16. Sun, X.; Huang, Z.; Peng, X.; Chen, Y.; Liu, Y. Building a model-based personalised recommendation approach for tourist attractions from geotagged social media data. *Int. J. Digit. Earth* **2019**, *12*, 661–678. [[CrossRef](#)]
17. Xia, P.; Zhou, H. A novel popular tourist attraction discovering approach based on geo-tagged social media big data. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 216.
18. Jiang, K.; Yin, H.; Wang, P.; Yu, N. Learning from contextual information of geo-tagged web photos to rank personalized tourism attractions. *Neurocomputing* **2013**, *119*, 17–25. [[CrossRef](#)]
19. Memon, I.; Chen, J.; Majid, A.; Lv, M.; Hussain, I.; Chen, G. Travel recommendation using geo-tagged photos in social media for tourist. *Wirel. Pers. Commun.* **2015**, *80*, 1347–1362. [[CrossRef](#)]
20. Sun, Y.; Fan, H.; Bakillah, M.; Zipf, A. Road-based travel recommendation using geo-tagged images. *Comput. Environ. Urban Syst.* **2015**, *53*, 110–122. [[CrossRef](#)]
21. Zhou, X.; Xu, C.; Kimmons, B. Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Comput. Environ. Urban Syst.* **2015**, *54*, 144–153. [[CrossRef](#)]
22. Wu, X.; Huang, Z.; Peng, X.; Chen, Y.; Liu, Y. Building a spatially-embedded network of tourism hotspots from geotagged social media data. *IEEE Access* **2018**, *6*, 21945–21955. [[CrossRef](#)]
23. Orsi, F.; Geneletti, D. Using geotagged photographs and GIS analysis to estimate visitor flows in natural areas. *J. Nat. Conserv.* **2013**, *21*, 359–368. [[CrossRef](#)]
24. Chua, A.; Servillo, L.; Marcheggiani, E.; Vande Moere, A. Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tour. Manag.* **2016**, *57*, 295–310. [[CrossRef](#)]
25. Qin, J.; Li, L.; Tang, M.; Sun, Y.; Song, X. Exploring the spatial characteristics of Beijing inbound tourist flow based on geotagged photos. *Acta Geogr. Sin.* **2018**, *73*, 1556–1570.
26. Shao, H.; Zhang, Y.; Li, W. Extraction and analysis of city's tourism districts based on social media data. *Comput. Environ. Urban Syst.* **2017**, *65*, 66–78. [[CrossRef](#)]
27. Yang, Y.; Wong, K.F. Spatial distribution of tourist flows to China's cities. *Tour. Geogr.* **2013**, *15*, 338–363. [[CrossRef](#)]
28. Xia, J.; Zeephongsekul, P.; Arrowsmith, C. Modelling spatio-temporal movement of tourists using finite Markov chains. *Math. Comput. Simul.* **2009**, *79*, 1544–1553. [[CrossRef](#)]
29. Wang, Y.F.; Cheng, S.; Hsu, M.H. Incorporating the markov chain concept into fuzzy stochastic prediction of stock indexes. *Appl. Soft Comput.* **2010**, *10*, 613–617. [[CrossRef](#)]
30. Qiao, Y.; Si, Z.; Zhang, Y.; Abdesslem, F.B.; Zhang, X.; Yang, J. A hybrid markov-based model for human mobility prediction. *Neurocomputing* **2018**, *278*, 99–109. [[CrossRef](#)]
31. Tseng, Y.T.; Kawashima, S.; Kobayashi, S.; Takeuchi, S.; Nakamura, K. Forecasting the seasonal pollen index by using a hidden Markov model combining meteorological and biological factors. *Sci. Total Environ.* **2019**. [[CrossRef](#)]
32. Ali, M.; Deo, R.C.; Downs, N.J.; Maraseni, T. Multi-stage hybridized online sequential extreme learning machine integrated with Markov Chain Monte Carlo copula-Bat algorithm for rainfall forecasting. *Atmos. Res.* **2018**, *213*, 450–464. [[CrossRef](#)]
33. Yun, U. Mining lossless closed frequent patterns with weight constraints. *Knowl. Based Syst.* **2017**, *20*, 86–97. [[CrossRef](#)]
34. Lee, A.J.; Chen, Y.A.; Ip, W.C. Mining frequent trajectory patterns in spatial-temporal databases. *Inf. Sci.* **2009**, *179*, 2218–2231. [[CrossRef](#)]
35. Shaw, A.A.; Gopalan, N.P. Finding longest frequent trajectory of dynamic objects using association approaches. *Intell. Data Anal.* **2014**, *18*, 637–651. [[CrossRef](#)]
36. Lancichinetti, A.; Fortunato, S. Community detection algorithms: A comparative analysis. *Phys. Rev. E* **2009**, *80*, 056117. [[CrossRef](#)] [[PubMed](#)]
37. Zhong, C.; Arisona, S.M.; Huang, X.; Batty, M.; Schmitt, G. Detecting the dynamics of urban structure through spatial network analysis. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 2178–2199. [[CrossRef](#)]

38. Agrawal, R.; Srikant, R. Fast algorithms for mining association rules. In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), Santiago, Chile, 12–15 September 1994; pp. 487–499.
39. Sobolevsky, S.; Campari, R.; Belyi, A.; Ratti, C. General optimization technique for high-quality community detection in complex networks. *Phys. Rev. E* **2014**, *90*, 012811. [[CrossRef](#)]
40. Newman, M.E. Detecting community structure in networks. *Eur. Phys. J. B* **2004**, *38*, 321–330. [[CrossRef](#)]
41. Chen, J.; Yang, S.T.; Li, H.W.; Zhang, B.; Lv, J.R. Research on geographical environment unit division based on the method of natural breaks (jenks). *ISPR Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *XL-4/W3*, 47–50. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).