




Article

Extracting Spatial Patterns of Intercity Tourist Movements from Online Travel Blogs

Yong Gao ^{*}, Chao Ye , Xiang Zhong, Lun Wu and Yu Liu 

Institute of Remote Sensing and Geographic Information System, School of Earth and Space Sciences, Peking University, Beijing 100871, China

* Correspondence: gaoyong@pku.edu.cn; Tel.: +86-136-1116-1215

Received: 17 May 2019; Accepted: 21 June 2019; Published: 27 June 2019



Abstract: Spatial patterns of tourist mobility are important for tourism management and planning. A large number of traveler-generated content accumulated on the internet provide a unique opportunity for revealing comprehensive spatial patterns of tourist movements. Instead of concentrating on a single city or attraction in previous research, this work investigates the intercity travel flows extracted from the online travel blogs in China from 2012 to 2016. The descriptive statistics of travel flows are first analyzed. The distribution of travel volume is found to satisfy the power-law distribution. Based on the intercity travel flows, a network structure is then constructed to investigate tourism interactions between cities. After four communities and 14 sub-communities being detected from the network, a tourism spatial layout with regional agglomeration effects are recognized. This research concludes that distance is essential in determining tourist movements based on a spatial interaction model. Intercity travel flows decline with distance under a power-law function. These results reveal the spatial patterns of tourist movements at an intercity scale. It will be helpful for arranging tourism resources, predicting tourist flows, and maintaining sustainable tourism.

Keywords: spatial pattern; tourist movement; complex network; spatial interaction

1. Introduction

The tourism market in the world has been booming in recent years. The consequent influence on the environment drives the focus of sustainable tourism. Tourism mobility is an important component of sustainable strategies for tourism [1,2]. The destination choices, travel trips and transportation in tourists' mobility are essential for effective policy interventions to address the impacts of tourism on the environment and destinations [3,4]. Supported by emerging information technologies, people enjoy sharing their travel experiences at tourism websites. Many high-quality travel blogs and reviews that have accumulated online provide new approaches and abundant resources for investigating tourism mobility. By extracting information from tourists' experiences, their travel flows can be reconstructed. The tourist movements are not only the direct manifestation of their behavior and perception, but also the interaction between travelers and the attractions. An in-depth understanding of the inherent mechanism and spatial patterns will benefit tourism planning and sustainable development.

There has been research focused on tourist movements. The concept of tourism flow was first proposed by Mercer [5] and Rajotte [6]. They concluded that the spatial scale of tourism flows was related to the leisure time of tourists. Distance is found as another essential factor to determine travel activities. Using questionnaires of self-driving travelers in Nanning, Liu et al. found that all spatial characteristics of the travel flows were consistent with the general distance decay pattern, but some attractive attractions caused irregular changes [7]. Tourist experience is determined by various factors [8]. What affects the selection of tourist destinations, such as travelers' age, cognition psychology, price, spatial configuration and urban soundscape, were also investigated in detail [9–15].

Benefiting from big data in recent years, researchers have gained deeper insight into spatial activities of tourists. GPS traces, transit smart card data, social media, and other online contents generated by tourists have become the most popular and convenient resources. By integrating GPS technologies and surveys, movement patterns of various tourists were found in a city or around a tourist attraction [16–18]. It was found from their mobile phone positioning data that the distribution of tourist volume satisfied the attenuation effect of distance [19]. Survey questionnaires and travelers' active participation (in-vehicle GPS data and transit smart card data) have improved the precision of travel models [20]. Social media, tourism websites, and other online platforms provide numerous opportunities for tourists to present their travel experience. People also create their expectations about the tourism destination by trusting received social media contents [21]. These online user-generated contents (UGC) contain rich information for discovering tourist movement patterns. Wise et al. assessed UGC using an interpretative framework when the Facebook page "See You in Iran" is used to promote the tourism of Iran [22]. Hu et al. crawled tourists' tweets with geo-tags, and then applied DBSCAN and network analysis methods to detect tourist movement patterns in New York [23]. Ahani et al. developed a method to predict spa hotel segmentation and travel choice by applying machine learning approaches based on online reviews and ratings [24]. By collecting information from an open tourism web service, the temporal heterogeneity in intracity tourist movements were explored, and the power law of distance decay of tourist mobility was confirmed in Nanjing, China [25]. From posts on the NAVER blog in Korea, it was found that London, Paris, Venezia, and Firenze were key cities where Korean backpackers tended to enter Europe [26]. Tourist trajectories were extracted from Flickr's geo-tagged photos and then motif tourist mobility patterns were detected [27]. Wu et al. introduced a tourism hotspot network approach to investigate travel patterns from social media data for tourism resources planning [28]. Moreover, Travel topics or sentiments were discovered from online traveler-generated content, such as Flickr's geo-tagged photos [29,30], tweets [31], or multi-source travelogues [32]. Based on various data sources, travelers' profiles and movement patterns were depicted, tourism market segmentation and travel choice were predicted, and finally the tourism was promoted. Multi-source data provide power for tourism researches and applications.

Most of the previous works focused on individual travel mobility at small scales for attraction descriptions or tourism recommendations. The investigation of tourist movements is mainly confined to a single city or attraction. However, tourists usually travel more often among cities within their country. They travel from the city where they live to the destination city, constructing a flow between the two cities. Intercity tourism is more valuable for the national tourism market. Furthermore, collective intercity travels imply specific tourist mobility at a large spatial scale, which is different to intracity movements. Intercity tourism involves more long-distance transportation, accommodation, entertainment, shopping and so on. These are all primary elements of sustainable tourism. Characteristics of the departure and destination cities, travel modes and travel distances affect the activities of tourists which in turn make impacts on the local environment and economy. Revealing the spatial arrangement and patterns of intercity travels will provide the local and national governments with necessary insights into the tourism mobility. It can help planners assess the attractiveness and carrying capacity of tourism destinations and optimize tourism policies based on collective travel movements. Transportation can also be improved to reduce environmental impacts according to the tourist distributions and travel distances, and thus maintaining the sustainable development.

Under the framework of social sensing [33], online travelogues are utilized to extract intercity tourism flows. These tourist-generated contents provide a new perspective to investigate tourist spatial behaviors which are different from what were found in other data sources. Disciplines of statistical physics, complex network approaches, and spatial interaction theory are integrated to provide the theoretical basis and methods for investigating the spatial structures, interactions and patterns of travel mobilities [34–36]. By exploring intercity tourism flows, city-level movement patterns of tourists are revealed from multiple perspectives and the community structure of tourist cities are discovered

in multi-scale. These knowledge can contribute to tourism resources arrangement, tourist flows prediction, and tourism recommendation.

2. Data

Baidu Travel (<https://lvyou.baidu.com/>) is one of the most popular tourism websites in China to share travel experiences through blogs. For each trip, the origin city and the destination city described in a blog can be extracted to the form with triple attributes (user, origin city, destination city). Each travel entry represents one movement from one city to another. Domestic tourism is different from the overseas because of the tourism motives and travel modes. To simplify the problem, only the travels in mainland China were considered. The travel entries whose departures or destinations are beyond the research area were filtered out. Finally, 1,105,928 travel entries of 72,999 users from 2012 to 2016 in China were collected. The valid travel flows between two cities in mainland China are represented as bright red lines in Figure 1. The brighter the color in a location, the more flows are from and to the corresponding city.

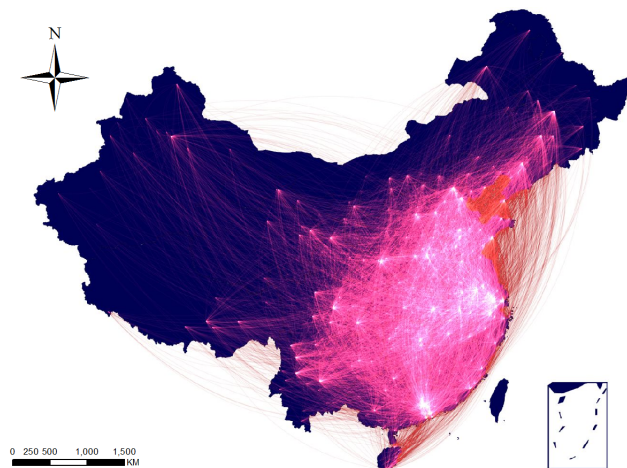


Figure 1. Intercity travel flows extracted from Baidu blogs.

Note that the amount of users cannot exactly represent the tourist distribution in cities due to their uneven distribution in different cities. Penetration, which is proposed for human mobility in Twitter [37], was employed for the representativeness of users in a city. The travel penetration was defined as the proportion of users to the total population of a city. As expected, the more developed the economy of a city, the greater the penetration was. It was similar to the popularity diffusion of social media such as Weibo and Twitter. For the representativeness, the cities with fewer than 20 users or less than 0.001% penetration were filtered out. Finally, 259 cities were retained.

3. Methods

Cities are connected by a large number of travel flows to form a network structure. This structure contains rich information about the relationship between cities, revealing the topological property and spatial arrangement of tourism movements. At the same time, the flows and structures are distributed in space, so spatial effects, especially distance, have impacts on them. Therefore, quantitative, structural and spatial factors were the three primary aspects of investigating tourism movements. Statistical methods were first utilized to explore the distributions and disparities of tourists, travel inflows and outflows. Then, a complex network structure of intercity travels was constructed based on the travel origin and destination pairs to discover the topological and structural properties of flow patterns. Finally, the distance decay of intercity tourism interaction was modeled to investigate the spatial effects on tourist mobility.

3.1. Statistical Analysis

Statistical analysis was used to investigate the distributions of travels described in travel blogs, thus discovering the macroscopic patterns of collective tourist mobility between cities. The correlation between the effective users and the rank of a city was calculated by Equation (1) using log-log transformations,

$$P_r = P_1 r^{-q}, \quad (1)$$

where r represents the volume rank of a city, P_r is the user volume of the r -th city, P_1 is the theoretical value of the top-ranked city, and q represents the degree of decline trend in the city's user volume with its rank increase. In the same way, the accumulative travels of all users in a city against its rank were fitted by log-log transformations as well. Equation (1) verifies whether the correlations satisfy the rank-size law that is subjected to a power law distribution. It characterizes the disparity that a small number of cities contribute or attract most tourists and most cities have a small number of tourists.

Additionally, the average gyration radius of all users in each city was calculated to measure the collective travel pattern. The gyration radius indicates the average movement extent of all travelers from a city. The larger the gyration radius is, the longer distance the preferred travels are.

3.2. Complex Network Analysis

Complex network analysis was used to analyze human mobility or behavior to reveal spatial structures or interactions. Two cities are considered to interact with each other if users travel from one to the other. All cities are connected by travel flows to construct a network structure in which a vertex denotes a city and an edge denotes the interaction relationship. Then, the intercity interaction characteristics were investigated by complex network methods.

Generally, a city interaction network of tourism is an undirected weighted graph represented by a triple $G = (V, E, W)$. A vertex in the vertex set V denotes a city. The edge set E is the subset of $V \times V$. An edge in E is represented as $E_{ij} = (V_i, V_j, W_{ij})$, where V_i and V_j are two vertices in V connected by the edge, and $W_{ij} \in W$ is the weight of the edge. W_{ij} is defined by Equation (2),

$$W_{ij} = F_{ij} + F_{ji}, \quad (2)$$

where F_{ij} represents the travel flows from city i to city j . The number of vertices is $N = |V|$ and the number of edges is $M = |E|$.

The structural characteristics of the city interaction network were evaluated by statistical parameters including centrality, small-world property, degree distribution and assortativeness. Degree distribution can be described by the distribution function $P(k)$, which is the probability that the degree of a random node is exactly k . In the proposed network of intercity travel flows, however, the degree spectrum is discrete with $P(k) = 0$ at some k s, so the cumulative degree distribution function [38], which is the probability distribution of vertices with degrees no less than k , is employed instead here as Equation (3),

$$P(k) = \sum_{k'=k}^{\infty} P(k'). \quad (3)$$

If the degree distribution satisfies the power law $P(k) \propto k^{-\gamma}$, the cumulative distribution follows the power law with an exponent of $\gamma - 1$. If the degree distribution satisfies an exponential distribution which is $P(k) \propto e^{-\gamma*k}$ where $\gamma > 0$ is a constant, the cumulative distribution also satisfies the exponential distribution with the same exponent.

Network assortativity describes the tendencies of nodes in a network to connect to nodes with the same degree, measured by assortativity coefficient Γ and $k_{nn}(k)$. The assortativity coefficient Γ is represented by Equation (4):

$$\Gamma = \frac{|E|^{-1} \sum_i j_i k_i - [|E|^{-1} \sum_i \frac{1}{2} (j_i + k_i)]^2}{|E|^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - [|E|^{-1} \sum_i \frac{1}{2} (j_i + k_i)]^2} \quad (4)$$

where j and k represent the degrees of nodes at both ends of the edge, respectively. The range of Γ is $[-1, 1]$. If $\Gamma > 0$, the network is assortative, i.e., the nodes with larger degrees tend to connect to the nodes with larger degrees in the network. On the contrary, the network is disassortative, indicating that the nodes with larger degrees tend to connect to the nodes with smaller degrees. $k_{nn}(k)$ is the average degree of the adjacent nodes of the nodes with a degree of k , as calculated by Equation (5):

$$k_{nn}(k) = \sum_{k'} k' p(k'|k) \quad (5)$$

If $k_{nn}(k) \propto k^{-\mu}$, $\mu > 0$, is satisfied for any k , the network is disassortative and μ is disassortativity index. The disassortativity of a network can be proved by Γ and μ .

The city interaction network was partitioned by a community detection method to find clustering structures. Then, the spin-glass model based on modularity optimization [39] was employed to investigate the community structure of the city interaction network.

3.3. Spatial Interaction Model

Intercity interactions are described by travel flows of tourists, in which some cities are primarily determined as sources with more outflows and less inflows and others are sinks with more inflows and less outflows. The interactions reveal not only the travel patterns of each city but also the relations between cities. The former were investigated by incoming or outgoing flows of a city as the travel origin or destination, respectively. The latter were evaluated by the interaction model that measures distance decay.

Intercity travel patterns were discovered from the similarity of travel flows between cities. Let F_{ij} denote the frequency of travels from city i to city j , and N is the number of cities, then the outflows of city i are represented as a vector $outflow_i = \langle F_{i1}, F_{i2}, \dots, F_{iN} \rangle$, and its inflows are $inflow_i = \langle F_{1i}, F_{2i}, \dots, F_{Ni} \rangle$. The spatial distributions of the outflows or inflows indicate the choice patterns of tourist destinations from a city or tourist origins to a city, respectively. Pearson's correlation coefficient was employed to measure the tourist similarity of two cities, as Equation (4),

$$R_{ij} = \frac{\sum_t (F_{it} - \bar{F}_i)(F_{jt} - \bar{F}_j)}{\sqrt{\sum_t (F_{it} - \bar{F}_i)^2} \sqrt{\sum_t (F_{jt} - \bar{F}_j)^2}}, t \neq i, j, \quad (6)$$

where R_{ij} is the correlation coefficient between city i and city j , F_{it} is the flow from i to t , and \bar{F}_i is the average flow of city i . From the perspective of the tourism source, R_{ij} measures the similarity of the destination choices of travelers from the two cities if both \bar{F}_i and \bar{F}_j are average outflows. From the perspective of a sink, on the contrary, R_{ij} measures the similarity of tourist origins to the two cities if both \bar{F}_i and \bar{F}_j are average inflows. Based on the similarity measurement, all cities were clustered by a hierarchical method to discover the spatial patterns of the tourist flow distributions.

The intercity interaction of tourist mobility was modeled by the gravity model [40]. Constrained by the distance decay effect, generally, the spatial interaction intensity of two cities was negatively

correlated with the distance. The negative power-law function was employed as the distance decay function to fit the following gravity model,

$$F_{ij} = k \frac{p_i^\alpha p_j^\beta}{d_{ij}^\gamma}, \quad (7)$$

where F_{ij} represents the flow frequency from city i to city j , p_i represents the volume of users traveling from city i , p_j represents the number of users traveling to city j , d_{ij} is the geographic distance between the two cities, and k is a constant factor. α , β and γ are parameters to be estimated, in which γ is the distance friction coefficient evaluating the impact of distance on flows.

This model was solved by linear regression of the inverse gravity model. Equation (7) is transformed to Equation (8) by logarithmic transformation as

$$\ln F_{ij} = \ln k + \alpha \ln p_i + \beta \ln p_j - \gamma \ln d_{ij}. \quad (8)$$

Finally, the optimal solution of the parameters α , β and γ are solved under a certain condition by the least square method.

4. Results

4.1. Statistical Features

4.1.1. Tourist Distributions

By fitting the effective tourist volumes of cities to Equation (1), it was found that user volumes satisfied the power-law distribution against their rank with $P_1 = 8259$ and $q = -0.84$, as shown in Figure 2a. The amounts of users in different cities were significantly different. A small number of cities concentrated most of the travelers, while most cities accounted for a small proportion of travelers. The first-ranked city was Beijing with 8143 users which was very close to the estimated P_1 . The R^2 reached 0.981 under the 0.01 significance level, indicating that the fitting result could explain almost all the variance of the traveler counts from all cities. The spatial disparity of tourists was significant.

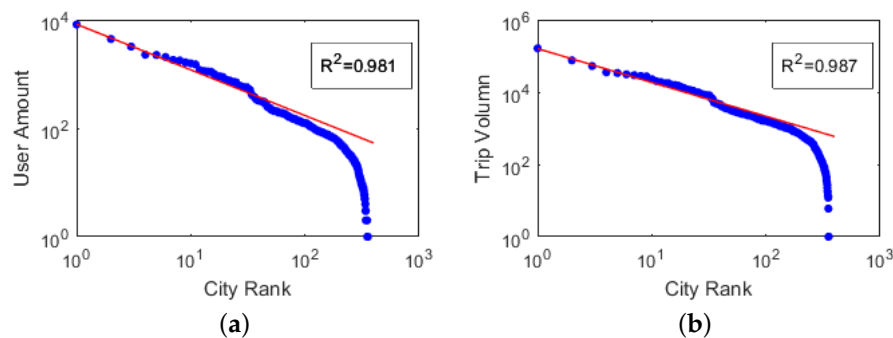


Figure 2. Log–log distributions of (a) effective users and (b) travels over the city ranks.

The total travels of all users in each city were accumulated. Then, in the same way, the travel volumes and ranks of the cities were fitted by Equation (1) after log-log transformations, as shown in Figure 2b. It was estimated that P_1 and q were 168,700 and 0.93, respectively, with $R^2 = 0.987$ under the 0.01 significance level. The travel volume in a city presented a negative power-law decay with its rank. The spatial disparity of travel volumes was significant, too.

4.1.2. Flow Distributions

To reveal the tourism sources and sinks, all travel inflows and outflows that were extracted from the travel blogs in every city were quantified. The top 20 cities are shown in Figure 3. The cities with large inflows indicated tourism sources, and the cities with large outflows were usually sinks.

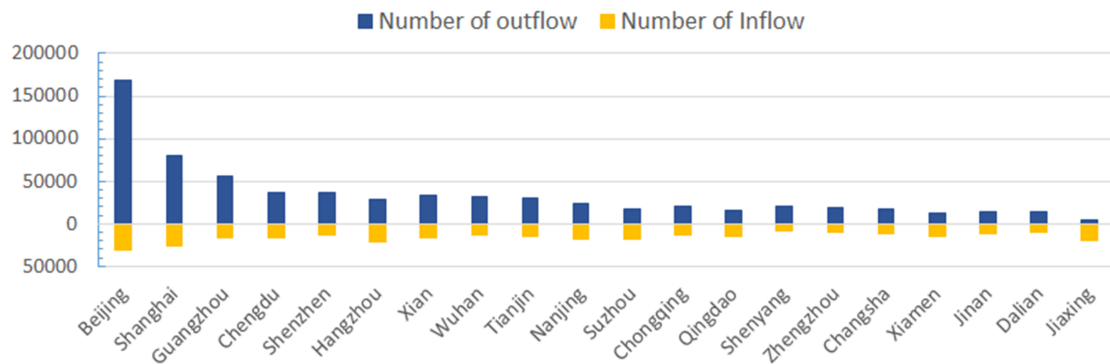


Figure 3. Top 20 cities of inflows and outflows.

Among the top 20 cities of outflows, most were provincial capitals or in eastern China with developed economic, for example, Beijing, Shanghai, Guangzhou, and Xi'an. In these cities, the residents had relatively higher incomes, which led to more tourists. Because of more mature social networks, people were more willing to share their travel experience on the Internet. The top 30 travel destinations from four representative tourism source cities are shown in Figure 4. It was obvious that most travelers preferred big tourism cities or surrounding cities. Short-distance travel was favored as well when time and cost were taken into account.

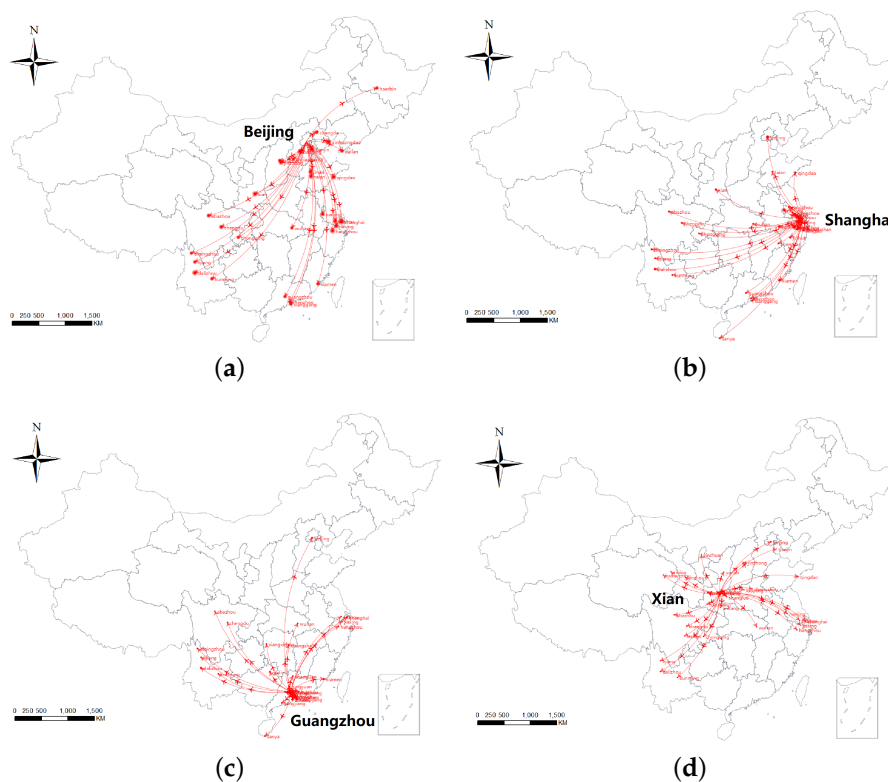


Figure 4. The top 30 destination cities of travelers from (a) Beijing, (b) Shanghai, (c) Guangzhou, and (d) Xi'an.

Among the top 30 cities of inflows, many well-known tourism cities ranked relatively high, for instance, Guilin, Sanya, Aba Prefecture, and Qingdao. Rich tourist resources in these cities attracted a large number of visitors all over the country, thus forming representative tourist sinks. Top cities where travelers went to four tourist-sink cities are shown in Figure 5. Distance was no longer a constraint factor for travels in inflows than in outflows.

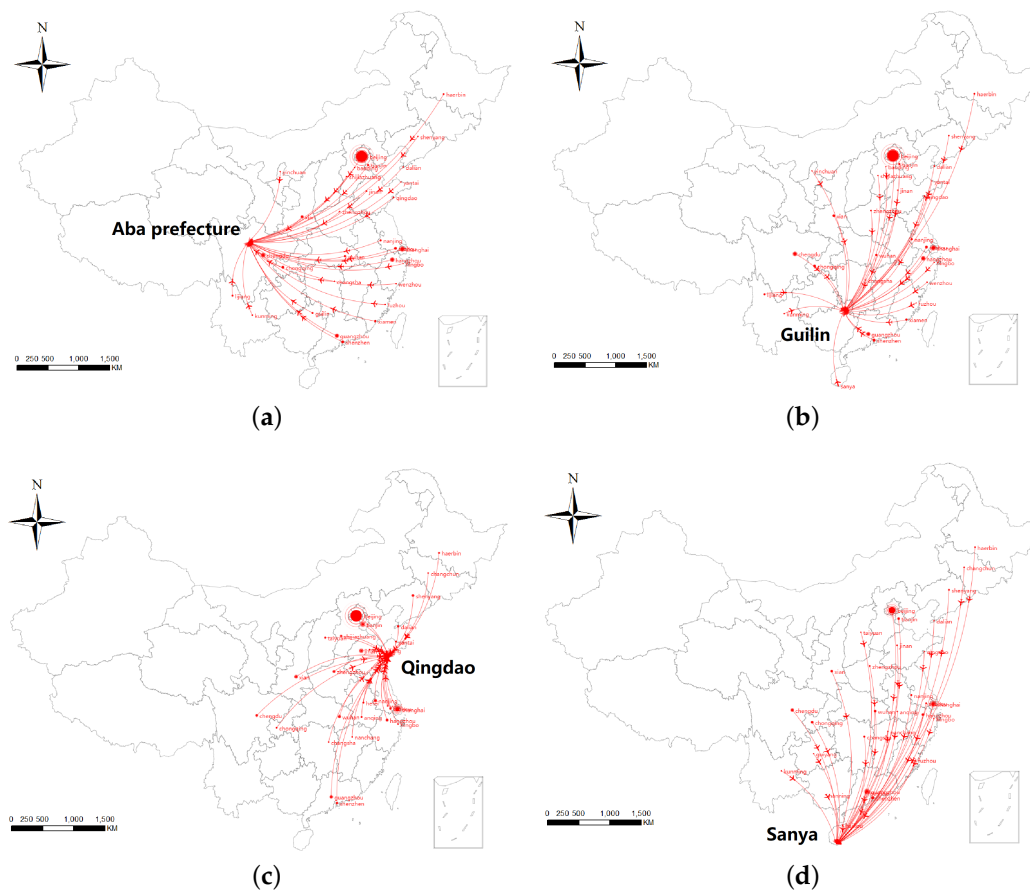


Figure 5. The top 30 origin cities of travelers to (a) Aba Prefecture, (b) Guilin, (c) Qingdao, and (d) Sanya.

4.2. Interaction Network Features

The city interaction network of tourism was constructed as shown in Figure 6. The network contains 259 city nodes and 9283 interaction edges. The colors of the edges are the same as the colors of the departure cities. The cities with more connections are closer to the center, while the cities with less connections are more likely on the periphery. Table 1 shows the basic statistics of the network.

The average degree of nodes in the network was 71.7, which represented the average number of interactions between cities. The diameter of the network was 2. Almost all cities connected to Beijing that was the main hub city. Subsequently, any two cities could be associated indirectly through Beijing.

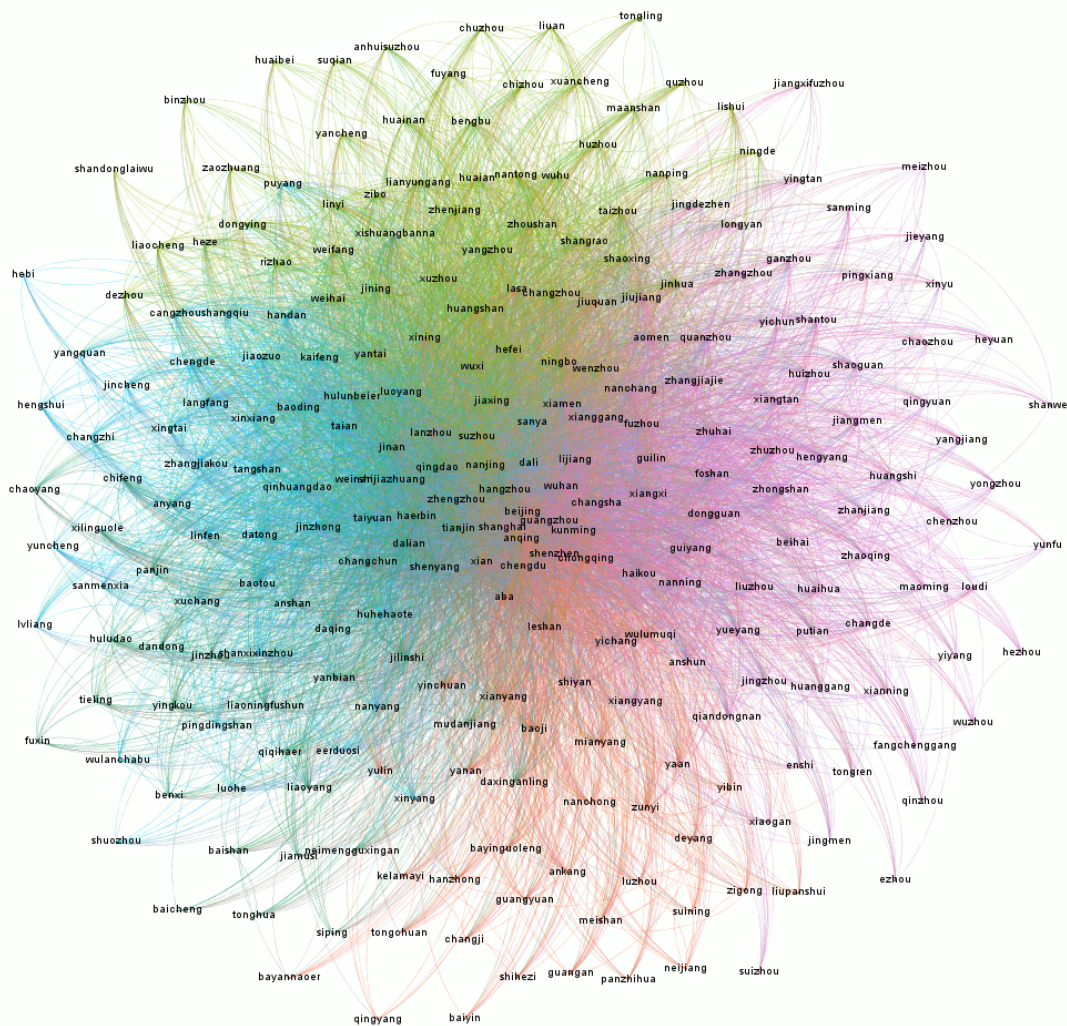


Figure 6. City interaction network of tourism.

Table 1. The basic statistical characteristic of the city interaction network of tourism.

Metrics	Values
number of nodes	259
number of edges	9283
average degree $\langle k \rangle$	71.7
network diameter D	2
average path length L	1.72
aggregation coefficient C	0.812
degree distribution index γ	0.016
assortativity coefficient Γ	-0.432

4.2.1. Centrality

Centrality indicates the importance and influence of a city in the tourism network. A city’s prominence, transit capacity, and accessibility are measured by degree centrality, betweenness centrality, and closeness centrality respectively. The top 20 cities of degree centrality, betweenness centrality and closeness centrality in the city interaction network of tourism are shown in Table 2. The top 20 cities mostly the first or second-tier cities in China, including state capital (Beijing), municipality

directly under the central government (e.g., Shanghai, Shenzhen and Chongqing) and provincial capitals (e.g., Guangzhou, Chengdu and Wuhan) with developed economic. The correlations between these indicators and the ranks of all cities are shown in Figure 7. They are all subjected to exponential distributions, which are plotted as the red lines. The values of R^2 , which equal to 0.97, 0.99 and 0.97 under the 0.01 significance level, indicate that the correlations are strong and significant. The degree centrality satisfied the exponential decay with its rank. Beijing ranked first with the normalized degree centrality equaling 1. All other cities connected to Beijing directly, which was also the reason that the diameter of the network was 2. Forty-four cities had values greater than 0.5, which meant that these cities interacted directly with more than half of the nodes in the network. Cities with a better economy and developed tourism industry had higher degree centrality. The betweenness centrality satisfied the exponential decay and heavy-tailed distribution. It was positively related with degree centrality. The descending trend of the closeness centrality satisfied the exponential decay. Cities with a more developed economy and tourism industry had relatively higher closeness centralities as well.

Table 2. The top 20 cities of centralities.

Rank	Degree Centrality	Betweenness Centrality	Closeness Centrality
1	Beijing	Beijing	Beijing
2	Shanghai	Shanghai	Shanghai
3	Guangzhou	Guangzhou	Guangzhou
4	Shenzhen	Shenzhen	Shenzhen
5	Hangzhou	Hangzhou	Hangzhou
6	Xi'an	Chengdu	Xi'an
7	Chengdu	Xi'an	Chengdu
8	Wuhan	Wuhan	Wuhan
9	Tianjin	Tianjin	Tianjin
10	Nanjing	Nanjing	Nanjing
11	Suzhou	Suzhou	Suzhou
12	Qingdao	Dali	Qingdao
13	Dali	Qingdao	Dali
14	Zhengzhou	Chongqing	Zhengzhou
15	Chongqing	Changsha	Chongqing
16	Changsha	Zhengzhou	Changsha
17	Kunming	Kunming	Kunming
18	Xiamen	Xiamen	Xiamen
19	Jinan	Dalian	Jinan
20	Jiaxing	Shenyang	Jiaxing

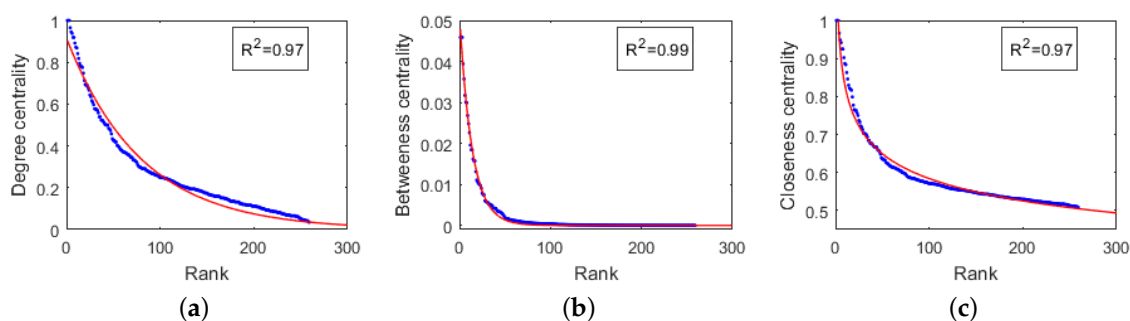


Figure 7. Distribution of (a) degree centrality, (b) betweenness centrality, and (c) closeness centrality.

4.2.2. Small-World Property

Small world property measures the interconnectivity between cities. The average shortest path length of the interaction network was 1.72, which was very close to 1.71 of the random network. The aggregation coefficient was 0.812, which was much higher than 0.277 for the random network.

Therefore, the city interaction network of tourism had small-world network properties and was relatively compact. A city with more tourism resources would drive surrounding small cities to form a large tourism region.

4.2.3. Degree Distribution

Degree distribution measures the disparity of the connections between cities. The degree of the interaction network satisfied the exponential distribution as shown in Figure 8a. The disparity of travel flows was obvious, although it was less strong than that under a power-law distribution that was more common in complex networks. A small number of famous cities attracted plenty of tourists from most cities, while most cities only had tourists from a few cities. Less than 13% of cities interacted more than half of other cities.

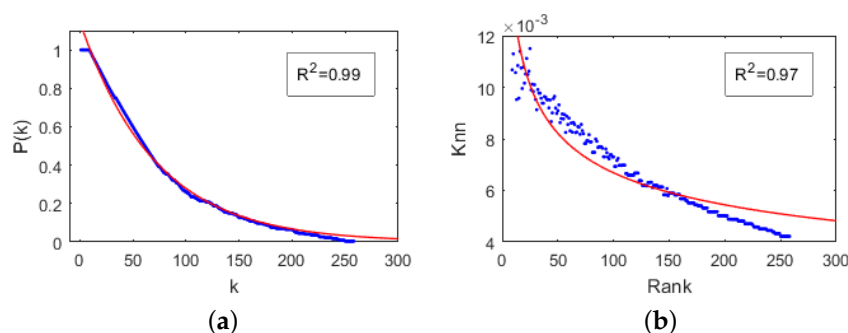


Figure 8. (a) The cumulative degree distribution; (b) the correlation between k and $k_{nn}(k)$.

4.2.4. Disassortativity

Disassortativity indicates the connectivity between dissimilar cities. The correlation between k and $k_{nn}(k)$ is plotted in Figure 8b. The network conformed to heterozygosity because the index satisfied the power-law decay, i.e., $k_{nn}(k) \propto k^{-\mu}$ with the heterogeneity index $\mu = 0.297$. As a result, cities with high degrees were intermediary hub nodes, and they were dispersed in the network rather than clustered together. Conversely, cities with low degrees were usually at the periphery of the network. On the other side, the assortativity coefficient Γ was -0.43 . The negative exponent verified the disassortativity of the network. Cities with high degrees tended to connect the low ones, and vice versa. Therefore, travels were more popular between big cities and small cities. Short tours around cities were the main choice on short holidays because of the convenience and limited time. Cities with more tourism resources promoted the tourism of the adjacent small cities.

4.2.5. Community

Modularity is designed to measure the strength of division of a network into communities. Higher modularity represents a closer connection between nodes in a community and sparser connection between communities. By iteratively setting the classification numbers of the spin-glass model, the modularity of the city interaction network peaked at four, so the network was partitioned into four communities as shown in Figure 9a. There were more travel flows between cities inside a community.

The four communities formed four significantly clustered regions in China mainland. Community #1 was the northeastern region centered on Beijing, mainly including the Beijing–Tianjin–Hebei–Shanxi–Inner Mongolia region, the three northeastern provinces, and Shandong Peninsula. Community #2 was the eastern region centered on Shanghai, mainly including Jiangsu, Zhejiang, Shanghai, and Fujian. Community #3 was the southern region mainly consisting of Hunan, Hubei, Guangdong, Guangxi, and Hainan. Community #4 was the western region including Xinjiang, Qinghai, and some northwestern and southwestern cities.

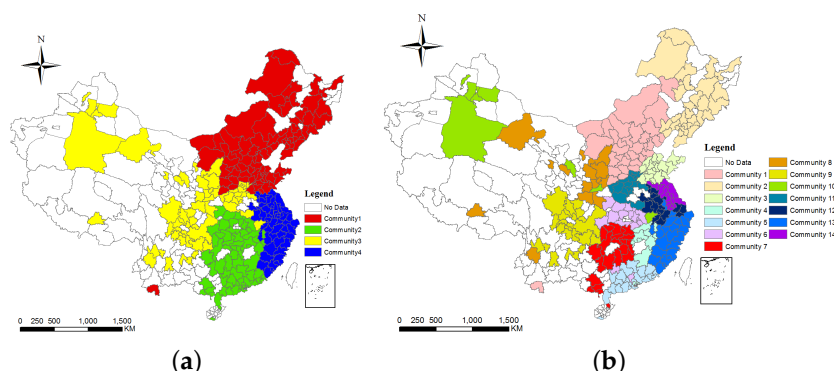


Figure 9. Spatial distributions of the communities detected of the network: (a) four communities, (b) 14 sub-communities.

To reveal more details, the network was further partitioned into 14 sub-communities as shown in Figure 9b based on the previous four communities. For example, community #1 was divided into three sub-communities including the Shandong Peninsula, the three northeastern provinces, and Beijing–Tianjin–Hebei–Shanxi–Inner Mongolia. Each sub-community contained a core city with higher degree centrality, betweenness centrality, and close centrality. The core city drove the tourism development of its adjacent cities together.

Regional boundaries of the communities were consistent with the administrative boundaries. For example, the Shandong Peninsula region was completely consistent with the provincial boundary of Shandong Province. It implied provincial administrations had a great influence on tourism arrangement. As the city interaction network of tourism was a small-world network, cities within a community had more travel flows between each other, but less travel flows occurred between communities. Actually, the communities consisted of the tourism regions resulted in by its small-world property.

4.3. Interaction Features

4.3.1. Similarity of Travel Flows

Travel flows of a city can reveal its tourism patterns. The outflows indicate a source city and its choices of tourist destination. Similarly, the inflows manifest a sink city and its tourist origins. Through the outflow and inflow vectors of cities, tourism choices about destinations or origins were discovered.

For the tourism sources, the similarity between a given city and others was measured by the Pearson's correlation coefficients of their outflows. Four representative source cities, including Beijing, Shanghai, Guangzhou, and Xi'an, were selected to illustrate the spatial distribution of outflow similarities as shown in Figure 10. In each map, the green point represents the location of the specific city, and the gradient colors denote the similarities between this specific city and the corresponding regions. Cities with shorter distances usually exhibit higher travel similarities with more similar travel destinations.

Hierarchical clustering was employed to investigate the global travel similarities of all cities. In the hierarchical clustering dendrogram, several cutting levels were artificially chosen from the bottom to the top according to its tree structure. After trials and errors to obtain good cluster separation and visualization, four cluster numbers, 29, 14, 7 and 2, were selected to explore spatial aggregation trends in near cities. Four level hierarchical results, i.e., 29, 14, 7, and 2 clusters, are shown in Figure 11. The results exhibited the similarity of travel destination choices of all cities. In the process of bottom-up aggregation, cities with a shorter distance tended to merge into one region. In the end, two regions of the north and the south separated by the Yangtze River were distinguished.

In the same way, the similarities of inflows, as well as of tourist origins, were measured to investigate the sink cities. Four famous tourism cities including Aba, Chengdu, Dali, and Guilin are selected to plot the spatial distribution of inflow similarities in Figure 12, in which the green points represent the four cities. The clustering results are shown in Figure 13. It was found that distance

had less effect on tourist origins than on destination choices. Even though the trend of geographical aggregation still existed, similar cities usually attracted similar tourists. For example, developed cities and tourism cities were usually clustered into their respective groups.

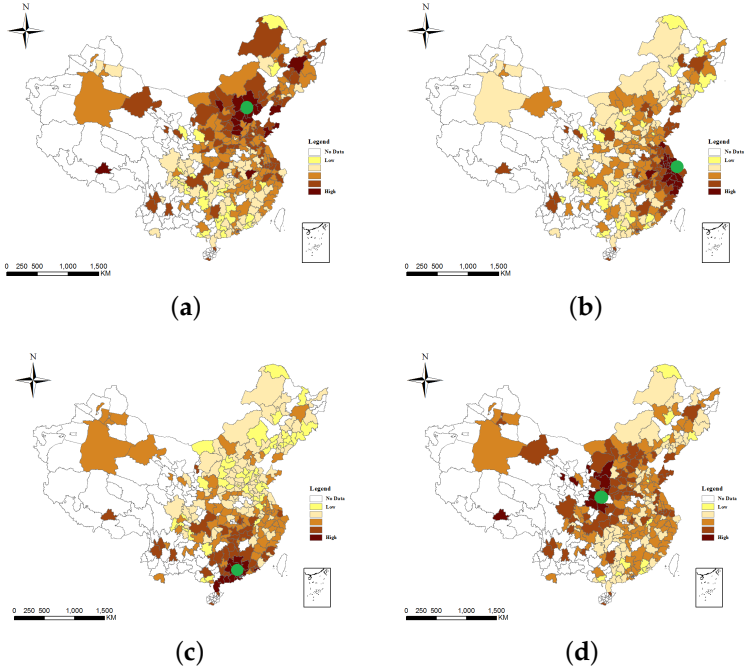


Figure 10. The spatial distribution of outflow similarities of (a) Beijing, (b) Shanghai, (c) Guangzhou, and (d) Xi'an.

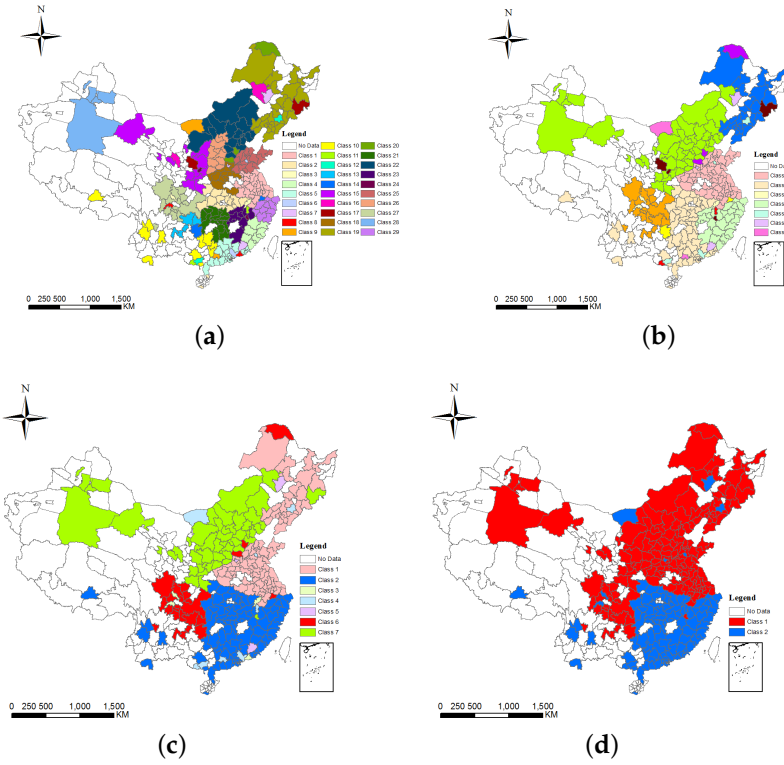


Figure 11. Hierarchical clustering by travel similarities. (a) 29 clusters, (b) 14 clusters, (c) seven clusters, and (d) two clusters.

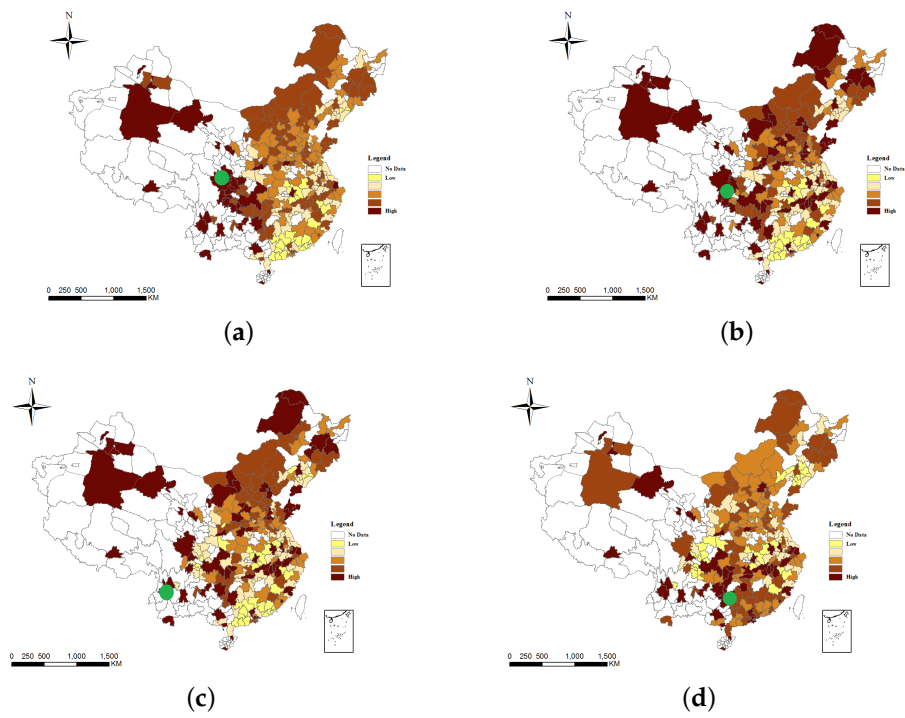


Figure 12. The spatial distribution of inflow similarities of (a) Aba, (b) Chengdu, (c) Dali, and (d) Guilin.

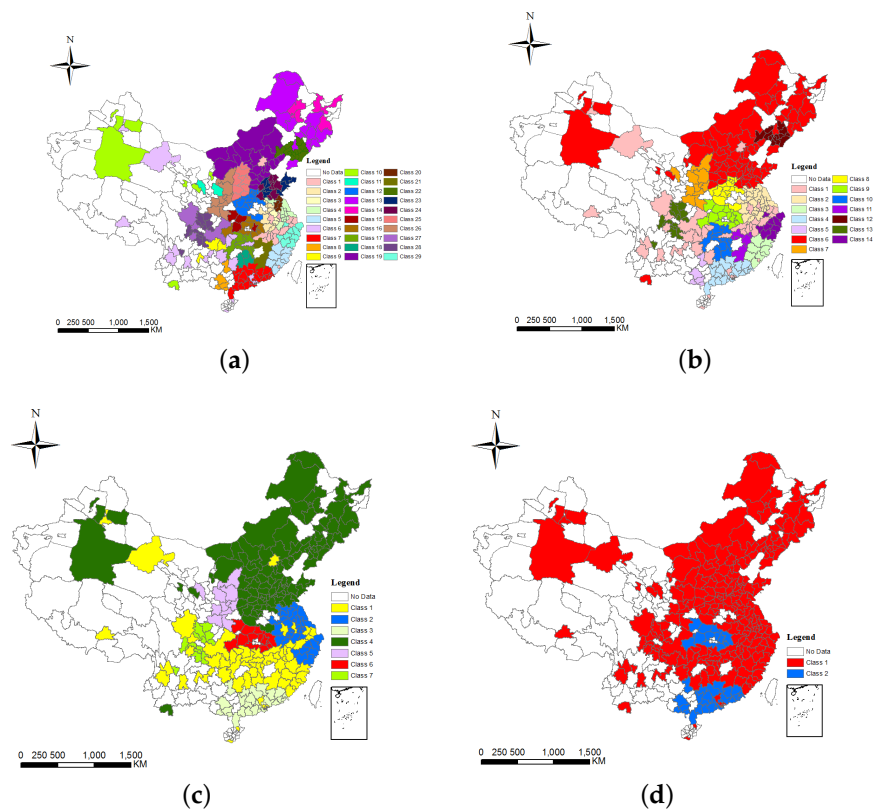


Figure 13. Hierarchical clustering by similarities of tourist origins. (a) 29 clusters, (b) 14 clusters, (c) seven clusters, and (d) two clusters.

4.3.2. Distance Decay Effect

The correlation between the travel frequency and its distance satisfied the negative power-law distribution as shown in Figure 14a with exponent $\gamma = 0.585$. The correlation between the travel frequency and the gyration radius of travelers is shown in Figure 14b. It was not a one-way declining trend as expected but rather divided into two parts. When the gyration radius was less than 800 m, the travel frequency increased linearly. After that, an increase in the gyration radius led to a sharp drop in the travel frequency. By fitting the second part individually, the correlation approximately satisfied the negative power-law distribution with exponent $\gamma = 1.60$. The exponent was in a similar value range as the results obtained from other mobility datasets that also obey the negative power-law distributions, such as the cell phone call records ($\gamma = 1.75$) [41], the dispersal of bank notes in the United States ($\gamma = 1.59$) [42], Foursquare check-ins ($\gamma = 1.88$) [43], and geo-located tweets across the world in 2012 ($\gamma = 1.62$) [37].

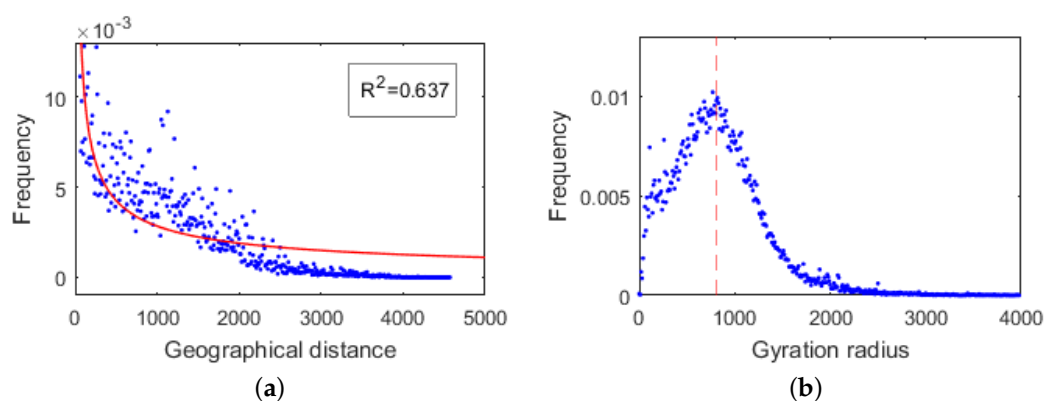


Figure 14. Distance impact on travel frequencies: (a) geographical distance; (b) radius of gyration.

4.3.3. Gravity Law

By solving the gravity model defined by Equation (5), the parameters α , β and γ were estimated as 0.716, 0.275 and 0.48, respectively. As a result, the tourist volume, no matter in origin or destination cities, had a sublinear correlation with the travel flows, but the former had a stronger influence than the latter. The declining distance friction coefficient was slightly smaller ($\gamma = 0.48$), so the tourist movement between cities had a slight distance decay effect. The exponent was essentially consistent with what was obtained by Xiao et al. ($0.4 < \gamma < 0.6$) from the air passenger flows in China [44]. The exponent that was approximately 0.5 indicated a hybrid tourist pattern of both long-distance travels to determinate destinations and short-haul trips of random choices.

5. Conclusions

Social media and online content generated by travelers provide a good way to investigate tourist activities and experiences. Using travel blogs from online tourism websites, the collective spatial patterns of intercity tourist movement are discovered from multiple perspectives. The rank of travel volume satisfies the power-law distribution. Developed cities generate more tourists, and more travel flows occur between developed cities. To investigate intercity travels, an interaction network is constructed based on intercity travel flows. The network is found to have an exponential degree distribution, disassortativity, and small-world property. The spatial arrangement of tourism in mainland China is also recognized after four communities and 14 sub-communities are detected. Intercity tourism presents a regional agglomeration effect because travel flows exist more in one community but less between communities. By distinguishing tourist sinks and sources, it is found that tourists from similar cities usually have similar tourism choices. Specifically, distance is essential

in determining tourist movements. Intercity travel flows decline with distance under a negative power-law distribution.

These results reveal the spatial patterns of tourist movements at an intercity scale. The spatial disparity of tourist sources is significant in China. Developed cities contribute most of the tourists because of their developed economy and large population. Tourists from similar cities have similar travel choices, and similar attractions have similar tourist sources. Spatial factors, especially the travel distance, make a large impact on tourist mobility. Tourism intentions have significant regionality. The spatial mobility, arrangement and patterns of tourism discovered in the results will be helpful for arranging tourism resources, predicting tourist flows, and understanding tourist activities. They can provide a basis for local, regional and national governments for tourism planning, city management, and sustainable development.

More efforts can be made for further expansion and deepening. The textual contents of online travel blogs are not included in this research. Text mining can be further conducted to discover detailed thematic information about tourist attractions. Moreover, the representativeness of travel blogs for general tourist mobility is still a worthwhile discussion because of the sparsity, incompleteness, and possible bias.

Author Contributions: conceptualization, Y.G. and X.Z.; methodology, Y.G., X.Z. and C.Y.; formal analysis, C.Y. and X.Z.; writing—original draft preparation, Y.G. and C.Y.; writing—review and editing, L.W. and Y.L.; funding acquisition, Y.L.

Funding: This research was funded by the National Key R&D Program of China under Grant 2017YFB0503602; and the National Natural Science Foundation of China, grant number 41830645.

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

References

1. Cohen, S.A.; Higham, J.E.; Peeters, P.; Gössling, S. Why tourism mobility behaviours must change. In *Understanding and Governing Sustainable Tourism Mobility*; Abingdon: ROUTLEDGE in Association with GSE Research; Routledge: Abingdon, UK, 2014; Volume 1, pp. 1–11.
2. Scuttari, A.; Lucia, M.D.; Martini, U. Integrated planning for sustainable tourism and mobility. A tourism traffic analysis in Italy's south tyrol region. *J. Sustain. Tour.* **2013**, *21*, 614–637.
3. Wheeler, B. Sustainable mass tourism: More smudge than nudge the canard continues. *Tour. Recreat. Res.* **2007**, *32*, 73–75.
4. Larsen, G.R.; Guiver, J.W. Understanding tourists' perceptions of distance: A key to reducing the environmental impacts of tourism mobility. *J. Sustain. Tour.* **2013**, *21*, 968–998.
5. Mercer, D. The geography of leisure—A contemporary growth-point. *Geography* **1970**, *55*, 261–273.
6. Rajotte, F. The different travel patterns and spatial framework of recreation and tourism. *Tour. Factor Nalt. Reg. Dev.* **1975**, 43–52.
7. Liu, Y.; Zhang, Y.; Nie, L. Patterns of self-drive tourists: The case of Nanning city, China. *Tour. Manag.* **2012**, *33*, 225–227.
8. Cohen, E. A phenomenology of tourist experiences. *Sociology* **1979**, *13*, 179–201.
9. Lew, A.; McKercher, B. Modeling tourist movements: A local destination analysis. *Ann. Tour. Res.* **2006**, *33*, 403–423.
10. Lau, G.; McKercher, B. Understanding tourist movement patterns in a destination: A gis approach. *Tour. Hosp. Res.* **2006**, *7*, 39–49.
11. Mckercher, B.; Lau, G. Movement patterns of tourists within a destination. *Tour. Geogr.* **2008**, *10*, 355–374.
12. Aletta, F.; Brambilla, G.; Maffei, L.; Masullo, M. Urban soundscapes: Characterization of a pedestrian tourist route in Sorrento (Italy). *Urban Sci.* **2016**, *1*, 4.
13. Votsi, N.-E.P.; Mazaris, A.D.; Kallimanis, A.S.; Pantis, J.D. Natural quiet: An additional feature reflecting green tourism development in conservation areas of Greece. *Tour. Manag. Perspect.* **2014**, *11*, 10–17.
14. Yang, Y.; Fik, T.; Zhang, J. Modeling sequential tourist flows: Where is the next destination? *Ann. Tour. Res.* **2013**, *43*, 297–320.

15. Benkhard, B. Determination of tourist flow patterns in a low mountain study area. *Tour. Manag. Stud.* **2018**, *14*, 19–31.
16. Edwards, D.; Griffin, T. Understanding tourists' spatial behaviour: Gps tracking as an aid to sustainable destination management. *J. Sustain. Tour.* **2013**, *21*, 580–595. [[CrossRef](#)]
17. De Cantis, S.; Ferrante, M.; Kahani, A.; Shoval, N. Cruise passengers' behavior at the destination: Investigation using gps technology. *Tour. Manag.* **2016**, *52*, 133–150.
18. East, D.; Osborne, P.; Kemp, S.; Woodfine, T. Combining gps & survey data improves understanding of visitor behaviour. *Tour. Manag.* **2017**, *61*, 307–320.
19. Nilbe, K.; Ahas, R.; Silm, S. Evaluating the travel distances of events visitors and regular visitors using mobile positioning data: The case of estonia. *J. Urban Technol.* **2014**, *21*, 91–107.
20. Zhang, R.; Ye, X.; Wang, K.; Li, D.; Zhu, J. Development of commute mode choice model by integrating actively and passively collected travel data. *Sustainability* **2019**, *11*, 2730.
21. Narangajavana, Y.; Fiol, L.J.C.; Tena, M.Á.M.; Artola, R.M.R.; García, J.S. The influence of social media in creating expectations. An empirical study for a tourist destination. *Ann. Tour. Res.* **2017**, *65*, 60–70.
22. Wise, N.; Farzin, F. "See you in iran" on facebook: Assessing user-generated authenticity. *Emerald Publ. Ltd.* **2018**, *24*, 33–52.
23. Hu, F.; Li, Z.; Yang, C.; Jiang, Y. A graph-based approach to detecting tourist movement patterns using social media data. *Cartogr. Geogr. Inf. Sci.* **2018**, *46*, 1–15.
24. Ahani, A.; Nilashi, M.; Ibrahim, O.; Sanzogni, L.; Weaven, S. Market segmentation and travel choice prediction in spa hotels through tripadvisor's online reviews. *Int. J. Hosp. Manag.* **2019**, *80*, 52–77.
25. Jin, C.; Cheng, J.; Xu, J. Using user-generated content to explore the temporal heterogeneity in tourist mobility. *J. Travel Res.* **2018**, *57*, 779–791.
26. Chung, H.C.; Chung, N.; Nam, Y. A social network analysis of tourist movement patterns in blogs: Korean backpackers in Europe. *Sustainability* **2017**, *9*, 2251.
27. Yang, L.; Wu, L.; Liu, Y.; Kang, C. Quantifying tourist behavior patterns by travel motifs and geo-tagged photos from flickr. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 345.
28. 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.
29. Min, W.; Bao, B.-K.; Xu, C. Multimodal spatio-temporal theme modeling for landmark analysis. *IEEE Multimed.* **2014**, *21*, 20–29.
30. Kou, N.M.; Yang, Y.; Gong, Z. Travel topic analysis: A mutually reinforcing method for geo-tagged photos. *Geoinformatica* **2015**, *19*, 693–721.
31. Padilla, J.J.; Kavak, H.; Lynch, C.J.; Gore, R.J.; Diallo, S.Y. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on twitter. *PLoS ONE* **2018**, *13*, e0198857.
32. Hao, Q.; Cai, R.; Wang, C.; Xiao, R.; Yang, J.-M.; Pang, Y.; Zhang, L. Equip tourists with knowledge mined from travelogues. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; pp. 401–410.
33. 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.
34. Csáji, B.C.; Browet, A.; Traag, V.A.; Delvenne, J.-C.; Huens, E.; Van Dooren, P.; Smoreda, Z.; Blondel, V.D. Exploring the mobility of mobile phone users. *Phys. A Stat. Mech. Appl.* **2013**, *392*, 1459–1473.
35. Wang, L.; Wang, Q.; Jing, Y.; Yu, H. Enhancing complex network synchronization based on the node betweenness. *IFAC Proc. Vol.* **2008**, *41*, 13282–13286.
36. de Vries, J.J.; Nijkamp, P.; Rietveld, P. Alonso's theory of movements: developments in spatial interaction modeling. *J. Geogr. Syst.* **2001**, *3*, 233–256. [[CrossRef](#)]
37. Hawelka, B.; Sitko, I.; Beinart, E.; Sobolevsky, S.; Kazakopoulos, P.; Ratti, C. Geo-located twitter as proxy for global mobility patterns. *Cartogr. Geogr. Inf. Sci.* **2014**, *14*, 260–271.
38. Dorogovtsev, S.N.; Goltsev, A.G.; Mendes, J.F.F. Pseudofractal scale-free web. *Phys. Rev. E* **2002**, *65*, 066122. [[CrossRef](#)]
39. Reichardt, J.; Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* **2006**, *74*, 016110.
40. Erlander, S.; Stewart, N.F. *The Gravity Model in Transportation Analysis: Theory and Extensions*; VSP Publishing: Wakefield, UK, 1990; Volume 3.

41. Gonzalez, M.C.; Hidalgo, C.A.; Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779.
42. Brockmann, D.; Hufnagel, L.; Geisel, T. The scaling laws of human travel. *Nature* **2006**, *439*, 462. [[CrossRef](#)]
43. Cheng, Z.; Caverlee, J.; Lee, K.; Sui, D.Z. Exploring millions of footprints in location sharing services. *ICWSM 2011, 2011*, 81–88.
44. Xiao, Y.; Wang, F.; Liu, Y.; Wang, J. Reconstructing gravitational attractions of major cities in china from air passenger flow data, 2001–2008: A particle swarm optimization approach. *Prof. Geogr.* **2013**, *65*, 265–282.



© 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/>).