Accepted manuscript. Please cite this article as: Mou, N. et al. (2020) Tourists' Digital Footprint: The Spatial Patterns of Tourist Flows in Qingdao, China. Tourism Management 81: 104151.
https://doi.org/10.1016/j.tourman.2020.104151

Tourists' Digital Footprint: The Spatial Patterns of Tourist Flows in Qingdao, China

Mou, Naixia; Zheng, Yunhao; Makkonen, Teemu; Yang, Tengfei; Tang, Jinwen (Jimmy); Song, Yan

Abstract: Spatial patterns of tourist flows represent the movement of tourists and show differences in tourism resources giving advice for promoting balanced and sustainable tourism development. This paper proposes a novel framework for analyzing these patterns based on tourists' digital footprint data collected from online travel diaries. Based on illustrative case study data from Qingdao (China), the framework, combining traditional quantitative and social network analysis, is able to pinpoint: (1) The influence of distance decay and attractions' popularity on the spatial patterns of tourist flows; (2) The uneven distribution of the core tourist nodes and the existence of the structural hole phenomenon, which form a network pattern with unbalanced power and intense internal competition; (3) The formation of the core area for tourism along the coastline – as is typical for coastal tourism cities. This difference of tourism resources between coastal and inland areas, thus, remains a challenge for future tourism development in Qingdao.

Keywords: tourist flow; spatial pattern; digital footprint; Qingdao; online travel diary; social network analysis

1 Introduction

With the change in urban functions and the development of tourism, the analysis of spatial patterns of tourist flows in cities has become increasingly important (Hu et al., 2015; Liu et al., 2012). Studying them is of great significance for the design of urban leisure and recreation sites, public service facilities, transport, tourism development and other aspects of the urban structure of cities.

Most of the "traditional" research on tourist flows is based on quantitative data from statistical offices or accommodation registers (Yang & Wang, 2014), questionnaires (Liu, Shi & Jian, 2017), etc. Such data sources, however, face difficulties in accurately reflecting the spatial and temporal distribution and flow characteristics of tourists. With the increasing popularity of social media sharing platforms and travel websites (where tourists exchange photos and comments while travelling), unique spatial and temporal data on tourist mobility has recently become available. This kind of data is commonly referred to as tourists' "digital footprint" (Girardin, Calabrese, Fiore, Ratti, & Blat, 2008). Tourists' digital footprint, as an electronic trace, not only offers a novel way of collecting data on tourist flows but also provides new research perspectives for tourist mobility research. With the rapid development of Information and Communication Technology (ICT) and the rising popularity of the global ubiquitous network, utilizing tourists' digital footprint data has become a dominant orientation in tourism development and research.

Digital footprints can be collected from many different types of data sources, such as GPS trajectory data (Li, Yang, Shen, & Wu, 2019), mobile phone signaling data (Zhu, Sun, Yuan, Hu,

& Miao, 2019), geotagged photos (Wood, Guerry, Silver, & Lacayo, 2013), and online travel diaries like Qunar.com (Jin, Cheng & Xu, 2018) or Ctrip.com (Ma, Wang, Xu, & Tai, 2018). With the rapid development of Location Based Service (LBS) technology in recent years, the use of online travel diaries is becoming increasingly popular among tourists. As a new type of tourism data that tourists voluntarily share, online travel diaries contain both the information of traditional texts and images but also record the information of the location of the tourists. This greatly enhances the accuracy of depicting tourist flows: online travel diaries have gradually become one of the most popular data sources for studies on tourists' digital footprint and tourist mobility.

The spatial and temporal characteristics of tourist mobility and the subsequent various socio-economic impacts are at the core of tourist flow research. As such, scholars have tried to explore the patterns of tourist flows but have, thus far, tended to focus mainly on the movement of tourists between attractions. As more and more tourists are willing to share their digital footprints on social media sharing platforms and travel websites, digital footprint data of tourists have gradually contributed to a better understanding of the mobility behaviour of city tourists (Önder, Koerbitz & Hubmann-Haidvogel, 2014). This has allowed researchers to construct more detailed depictions and to gain a deeper understanding of the spatial patterns of tourist flows. However, the most commonly used analysis methods, for example, Markov chains (Vu, Li, Law, & Ye, 2015), association rules mining (Versichele et al., 2014) and centrality analysis (Leung et al., 2012), often lack breadth or depth in pattern detection. Additionally, combining them with traditional quantitative methods is problematic.

To tackle the above shortcomings, we propose a novel research framework, combining traditional quantitative methods and social network analysis, for studying the spatial patterns of tourist flows, of which the framework offers a comprehensive overview. Tourists' digital footprint data collected from travel diaries of tourists (including same-day visitors; see Section 3.3) visiting Qingdao (China), our case study location, is used here to verify the proposed research framework.

2 Conceptual Backgrounds

Digital footprint data, as a "Big data" source widely recognized in tourism research (Li, Xu, Tang, Wang, & Li, 2018), can effectively reflect the spatial and temporal behavioral patterns of tourists. Different digital footprint data sources have different characterization capabilities for tourists. Salas-Olmedo et al. (2018) have compared a variety of common tourists' digital footprint data sources, pointing out that the characteristics of tourist flows reflected by geo-tagged photos, check-in data and other location-based data sources are often unavoidably inconsistent with the actual tourist behavior due to regulatory issues (such as prohibiting photographing, ethics, signal shielding of position sensors, etc.), and often also report redundant locational information. In order to enhance the reading experience, social media sharing platforms have started to add LBS modules to their editing tools to provide tourists with a new way to share their digital footprints. As such, contemporary online travel diaries data (edited by the tourists in relatively concise format) avoids the recording of deviant and redundant information by using position sensors that record locational data in real time.

Traditionally, research of online travel diaries has focused on the analysis of text data, mainly by exploring the implicit emotional information of tourists to study the image of destinations

(Choi, Lehto & Morrison, 2007; Lian & Yu, 2017). With the enhancement of the information on locational characteristics inherent in online travel diaries, scholars have begun to use this data to analyze the spatial patterns of tourist flows. Recent studies (Gao, Ye, Zhong, Wu, & Liu, 2019; Zeng & He, 2019) have highlighted the feasibility and practicability of online travel diaries as a source for digital footprint data to study the spatial patterns of tourist flows. This data not only overcomes the shortcomings of traditional data (statistical yearbooks, questionnaires, etc.), but also avoids the problems of information redundancy and record deviation common to other tourists' digital footprint data sources.

Economic development and the enhancements made in transportation and communication technologies have increased the number and frequency of people moving between places. These "flows of people" reinforce power relations between places and create urban systems (Limtanakool, Dijst & Schwanen, 2016). Tourist flows are a typical example of these flows. Information on spatial patterns of tourist flows, thus, gives valuable insights into decision-making processes related to, for example, resource allocation, planning, construction, etc. Providing tools to depict tourist flows should, therefore, be of high interest to policy-makers, tourism planners and city officials. Scholars have been trying to define the spatial patterns of tourist flows. At the conceptual level, studies by Lue et al. (1993), Lau and Mckercher (2006) and Zeng (2018) have classified these flows into different categories depending on tourists' movement patterns (e.g. single destination, base camp, stopover, regional tour, etc.) and attractions visited (e.g. single- or multi-center agglomerations). From a management perspective, based on the spatial patterns of tourist flows, scholars have provided proposals for improved tourism infrastructure (Smallwood, Beckley & Moore, 2012), travel guides (Zheng, Zha & Chua, 2012), tourism management (Liu, Zhang, Zhang, Sun, & Qiu, 2019) and marketing (Asakura & Iryo, 2007), etc. However, the existing research on the spatial patterns of tourist flows often lacks the perspective of spatial relationships. That is, it does not consider that travel processes are not solely about the movement of people but also entail interactions between attractions (and people and attractions) resulting in spatial effects lying beyond the reach of "traditional" research methods.

Regarding measurement, the most commonly utilized contemporary tools for analyzing the characteristics of tourist flows are statistical methods and models: for example, Markov chains (Vu et al., 2015; Zheng et al., 2012), regression models (Xia et al., 2010), as well as correlation (Kádár, 2014) and clustering analysis (Asakura & Iryo, 2007). The limitations of these research methods have led researchers to focus more on the statistical significance of different factors in explaining tourism mobility, while at the same time, ignoring the structure of spatial relationships among relevant actors of the tourism system. In other words, the gap in the literature, mentioned above, is often caused by limitations related to the utilized research methods. Physical theories, such as the gravity center model (Morley, Rosselló & Santana-Gallego, 2014), have paved the way for investigating various aspects of the spatial elements related to tourist flows, but it was the introduction of the social network theory that really provided scholars the tools to address this gap in the knowledge on spatial patterns of tourist flows.

Tourist flows do not solely reflect the characteristics of tourist mobility, but also reflect the connections between tourist destinations. Tourist flows, thus, form a relationship network with certain structural characteristics (Baggio, Scott & Cooper, 2010). There have been recent attempts to introduce social network theory into the study of spatial patterns of tourist flows via social network analysis. Notwithstanding, there are some important caveats in the existing literature on tourist flows using social network theory. The first is the lack of application depth of social network theory. In the theory of social networks, the evaluation of the "network structure" involves a variety of indicators, among which centrality indicators are the most widely used. Although the centrality indicators can effectively measure the concentration and dispersal of tourist flows (Leung et al., 2012; Liu, Huang & Fu, 2017), scholars generally limit their analysis on these centrality measures and ignore indicators reflecting the "structural holes" phenomenon in the tourist flow network with such metrics as "effective size" and "constraints" which measure disconnections and breaks between and among network nodes. These measures depicting "structural holes" can pinpoint uneven distributions of nodes in social networks (Brass, Butterfield & Skaggs, 1998). The measurement of "structural holes" can provide deeper insights on tourist flows than analyses based solely on centrality indicators. The existing literature has, however, with only a few exceptions (Leung et al., 2012; Zeng, 2018), focused on the analysis of the structural characteristics of the networks (Shih, 2006; Peng, Zhang, Liu, Lu, & Yang, 2016; Kang, Lee, Kim, & Park, 2018) rather than on the spatial patterns of tourist flows.

To address the above limitations, this paper proposes a novel research framework – combining traditional quantitative methods of spatial analysis with improved social network analysis tools – of the spatial patterns of tourist flows by using online travel diaries as the source for tourists' digital footprint data. The city of Qingdao (China) was chosen as a case study example to verify the practicability of the research framework (see Section 3.2).

3 Methodology

3.1 Research Framework for Analyzing the Spatial Patterns of Tourist Flows

We propose a novel research framework, shown in Fig. 1, for analyzing the spatial patterns of tourist flows.

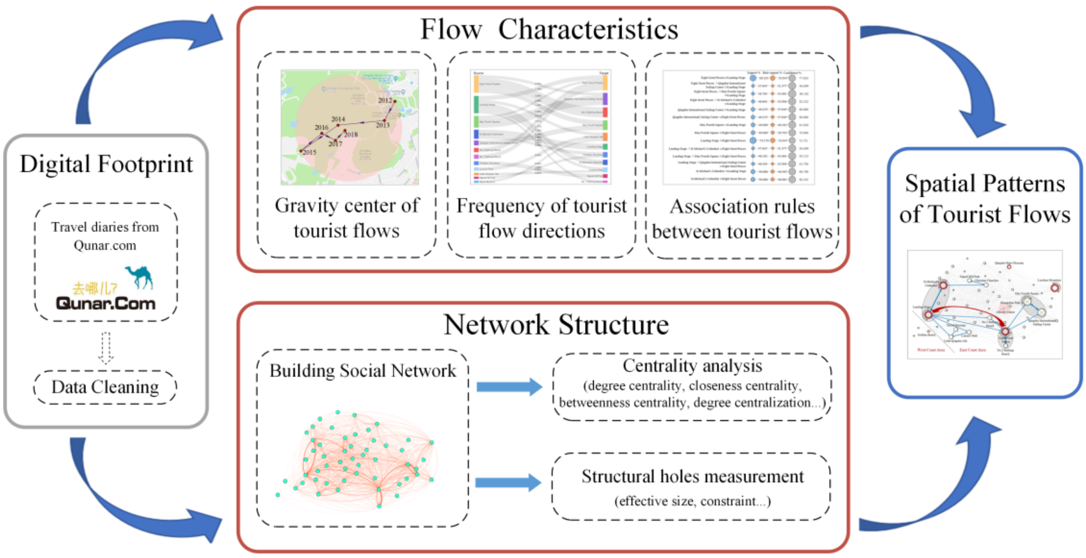


Figure 1. Research framework for analyzing the spatial patterns of tourist flows

The research framework is based on digital footprint data. We choose online travel diaries as our data source and designed intelligent data collection and processing rules to realize the dataset construction. The analysis of tourist flows is at the core of the research framework, which combines traditional quantitative and improved social network analysis. In order to analyze the spatial patterns of tourist flows comprehensively, we divided the research contents of several detection methods (gravity center model, statistics of flow directions, association rules mining, and social network analysis) into two sections: (1) "flow characteristics" and (2) "network structure". These sections are discussed in greater detail below to provide a reference point for those interested in spatial layout optimization and precision marketing of urban tourism.

We analyze the characteristics of tourist flows from three perspectives: (1) gravity center (the gravity center model is used to analyze the overall tourist flows); (2) frequent flow directions (the "explicit" flow characteristics of tourist flows between attractions are analyzed by statistics of flow directions); (3) association rules (the "implicit" flow characteristics of tourist flows between attractions are analyzed through the association rules mining algorithm, as a supplement to the analysis of the second perspective).

(1) Gravity center model. The gravity center model is an important tool for studying the variation of the spatial features of a city in the process of regional development. It calculates the gravity center of different activities in a city by simulating the balance center of the traction force between points or areas with different weights (Hilgard, 1872). The gravity center in the study area, taking into account the location and intensity of flows in all directions, is used to analyze the flow rule of tourist movement (Li, Jiang, Wang, Lei, & Deng, 2019). Our gravity center model of tourist flows takes the attractions (points of interest) in Qingdao as the basic calculation units. Therefore, it is necessary to set the tourist flow intensity of each attraction as the weight index to construct the gravity center calculation model. It is expressed as:

$$X=\frac{\sum_{i=1}^{n}\omega_i x_i}{\sum_{i=1}^{n}\omega_i} \qquad Y=\frac{\sum_{i=1}^{n}\omega_i y_i}{\sum_{i=1}^{n}\omega_i} \tag{1}$$

where (X,Y) is the gravity center coordinate; n is the total number of attractions; (x_i,y_i) is the geographic coordinates of the attraction i, expressed by the latitude and longitude coordinates; ω_i is the weight of the attraction i, expressed by the intensity of the tourist flows, i.e., how many times the attraction appears in the online travel diaries data.

(2) Statistics of flow directions. The direction of tourist flows is a movement sequence of tourists between two attractions, e.g., "from St. Michael's Cathedral to the May Fourth Square". The formula of the frequency of flow direction is as follows:

$$P_i=\frac{v_i}{V}\times 100\% \tag{2}$$

where P_i is the frequency of the flow direction i; v_i is the volume of the flow direction i, i.e., the occurrence number of the flow direction i in the sequence data; V is the total volume of all flow directions i, i.e., number of movements of tourists between two attractions in the data. By calculating and sorting the frequency of each flow direction in the data, the frequent directions of tourist flows can be obtained.

(3) Association rules mining. Association rules mining is a data mining method that searches for the correlations between different items in the same event through historical data. It is widely used in tourism research, for example, to investigate travel route recommendations (Xi & Yuan, 2017), tourist behavior (Versichele et al., 2014) as well as in tourism market analysis (Pyo, 2015). We apply the association rules mining method to analyze the characteristics of tourist flows, by finding the implicit relationships between the attractions. The CARMA algorithm (Hidber, 1999) is used to implement the association rules mining.

For analyzing the network structure of tourist flows, we use the tools of social network analysis. As an important means, based on graph theory, to study complex social systems, social network analysis has gradually become a popular paradigm for the study of the network structure of tourist flows (Casanueva, Gallego & García-Sánchez, 2016; Mou et al., 2020). Social network analysis is used to observe the connections between social entities (individuals, social organizations, etc.) and their structural characteristics from the perspective of group dynamics. Using social network analysis tools to analyze the network structure of tourist flows allows highlighting of the spatial patterns of the "overall" tourist flow and also the discovery of various intrinsic relationship characteristics of tourist flows. To this end, centrality analysis (measuring the power and the status of the nodes in the network) and structural holes measurement (reflecting connection breakdown among nodes in the network) are employed.

(1) Centrality analysis. Centrality analysis includes two types of indicators: node centrality and network centralization. Node centrality reflects the status of the node in the network, while network centralization reflects the concentration of the whole network. Our centrality analysis of the network structure of tourist flows is carried out by utilizing four relevant metrics: degree centrality, closeness centrality, betweenness centrality and degree centralization (Table 1).

Table 1: Metrics of centrality analysis

Name	Definition	Formula
Degree Centrality	Indicating the direct association between the target node and other nodes. It can be divided into in-degree centrality and out-degree centrality.	In-degree centrality: $C_{AD,in}(i)=\sum_{j}^{n} r_{ij,in}$ (3) Out-degree centrality: $C_{AD,out}(i)=\sum_{j}^{n} r_{ij,out}$ (4) where $r_{ij,in}$ and $r_{ij,out}$ represent the directional relationship between the nodes *i* and *j*, i.e., the number of tourists from attraction *i* to attraction *j* or the opposite, the former indicates that *j* flows to *i*, and the latter indicates that *i* flows to *j*, and *n* is the number of nodes in the tourist flow network.
Betweenness Centrality	Indicating the degree of the target node's control over	$C_{AB}(i)=\sum_{j}^{n}\sum_{k}^{n}\frac{g_{jk}(i)}{g_{jk}},\ j\neq k\neq i,\ j<k$ (5)

	other nodes. The higher the betweenness centrality, the higher the irreplaceability of the node in the tourist flow network.	where g_{jk} is the number of paths that the traveler reaches the node k from the node j, $g_{jk}(i)$ is the number of paths by node i in the paths from node j to node k, and n is the number of nodes in the tourist flow network.
Closeness Centrality	Indicating how close the target node is to other nodes.	$C_{AP}^{-1}(i)=\sum_{j}^{n} g_{ij}$ (6) where g_{ij} is the number of paths from node i to node j, and n is the number of nodes in the tourist flow network.
Degree Centralization	Indicating the concentration of the whole network. It is divided into in-degree centralization and out-degree centralization. The higher the degree centralization, the more obviously there is only one (or very few) central node(s) in the network.	$C_D=\frac{\left[\sum_{i=1}^{n}\left(C_{ADmax}-C_{AD}(i)\right)\right]}{n^2-3n+2}$ (7) where the numerator represents the sum of the difference value between the degree centrality $C_{AD}(i)$ of the nodes in the evaluated network and the maximum degree centrality C_{ADmax}, and n is the number of nodes in the tourist flow network

(2) Structural holes measurement. Structural holes are an indicator for judging whether or not relationships are easily broken between nodes (Burt, 1992). The nodes with structural hole advantages generally have strong regional competitive advantages and are less affected by the tourist flows of the surrounding nodes. They are irreplaceable and there are large differences in terms of accessibility between them and the surrounding nodes. These nodes often become “lone nodes”. The measurement of the structural holes helps a region to identify or detect potential bottleneck problems of their tourist flow networks. For this end, Burt (1992) has proposed the use of “effective size” and “constraint” as metrics of structural holes in social networks. These two metrics are widely used and, thus, employed also here (Table 2).

Table 2: Metrics of structural holes measurement

Name	Definition	Formula
Effective Size	Measuring the non-redundant part of the target node connected to all other nodes. The higher the effective size, the more obvious the competitive advantage of the target node.	$ES_i=\sum_{j}^{n}\left(1-\sum_{q}^{n} p_{iq} m_{jq}\right), \quad q \neq i, j$, where $p_{iq}=\frac{\left(z_{iq}+z_{qi}\right)}{\sum_{j}^{n}\left(z_{ij}+z_{ji}\right)}, \quad i \neq j$

		$m_{jq} = \frac{\left(z_{jq} + z_{qj}\right)}{\max\left(z_{jk} + z_{kj}\right)}, \quad j \neq k \quad (8)$ where z_{iq} is the number of connections from node i to node q, p_{iq} is the proportional relationship between the tourist node i and node q, i.e., the number of connections between node i and node q divided by the number of all the connections of nodes i; m_{jq} is the marginal strength between nodes j and q, which is the number of connections between node j and node q divided by the maximum number of connections between node j and other nodes; and n is the number of nodes in the tourist flow network.
Constraint	Reflecting the degree of direct and indirect dependence of the target node on other nodes. The smaller the constraint, the higher the status of the target node in the region, with a competitive advantage. On the contrary, the greater the constraint, the greater the impact of other nodes on the target node (a disadvantage in competition).	$CT_i = \sum_{j}^{n}\left(p_{ij} + \sum_{q}^{n} p_{iq} p_{qj}\right)^2, \quad q \neq i, j \quad (9)$ where p_{ij} is the proportional relationship between node i and node j; p_{iq} is the proportional relationship between node i and node q; p_{qj} is the proportional relationship between node q and node j; The calculation method of the proportional relationship between nodes is the same as equation (8). n is the number of nodes in the tourist flow network.

3.2 Case Study Region: Qingdao, China

Located in the south of Shandong Peninsula in China (see Fig. 2), Qingdao has a total area of 10,654 square kilometers and a population of 9,394,800. Backed by Laoshan Mountain and surrounding the “inner sea” Jiaozhou Bay, it is not only an important international port but also a popular site for coastal tourism and the venue for the sailing competitions of the 2008 Olympic Summer Games. In 2018, the city attracted 100 million visitors, while visitor expenditures amounted to a total of 186.71 billion yuan[1]. Therefore, Qingdao’s tourism industry has a pivotal

[1] http://qdtj.qingdao.gov.cn/n28356045/n32561056/n32561071/n32562222/190506094225533582.html

role in the local economy and thus was chosen as a representative case to study the spatial patterns of tourist flows. The applied methods can, naturally, be expanded to analyze any other city with similar data availability.

3.3 Digital Footprint Data

We chose the online travel diaries from Qunar.com as our digital footprint data source. Qunar.com (https://www.qunar.com/) is the leading travel search engine in China. It is currently the largest Chinese social media sharing platform for online travel diaries. The data from Qunar.com has been applied to the study of destination image (Lian & Yu, 2017), tourists' rating behavior (Zhang, Zhang & Yang, 2016), and tourist mobility (Jin et al., 2018), etc.

Qunar.com provides a smart travel editorial program: users can set the point of interest (POI) of the attractions involved in their travel diary entries to generate visual travel routes with the LBS module. Thus far, this aspect of Qunar.com has, however, been overlooked in previous studies using data collected from the social media sharing platform (Jin et al., 2018; Lian & Yu, 2017). The POI is recorded, as numbers, in the source code of the travel diary. The POI number can then be matched with the Qunar.com's database for attractions to obtain the details of the POIs visited. Here, information (user ID, diary ID, departure date, travel time and the sequence of the POIs visited) from 1,215 online travel diaries between 2012–2018 that were shared by tourists on the website were collected as the initial data (Table 3).

Table 3: Sample Records of Online Travel Diaries

User ID	Diary ID	Departure date	Travel time	The sequence of the POIs visited
926162@qunar	5903228	2015/9/16	5 day	722211;704711;702128
1148805@qunar	6611293	2016/4/1	5 day	702128;706245;710128;715470;5740219…
273748641@qunar	7087412	2018/8/6	7 day	702128;7561789;7644128;7525684;702719…
219875301@qunar	7063379	2017/7/2	4 day	716176;710128;713204;706645;7561776

Fig. 2 shows the distribution of the attractions in the online travel diaries data. The figure shows that the attractions recorded in the online travel diaries are mostly distributed along the southern coastal areas of Qingdao. Affected by the semi-closed bay of Jiaozhou Bay, the region is divided into two parts: "East coast area" and "West coast area". Due to its unique natural landscape and transportation hub advantages, the Qingdao East coast area is a prime destination for tourists in the city, compared to the less visited West coast area with sparse travel diaries data.

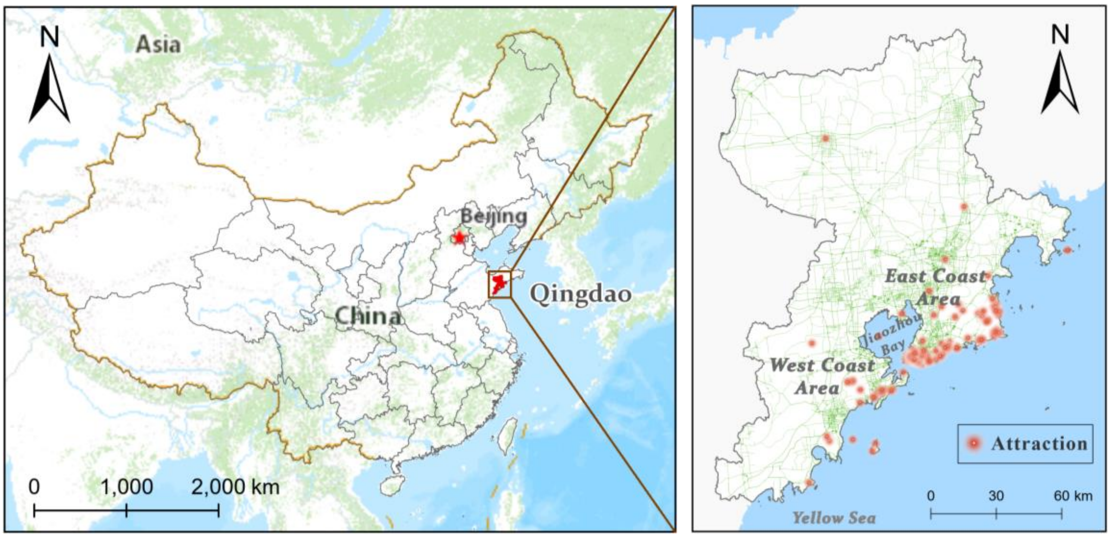


Figure 2. Distribution of online travel diaries data in Qingdao

It has to be noted that online travel diaries data often have information errors and logic problems. We use the following rules to clean the data:

(1) **Regional clipping.** Because in some online travel diaries, Qingdao is only one out of several cities visited by the tourists during their journey, or due to the negligence of the users in writing travel diaries, there are some out-of-town attractions in the travel diaries. Thus, the POI records of attractions outside the study location were cleaned from the data.

(2) **Attractions merging.** In online travel diaries, the users often record some small attractions inside larger attractions, such as “White Cloud Cave” in “Laoshan Mountain”. The POIs of small attractions were, therefore, merged according to their “affiliation” to larger ones.

(3) **Data deduplication.** If consecutive POI numbers appear in online travel diaries data, it is considered that the user has not moved between attractions, and the redundant POI records were deleted from the POI sequence.

(4) **Removing “lone point” data.** For the purpose of analyzing tourist flow networks, online travel diaries with less than two visited (recorded) attractions were filtered out.

After applying the above data processing rules, the final data consists of a total of 987 travel diaries and 7,657 visits to 53 attractions. As shown in Table 4, the Landing Stage and the Eight Great Passes are the most frequently occurring attractions in the data, followed by St. Michael’s Cathedral, May Fourth Square and Qingdao International Sailing Center.

Table 4: Popular attractions of online travel diaries data in Qingdao

Attractions	Count	Percent
Landing Stage	770	10.06%
Eight Great Passes	764	9.98%
St.Michael's Cathedral	529	6.91%
May Fourth Square	498	6.50%
Qingdao International Sailing Center	468	6.11%
Laoshan Mountain	317	4.14%

Signal Hill Park	287	3.75%
Little Qingdao Isle	282	3.68%
Luxun Park	275	3.59%
Christian Churches	273	3.57%

In our online travel diaries data, visitors staying in Qingdao for only one day account for only 4.35% of the total number of users. Thus, the vast majority of the visitors can be counted as tourists: for the sake of brevity, in the following we use the term “tourist” to collectively describe all the users (both same-day visitors and tourists) in our data.

4 Results

4.1 Flow Characteristics

4.1.1 Gravity Center of Tourist Flows

We collected and use online travel diaries data from 2012 to 2018. The tourist flow intensity and geographic location information of each attraction in the data are used to construct the gravity center model based on equation (1) to identify the gravity center of Qingdao tourist flows from 2012 to 2018 (Fig. 3).

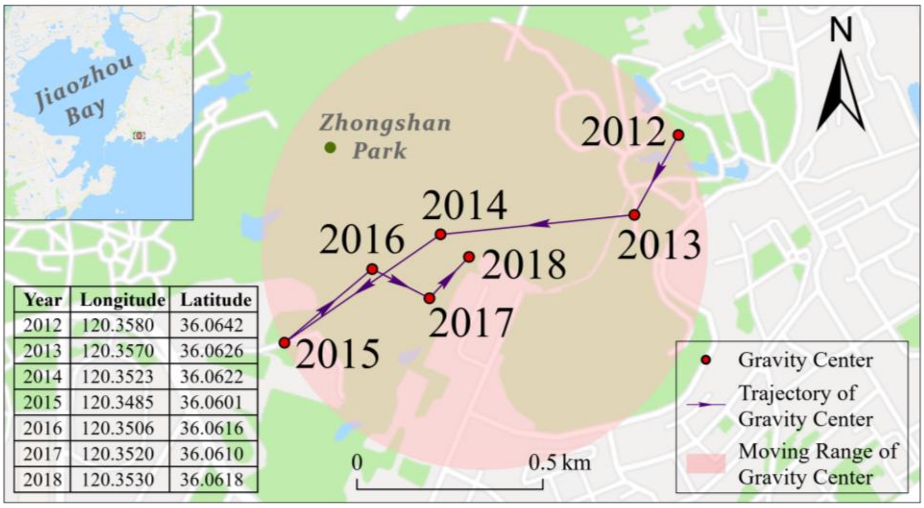


Figure 3. Evolution of gravity center of tourist flows in Qingdao from 2012 to 2018

As Fig. 3 shows, the gravity center of tourist flows in Qingdao has remained relatively stable over the years: the cumulative offset distance is only about 1.9 km. The gravity center of tourist flows is concentrated in an area of about 1.2 km in diameter in the East coast area adjacent to the Eight Great Passes in the north.

The gravity center of tourist flows has shifted “from northeast to southwest” in 2012–2015.

This is likely due to the opening of the Qingdao Jiaozhou Bay Subsea Tunnel (in June 2011). While the East coast area of Qingdao is the historical and cultural center and transportation hub of Qingdao, the improvement of traffic conditions facilitated the rapid development of Qingdao's tourism industry also in the West coast area (attracting tourist flows from the East coast area). However, after the initial growth in the development of tourism in the West coast area, problems in tourism management and service provision have emerged due to the rapid increase in the number of tourists. Moreover, the attractions in the West coast area are mainly related only to natural scenery, making it somewhat difficult for the West coast area to maintain long-tern attractiveness to tourists compared to the East coast area with its rich combination of both cultural and natural landscapes. The "pricey prawn" scandal in Qingdao[2] in 2015 has also had a likely impact in reducing Qingdao's tourism popularity after 2015 and in keeping tourists' traveling behavior conservative (centered around the main attractions and the transportation hub of the city). As a result, the gravity center of Qingdao's tourist flows has started to shift back to the northeast from 2016–2018.

The gravity center of tourist flows in Qingdao reflects the flow rule of overall tourist movement and the unbalanced development of tourism between the East and West coast areas. As such, developing the tourism industry of the West coast area is of great significance to narrow the development gap between the East and West coast areas.

4.1.2 Frequent Directions of Tourist Flows

A total of 6,669 movements of tourists between attractions can be observed from the online travel diaries data. These movements are sorted according to the frequency of tourist flow directions of each attraction, based on equation (2), in the data leading to a ranking of tourist flows in Qingdao. The top 20 tourist flow directions are presented in Fig. 4.

The most frequent tourist flow direction in the data is from the Eight Great Passes to the No.2 Bathing Beach, accounting for 2.80% of the total tourist flow volume, followed by the tourist flow direction from the May Fourth Square to Qingdao International Sailing Center, accounting for 2.65% of the total volume. The frequent tourist flow directions reflect the linkage relationships between the attractions. The tourist flows between the May Fourth Square and Qingdao International Sailing Center represent the most significant linkage relationship (the bi-directional tourist flow volume between them accounts for 4.03% of the total volume of tourist flows), followed by the tourist flows between the Eight Great Passes and the No.2 Bathing Beach (accounting for 4.02%), and the tourist flows between the Landing Stage and St. Michael's Cathedral (accounting for 3.16%). The routes between these attractions are therefore the core paths of tourist flows in Qingdao.

[2] https://www.thechairmansbao.com/chinas-38-yuan-large-prawn-scandal/

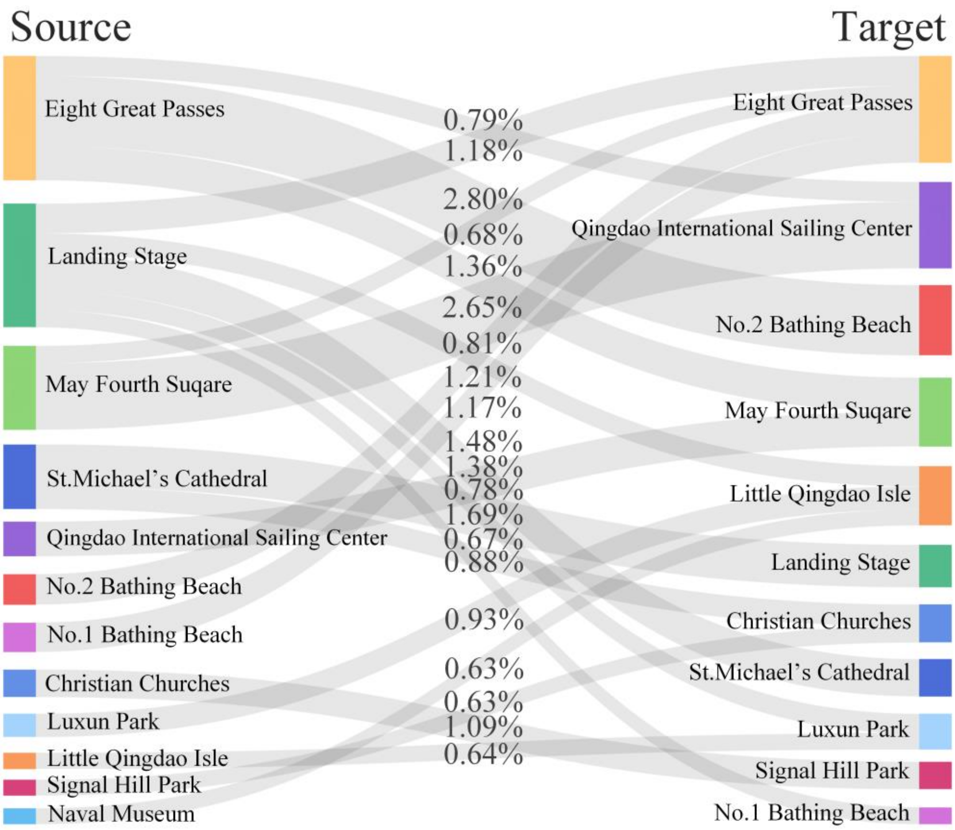


Figure 4. Frequencies of tourist flow directions in Qingdao

Figure 5 shows the distribution of the distance and the occurrence frequency between the attractions of each tourist flow direction, the tourist flow directions with larger volumes are often relatively short (less than 10km). This indicates that tourist flows are (unsurprisingly) affected by distance decay. It is very likely that, because of the congestion problems near the most popular attractions (particularly during the tourist season), the (rational) tourists will take the convenience of transportation into consideration and will choose to visit a bundle of attractions close to each other rather than spend their holiday in traffic congestions.

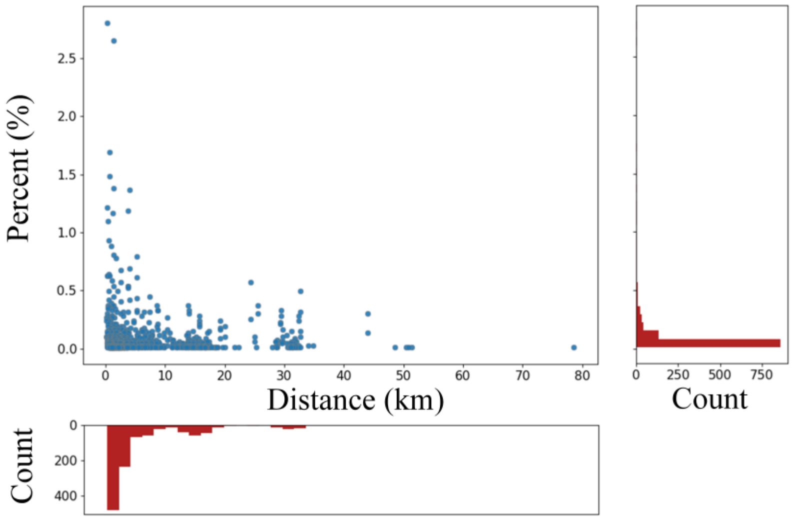


Figure 5. Distance-frequency distribution of tourist flow directions in Qingdao

4.1.3 Association Rules between Tourist Flows

The CARMA algorithm was employed to explore association rules in the online travel diaries data. After several experiments, a total of 14 association rule records were obtained (Fig. 6).

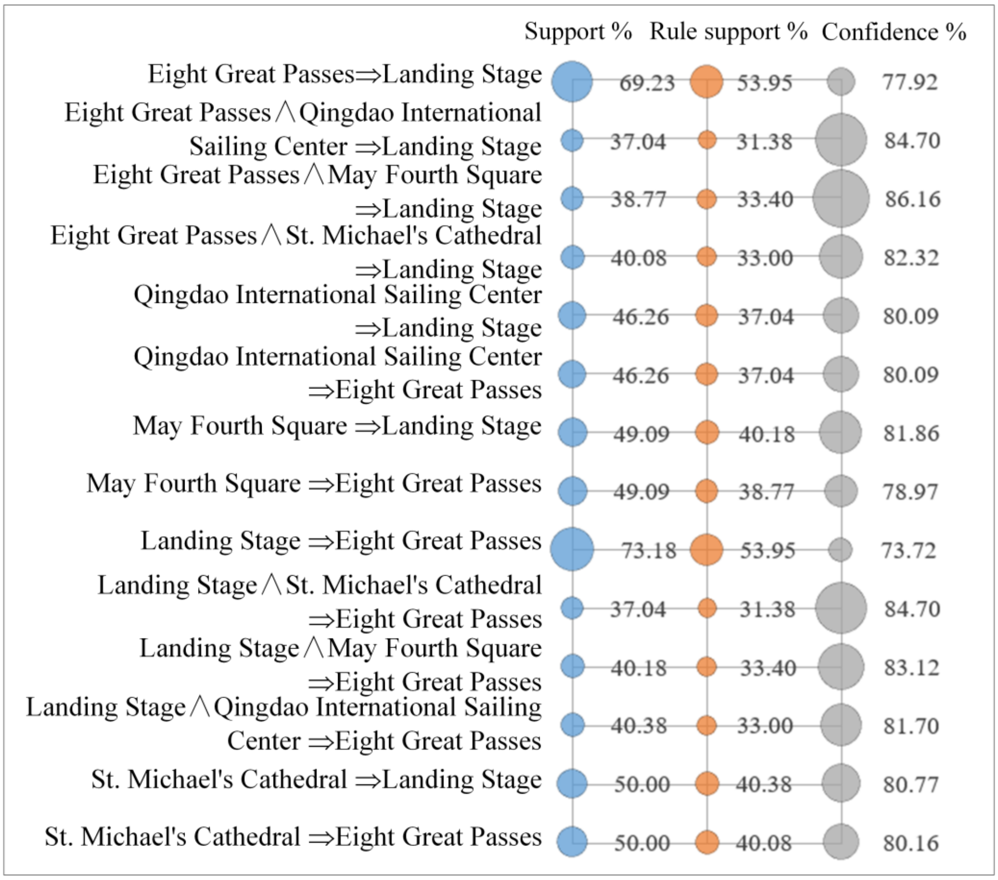


Figure 6. Results of association rule analysis

The analysis results of the CARMA algorithm are expressed as "[$X \Rightarrow Y$ *support*, *rule support*, *confidence*]", where X and Y are the antecedent and the consequent of the association rule, respectively, each of which can be composed of multiple sub-items connected by "$\wedge$ (=and)". The antecedent and the consequent represent the conditions and the results of the event, respectively. They are connected by "$\Rightarrow$": "*support*", which refers to the percentage of transactions containing X in the total transactions of the dataset; "*rule support"* indicating the percentage of transactions containing X and Y in the total transactions of the dataset (i.e. it is a measure of the accuracy of the association rule); "*confidence*" meaning the ratio of the number of transactions containing X and Y to the number of transactions containing X in the dataset (i.e., the ratio of the "*support*" to the "*rule support*") denoting the probability of the occurrence of Y in the presence of X (i.e., it is a measure of the importance of the association rule).

Figure 6 illustrates that the highest support is obtained when the antecedent of the association rule is either the Landing Stage, the Eight Great Passes, St. Michael's Cathedral, the May Fourth Square or Qingdao International Sailing Center (73.18%, 69.23%, 50.00%, 49.09% and 46.26% support respectively) indicating that these five attractions have the highest frequency of occurrence in the online travel diaries data. That is, they are also the most popular nodes of the tourist flows. Furthermore, the Landing Stage and the Eight Great Passes are not only the two attractions with the highest support, but also closely related to the results of the association rules:

(1) The 14 association rules obtained by the analysis are all related to the Landing Stage or the Eight Great Passes (containing at least one of them) including eight association rules containing both of them.

(2) The confidence of the association rules will be further (and significantly) improved if the Landing Stage and the Eight Great Passes are combined with surrounding popular attractions. For example, when tourists visit the Landing Stage and St. Michael's Cathedral, the confidence that they will also visit the Eight Great Passes can reach 84.70%. Further, when tourists visit the Eight Great Passes and the May Fourth Square, the confidence that they will also go to the Landing Stage can reach 86.16%.

Although no explicit association between the Landing Stage and the Eight Great Passes was detected in the analysis of frequent directions of tourist flows (Fig. 4), the results of the association rules exploration show that if a third attraction is added between the Landing Stage and the Eight Great Passes, the two will show a strong correlation. These results can be explained by the following: first, by the long distance of the nearly 60-minute walk, or more than 20-minute drive due to the heavy vehicle traffic along the coast in Qingdao between the two attractions (direct tourist flow between the two attractions is rare); second, and in relation to the tourism industry, by the geographically superior position of the two attractions, such as being at the core of transportation links and having a bundle of tourist attractions. Additionally, the Landing Stage is close to the transportation hub of the city, including Qingdao Railway Station, entrance of the Jiaozhou Bay Subsea Tunnel, Center of Qingdao Metro Line, etc., while there are many popular attractions around the Eight Great Passes. Therefore, there is an implicit connection between these two popular attractions.

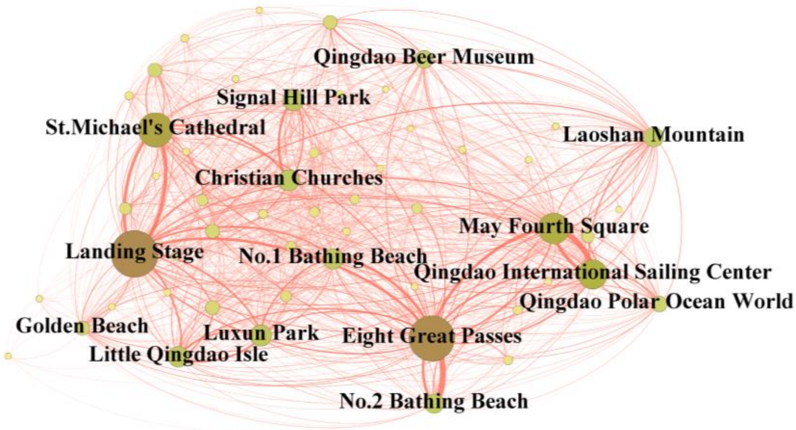
4.2 Network Structure

4.2.1 Centrality Analysis of Tourist Flow Network

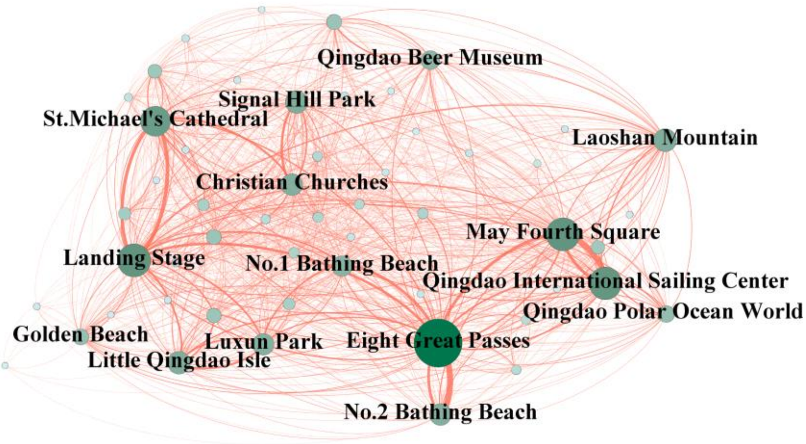
For the purposes of the social network analysis, a flow matrix between each attraction was constructed, according to the online travel diaries data, to illustrate the network structure of tourist flows in Qingdao. Node centrality metrics (in-degree centrality, out-degree centrality, betweenness centrality and closeness centrality) were calculated according to equations 3–6 using UCINET6. The results for a sample of the most relevant attractions are shown in Table 5. The tourist flow network is visualized according to these node centrality values (Fig. 7). The size of the nodes in Fig. 7 represents the level of the node centrality, while the thickness of the connections between the nodes indicates the volumes of the tourist flows.

Table 5: Calculation results of the node centrality

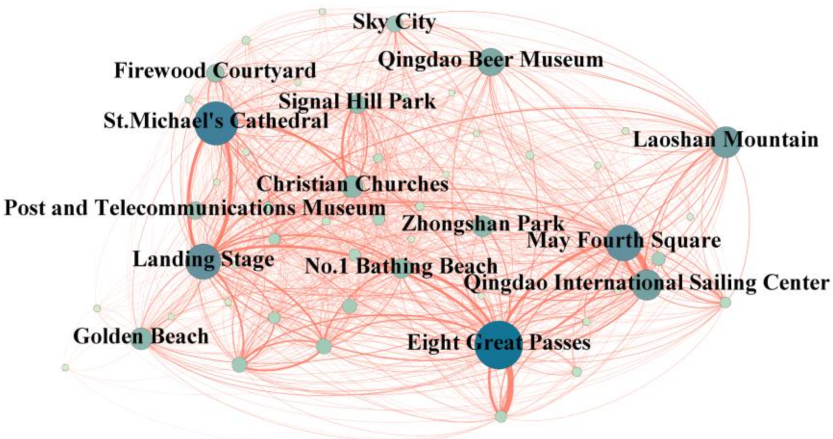
Attractions	Out-degree centrality	In-degree centrality	Betweenness centrality	Closeness centrality
Eight Great Passes	698	675	171.75	83.87
St. Michael’s Cathedral	487	395	152.42	83.87
May Fourth Square	428	448	122.64	81.25
Landing Stage	699	466	119.13	81.25
Laoshan Mountain	229	277	100.08	77.61
Qingdao International Sailing Center	397	436	98.50	80.00
Qingdao Beer Museum	199	211	85.80	75.36
Golden Beach	140	163	63.64	71.23
Christian Churches	255	257	62.51	76.47
Zhongshan Park	73	77	60.22	65.82
No.1 Bathing Beach	244	235	53.94	76.47
Signal Hill Park	261	270	50.41	75.36
Firewood Courtyard	137	126	45.10	70.27
Qingdao Post and Telecommunications Museum	102	106	44.51	68.42
Sky City	128	150	40.34	72.22
Luxun Park	259	253	36.86	72.22
Little Qingdao Isle	247	273	35.61	72.22
Naval Museum	133	137	22.15	67.53
No.2 Bathing Beach	226	256	22.08	70.27
Qingdao Polar Ocean World	163	175	17.72	66.67



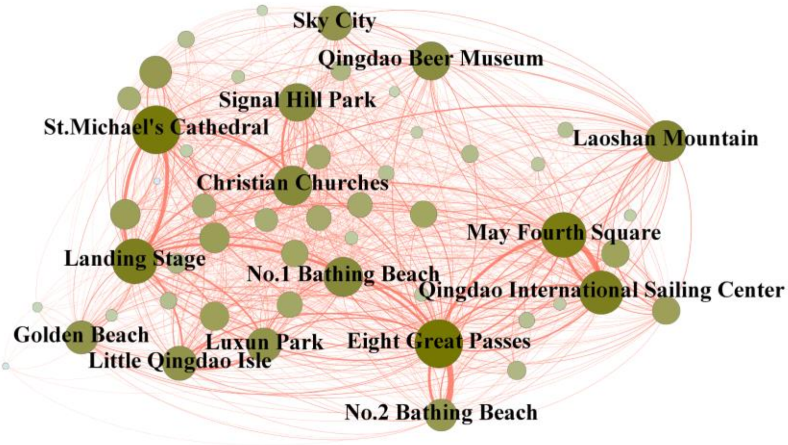
(a) Out-degree centrality



(b) In-degree centrality



(c) Betweenness centrality



(d) Closeness centrality

Figure 7. Network structure of tourist flows in Qingdao based on the calculated node centrality (nodes with centrality values among the top 15 are indicated with names)

A comparison between Fig. 7(a) and Fig. 7(b) shows that the calculation results of the out-degree centrality and in-degree centrality of the nodes of the Qingdao tourist flow network are generally consistent. The degree centrality values of the Landing Stage, the Eight Great Passes, St. Michael’s Cathedral, the May Fourth Square and Qingdao International Sailing Center are among the top five in both figures. Among them, the Landing Stage has the highest out-degree centrality value due to its vicinity to the transportation hub of Qingdao (which is often the starting point of tourists in Qingdao), while the Eight Great Passes has the highest in-degree centrality value, indicating that the Eight Great Passes is the most popular attraction in Qingdao due to its unique architectural landscape and cultural characteristics.

Based on Fig. 7(c) the Landing Stage, the Eight Great Passes, St. Michael’s Cathedral, the May Fourth Square and Laoshan Mountain are the most important attractions in terms of their betweenness centrality values, indicating that these five attractions have the highest “irreplaceability” and dispersion ability of tourist flows in the tourist flow network. The attraction with the highest betweenness centrality value, the Eight Great Passes, thus, plays the role of a “core intermediary” in the Qingdao tourist flow network. As shown in Fig. 7(d), the closeness centrality values of the nodes of the Qingdao tourist flow network are more evenly distributed than in the case of the degree and the betweenness centrality values, indicating that generally there is high accessibility between most of the attractions in Qingdao.

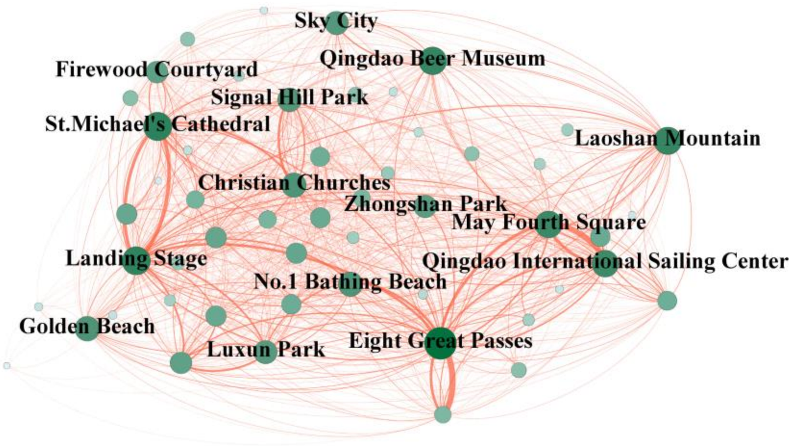
To summarize, the attractions of the Landing Stage, the Eight Great Passes, St. Michael’s Cathedral, the May Fourth Square, Qingdao International Sailing Center and Laoshan Mountain have become the core nodes of the Qingdao tourist flow network (i.e. they have high node centrality values). The values of the out-degree centralization and the in-degree centralization of the tourist flow network, calculated on the basis of equation 7, are also high, reaching 27.97% and 29.78% respectively, indicating that: (1) the distribution of node power in the tourist flow network in Qingdao is not balanced; (2) the circulation of the tourist flow network is mainly “controlled” by these few core nodes, and; (3) most other nodes are highly “dependent” on these few core nodes. Moreover, the node centrality value of the neighboring nodes of the core nodes have lower centrality values. Therefore, it can be assumed that the core nodes impose certain restrictions on the tourist flows of the surrounding attractions (i.e. structural holes in the tourist flow network are likely).

4.2.2 Structural Holes Measurement of Tourist Flow Network

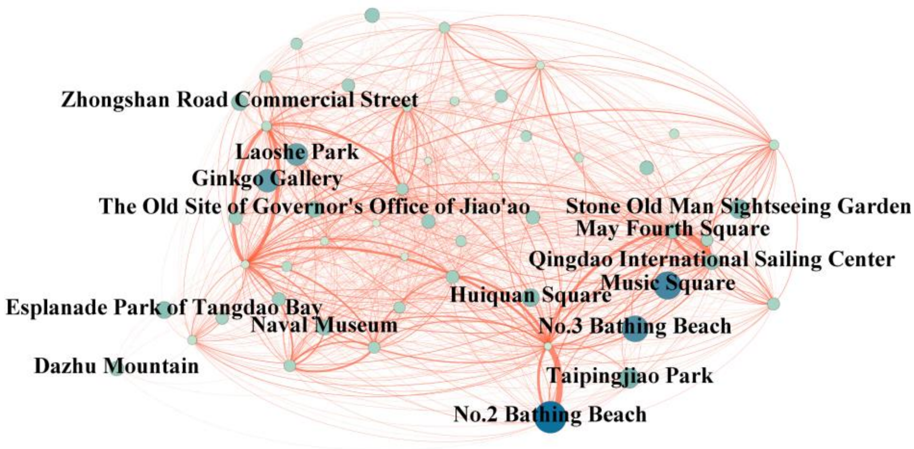
The effective size and constraint selected to measure the structural holes were calculated based on equations 8 and 9 using UCINET6. The results of a sample of the most relevant attractions are presented in Table 6, and a visualization of the tourist flow network, according to the effective size and constraint values, is shown in Fig. 8. The size of the nodes in the figure represents the level of the structural hole indicators and the thickness of the connection between the nodes indicates the volume of the tourist flow.

Table 6: Calculation results of the structural hole indicators

Attractions	Effective size	Constraint	Attractions	Effective size	Constraint
Eight Great Passes	40.96	0.18	Zhongshan Park	28.03	0.25
St. Michael’s Cathedral	36.06	0.21	Naval Museum	24.97	0.27
Landing Stage	35.65	0.18	The Old Site of Governor’s Office of Jiao’ao	21.53	0.28
Qingdao Beer Museum	35.06	0.18	No.2 Bathing Beach	20.11	0.47
Laoshan Mountain	34.85	0.20	Taipingjiao Park	17.78	0.32
May Fourth Square	33.99	0.27	Zhongshan Road Commercial Street	17.78	0.28
Qingdao International Sailing Center	33.58	0.27	No. 3 Bathing Beach	12.67	0.40
Golden Beach	30.75	0.20	Huiquan Square	10.22	0.30
Christian Churches	30.54	0.23	Laoshe Park	8.85	0.35
NO.1 Bathing Beach	30.27	0.24	Eplanade Park of Tangdao Bay	8.44	0.29
Signal Hill Park	29.54	0.20	Music Square	7.91	0.42
Sky City	29.23	0.21	Stone Old Man Sightseeing Garden	7.65	0.31
Luxun Park	28.78	0.23	Ginkgo Gallery	6.76	0.36



(a) Effective size



(b) Constraint

Figure 8. Network structure of tourist flows in Qingdao based on the calculated structural hole indicators (nodes with structural hole indicator values among the top 15 are indicated with names)

When comparing the results of structural hole indicators in Fig. 8(a) and Fig. 8(b), it becomes evident that the distribution of the effective size and constraint of the nodes in the Qingdao tourist flow network possess certain regular characteristics. For example, the Eight Great Passes, St. Michael’s Cathedral, the Landing Stage, Qingdao Beer Museum, Laoshan Mountain, and Golden Beach have the highest effective size values but lower constraint values, indicating that these attractions have obvious competitive advantages and are less affected by the tourist flows of their surrounding attractions. At the same time, their high effective size values and low constraint values have formed a stark contrast with the low effective size values and high constraint values of their surrounding nodes such as the No.2 Bathing Beach, Laoshe Park, the Stone Old Man Sightseeing Garden, the Esplanade Park of Tangdao Bay and Dazhu Mountain. This indicates that the development of tourist flows to and from these surrounding attractions has been limited and that the attractions are at a disadvantage in the competition for tourists. Additionally, both the effective size and constraint values of the May Fourth Square and Qingdao International Sailing Center are high, indicating that although they all have the competitive advantage in tourist flows, they are still greatly affected by the surrounding attractions. This is due to both of the two attractions being famous in Qingdao but being located in the “hot spot” of coastal tourism in Qingdao and, thus, affected by other popular, surrounding and close-by attractions (especially by each other). This indicates that there is a certain level of “competition” for tourist flows between the May Fourth Square and Qingdao International Sailing Center.

Consequently, it is advisable to strengthen the guidance and control of the tourist flows around the core nodes and to strengthen the traffic connections between them and their surrounding nodes to avoid the negative impacts of the structural hole phenomenon on the volume and distribution of tourist flows in Qingdao.

4.3 Spatial Patterns of Tourist Flows

Based on the above analysis of the characteristics and network structure of tourist flows, the spatial patterns of Qingdao tourist flows are summarized below according to our research framework (Fig. 1). A schematic diagram of the spatial patterns of tourist flows is presented in Fig. 9.

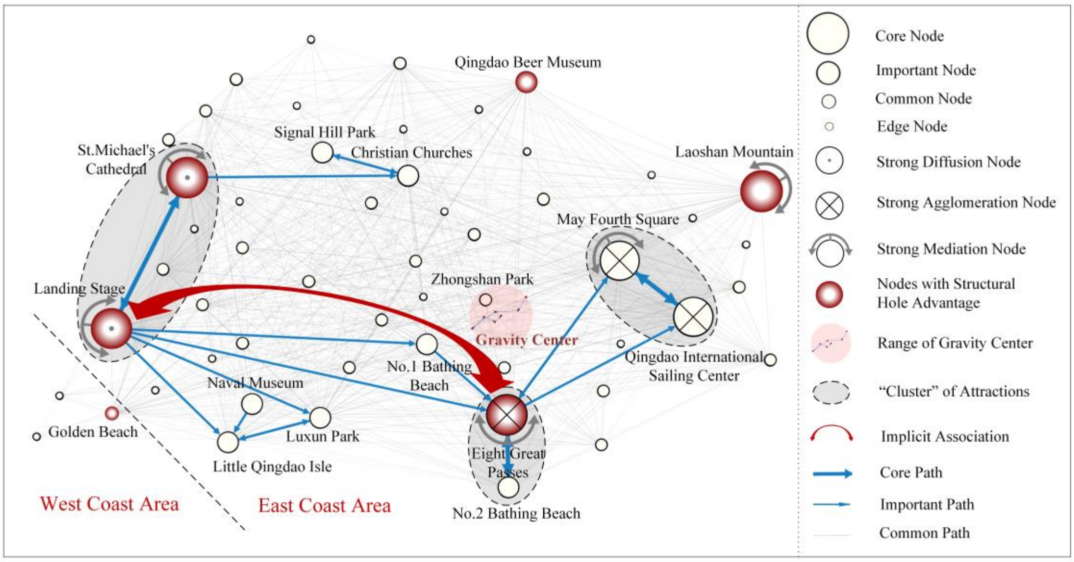


Figure 9. Spatial patterns of tourist flows in Qingdao

The spatial pattern of tourist flows in Qingdao is generally relatively “loose”, forming a network pattern with fierce competition and power imbalances between the destinations. The main characteristics of spatial patterns of tourist flows in Qingdao can be summarized as follows (Fig. 9):

(1) The Eight Great Passes, St. Michael’s Cathedral, the Landing Stage, Laoshan Mountain, the May Fourth Square and Qingdao International Sailing Center are the core destinations of tourist flows in Qingdao. Among them, the Landing Stage and the Eight Great Passes play the most central roles. There is also an implicit synergy between these two attractions. While the Landing Stage and the Eight Great Passes

have strong agglomeration (have high in-degree centrality values) and diffusion (have high out-degree centrality values) effects in the spatial pattern of tourist flows, they also connect other destinations in their surroundings to the network, thus, promoting the circulation efficiency of the tourist flows in Qingdao.

(2) Tourist flows in Qingdao are subject to distance decay: short-distance movements are much more common than long-distance ones.

(3) There are three prominent "clusters" of attractions with close internal connections but different internal relationships, which are revealed by structural holes measurement: affiliation in the case of the Eight Great Passes and the No.2 Bathing Beach, because the Eight Great Passes has high effective size value and low constraint value while No.2 Bathing Beach has low effective size value and high constraint value; competition in the case of the May Fourth Square and Qingdao International Sailing Center because of their high effective size and constraint values; and synergy in the case of the Landing Stage and St. Michael's Cathedral because of their high effective size values and low constraint values.

(4) The changes during 2012–2018 in the location of the gravity center (in the southern part of the East costa area) of tourist flows in Qingdao are small. This underlines that the spatial pattern of tourist flows in Qingdao is relatively stable, and that it is dominated by attractions in the south of East coast area, especially those along the coastline. Besides, the core destinations of the tourist flow network in Qingdao – the Eight Great Passes, the Landing Stage, the May Fourth Square and Qingdao International Sailing Center – are all well-connected to each other: together they form the "core area" of tourist flows along the Qingdao coastline. All of the above show that the "coastline" is a key feature in the spatial pattern of tourist flows in Qingdao, and, to a certain extent, has formed a "circulation obstacle" to the tourist flows to inland attractions, causing a huge difference in the scale of tourist flows between inland and coastal attractions. In short, Qingdao showcases the characteristics of a typical coastal tourist city (cf. Oppermann, 1992).

5 Discussion: Practical Implications

Based on the above results, it can be stated that the high concentration of tourism in coastal attractions is a constraint for developing tourism in inland attractions. Therefore, it is especially important for the future construction of Qingdao as a tourism city, to consider how to correctly handle the differences in tourism resources between coastal and inland areas, and how to settle the problem of uneven tourism development in coastal and inland attractions. From this perspective, stimulating the tourism potential of the western-style architectural attractions located in the inland area (e.g. St. Michael's Cathedral, the Old Site of Governor's Office of Jiao'ao and St. Paul's Church) might offer potential solutions – because such attractions are often less vulnerable to the negative impact of nearby attractions (i.e. they have lower constraint values) – to alleviate this mismatch. However, at present, only St. Michael's Cathedral occupies a dominant position in the spatial pattern of tourist flows in Qingdao. In the future, if the tourist flow connections between St. Michael's Cathedral and other similar attractions are strengthened (e.g. by combined marketing), it will be able to promote the overall vitality of Qingdao's inland tourism market.

The analysis of spatial patterns of tourist flows further reveals the differences in tourism resources between the East and West coast areas of Qingdao. Both the scale of the tourist flows, and the network structure are dominated by the attractions in the East coast area. The opening of the Jiaozhou Bay Subsea Tunnel has helped the development of tourism in the West coast area, at least to some degree, but the area still lacks strong attractiveness for tourists. Although the Golden Beach is the most prominent tourist flow node in the West coast area of Qingdao, its geographical advantage (only a 10-minute drive from the exit of the Jiaozhou Bay Subsea Tunnel) and the high structural hole level in the tourist flow network, to some extent, hinders the tourism development of the other attractions in this area. Thus, the overall development of tourism in the West coast area has been relatively slow. Measures to guide the tourists visiting Golden Beach also to other attractions in the West coast area, such as setting up sign-posts directing tourists to other attractions and a tourist publicity center focusing on the West coast area at Golden Beach, as well as providing convenient tour-buses, could potentially diminish Golden Beach's structural hole advantages in the West coast area. This would reduce the differences between the East and the West coast areas for more balanced and sustainable tourism development in Qingdao.

Judging from the evolution of the gravity center of tourist flows, the regional development imbalance of tourism in Qingdao has improved in recent years, but the change has been minimal. The tourism management department of Qingdao has formulated relevant policies for the balanced and sustainable development of tourism but considering that excessively biased policies may have a negative impact on the overall tourism development, the policy contents have understandably remained rather conservative[3]. Nevertheless, the tourism management department of Qingdao should still consider developing differentiated marketing strategies to meet the demands of tourists with varying preferences in the future to maximize the market potential that the diverse natural and cultural landscapes of the inland and coastal and the East and the West coast areas can offer.

As shown above (Fig. 3 and Fig. 9), our case study has (among other issues) revealed consistent regional imbalances in tourism resources. As such, tourism development needs to be considered as a long-term process that requires regular monitoring and evaluation. Our framework provides valuable tools to accomplish this.

6 Conclusion and Directions for Further Research

This paper proposed a novel research framework for analyzing the spatial patterns of tourist flows with tourists' digital footprint data collected from online travel diaries. The framework combines traditional quantitative methods of spatial analysis with social network analysis to examine the characteristics and network structure of tourist flows. It offers a comprehensive overview of the spatial patterns of tourist flows. We applied the proposed research framework to a case study city, Qingdao (China), and selected the online travel diaries data of Qunar.com as the digital footprint data source to explore the spatial patterns of tourist flows in 2012–2018. The conclusions can be summarized as follows:

(1) The spatial pattern of tourist flows is influenced by distance decay and attractions'

[3] http://qdsf.qingdao.gov.cn/n3707475/n32567547/n32567561/n32567606/190422144013034632.html

popularity, presenting a flow pattern with an identifiable gravity center and several core paths.

(2) In the spatial pattern of tourist flows, core nodes are unevenly distributed and the structural hole phenomenon is obvious, thus forming a network pattern with unbalanced power and intense internal competition. Tourism management departments should consider developing differentiated marketing strategies to meet the demands of tourists with varying preferences in the future to alleviate unhealthy internal rivalry and to maximize the overall market potential.

(3) In the spatial pattern of tourist flows, the core area is formed by important nodes and paths along the coastline, as is common in coastal tourism cities. This can lead to congestion problems. Traffic guidance around the core attractions in the coastal area still needs attention.

(4) The spatial patterns of tourist flows show differences in tourism resources within the city, such as the difference between coastal and inland areas. Concentrating efforts to reduce these differences for more balanced and sustainable tourism development is among the key point for future tourism management.

In summary, the framework is highly feasible and can be applied to other tourists' digital footprint data sources and in other case study locations. The limitations of the research framework and our subsequent suggestions for further research are as follows:

First, the digital footprint data source we use has some shortcomings. Online travel diaries data are mainly shared by tourists, but it is inevitable that some diaries are written by professional advertisers. Designing recognition rules to remove such travel diaries are urgently needed to improve the quality of the data. In addition, some of the less popular attractions in the inland areas, such as Daze Mountain, are not recorded in the online travel diaries data we obtained. This might be because online travel diaries are mostly written by individuals who take the cost of travel into account and, therefore, might not want to venture too far into the inland areas, which indicates that the coverage of the data still needs to be improved. Duggan (2015) pointed out that young and educated travelers are more likely to use these online travel sites. In the future, attempts to combine online travel data with official survey data – because the latter is based on a stratified random sample of the total population (LaMondia, Snell & Bhat, 2010; Yang, Tan & Li, 2019) – could significantly improve the precision of the data.

Second, some parts of the proposed research framework need to be strengthened. For instance, the CARMA algorithm we use was able to pinpoint only one type of implicit interaction between the Eight Great Passes and the Landing Stage: the algorithm may ignore similar phenomena between other attractions (especially popular attractions). Therefore, it is necessary to consider the improvement of the algorithm itself in the future to obtain more abundant and precise data mining information.

Finally, the current research framework focuses on the "static" analysis of tourist flows. In further analyses, it is necessary to incorporate the potential influence of the temporal aspects of tourist flows into the research framework. In addition, the research framework also lacks the richness that the non-spatial behavior information of tourists could offer. That is, the implicit information in the text data of online travel diaries, such as reviews closely related to tourists' choice of the attractions visited (Zhang et al., 2016), was not utilized here. This information can help to extend the geographical analysis in our framework to implement more detailed tourist flow analysis (e.g. analysis of the drivers of tourist flows).

References

Asakura, Y., & Iryo, T. (2007). Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A-Policy and Practice, 41*(7), 684-690. doi: https://doi.org/10.1016/j.tra.2006.07.003

Baggio, R., Scott, N., & Cooper, C. (2010). Network Science - A Review Focused on Tourism. *Annals of Tourism Research*, *37*(3), 802-827. doi: https://doi.org/10.1016/j.annals.2010.02.008

Brass, D., Butterfield, K., & Skaggs, B. (1998). Relationships and Unethical Behavior: A Social Network Perspective. *Academy of Management Review, 23*, 14-31. doi: https://doi.org/10.2307/259097

Burt, R. (1992). *Structural Holes: The social structure of competition*. Cambridge: Harvard University Press.

Casanueva, C., Gallego, Á., & García-Sánchez, M. (2016). Social network analysis in tourism. *Current Issues in Tourism*, *19*(12), 1190-1209. doi: https://doi.org/10.1080/13683500.2014.990422

Choi, S., Lehto, X. Y., & Morrison, A. M. (2007). Destination image representation on the web: Content analysis of Macau travel related websites. *Tourism Management*, *28*(1), 118-129. doi: https://doi.org/10.1016/j.tourman.2006.03.002

Duggan, M. (2015). The Demographics of Social Media Users. Pew Research Center: Internet, Science & Technology. Retrieved 2016-02-17 from: https://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/

Gao, Y., Ye, C., Zhong, X., Wu, L., & Liu, Y. (2019). Extracting Spatial Patterns of Intercity Tourist Movements from Online Travel Blogs. *Sustainability, 11*(13), 3526. doi: https://doi.org/10.3390/su11133526

Girardin, F., Calabrese, F., Fiore, F. D., Ratti, C., & Blat, J. (2008). Digital Footprinting: Uncovering Tourists with User-Generated Content. *IEEE Pervasive Computing, 7*(4), 36-43. doi: https://doi.org/10.1109/MPRV.2008.71

Hidber, C. (1999). Online association rule mining. *Sigmod Record, 28*(2), 145-156. doi: https://doi.org/10.1145/304181.304195

Hilgard, J. E. (1872). The advance of population in the United States. *Scribner's Monthly, 4*, 214-218.

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems, 54*, 240-254. doi: https://doi.org/10.1016/j.compenvurbsys.2015.09.001

Jin, C., Cheng, J., & Xu, J. (2018). Using User-Generated Content to Explore the Temporal Heterogeneity in Tourist Mobility. *Journal of Travel Research*, *57*(6), 779-791. doi: https://doi.org/10.1177/0047287517714906

Kádár, B. (2014). Measuring tourist activities in cities using geotagged photography. *Tourism Geographies, 16*(1), 88-104. doi: https://doi.org/10.1080/14616688.2013.868029

Kang, S., Lee, G., Kim, J., & Park, D. (2018). Identifying the spatial structure of the tourist attraction system in South Korea using GIS and network analysis: An application of anchor-point theory. *Journal of Destination Marketing & Management*, *9*, 358-370. doi: https://doi.org/10.1016/j.jdmm.2018.04.001

LaMondia, J., Snell, T., & Bhat, C. R. (2010). Traveler Behavior and Values Analysis in the Context of Vacation Destination and Travel Mode Choices. *Transportation Research Record, 2156*(1), 140-149. doi: https://doi.org/10.3141/2156-16

Lau, G., & Mckercher, B. (2006). Understanding Tourist Movement Patterns in a Destination: A GIS Approach. *Tourism and Hospitality Research*, *7*(1), 39-49. doi: https://doi.org/10.1057/palgrave.thr.6050027

Leung, X., Wang, F., Wu, B., Bai, B., A. Stahura, K., & Xie, Z. (2012). A Social Network Analysis of Overseas Tourist Movement Patterns in Beijing: The Impact of the Olympic Games. *International Journal of Tourism Research, 14*(5), 469-484. doi: https://doi.org/10.1002/jtr.876

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, *68*, 301-323. doi: https://doi.org/10.1016/j.tourman.2018.03.009

Li, Y., Yang, L., Shen, H., & Wu, Z. (2019). Modeling intra-destination travel behavior of tourists through spatio-temporal analysis. *Journal of Destination Marketing & Management, 11*, 260-269. doi: https://doi.org/10.1016/j.jdmm.2018.05.002

Li, Z., Jiang, W., Wang, W., Lei, X., & Deng, Y. (2019). Exploring spatial-temporal change and gravity center movement of construction land in the Chang-Zhu-Tan urban agglomeration. *Journal of Geographical Sciences, 29*(8), 1363-1380. doi: https://doi.org/10.1007/s11442-019-1664-5

Lian, T., & Yu, C. (2017). Representation of online image of tourist destination: a content analysis of Huangshan. *Asia Pacific Journal of Tourism Research, 22*(10), 1063-1082. doi: https://doi.org/10.1080/10941665.2017.1368678

Limtanakool, N., Dijst, M., & Schwanen, T. (2016). A Theoretical Framework and Methodology for Characterising National Urban Systems on the Basis of Flows of People: Empirical Evidence for France and Germany. *Urban Studies, 44*(11), 2123-2145. doi: https://doi.org/10.1080/00420980701518990

Liu, B., Huang, S. S., & Fu, H. (2017). An application of network analysis on tourist attractions: The case of Xinjiang, China. *Tourism Management, 58*, 132-141. doi: https://doi.org/10.1016/j.tourman.2016.10.009

Liu, F., Zhang, J., Zhang, J., Chen, D., Liu, Z., & Lu, S. (2012). Roles and functions of tourism destinations in tourism region of south anhui: A tourist flow network perspective. *Chinese Geographical Science, 22*(6), 755-764. doi: https://doi.org/10.1007/s11769-012-0557-6

Liu, P., Zhang, H., Zhang, J., Sun, Y., & Qiu, M. (2019). Spatial-temporal response patterns of tourist flow under impulse pre-trip information search: From online to arrival. *Tourism Management, 73*, 105-114. doi: https://doi.org/10.1016/j.tourman.2019.01.021

Liu, Y., Shi, J., & Jian, M. (2017). Understanding Visitors' Responses to Intelligent Transportation

System in a Tourist City with a Mixed Ranked Logit Model. *Journal of Advanced Transportation, 2017*(1), 1-13. doi: https://doi.org/10.1155/2017/8652053

Lue, C., Crompton, J. L., & Fesenmaier, D. R. (1993). Conceptualization of multi-destination pleasure trips. *Annals of Tourism Research*, *20*(2), 289-301. doi: https://doi.org/10.1016/0160-7383(93)90056-9

Ma, Y., Wang, Y., Xu, G., & Tai, X. (2018). Multilevel Visualization of Travelogue Trajectory Data. *ISPRS International Journal of Geo-Information*, *7*(1), 1-24. doi: https://doi.org/10.3390/ijgi7010012

Morley, C., Rosselló, J., & Santana-Gallego, M. (2014). Gravity models for tourism demand: theory and use. *Annals of Tourism Research*, *48*, 1-10. doi: https://doi.org/10.1016/j.annals.2014.05.008

Mou, N., Yuan, R., Yang, T., Zhang, H., Tang, J. J., & Makkonen, T. (2020). Exploring spatio-temporal changes of city inbound tourism flow: The case of Shanghai, China. *Tourism Management, 76*, 103955. doi: https://doi.org/10.1016/j.tourman.2019.103955

Önder, I., Koerbitz, W., & Hubmann-Haidvogel, A. (2014). Tracing Tourists by Their Digital Footprints: The Case of Austria. *Journal of Travel Research*, *55*(5), 566-573. doi: https://doi.org/10.1177/0047287514563985

Oppermann, M. (1992). Intranational tourist flows in Malaysia. *Annals of Tourism Research*, *19*(3), 482-500. doi: https://doi.org/10.1016/0160-7383(92)90132-9

Peng, H., Zhang, J., Liu, Z., Lu, L., & Yang, L. (2016). Network analysis of tourist flows: a cross-provincial boundary perspective. *Tourism Geographies*, *18*(5), 561-586. doi: https://doi.org/10.1080/14616688.2016.1221443

Pyo, S. (2015). Integrating tourist market segmentation, targeting, and positioning using association rules. *Information Technology & Tourism*, *15*(3), 253-281. doi: https://doi.org/10.1007/s40558-015-0029-9

Salas-Olmedo, H. M., Moya-Gomez, B., Garcia-Palomares, J. C., & Gutierrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*, *66*, 13-25. doi: https://doi.org/10.1016/j.tourman.2017.11.001

Shih, H. (2006). Network characteristics of drive tourism destinations: An application of network analysis in tourism. *Tourism Management*, *27*(5), 1029-1039. doi: https://doi.org/10.1016/j.tourman.2005.08.002

Smallwood, C., Beckley, L. E., & Moore, S. (2012). An analysis of visitor movement patterns using travel networks in a large marine park, north-western Australia. *Tourism Management*, *33*(3), 517-528. doi: https://doi.org/10.1016/j.tourman.2011.06.001

Versichele, M., de Groote, L., Bouuaert, M. C., Neutens, T., Moerman, I., & Van de Weghe, N. (2014). Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium. *Tourism Management*, *44*, 67-81. doi: https://doi.org/10.1016/j.tourman.2014.02.009

Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management*, *46*, 222-232.

doi: https://doi.org/10.1016/j.tourman.2014.07.003
Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports, 3*(1), 2976.
doi: https://doi.org/10.1038/srep02976
Xi, Y., & Yuan, Q. (2017). *Intelligent Recommendation Scheme of Scenic Spots Based on Association Rule Mining Algorithm.* Paper presented at the International Conference on Robots & Intelligent System, Los Alamitos, CA, USA.
doi: https://doi.org/10.1109/ICRIS.2017.53
Xia, J. C., Evans, F., Spilsbury, K., Ciesielski, V., Arrowsmith, C., & Wright, G. (2010). Market segments based on the dominant movement patterns of tourists. *Tourism Management, 31*, 464-469.
doi: https://doi.org/10.1016/j.tourman.2009.04.013
Yang, X., & Wang, Q. (2014). Exploratory Space-time Analysis of Inbound Tourism Flows to China Cities. *International Journal of Tourism Research, 16*(3), 303-312.
doi: https://doi.org/10.1002/jtr.1932
Yang, Y., Tan, K. P., & Li, X. R. (2019). Antecedents and consequences of home-sharing stays: Evidence from a nationwide household tourism survey. *Tourism Management*, *70*, 15-28.
doi: https://doi.org/10.1016/j.tourman.2018.06.004
Zeng, B. (2018). Pattern of Chinese tourist flows in Japan: a Social Network Analysis perspective. *Tourism Geographies*, *20*(5), 810-832.
doi: https://doi.org/10.1080/14616688.2018.1496470
Zeng, B., & He, Y. (2019). Factors influencing Chinese tourist flow in Japan - a grounded theory approach. *Asia Pacific Journal of Tourism Research*, *24*(1), 56-69.
doi: https://doi.org/10.1080/10941665.2018.1541185
Zhang, Z., Zhang, Z., & Yang, Y. (2016). The power of expert identity: How website-recognized expert reviews influence travelers' online rating behavior. *Tourism Management*, *55*, 15-24.
doi: https://doi.org/10.1016/j.tourman.2016.01.004
Zheng, Y., Zha, Z., & Chua, T. (2012). Mining Travel Patterns from Geotagged Photos. *ACM Transactions On Intelligent Systems and Technology, 3*(3), 56.
doi: https://doi.org/10.1145/2168752.2168770
Zhu, X., Sun, T., Yuan, H., Hu, Z., & Miao, J. (2019). Exploring Group Movement Pattern through Cellular Data: A Case Study of Tourists in Hainan. *ISPRS International Journal of Geo-Information, 8*(2), 74.
doi: https://doi.org/10.3390/ijgi8020074