Tourism Management 76 (2020) 103955

Contents lists available at ScienceDirect

Tourism Management

journal homepage: www.elsevier.com/locate/tourman

Exploring spatio-temporal changes of city inbound tourism flow: The case of Shanghai, China

Naixia Mou[a,b], Rongzheng Yuan[a], Tengfei Yang[c], Hengcai Zhang[b,*], Jinwen(Jimmy) Tang[d,e], Teemu Makkonen[f]

[a] College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China
[b] State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[c] Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China
[d] School of Economics and Commerce, South China University of Technology, Guangzhou, 510006, China
[e] School of Hospitality and Tourism Management, University of Surrey, Guildford, GU2 7XH, UK
[f] Karelian Institute, University of Eastern Finland, Joensuu, FI-80101, Finland

ARTICLE INFO

Keywords:
Inbound tourism flow
Spatio-temporal changes
Shanghai
Geotagged photo
Complex network

ABSTRACT

Knowledge on spatio-temporal changes of inbound tourism flow is important for destination economy, cultural communication and city image. This paper proposes a novel research framework for the spatio-temporal distribution and changes of inbound tourism flow by, first, using R-HDBSCAN clustering algorithm to extract tourism area of interest (AOI), second, by utilizing several key indicators adopted from the complex network theory literature to study the structure of inbound tourism flow with a case study example from Shanghai, China. The results show, first, that tourism in Shanghai is highly concentrated on the most popular AOI clustered in the city center relatively close to each other and, second, that, the inbound tourism flow network of Shanghai has small-world characteristics, while the distribution of its AOI (nodes) and tourist routes (edges) has general power law features, which has been influenced by the World Expo.

1. Introduction

Inbound tourism is a good indicator for assessing regional tourism competitiveness as it constitutes a significant part of local economic systems. The inbound tourism flow shows the spatio-temporal distribution characteristics of inbound tourists and reflects the dynamic changes of the attraction of tourist destinations to international tourists. Therefore, it has important guiding significance for the destination cities to develop their international tourism market and new tourism products.

In terms of data sources, the study of inbound tourism flow can be divided into three time periods: 1): the static, statistical data phase, 2) the network travel notes phase, and 3) the social media phase with geotagging. Most of the early studies on inbound tourism flow have been conducted with static statistical data, for example, by using statistical yearbooks (Asakura & Iryo, 2007) and surveys (Lew & McKercher, 2006) as data sources. Statistical yearbook data reflects the change in the number of inbound tourists in a certain area. However, it cannot reflect the spatial distribution and flow characteristics of inbound tourists. Moreover, it is impossible to count the no-visit spots with statistical yearbook data (Girardin, Fiore, Ratti, & Blat, 2008). In case of surveys, the samples are commonly poorly representative and, thus, it is difficult to accurately and comprehensively generalize and reflect the behavioral characteristics of the total population of inbound tourists (Onder, Koerbitz, & Hubmann-Haidvogel, 2014).

More recently, due to the emergence of blogs, (visitors often post trip diaries on the Internet and comment on scenic spots to share their travel experience), many scholars have utilized online trip diaries to study the spatio-temporal distribution characteristics of inbound tourists and the attractiveness of tourist destinations. For example, Leung, et al. (2012) utilize 500 travel notes on six travel websites to study the changes in the attractiveness of Beijing tourist attractions to inbound tourists, Manrique-Sancho, Avelar, Iturrioz-Aguirre, and Manso-Callejo (2018) study the urban spatial cognition of inbound tourists in Madrid and Ma, Yang, Xu, and Tai (2018) examine, with Ctrip travel notes data, the tourist movement patterns in travelogues of Taiwan's inbound tourists. However, there are two major shortcomings in the study of inbound tourism flow by using trip diaries: 1) the sparseness of the

* Corresponding author.
E-mail address: zhanghc@lreis.ac.cn (H. Zhang).

https://doi.org/10.1016/j.tourman.2019.103955
Received 9 April 2019; Received in revised form 25 June 2019; Accepted 22 July 2019

sample leads to an incomplete and inaccurate depiction of the inbound tourism flow; 2) the integrity and accuracy of the trip diaries can also potentially mislead the results of the analysis, since, for example, travel notes are often written by the author after the trip from memory.

Due to the recent popularity of social media, researchers have gained access to new data sources on inbound tourism flow. Visitors are more willing to use social media software (such as Flickr) to sign in and share photos of the scenery in their journey. For example, Flickr (the data source applied in this paper) is a popular geotagged photo site. Via Flickr data researchers can accurately get the check-in time information and geographical location information of the users' photos. Therefore, data mined from social media can overcome many of the shortcomings of earlier data sources (statistical yearbooks, surveys and online trip diaries) on inbound tourism flow. In this paper we propose a novel research framework, which utilizes these social media data to analyze the spatio-temporal distribution and change of inbound tourism flow in Shanghai (China) with Flickr data.

2. Literature review

There have been a number of studies applying Flickr data to investigate tourism (Li, Xu, Tang, Wang, & Li, 2018), since the increasing numbers of digital photos with geo-tags in Flickr offers great opportunities to study people's travel experiences, behavior and preferences (Yuan & Medel, 2016; Zhou, Xu, & Kimmons, 2015) and opens new possibilities for location based analysis (Sun, Fan, Helbich, & Zipf, 2013). Because Flickr data is freely available and has good spatial coverage, it solves several of the shortcomings (small sample sizes, static nature of yearbook data, etc.) associated with alternative data sources such as surveys, statistical yearbooks and online travel notes. Therefore, Flickr data has been utilized, for example, as a reliable indicator of the perceived image of destinations (Donaire, Camprubí, & Galí, 2014), to investigate the seasonality of tourism (Sun et al., 2013) and as a reliable source of data for market studies of tourism consumption (Iñiguez, Plumed, & Martinez, 2014; Latorre-Martínez, Iñíguez-Berrozpe, & Plumed-Lasarte, 2014). Further, with regard to data quality, for example, Wood, Guerry, Silver, and Lacayo (2013) demonstrate that Flickr data is a reliable data source for estimating tourist attraction visit rates, Girardin et al. (2008) compare Flickr and actual ticket data to show that Flickr data can be used as an index to reflect actual travel demand and number of tourists, Tenkanen et al. (2017) utilize Flickr data to show that visitation patterns based on social media data match relatively well with the official visitor counts in Finnish and South African nature reserves, while Cao et al. (2010) use Flickr data to build a global travel recommendation system to find tourism destinations of interests. As a point of departure, this paper aims to go (methodologically) beyond the existing literature on Flickr data and tourism (as discussed in greater detail below) by solving the problem of multiple sightseeing points in the tourism **Area of Interest (AOI)** extraction, and by incorporating perspectives from complex networks literature to the study of inbound tourism flow.

To use Flickr data to research inbound tourism flow in a regional scale it is necessary to extract the AOI of the destination city. Different from administrative divisions with clear boundaries, usually, an AOI contains multiple Point of Interest (POI). Therefore, the boundaries of AOI are often ambiguous: they may contain different types of areas, such as urban landmarks, commercial centers, recreational areas and scenic spots. In short, AOI is a complex of multiple functional types (Kuo, Chan, Fan, & Zipf, 2018) and, therefore, it is necessary to extract AOI through spatial clustering (Han & Lee, 2015). The most commonly used spatial clustering algorithms are K-Means Macqueen (1965), Mean-Shift (Comaniciu & Meer, 2002), Grid clustering (Wang, Yang, & Muntz, 1997), and density clustering (Ester, Kriegel, Sander, & Xu, 1996). For example, perform spatial clustering using the Flickr global data set and K-Means algorithm, Cao et al. (2010) use the Flickr global data set to recommend global travel-related AOI using the Mean-Shift clustering method, while Feick and Robertson (2015) utilize Flickr data to extract AOI using hexagonal mesh clustering.

However, because the shape of AOI is irregular, clustering methods based on centroid and mesh are not suitable for AOI extraction. Therefore, most scholars use density clustering algorithm to extract AOI. For example, Hu et al. (2015) use Flickr data to extract AOI from different cities using DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm, Kisilevich, Mansmann, and Keim (2010) use Flickr data to solve the nonidentity of numbers between photos and visitors using P-DBSCAN algorithm, and Du, Dong, Huang, and Ren (2016) use Flickr data to cluster AOI using the C-DBSCAN algorithm based on geographic knowledge background constraints. However, since density clustering algorithms cannot explain the clustering of different densities through fixed neighborhood radius, researchers need to adjust the threshold of neighborhood radius and set minimum number of samples. Campello, Moulavi, and Sander (2013) have solved the multi-density clustering of DBSCAN by using HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) algorithm. By converting DBSCAN into a hierarchical clustering algorithm, the user input is reduced to only one minimum cluster size. However, when using the HDBSCAN algorithm to extract the tourism AOI, there will still be clustering results that split the same AOI into multiple sightseeing points. Plasacristan, Gonzalezdiaz, Martinezcortes, and Diazdemaria (2019) used Gauss-Kernel-Density-based K + V-DBSCAN to solve this multi-density clustering problem by utilizing an improved DBSCAN algorithm to investigate tourist destination landmarks. However, the same AOI has many photo gathering areas, i.e. sightseeing points; if extracting tourism AOI through HDBSCAN and K + V-DBSCAN algorithms, the same tourism AOI could be divided into several sightseeing points. Therefore, other factors (e.g. urban road network) need to be considered to improve clustering accuracy.

When studying flow patterns of tourists between AOI, most researchers choose to investigate tourist routes between AOI by adopting such methods as Markov chains (Xia, Zeephongsekul, & Arrowsmith, 2009), cluster analysis (Asakura & Iryo, 2007), and sequential pattern mining (Jankowski, Andrienko, Andrienko, & Kisilevich, 2010). Inbound tourism flow has been classified, for example, by Önder (2017) who apply the improved LCF (Lue-Crompton-Fesenmaier) model with data from AustriaÖnder (2017), Vu, Li, Law, and Ye (2015) who use the Markov chain with data from Hong Kong and Qin et al. who utilize Markov random chain with data from Beijing (Qin et al., 2018). However, most of the above studies are qualitative: they analyze the types of tourism flows between AOI in a static setting, but do not quantitatively verify the dynamic network characteristics of tourism flows.

Consequently, **complex network theory** can provide a new perspective for inbound tourism flow research (Wu, Zhou, Xia, Chen, & Yu, 2018). A complex network can be described as “a network with nontrivial topological features and patterns of connections between their elements that are neither purely regular nor purely random” (Iñiguez et al., 2014). While there are some related examples – by applying complex networks theory, Miguéns and Mendes (2008) build a global travel network and discuss the importance of tourist area weights to global travel network connections, Baggio and Cooper (2010) use complex network analysis methods to study the competitiveness of tourism destinations, while Wu et al. (2018) introduce a complex network to construct an inbound tourism flow network for the city of Beijing for tourist route prediction and bus route design – the opportunities offered by the complex networks approach (see Baggio, Scott, & Cooper, 2010 for a review of network methods in the field of tourism) for the study of tourism flows have thus far mostly been restricted to related studies on transport systems (Tsiotas & Polyzos, 2015; Wang, Mo, Wang, & Jin, 2011). Moreover, the above research has not yet analyzed the dynamic evolution process of the inbound tourism flow network by showing how inbound passenger flow change.

Therefore, this study proposes a research framework for the spatio-temporal distribution and change of inbound tourism flow by adopting

R-HDBSCAN (Road network constraint Hierarchical Density-Based Spatial Clustering of Applications with Noise) algorithm to extracts the tourism AOI and by using the complex network studies as a reference for the spatio-temporal evolution characteristics of the inbound tourism flow network. The study, thus, has two main contributions to the existing literature. First, it solves the problem of multiple sightseeing points in the tourism AOI extraction, and, second, it also studies the inbound tourism flow changes from the perspective of complex networks.

3. Methodology

3.1. Sample

Shanghai is one of the four municipalities in China, which act as centers for China's technology, trade, information, finance and shipping. In 2017, Shanghai's GDP was ranked first in China and second among all Asian cities. It is a world-famous international metropolis (Deng, Liu, Dai, & Li, 2019) and a tourist destination with many famous attractions, such as Oriental Pearl, Nanjing East Road and the Bund. In 2018, the number of foreign tourists in Shanghai was 6,859,000.[1]

3.2. Data sources

Flickr is a widely used Internet photos sharing website (https://www.flickr.com/). It was founded in 2004, and the openness of its massive photos data has made it a recognized data source for social science research (Thomee et al., 2016). In addition to photo content data, Flickr photos usually contain descriptions of the photos themselves, that is, Metadata, which records auxiliary information such as photos number (Pid), photographer (Oid), shooting time (P_date), longitude (Lon), latitude (Lat), and user information (User_info). By utilizing Flickr API interface, this paper obtains the geotagged Flickr photos and their attribute data from January 1st, 2004 to December 31st, 2018 in Shanghai area (120.9327°E~122.2339°E, 30.6014°N~31.8480°N) resulting in a total of 639,895 data points of inbound tourists by screening the nationality of foreign visitors from the available user information (see also Appendix). The distribution of Flickr photos of all inbound tourists in Shanghai is shown in Fig. 1.

3.3. Research framework on inbound tourism flow

This paper proposes a research framework based on Flickr data for analyzing the spatio-temporal distribution and change of inbound tourism flow. Our framework includes, first, the adoption of R-HDBSCAN to obtain urban tourism AOI, extract AOI clustering center and construct inbound tourism flow network. Second, complex network indicators (such as node degree, average shortest path length and clustering coefficient) are used to calculate the topology and the characteristics of evolution of inbound tourism flow network. An overview of the research framework is shown in Fig. 2. The steps (data cleaning, the extraction of the AOI with R-HDBSCAN, building the network of inbound tourism flow and calculating network indexes) included in our research framework are discussed in greater detail below.

3.3.1. Data cleaning

Flickr data often has some errors in location and time. Therefore, this paper applies specific rules to clean the data. An overview of the data cleaning procedure (five rules) and the subsequent loss of data after each step of data cleaning are presented in Table 1. After implementing all our five cleaning rules, we have obtained 666,970 valid records (photos) from 6589 foreign tourists.

To be more precise our data cleaning rules are as follows: 1) When acquiring data through the API interface, this paper divides Shanghai into 50*50 grids according to the longitude and latitude range and obtains 639,895 original points. However, there is repeated data at the junction of each block of data, that is, data repeated by Pid are removed. 2) For an individual visitor, data recorded at same point in time but at different location is recorded as incorrect, because the same visitor cannot be present in different locations at the same time. 3) A considerable part of the photos may be taken by resident foreigners such as international students (Ma et al., 2018; Manrique-Sancho et al., 2018). If the time interval between the first and last photos taken by a user is more than one month, the photographer is determined to be a resident foreigner and excluded from our sample. Otherwise the photographer is determined to be a tourist and included in our sample. 4) If the same user is in the same position and the photos are repeatedly taken within one minute, it is concluded that the user repeats the shooting data, which are removed from the sample. 5) For the purposes of establishing an inbound tourism flow network, only locations with photos from at least two visitors are chosen, while removing "lone point" data.

3.3.2. The extraction of AOI tourism

(1) HDBSCAN

HDBSCAN algorithm is a combination of hierarchical clustering and density clustering, which uses variable mutual reach distance to separate clustering results of different densities from sparse noise points (Campello et al., 2013; Mcinnes & Healy, 2017; McInnes et al., 2017). The mutual reach distance $d_{mreach-k}(a, b)$ is defined:

$$d_{mreach-k(a,b)} = \max\{core_k(a), core_k(b), core_k(a, b)\} \tag{1}$$

Here, $core_k(a)$ represents the cluster radius in the DBSCAN algorithm of a under the conditions of minimum clustering parameter k. while $d(a, b)$ indicates the Euclidean distance between a and b. Through using $d_{mreach-k}(a, b)$ distance matrix, HDBSCAN generates a tree diagram, as shown in Fig. 3.

By introducing the concepts of cluster stability, which measure the persistence of a cluster at each level of the tree diagram construction, the cluster extraction task is regarded as an optimization problem. For a given cluster, we can define values λ_{birth} and λ_{death} to be the value when the cluster split off and became its own cluster; λ_p is the value at which that point fell out of the cluster. This cluster stability is defined as:

$$\sum_{p \in cluster} (\lambda_p - \lambda_{birth}) \tag{2}$$

When extracting the HDBSCAN cluster center we can see that some of the identified tourism AOI contain multiple sightseeing points. When clustered, a tourism AOI will, thus, be divided into several parts, as shown by the example of the clustering results of Nanjing East Road, Lujiazui, Expo Park and Yu Garden (Fig. 4): the multiple sightseeing points on each map actually present the same tourism AOI surrounded by urban secondary roads. Therefore, this paper adopts the R-HDBSCAN algorithm to solve the above problems.

(2) R-HDBSCAN

The specific description of the R-HDBSCAN algorithm is as follows (see also Appendix):

(1) Obtain the original cluster center by using the HDBCAN algorithm.
(2) Obtain the urban block plot by utilizing the urban secondary road network.
(3) Obtain the final cluster center by spatially associating the urban block plot with the original cluster center, and constraining the

[1] http://lyw.sh.gov.cn/lyj_website/HTML/DefaultSite/lyj_xxgk_lytj_2018/2019-03-25/Detail_141455.htm.

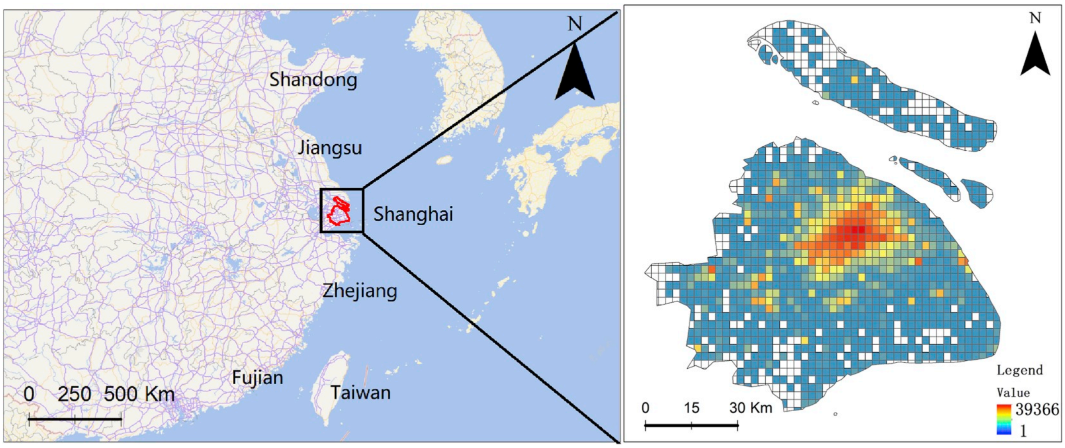


Fig. 1. The distribution of Flickr photos in Shanghai (source: authors' elaboration from Google Map).

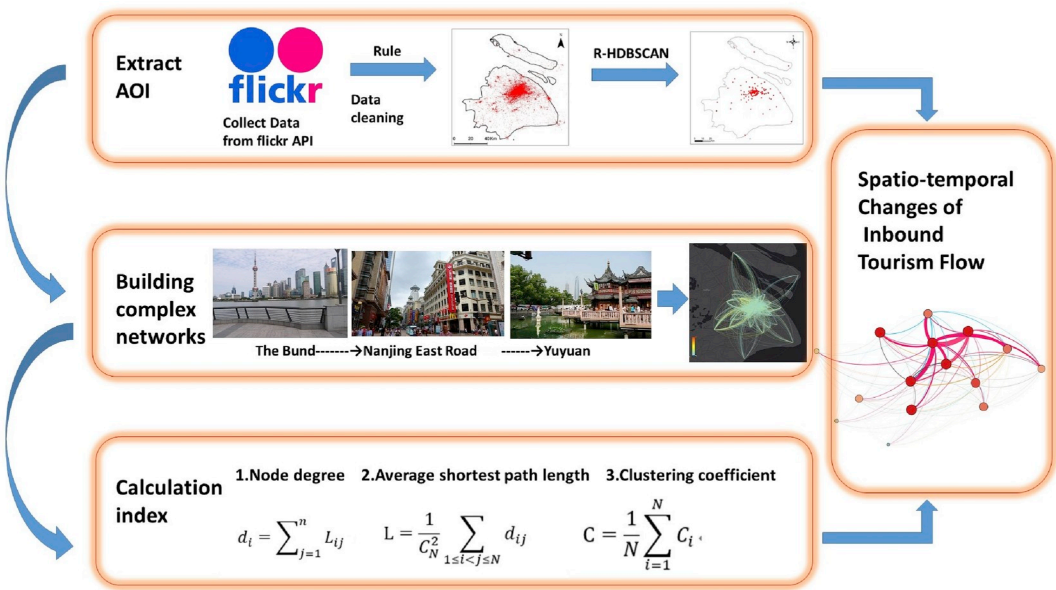


Fig. 2. The research framework on inbound tourism flow based on Flickr data (source: authors' elaboration).

Table 1
The results of data cleaning.
source: authors' elaboration.

Cleaning condition	Field selection condition	Remaining amount of data points	Number of foreign tourists the photos belong to
Repeated obtaining	Pid is the same	632,684	26,895
Error data of the same user	Lat, Lon, Oid, and P_date are the same	610,447	26,895
Excluding international students	Staying longer than 1 month	320,329	25,367
Repeated data captured by the same user	Lat, Lon, Oid are the same, and the interval is less than 1 min	239,620	16,672
Solitary point data	Oid is the same and the total number of Pid is less than 2	66,970	6,589

original cluster center based on the urban secondary road network.

Definition 1. City block plot:

The urban block plot is generated according to the urban secondary road network segmentation, A city block plot can be expressed as $B_i(i = 1,2..., N)$. Here, N is the serial number of the parcel. The steps for generating it are as follows: under the premise of retaining the connectivity of the road, the center line of the city's secondary road network is extracted to simplify the interpretation. The city's secondary road network is used to divide the land parcels and the simplified land elements are processed.

Definition 2. Original cluster center

The original cluster center obtained using HDBSCAN can be expressed as $C_i(i = 1,2..., N)$. Here, N is the number of the original cluster center.

Definition 3. Spatial association rules:

Each city block B_i may contain 0, 1 or more original cluster centers

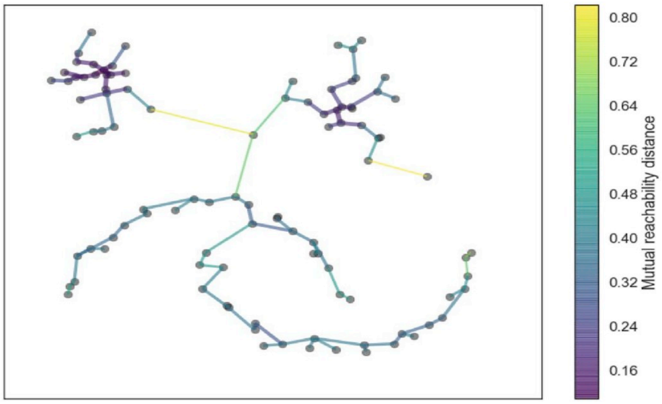


Fig. 3. The tree diagram generated by distance matrix (source: McInnes, Healy, & Astels, 2017).

3.3.3. Network of inbound tourism flow

(1) Construction of complex network

In addition to tourist attractions, this paper also retains some hot spots closely related to tourism, such as airports, hotels, shopping malls, etc. to fully present the distribution of tourism flow, and to regard the cluster centers of these AOI as nodes in complex networks. According to the time sequence of inbound tourists visiting AOI, the travel route of each inbound tourist is generated. If the travel route of a tourist is {Bund→Nanjing East Road→Yu Garden→Oriental Pearl→Expo Park}, the travel route between any two tourism AOI is regarded as the undirected edge between the pairs of nodes in the network, and the number of the tour paths is the weight of the edges in the network.

(2) Computation of complex network indicators

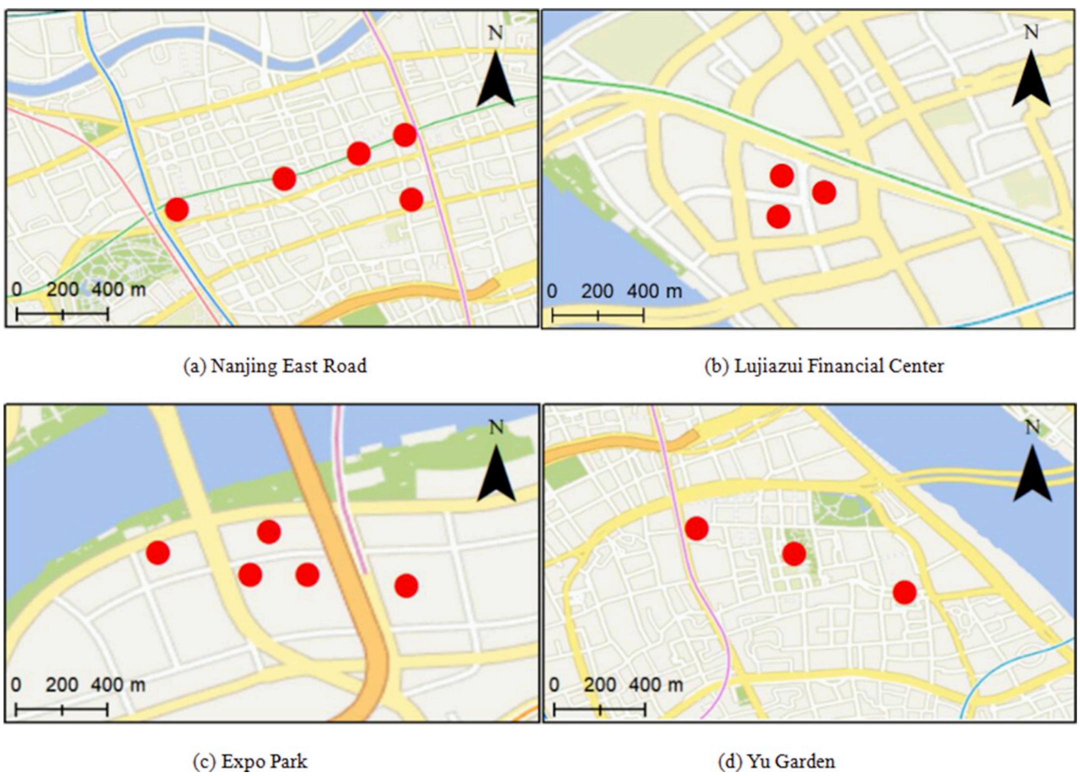


Fig. 4. An example of HDBSCAN clustering results for Fkickr data (source: authors' elaboration).

C_iC_i. When B_i contains more than two C_i, these C_i will be treated as different sightseeing points for the same AOI. According to the four-intersection model (Egenhofer & Franzosa, 1991), city block plot B_i and original clustering center C_i can be regarded as a polygon and a point entity in a two-dimensional space respectively, and the spatial relationship between the two entities can be determined by whether the intersection of the boundary and the internal space of the two is empty (ϕ) or non-empty (-ϕ). If ∂K and K^0 represents the boundary and interior space respectively, then the boundary and internal space of B_i and C_i can be expressed as ∂B, B^0 and ∂C, C^0, and the relationship between the four groups is expressed as:

$$SR_4(B_i, C_i) = \begin{pmatrix} \partial B \cap \partial C & \partial B \cap C^0 \\ B^0 \cap \partial C & B^0 \cap C^0 \end{pmatrix} \tag{3}$$

If $SR_4(B_i, C_i) = \begin{pmatrix} \Phi & \Phi \\ -\Phi & -\Phi \end{pmatrix}$, then it is considered that the block of the city block B_i contains the original cluster center C_i.

In a complex network, node degree reflects the importance of nodes in the network, the shortest path length reflects the average distance between any node and the overall efficiency of the network, and clustering coefficient reflects the tight connection of nodes in the network (Mou et al., 2018). These indicators are applied here to assess the status of tourism AOI and the structural characteristics of Shanghai inbound tourism flow network.

1) Node degree and weighting

The degree of a node is defined as the number of connections between a node and other nodes. The greater the degree of a node is, the more important this node is in the network. The calculation formula of the node degree is as follows (Barrat, Barthelemy, Pastor-Satorras, & Vespignani, 2004):

$$d_i = \sum_{j=1}^{n} L_{ij} \tag{4}$$

Here, L_{ij} is the number of edges between node i and j, and n is the total number of nodes.

In the inbound tourism flow network, the weighting degree can reflect the frequency of contact between various scenic spots in Shanghai. The calculation formula of the weighting degree is (Li, Xu, & Shi, 2015):

$$S_i = \sum_{i \in N_i} W_{ij} \tag{5}$$

Here, N_i is a collection of adjacent points of node i, and W_{ij} is the weight of the side for connecting node i and node j, that is, the number of inbound tourists. If the node i and node j have no connections, $W_{ij} = 0$.

2) Average shortest path length

The average shortest path length of the network is defined as the average of the distance between any two nodes. The formula is as follows (Yang & Knoke, 2001):

$$L = \frac{1}{C_N^2} \sum_{1 \le i < j \le N} d_{ij} \tag{6}$$

Here, d_{ij} indicates the number of sides of the node's shortest path connecting node i and node j. When there is no connection between node i and node j, d_{ij} is considered infinite. For inbound tourism flow networks, the smaller the L is, the smaller the number of tourism AOI that need transiting during the tour, and the higher the accessibility of the inbound tourism flow network.

3) Clustering coefficient

The ratio of the number of edges E_i that actually exist between the k_i neighbor nodes of node i and the total number of possible sides $C_{k_i}^2$ is defined as the clustering coefficient of the node. The formula is (Opsahl & Panzarasa, 2009):

$$C_i = \frac{E_i}{C_{k_i}^2} \tag{7}$$

The clustering coefficient of the node indicates the tightness of the interconnection between the nodes in the network. The larger the ratio, the higher the proportion of the nodes directly connected in the local area where the node is located, and higher the number of tourism AOI for the inbound tourists to choose from. Clustering coefficient of the entire network C represents the average of the clustering coefficients of all nodes in the network summary. The formula is as follows (Newman, 2003):

$$C = \frac{1}{N} \sum_{i=1}^{N} C_i \tag{8}$$

4. Results

4.1. Temporal characteristics

The Flickr data is counted separately by year and month to calculate the number of photos by inbound tourists. As shown in Fig. 5(a), the number of photos by inbound tourists reached its peak in 2010 (the World Expo was host from May to October of that year). During the period from 2004 to 2009, the number of photos by inbound tourists was very small but growing. From 2010 to 2013, the number of photos by inbound tourists increased after a short-term decline, but since 2013, the number has declined steadily. Therefore, we divide our observations on Shanghai tourism into three periods, namely, the pre-World Expo period (2004–2009), the World Expo and the subsequent "heat period" (2010–2013) and the post-World Expo period (after 2013). According to the monthly statistics of the total number of photos by inbound tourists from 2004 to 2018, we can see that inbound tourism in Shanghai has obvious seasonal characteristics. As shown in Fig. 5(b), inbound tourists choose to travel to Shanghai in March–May and September–November, because this period is the most suitable months in terms of the weather in Shanghai.

The overall staying time of inbound tourists from 2004 to 2018 in our Flickr data is shown in Fig. 6(a). Visitors staying for one day accounted for 69.79%, and tourists staying for two days accounted for 11.9%, with a distinct long-tailed distribution, which indicates that the vast majority of tourists are short-haul tourists (Shanghai with its two busy international airports is a well-known transfer place for inbound tourists). According to the three stages obtained from the characteristics of annual variation, the inbound visitors' stay time in the pre-Expo period, the World Expo and the subsequent "heat period" and the post-World Expo period are calculated respectively:as shown in Fig. 6(b), (c) and (d), the length of stay of inbound tourists do not change significantly.

4.2. Spatial characteristics

The urban secondary road network used for road network constraints is shown in Fig. 7(a). The obtained urban block plots are shown in Fig. 7(b), while Fig. 7(c) shows the urban plots on both banks of the Huangpu River. A total of 161 tourism AOI cluster centers were obtained by the R-HDBSCAN algorithm. The cluster centers are shown in Fig. 7(d).

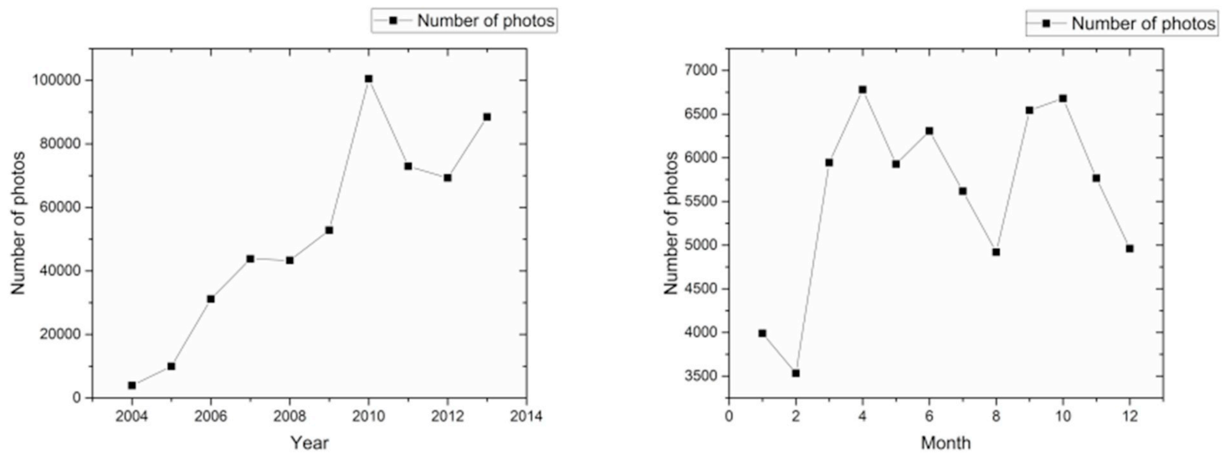
The results of spatial clustering reveal the spatial distribution, shown in Fig. 8, of the main tourism AOI in the Shanghai Third Ring Road area. The figure reveals that the tourism AOI are mainly distributed within the outer ring line. The tourism AOI within the inner ring line are highly concentrated, mainly on the banks of the Huangpu River, including AOI such as the Bund, Nanjing East Road, Yu Garden, Oriental Pearl, Nanjing West Road, Tianzifang, Jing'an Temple, Zhongshan Park and Xujiahui. Contrarily, the number of tourism AOI in the outskirts of the city is small and the few AOI are scattered. These AOI include, for example, Lushan National Forest Park, Drip Lake, Dongping National Forest Park on Chongming Island and Xisha Wetland. The number of inbound tourists in the AOI decreases when moving away from the city center (the Bund and Yu Garden) to more faraway AOI.

4.3. Network characteristics

4.3.1. Network characteristics of Shanghai inbound tourism flow

With the extracted tourism AOI data from Flickr, we can construct Shanghai inbound tourism flow network, which is shown in Fig. 9:

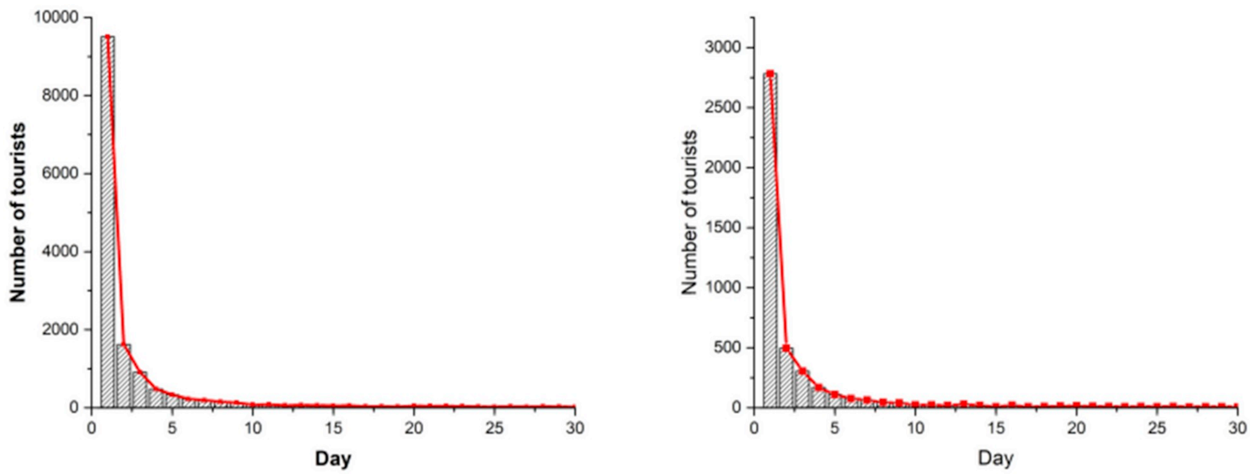
The basic characteristics, as informed by indicators derived from the complex network theory literature, of the inbound tourism flow network in Shanghai are shown in Table 2. Altogether, we identified 161 tourism AOI cluster centers and 601 sightseeing routes with our Flickr data. The average tourism AOI value of the entire inbound tourism flow network (based on equation (4)) is 6.642, indicating that the average number of tours through any of the AOI in the Shanghai inbound tourism network is about six; i.e. on average tourists visit six AOI. The average weighting degree (based on equation (5)) is 11.034; i.e. on average a tourist route, in our Flickr data, in Shanghai has 11 travelers. The average shortest path length of inbound tourism flow network (based on equation (6)) is 2.259, while the average clustering coefficient (based on equation (8)) of the network is 0.587 Thus, since the inbound tourism network has a small average shortest path length (2.259) and a large clustering coefficient (0.587), the Shanghai inbound tourism flow network can be seen as a small world network (Watts & Strogatz, 1998).



(a) Characteristics of annual variation

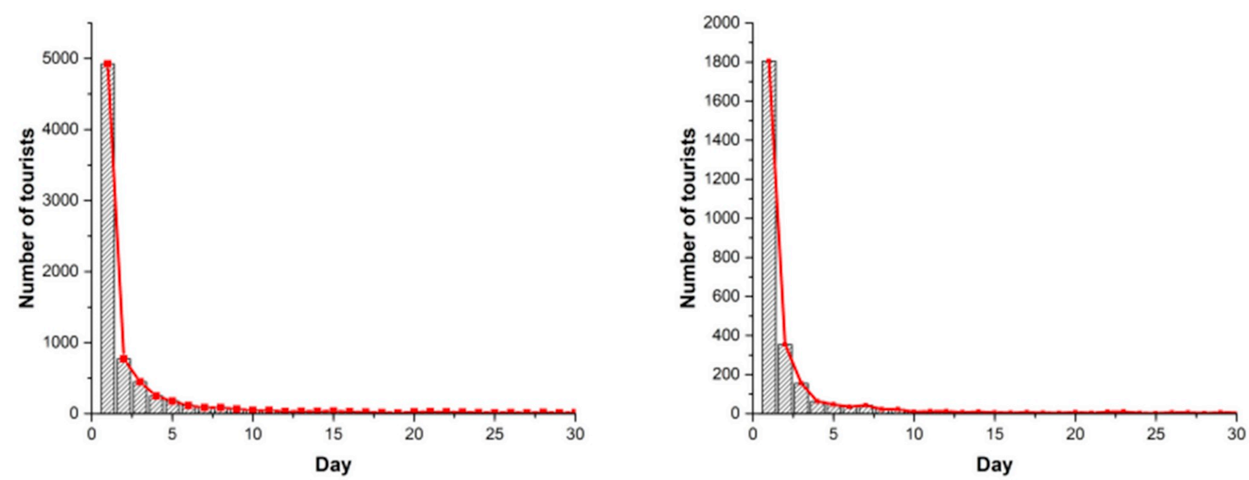
(b) Characteristics of monthly variation

Fig. 5. Number of photos by inbound tourists to Shanghai in 2004–2018 (source: authors' elaboration).



(a) Overall staying time

(b) Pre-World Expo period



(c) World Expo and subsequent "heat period"

(d) Post-World Expo period

Fig. 6. Length of stay of inbound visitors (source: authors' elaboration).

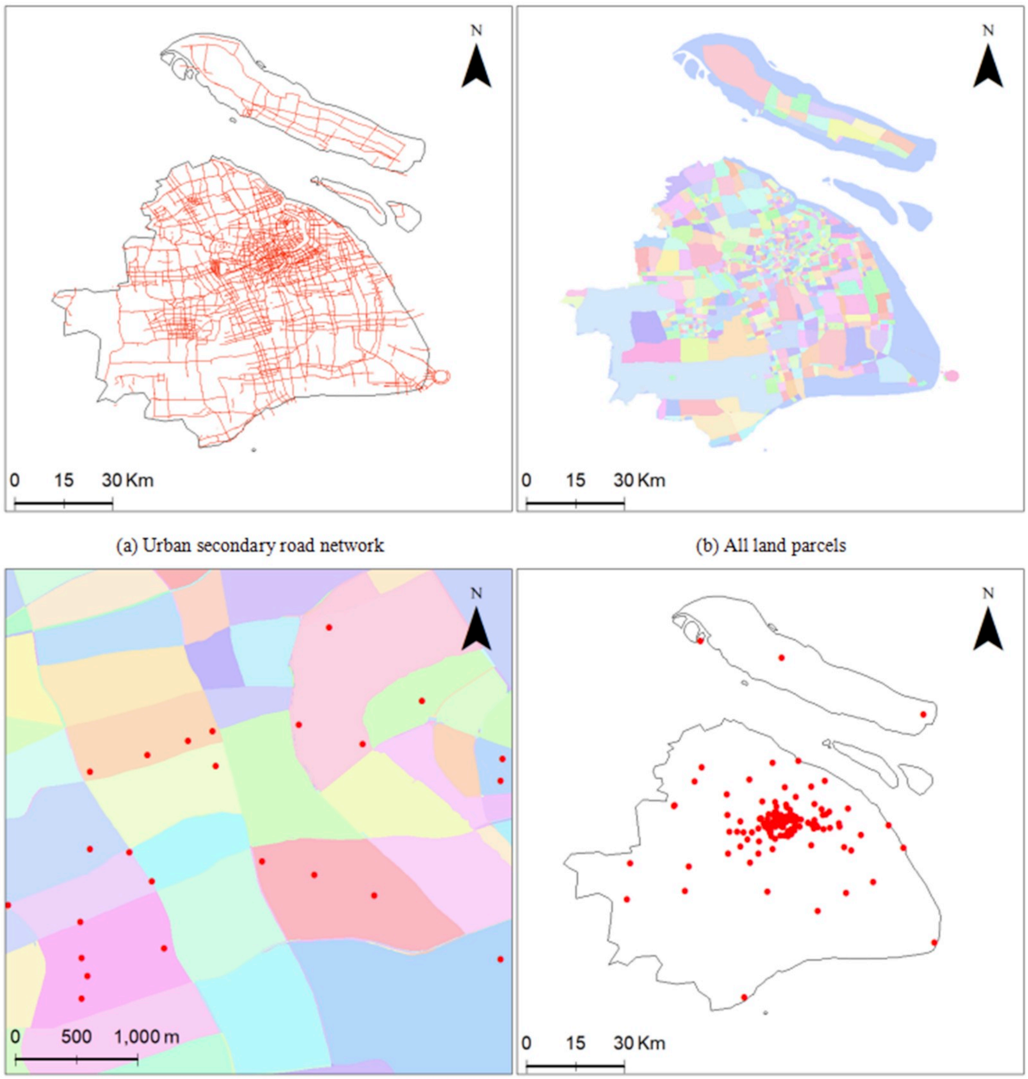


Fig. 7. R-HDBSCAN clustering results of Flickr data (source: authors' elaboration).

The degree value and degree distribution curve of each tourism AOI is shown in Fig. 10. The largest node degree value in the network is 83. The node degree distribution of the Shanghai inbound tourism flow network basically satisfies the power rate distribution. As such, the Shanghai inbound tourism flow network has scale-free characteristics, indicating that most tourists are concentrated in a few of the most popular tourism AOI.

4.3.2. Dynamic characteristics of inbound tourism flow in Shanghai

The changes in Shanghai's top-10 inbound tourism AOI and the main tour routes are shown in Tables 3 and 4.

The results of the visualization of the complex network structure of the inbound tourism flow network under the designated three time periods of the pre-World Expo period, the World Expo and the subsequent “heat period” and the post-World Expo period are shown in Figs. 11–13.

(1) During the pre-World Expo period, the preferred AOI for inbound tourists are the Bund, Nanjing East Road, Yu Garden and People's Square. The main sightseeing route were located between the Bund and Yu Garden, the Bund and Nanjing East Road, the Bund and the Oriental Pearl. The links between the AOI were weak, which shows that before the World Expo, Shanghai's inbound tourists was mainly concentrated in the Bund, Yu Garden and Nanjing East Road.
(2) During the World Expo and the subsequent “heat period”, the link between the Bund and the World Expo became the most important tourist route for Shanghai's inbound tourism. Thus, the hosting of the World Expo undoubtedly changed the structure of the Shanghai inbound tourism network. It can, thus, be concluded that World Expo played a positive role in promoting Shanghai tourism.
(3) In the post-World Expo period the Shanghai inbound tourism network becomes more complicated. In addition to the traditional “best performing” AOI such as Bund, Yu Garden, Nanjing East Road, Expo Park, Oriental Pearl, there are now also emerging tourism AOI such as Tianzifang, Lujiazui, Zhongshan Park and Jing'an Temple. This indicates that the scale of inbound tourism in Shanghai is constantly expanding, the number of tourist routes is increasing and that tourist have more AOI to choose from.

5. Discussion

The number of tourists counted according to Flickr data show a distinct peak during the World Expo and the subsequent “heat period”. However, these peak figures have started to gradually decrease during

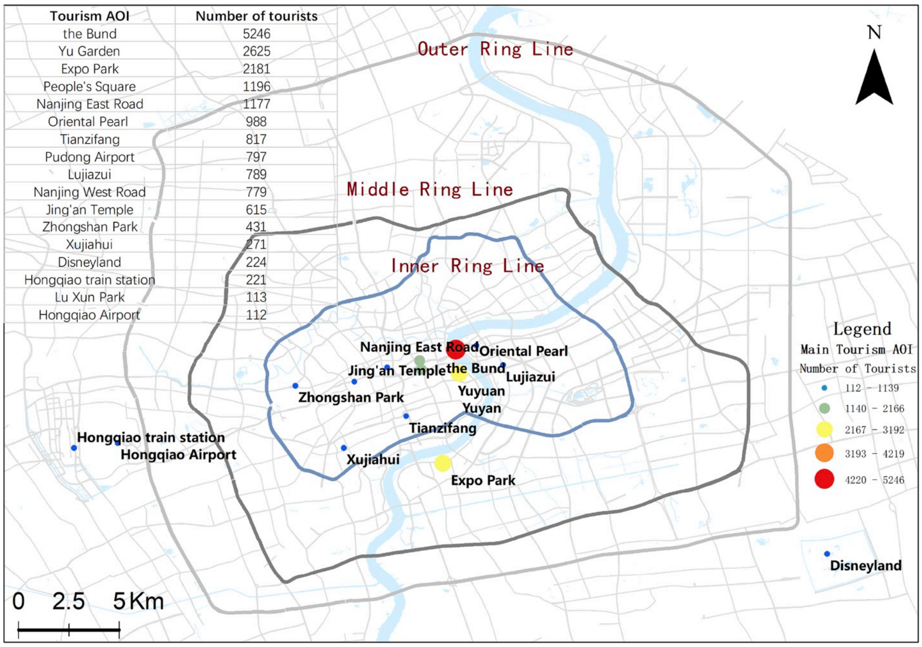


Fig. 8. Distribution of main tourism AOI in the Third Ring Road area (source: authors' elaboration).

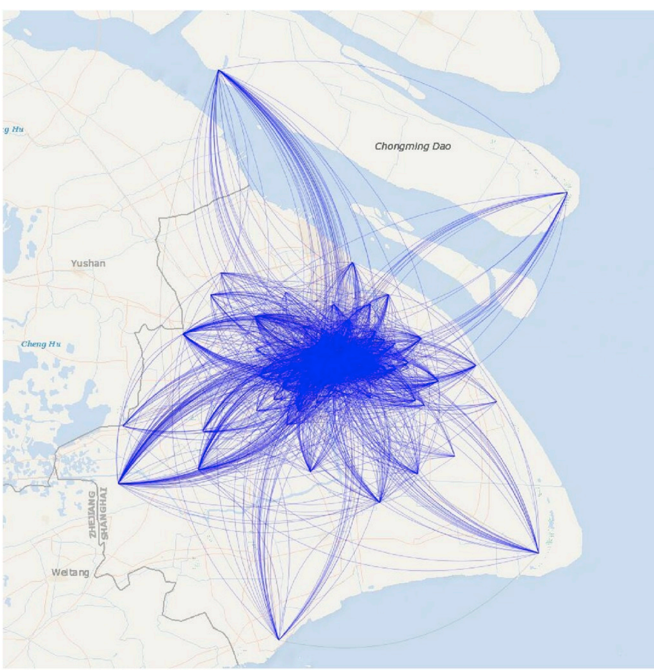


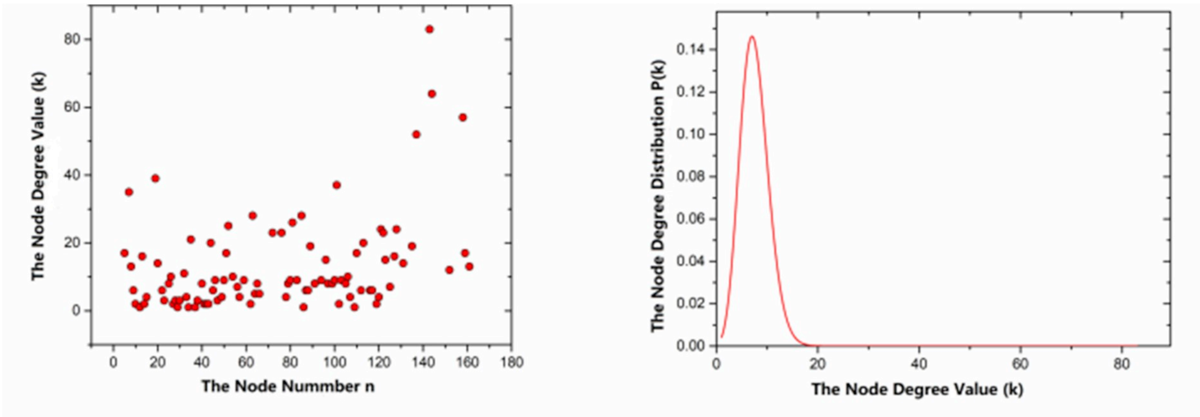
Fig. 9. Shanghai inbound tourism network (source: authors' elaboration).

the post-World Expo period (after 2013). Additionally, the World Expo influenced the tourism industry of Shanghai, by expanding the number of popular more AOI visited by inbound tourists. This is in line with earlier notions indicating that large-scale international events (at the Expo level) can help to shape the image of the host country and the government (Chen, 2012). However, the popularity of the Shanghai World Expo lasted for three years only, and thus, for example, Xue, Chen, and Yu (2012) believe that while the World Expo improved the attitude of overseas media towards Shanghai, a single event, like the World Expo, will not have long-term effects.

Notwithstanding, the World Expo was an excellent window and business card to introduce China and Shanghai to the world. It also promotes cultural exchanges between countries and attracted high numbers of overseas visitors. The Expo Pavilion itself also became a new popular tourism AOI. Still, inbound tourists are only interested in a few main AOI, which can be divided into two categories: 1) the traditional tourism AOI with high visibility in the destination city, such as the Bund, Nanjing East Road and Yu Garden, 2) the Expo Exhibition Hall and Expo Park developed for the World Expo. This indicates that inbound tourists might not know enough about Shanghai's tourism resources. At the same time tourist might also have time constraints: visitors staying only for a short period of time are more likely to visit (only) the most central AOI in Shanghai. Therefore, Shanghai should strengthen the links between the tourism AOI of the central city by developing a more suitable tourist route for short-term trips for maximizing the income from inbound tourists.

Table 2
Basic characteristics of Shanghai inbound tourism flow network.
source: authors' elaboration.

Network name	Network Type	Number of AOI	Tour path	Average	Average weighting	Average shortest path	Network diameter	Average clustering coefficient
Overall network	Undirected	161	601	6.642	11.034	2.259	3	0.587
Pre-World Expo period	Undirected	153	516	5.96	10.03	1.595	3	0.515
World Expo and subsequent “heat period”	Undirected	154	568	6.804	13.055	2.991	3	0.687
Post-World Expo period	Undirected	161	601	6.625	11.953	2.744	2	0.593



(a) Degree value of AOI (b) Degree distribution curve of AOI

Fig. 10. Degree value and degree distribution curve of AOI (source: authors' elaboration).

Table 3
The ranking of Shanghai's inbound tourism AOI.
source: authors' elaboration.

Ranking	Pre-World Expo period	World Expo and subsequent “heat period”	Post-World Expo period
1	The Bund	Expo Park	The Bund
2	Nanjing East Road	The Bund	Nanjing East Road
3	Yu Garden	Nanjing East Road	People's Square
4	People's Square	People's Square	Yu Garden
5	Pudong Airport	Yu Garden	Oriental Pearl
6	Nanjing West Road	Oriental Pearl	Expo Park
7	Tianzifang	Pudong Airport	Tianzifang
8	Oriental Pearl	Tianzifang	Pudong Airport
9	Jing'an Temple	Nanjing West Road	Lujiazui
10	Hongqiao train station	Hongqiao train station	Nanjing West Road

Shanghai's inbound tourism flow network is a small-world network. By changing a few key nodes, the characteristics of the network can be drastically changed. For example, owing to the hosting of the World Expo, the number of visiting routes increases by 52 and the average clustering coefficient increases from 0.515 to 0.687. The distribution of AOI (nodes) and tourist routes (edges) of the inbound tourism flow network follows a general power law distribution: in Shanghai 87.78% of foreign tourists concentrating in the top 20 popular AOI, while only 12.22% of tourists visited “non-hot” AOI. The structure of the inbound tourism flow network is relatively stable: the Bund, Nanjing East Road and Yu Garden have consistently been the most important tourist nodes during the time period analyzed in this paper. An interesting side notion is that the globally famous theme park, Shanghai Disneyland (opened in 2016) has not yet reached the status of a popular inbound tourism AOI. The main reason behind this observation may be that there are five other Disneyland in the world (California, Orlando, Hong Kong, Tokyo, and Paris). Therefore, Shanghai Disneyland does not have the unique charm of Shanghai, since inbound tourists want to experience a different culture from their own country and broaden their horizons and experience different customs. Shanghai should vigorously develop tourism AOI with Chinese elements and Shanghai characteristics and hold corresponding international tourism events.

In terms of the advantages of our proposed research framework, we can conclude that, first, our decision to adopt the R-HDBSCAN algorithm to investigate Flickr data and tourism solves the problem of multiple sightseeing points in the tourism AOI extraction. Second, the incorporation of indicators derived from the complex network theory literature into our analysis provides us with novel insights into the dynamics and characteristics of the inbound tourism flow network that can benefit:

1) Inbound tourists to understand the layout of tourism AOI for improved route planning.
2) Travel companies to develop travel routes and tourist products.
3) Tourism administration to understand the travel characteristics of foreign tourists for improved of tourism space layout design and more accurate destination marketing.

The limitations of this study are as follows: 1) Flickr is more popular in Western countries than in e.g. African or Asian countries, 2) data coverage of online photography is far from perfect, since not all tourists post photos online (Lo, Mckercher, Cheung,& Law, 2011), 3) visiting paths might be missing from the data, since if a tourist does not take a photo in a certain AOI, the generated path does not pass through that AOI resulting in a small deviation in the node and edge weight calculation. In response to the above issues, the next step for further and

Table 4
The ranking of Shanghai's main tourist routes.
source: authors' elaboration.

Ranking	Pre-World Expo period	World Expo and its continuous subsequent “heat period”	Post-World Expo period
1	The Bund—Yu Garden	The Bund—World Expo	The Bund—Yu Garden
2	The Bund—Nanjing East Road	The Bund—Yu Garden	The Bund—Nanjing East Road
3	The Bund - Oriental Pearl	The Bund-Nanjing East Road	The Bund — Oriental Pearl
4	The Bund— People's Square	The Bund—Oriental Pearl	The Bund—People's Square
5	The Bund—Pudong Airport	The Bund—People's Square	The Bund—Expo Park
6	Yu Garden—Nanjing East Road	The Bund—Pudong Airport	The Bund—Pudong Airport
7	Yu Garden—Oriental Pearl	Yu Garden—World Expo	Yu Garden—People's Square
8	Yu Garden—People's Square	Yu Garden—Tianzifang	Yu Garden—Nanjing East Road
9	Yu Garden—Pudong Airport	Yu Garden—Oriental Pearl	The Bund—Tianzifang
10	Nanjing East Road —Oriental Pearl	Yu Garden—Nanjing East Road	The Bund—Nanjing West Road

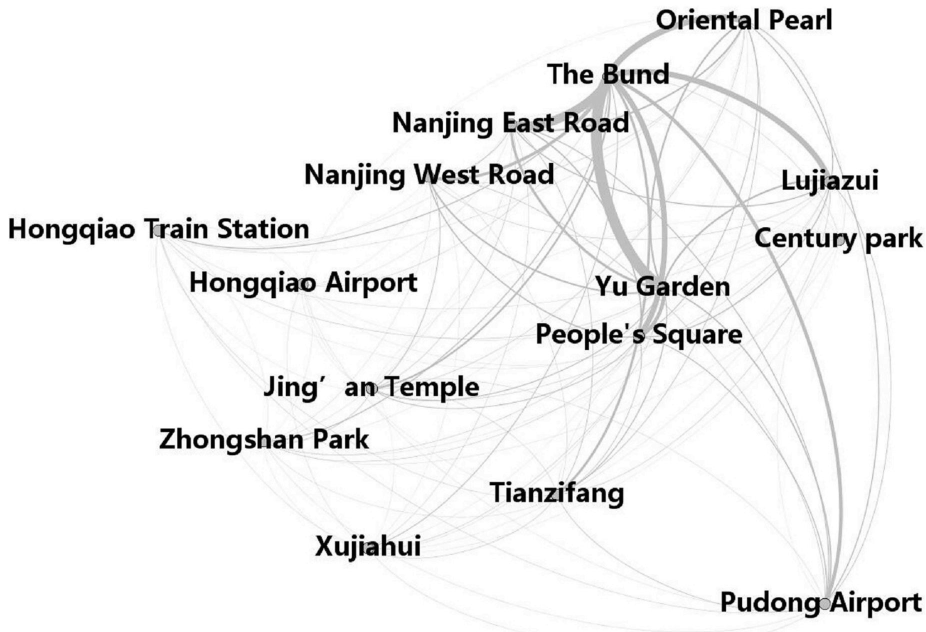


Fig. 11. The network structure in pre-World Expo period (source: authors' elaboration).

more accurate analysis would be to combine data from other social media sites (such as Twitter and Facebook), foreign travel evaluation website data (such as Tripadvisor) and mobile phone signaling data of inbound tourists into a single comprehensive dataset. This would greatly enhance the accuracy of analysis on spatio-temporal changes of inbound tourism flow networks.

6. Conclusion

This paper proposes a framework for studying spatio-tempora distribution and change analysis of inbound tourism flow with data gathered from Flickr. The framework combines the R-HDBSCAN algorithm with relevant complex network computing indicators. The R-HDBSCAN algorithm solves the problem of multi-sight points in the traditional density clustering algorithm for extracting AOI, and thus, improves the clustering accuracy. The application of indicators derived from the complex network theory literature further expand our knowledge on the spatio-temporal evolution processes of inbound tourism flow network, as shown here with a case study example from Shanghai, China. By applying the proposed framework to the case of Shanghai and utilizing Flickr data we obtained the following result: the World Expo was an important turning point in the spatio-temporal characteristics of Shanghai's inbound tourism flow network, since it expanded the number of popular AOI, and subsequently tourist routes, while it also signaled a peak in the number of inbound tourists visiting Shanghai.

Fig. 12. The network structure in World Expo and subsequent "heat period" (source: authors' elaboration).

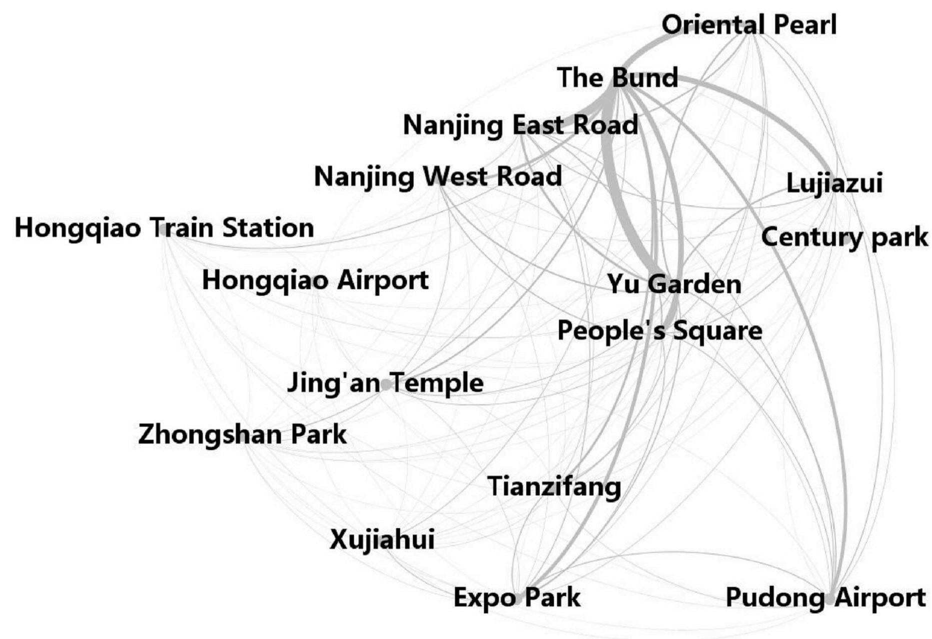


Fig. 13. The network structure in the post-World Expo period (source: authors' elaboration).

To conclude, it can be stated that the proposed research framework is feasible. Thus, it can be applied to other geotagged photo and social media data. The analysis can also be expanded from the intra-regional scale, adopted in this paper, into inter-regional contexts to investigate tourism flows between cities or regions.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

All of the authors contributed to the work in the paper. Naixia Mou and Rongzheng Yuan designed the research and wrote the paper. Tengfei Yang, Jinwen Tang, and Hengcai Zhang contributed to data collecting and preprocessing. Teemu Makkonen contributed to the final writing of the paper.

Acknowledgment

This research was supported in part by the National Natural Science Foundation of China (41771476) and the Natural Science Foundation of Shandong Province (ZR2016DM02).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tourman.2019.103955.

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. https://doi.org/10.1016/j.tra.2006.07.003.

Baggio, R., & Cooper, C. (2010). Knowledge transfer in a tourism destination: The effects of a network structure. *Service Industries Journal, 30*(10), 1757–1771. https://doi.org/10.1080/02642060903580649.

Baggio, R., Scott, N., & Cooper, C. (2010). Network science: A review focused on tourism. *Annals of Tourism Research, 37*(3), 802–827. https://doi.org/10.1016/j.annals.2010.02.008.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences, 101*(11), 3747–3752. https://doi.org/10.1073/pnas.0400087101.

Campello, R. J., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Pacific-Asia conference on knowledge discovery and data mining: Vol. 2013*, (pp. 160–172). . https://doi.org/10.1007/978-3-642-37456-2_14.

Cao, L., Luo, J., Gallagher, A. C., Jin, X., Han, J., & Huang, T. S. (2010). A worldwide tourism recommendation system based on geotagged web photos. *International conference on acoustics, speech, and signal processing*https://doi.org/10.1109/ICASSP.2010.5495905.

Chen, N. (2012). Branding national images: The 2008 Beijing Summer Olympics, 2010 Shanghai world Expo, and 2010 Guangzhou Asian Games. *Public Relations Review, 38*(5), 731–745. https://doi.org/10.1016/j.pubrev.2012.04.003.

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(5), 603–619. https://doi.org/10.1109/34.1000236.

Deng, N., Liu, J., Dai, Y., & Li, H. (2019). Different cultures, different photos: A comparison of Shanghai's pictorial destination image between East and west. *Tourism Management Perspectives, 30*, 182–192. https://doi.org/10.1016/j.tmp.2019.02.016.

Donaire, J. A., Camprubí, R., & Galí, N. (2014). Tourist clusters from Flickr travel photography. *Tourism Management Perspectives, 11*, 26–33. https://doi.org/10.1016/j.tmp.2014.02.003.

Du, Q., Dong, Z., Huang, C., & Ren, F. (2016). Density-based clustering with geographical background constraints using a semantic expression model. *ISPRS International Journal of Geo-Information, 5*(5), 72. https://doi.org/10.3390/ijgi5050072.

Egenhofer, M. J., & Franzosa, R. D. (1991). Point-set topological spatial relations. *International Journal of Geographical Information Systems, 5*(2), 161–174. https://doi.org/10.1080/02693799108927841.

Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Knowledge Discovery and Data Mining, 1996*, 226–231.

Feick, R., & Robertson, C. (2015). A multi-scale approach to exploring urban places in geotagged photographs. *Computers, Environment and Urban Systems, 53*, 96–109. https://doi.org/10.1016/j.compenvurbsys.2013.11.006.

Girardin, F., Fiore, F. D., Ratti, C., & Blat, J. (2008). Leveraging explicitly disclosed location information to understand tourist dynamics: A case study. *Journal of Location Based Services, 2*(1), 41–56. https://doi.org/10.1080/17489720802261138.

Han, J., & Lee, H. (2015). Adaptive landmark recommendations for travel planning: Personalizing and clustering landmarks using geo-tagged social media. *Pervasive and Mobile Computing, 18*, 4–17. https://doi.org/10.1016/j.pmcj.2014.08.002.

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, 53*, 240–254. https://doi.org/10.1016/j.compenvurbsys.2015.09.001.

Iñiguez, T., Plumed, M., & Martinez, M. P. (2014). Ryanair and Spain: Air connectivity and tourism from the perspective of complex networks. *Tourism & Management Studies, 10*(1), 46–52.

Jankowski, P., Andrienko, N., Andrienko, G., & Kisilevich, S. (2010). Discovering landmark preferences and movement patterns from photo postings. *Transactions in GIS, 14*(6), 833–852. https://doi.org/10.1111/j.1467-9671.2010.01235.x.

Kisilevich, S., Mansmann, F., & Keim, D. A. (2010). P-DBSCAN: A density-based clustering algorithm for exploration and analysis of attractive areas using collections of geotagged photos. *International conference and exhibition on computing for geospatial research application*https://doi.org/10.1145/1823854.1823897.

Kuo, C., Chan, T., Fan, I., & Zipf, A. (2018). Efficient method for POI/ROI discovery using

Flickr geotagged photos. *ISPRS International Journal of Geo-Information, 7*(1213) https://doi.org/10.3390/ijgi7030121.
Latorre-Martínez, M. P., Iñíguez-Berrozpe, T., & Plumed-Lasarte, M. (2014). Image-focused social media for a market analysis of tourism consumption. *International Journal of Technology Management, 64*(1), 17–30. https://doi.org/10.1504/IJTM.2014.059234.
Leung, X. Y., Wang, F., Wu, B., Bai, B., Stahura, K. A., & 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. https://doi.org/10.1002/jtr.876.
Lew, A., & McKercher, B. (2006). Modeling tourist movements: A local destination analysis. *Annals of Tourism Research, 33*(2), 403–423. https://doi.org/10.1016/j.annals.2005.12.002.
Li, Z., Xu, M., & Shi, Y. (2015). Centrality in global shipping network basing on worldwide shipping areas. *Geojournal, 80*(1), 47–60. https://doi.org/10.1007/s10708-014-9524-3.
Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management, 68*, 301–323. https://doi.org/10.1016/j.tourman.2018.03.009.
Lo, I. S., Mckercher, B., Lo, A., Cheung, C., & Law, R. (2011). Tourism and online photography. *Tourism Management, 32*(4), 725–731. https://doi.org/10.1016/j.tourman.2010.06.001.
Macqueen, J. (1965). Some methods for classification and analysis of multivariate observations. *Proc of Berkeley symposium on mathematical statistics & probability*.
Manrique-Sancho, M. T., Avelar, S., Iturrioz-Aguirre, T., & Manso-Callejo, M.Á. (2018). Using the spatial knowledge of map users to personalize city maps: A case study with tourists in Madrid, Spain. *ISPRS International Journal of Geo-Information, 7*(332) https://doi.org/10.3390/ijgi7080332.
Ma, Y., Yang, W., Xu, G., & Tai, X. (2018). Multilevel visualization of travelogue trajectory data. *ISPRS International Journal of Geo-Information, 7*(1), 12. https://doi.org/10.3390/ijgi7010012.
Mcinnes, L., & Healy, J. (2017). Accelerated hierarchical density-based clustering. *International Conference on Data Mining, 2017*, 33–42. https://doi.org/10.1109/icdmw.2017.12.
McInnes, L., Healy, J., & Astels, S. (2017). HDBscan: Hierarchical density-based clustering. *The Journal of Open Source Software, 2*https://doi.org/10.21105/joss.00205.
Miguéns, J. I. L., & Mendes, J. F. F. (2008). Travel and tourism: Into a complex network. *Physica A: Statistical Mechanics and its Applications, 387*(12), 2963–2971. https://doi.org/10.1016/j.physa.2008.01.058.
Mou, N., Liu, C., Zhang, L., Fu, X., Xie, Y., Li, Y., et al. (2018). Spatial pattern and regional relevance analysis of the maritime silk Road Shipping network. *Sustainability, 10*(4) https://doi.org/10.3390/su10040977.
Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review, 45*(2), 167–256. https://doi.org/10.1137/S003614450342480.
Önder, I. (2017). Classifying multi-destination trips in Austria with big data. *Tourism Management Perspectives, 21*, 54–58. https://doi.org/10.1016/j.tmp.2016.11.002.
Onder, 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. https://doi.org/10.1177/0047287514563985.
Opsahl, T., & Panzarasa, P. (2009). Clustering in weighted networks. *Social Networks, 31*(2), 155–163. https://doi.org/10.1016/j.socnet.2009.02.002.
Plasacristan, E., Gonzalezdiaz, I., Martinezcortes, T., & Diazdemaria, F. (2019). Finding landmarks within settled areas using hierarchical density-based clustering and metadata from publicly available images. *Expert Systems with Applications, 123*, 315–327. https://doi.org/10.1016/j.eswa.2019.01.046.
Qin, J., Li, L., Tang, M., Sun, Y., & Song, X. (2018). Exploring the spatial characteristics of Beijing inbound tourist flow based on geotagged photos. *Acta Geographica Sinica, 73*(8), 1556–1570 (in Chinese).
Sun, Y., Fan, H., Helbich, M., & Zipf, A. (2013). Analyzing human activities through volunteered geographic information: Using Flickr to analyze spatial and temporal pattern of tourist accommodation. *Progress in location-based services*Berlin Heidelberg: Springerhttps://doi.org/10.1007/978-3-642-34203-5_4.
Tenkanen, H., Minin, E. D., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., et al. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports, 7*(1), 17615. https://doi.org/10.1038/s41598-017-18007-4.
Thomee, B., Elizalde, B., Shamma, D. A., Ni, K., Friedland, G., Poland, D., et al. (2016). YFCC100M: The new data in multimedia research. *Communications of the ACM, 59*(2), 64–73. https://doi.org/10.1145/2812802.
Tsiotas, D., & Polyzos, S. (2015). Analyzing the maritime transportation system in Greece: A complex network approach. *Networks and Spatial Economics, 15*(4), 981–1010. https://doi.org/10.1007/s11067-014-9278-y.
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. https://doi.org/10.1016/j.tourman.2014.07.003.
Wang, J., Mo, H., Wang, F., & Jin, F. (2011). Exploring the network structure and nodal centrality of China's air transport network: A complex network approach. *Journal of Transport Geography, 19*(4), 712–721. https://doi.org/10.1016/j.jtrangeo.2010.08.012.
Wang, W., Yang, J., & Muntz, R. R. (1997). Sting: A statistical information grid approach to spatial data mining. In: *Proc. of the 23rd very large database conf. Vol. 1997*, (pp. 186–195).
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442. https://doi.org/10.1038/30918.
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*(2976), 2976. https://doi.org/10.1038/srep02976.
Wu, X., Zhou, H., Xia, P., Chen, Y., & Yu, L. (2018). Building a spatially embedded network of tourism hotspots from geotagged social media data. *IEEE Access, 6*(99), 21945–21955. https://doi.org/10.1109/access.2018.2828032.
Xia, J. C., Zeephongsekul, P., & Arrowsmith, C. (2009). Modelling spatio-temporal movement of tourists using finite Markov chains. *Mathematics and Computers in Simulation, 79*(5), 1544–1553. https://doi.org/10.1016/j.matcom.2008.06.007.
Xue, K., Chen, X., & Yu, M. (2012). Can the World Expo change a city's image through foreign media reports? *Public Relations Review, 38*(5), 746–754. https://doi.org/10.1016/j.pubrev.2012.06.013.
Yang, S., & Knoke, D. (2001). Optimal connections: Strength and distance in valued graphs. *Social Networks, 23*(4), 285–295. https://doi.org/10.1016/S0378-8733(01)00043-0.
Yuan, Y., & Medel, M. (2016). Characterizing international travel behavior from geotagged photos: A case study of Flickr. *PLoS One, 11*(5)https://doi.org/10.1371/journal.pone.0154885.
Zhou, X., Xu, C., & Kimmons, B. (2015). Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Computers, Environment and Urban Systems, 54*, 144–153. https://doi.org/10.1016/j.compenvurbsys.2015.07.006.

Dr. **Naixia Mou** is an associate professor at College of Geomatics, Shandong University of Science and Technology (China). His main research interests are data mining, big data analysis, tourism and transport geography. Email: mounaixia@163.com.

Rongzheng Yuan, 912917507@qq.com is a M.S. student of College of Geomatics, Shandong University of Science and Technology. His research focuses on tourism research and spatio-temporal data mining.

Tengfei Yang, PhD student (email: yangtf@radi.ac.cn). He is a PhD student of Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. His research focuses on natural language processing and spatio-temporal data mining. For this study, he contributed data collecting and preprocessing.

Dr. Hengcai ZHANG is an assistant professor of Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. He received his Ph.D. degree from Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. He is the member of the Theory and Methodology Committee of the Chinese Association of Geographic Information System, and member of Chinese Branch of ACM SIGSPATIAL. His interests focus on moving objects database and spatial-temporal data mining. During the past years, he has published over 26 referred journal and conference papers. Email: zhanghc@lreis.ac.cn.

Jinwen (Jimmy) Tang is a PhD candidate from South China University of Technology (China) and visiting PhD researcher from University of Surrey (UK). His research fields are innovation, tourism management and tourism geography. Email: tcz0519@163.com.

Teemu Makkonen is a Research Director at the Karelian Institute, University of Eastern Finalnd. His main research interests include border regions, economic geography, tourism and innovation. Email: teemu.makkonen@uef.fi.