Changes in the spatial and temporal characteristics of inbound tourism flows in Tibet based on geotagged photographs

HuaJian Gao^{1, a}, NaiXia Mou^{1, b*}

¹ College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China

Abstract: With the further advent of the era of big data, the scale of social media data containing geolocation information is exploding, providing a new source of big data information and perspective for an in-depth study of the changing spatio-temporal and geographical characteristics of the current tourist population. This paper extracts data on popular attractions in the Tibet Autonomous Region using the HDBSCAN algorithm combined with the TF-IDF algorithm based on information on images with geotags shared by users in the Flickr image sharing site from 2005-2018. Social network analysis was used to explore the changes in the spatial and temporal characteristics of inbound tourism flows in Tibet. The results show that: (1) in terms of temporal characteristics, the number of inbound tourists shows obvious off-peak seasons, with relatively high sensitivity to the influence of economic, policy and infrastructure construction factors; (2) in terms of spatial distribution characteristics, the inbound tourism flow in Tibet shows an "axis-scattered" distribution. The core area is centred on Lhasa and extends in three directions: west, north and east along important roads.

1 Introduction

Inbound tourism is an important part of China's tourism market and an important indicator of the country's tourism competitiveness^[1]. The flow pattern of inbound tourists between tourism hotspots can reflect the dynamic trend of the international tourism source market, and is also important for the development of tourism markets and tourism products^[2].

Research on inbound tourism flows usually revolves around traditional data, such as statistical yearbooks^[3] and questionnaires^[4], which can reflect changes in the number of inbound tourists in a certain region and time, but often fail to accurately and comprehensively reflect the spatial distribution and flow characteristics of tourists^[5]. The sample representativeness of survey questionnaires is poor, and it is also difficult to accurately and comprehensively reflect the behavioural characteristics of inbound tourists^[6]. With the increasing popularity and rapid industrial development of information technology of mobile application terminals such as modern mobile smart Internet and mobile phone intelligence, it has become possible and trendy to use social media data containing geographical location information to study tourism flows^[7].

Flickr, Panoramio, Instagram and a host of other photo social applications allow users to share photos to online communities and add location information to them using tags, hence the name geotagged photos. the Flickr website has over 49 million geotagged photos. This social media data contains not only descriptive textual information such as captions, messages and tags, but also spatial and

temporal information such as check-in times and latitude and longitude, making it an ideal data source for studying inbound tourism flows^[7].

In this paper, based on geotagged photo information from Flickr, we use the HDBSCAN algorithm to extract tourism AOI and use the TF-IDF algorithm to constrain the clustering results according to geotags, solving the problem of multiple tourism points in tourism AOI extraction. On this basis, social network theory is introduced to explore the structure of spatial relationships and the dynamic evolution process of inbound tourism flows.

2 Materials and Methods

2.1 Study area and dataset

Located in the west and south of China's Tibetan Plateau, with an average altitude of more than four kilometres above sea level and an area of 1,228,400 square kilometres, Tibet has a resident population of 3,348,200 as of 31 December 2018^[8] and is the richest and least developed region in China in terms of tourism resources, with the least resources being infringed upon and subject to any artificial damage^[9]. The geographical isolation has given Tibet a unique cultural history and tradition that has created a strong tourism attraction for modern travellers, especially from foreign countries. Between 2005 and 2018, Tibetan inbound tourists increased from 120,000 to 470,000, an average annual growth rate of 21.1%, and foreign exchange earnings from international tours increased from US\$44.43 million to US\$247.09 million,

^a729843607@qq.com

* Corresponding author: ^bmounaixia@163.com

an average annual growth rate of 90.8%^[10]

Flickr (<https://www.flickr.com/>) is a widely used web-based photo management and sharing site worldwide. officially founded in 2004, the openness of Flickr's vast amount of social photo data has also made it one of the more recognised data collections in social photo research today. the Flickr website provides a free API interface for people to Flickr provides a free API for accessing its data. This paper crawls geotagged Flickr images and metadata in the Tibet Autonomous Region (78.3598°E~99.1146E, 27.2744°N~36.4619°N) from 1 January 2005 to 30 December 2018 through the Flickr API interface. The total number of records is 95,321. As shown in Table 1

Table1. Flickr example data

UserID	PictureID	Lat	Lon	P_date
100034399@N08	1000343991	31.868139	88.759407	2011/10/19 9:04
100034399@N08	1000343992	31.868139	88.759407	2011/10/19 9:24
100082720@N07	1000827201	29.653363	91.116861	2016/3/27 10:22

2.2 Data Cleaning

Flickr photo data may have some errors due to user device accuracy problems. Therefore, the following rules were used to clean the data in this paper: (1) Removing duplicate data. When obtaining Flickr data through the API interface, the data was obtained by chunking according to latitude and longitude, and duplicate crawled data would inevitably appear at the junction of each chunk of data, so these duplicate data needed to be cleaned, and a total of 631 data were removed. (2) Removal of resident population data, part of the photos may be taken by residents based in Tibet^[11, 12], and the length of stay in a place can be used as a criterion for judging tourists or residents. girardin et al. set the threshold value to 30 days^[13], and this criterion has been adopted by most studies^[14], this paper also adopts the threshold value of thirty days, and the two photos taken more than thirty days apart users were excluded. A total of 7435 items were removed. (3) Removal of duplicate data. If the same user took multiple photos at the same location and within one hour, the data were excluded as duplicate photos taken by the same user. A total of 12,037 pieces of data were removed. (4) Removal of orphaned data, as the aim of this paper is to study inbound tourism flows, so tourists with at least 2 shooting locations were filtered in order to build a network of inbound tourism flows. A total of 3858 pieces of data were removed. After cleaning, a total of 71,360 records from 1576 tourists were collected. The distribution is shown in Figure 1.

Fig 1. Tibet Flickr photo distribution

2.3 Travel AOI Extraction

The extraction of tourism AOIs (Area of Interest) from geotagged social media data is a prerequisite for studying the spatial and temporal variation of tourism flows. An AOI usually consists of multiple POIs (Points of Interest), and unlike administrative divisions with clear boundaries, the boundaries of tourism AOIs are mostly fuzzy^[15] so tourism researchers often use clustering algorithms to calculate the boundaries of AOIs^[16]. Clustering algorithms are a method of classifying and categorizing data using similarities between objects. Traditional clustering algorithms mostly use density and distance-based methods to extract AOIs, such as the K-means algorithm and K-modes algorithm for distance-based clustering; the DBSCAN algorithm and Mean-Shift algorithm for density-based clustering. However, traditional clustering algorithms based on density and distance also have certain limitations. The core of these clustering algorithms is Single Linkage Clustering, which is very sensitive to noise, and a noisy data point that lies between two class clusters may cause the class clusters to stick together and thus be classified into one class. However, the distribution of social media data in geographic space is complex and often contains a large amount of noisy data. Carlos et al. use the HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) algorithm, which combines the idea of hierarchical clustering with density clustering to solve the problem of traditional clustering algorithms. that solves the class-cluster adhesion and multi-density clustering problems that exist in traditional clustering algorithms^[17].

2.3.1 HDBSCAN algorithm

The HDBSCAN algorithm idea is a combination of the ideas of hierarchical clustering and density clustering, which proposes the concept of mutual reachable distances and generates a tree diagram by means of a mutual reachable distance matrix, thus achieving the separation of clustering results of different densities from sparse noise. The mutual reachable distance $d_k(a, b)$ is defined as:

$$d_k(a, b) = \max\{core_k(a), core_k(b), d(a, b)\} \quad (1)$$

Where $core_k(a)$ denotes the radius of clustering of a in the DBSCAN algorithm subject to the minimum clustering parameter k , and $d(a, b)$ denotes the Euclidean distance between a and b . Under this metric, dense points

remain at the same distance from each other, but sparse points are pushed away from the core distance of other points, effectively separating clusters from sparse noisy points.

After calculating the mutual reachable distance, the data is considered as a weighted graph, the data points are considered as vertices, $d_k(a, b)$ between any two vertices is the weight of the edges to construct the minimum spanning tree, and the edges of the tree are sorted by distance to construct the cluster hierarchy, as shown in Figure 2.

Fig 2. Distance matrix dendrogram

Introducing cluster-like stability to compress the resulting cluster hierarchy, λ is the reciprocal of $d_k(a, b)$, and for a given cluster, for any point p in that cluster, the value of λ that removes that point from the cluster is defined as λ_p , then the cluster-like stability of that cluster is shown in equation (2).

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

Where λ_{birth} is the λ value for successful cluster splitting. If the cluster stability is greater than the sum of the sub-clusters, the sub-clusters under the cluster are merged until the stability of the whole system is highest, all the newly generated class clusters of the current set are regarded as the clustering result and returned as such, and the remaining points are grouped according to the extracted class cluster centres, thus achieving the initial clustering of the clusters. In this paper, the relatively microscopic tourist attractions are also extracted based on the principle of extracting the attraction information as carefully as possible, however, this may lead to the following problem: some of the tourist AOIs contain more than one tourist attraction, thus the situation that "different class clusters actually belong to the same attraction", in order to solve this problem, this paper uses the TF-IDF algorithm. In order to solve this problem, this paper uses the TF-IDF algorithm to calculate the cosine similarity of the geotagging information of the different clusters to merge the clusters that actually belong to the same attraction.

2.3.2 Geotagging-based merging of class clusters

In order to solve the problem that different clusters of classes actually belong to the same attraction, this paper uses the tagging information in the Flickr photo information to merge the clusters of classes. In this paper, we use the TF-IDF algorithm, a text similarity measure for attractions, which is based on term frequency (TF) and inverse document frequency (IDF) to represent text as a vector of n weighted lexical items appearing in the text [31]. The TF-IDF value of each lexical item ω_i can be calculated using equation (3).

$$TF - IDF(\omega_i) = tf_j(\omega_i) \times \log(N/df(\omega_i)) \quad (3)$$

$tf_j(\omega_i)$ denotes the frequency of the current lexical item ω_i denoted in text j ; N denotes the total number of all texts in the text collection; $df(\omega_i)$ denotes the number of texts in the text collection in which the current lexical item ω_i appears. through equation (3) an n -dimensional vector can be calculated for each additional text to replace the text, by calculating the cosine between the corresponding vectors of different texts similarity to characterise their corresponding degree of similarity. By setting a similarity threshold, the corresponding class clusters can be merged to different degrees.

2.4 Inbound tourism flow evaluation indicators

The migration and diffusion of tourists between tourist destinations creates certain connections between destinations, forming a system of evolving and changing networks. Social network theory and methods can accurately portray the various relationships within this system from the perspective of macro-structural relationships, highlighting the spatial characteristics of tourism flows as a whole as well as exploring the relational characteristics of nodes within tourism flows, and have gradually become a popular paradigm for the study of tourism flow network structures in recent years^[18, 19]. In this paper, three indicators, namely Relative centrality, average path length and average clustering coefficient, are selected to assess the structural characteristics of Tibetan inbound tourism flow networks.

2.4.1 Relative centrality

Centrality measures the degree to which nodes in a travel stream are directly related to other nodes. They are divided into in-degree centrality and out-degree centrality, which reflect the cohesion and dispersion of tourism flow nodes respectively. The node centrality is calculated by the formula:

$$D_{in}(i) = \sum_j^n L_{ij, in} \quad (4)$$

$$D_{out}(i) = \sum_j^n L_{ij, out} \quad (5)$$

$L_{ij, in}$ and $L_{ij, out}$ denote the directionality between node v_i and node v_j . n is the total number of nodes in the travel flow network.

In order to compare the changes in the degree

centrality of inbound tourism flows in different years horizontally, this paper introduces the relative degree centrality to balance the gap caused by the difference in the total number between different years, and the formula for calculating the relative degree centrality is:

$$S_i = \frac{D_i - \min_{1 \leq j \leq n} \{D_j\}}{\max_{1 \leq j \leq n} \{D_j\} - \min_{1 \leq j \leq n} \{D_j\}} \quad (6)$$

where $\max_{1 \leq j \leq n} \{D_j\}$ and $\min_{1 \leq j \leq n} \{D_j\}$ are the points with the lowest and highest degree centrality in the travel flow network, and D_i is the sum of the outward degree centrality and inward degree centrality of node v_i .

2.4.2 Average path length

The size of the average path length measures the connectivity of the network and the overall efficiency of the network. For inbound tourism flow networks, a smaller L means a smaller number of nodes travel AOI that need to transit during the tourist visit, then the higher the accessibility of the inbound tourism flow network. Defined as the average of the distance between any two nodes, the formula is.

$$L = \frac{1}{C_N^2} \sum_{1 \leq i < j \leq N} d_{ij} \quad (7)$$

d_{ij} denotes the number of edges of the shortest path connecting two nodes v_i and v_j . When v_i and v_j are not connected, d_{ij} is considered infinite, n is the total number of nodes, and C_N^2 denotes the total number of possible edges of the network composed of n nodes.

2.4.3 Average clustering coefficient

The average clustering coefficient reflects the average closeness of the links of the nodes in the tourism flow network. The higher the ratio, the higher the ratio of tourist attractions connected by two direct transport networks in the network, which also means that the accessibility of local transport and the availability of tourist attractions in the tourism flow network is higher, calculated by the formula:

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n \frac{E_i}{C_{k_i}^2} \quad (8)$$

E_i denotes the actual number of edges between the k_i neighbouring nodes of node v_i ; $C_{k_i}^2$ denotes the total number of possible edges between the k_i neighbouring nodes of node v_i ; and n is the total number of nodes.

3 Results & Discussion

This paper extracts the attractions through algorithms such as spatial clustering and geotag-based merging to obtain the corresponding class clusters of attractions. The data points of the class clusters are counted to obtain the clustering centres of 115 class clusters, and the inbound tourism flow network is constructed in this way, as shown in Figure 3. In this paper, the development of inbound tourism flow in Tibet will be analysed in the temporal and

spatial dimensions respectively, and the development pattern will be explored.

Fig 3. Distribution of inbound tourism flows in Tibet

3.1 Changes in the temporal characteristics of inbound tourism flows

The collected Flickr check-in data was decomposed to obtain 12 months as well as annual data in turn; the data was extracted from the temporal attributes of the 71,360 valid check-ins and the frequency of travel of inbound tourists by year and month was counted, as shown in Figure 4.

Fig 4. Changes in the temporal characteristics of inbound tourism flows in Tibet

Figure 4(a) shows a clear temporal heterogeneity in the tourism fervour of inbound tourists over a 12-month period. The peak season for inbound tourism in Tibet is from May to October each year, while November to April is the low season for tourism in Tibet, which coincides with the natural climatic characteristics of Tibet: due to its high altitude geographical features, Tibet has a temperate monsoonal semi-humid and semi-arid climate on the plateau, with the year divided into dry and wet seasons. From October to April each year, the Tibetan plateau is over the westerly rapids and the ground receives control from cold high pressure, which is dry and full of From May to September, the near-surface layer of the plateau is controlled by thermal low pressure, which increases rainfall, revives plants and increases the oxygen content of the air, helping to alleviate plateau reactions, making it the most suitable time for tourism in Tibet. It can be deduced that the climatic characteristics of the dry and wet seasons are an important factor in the temporal heterogeneity of inbound tourists to Tibet in terms of months.

From Figure 7(b), it can be seen that the tourism fever

of inbound tourists can be roughly divided into three phases during the period 2005-2018, using certain major events as cut-off points: a period of rapid development from 2005-2007, a period of rapid recovery from 2008-2011, and a period of downturn from 2012-2018. the period 2005-2007 saw a rapid increase in the tourism fever of inbound tourists in Tibet. The opening of the Qinghai-Tibet Railway on 1 July 2006 led to a rapid growth in inbound tourism in 2007, with a growth rate of 124.84%. 2008 saw a serious violent crime in Lhasa on 14 March causing a precipitous fall in inbound tourism that year, which took three years to recover and broke through to a new high of 9,856 annual arrivals in 2011. 2012 saw the financial crisis The outbreak of the financial crisis in 2012 saw tourism around the world take a huge hit, with the global economy and tourism market continuing to decline and inbound tourism flows to Tibet continuing to slump between 2012 and 2018, maintaining an annual decline of 10-30% except for a small rebound in 2017. It is evident that the tourism industry is highly sensitive to both the political and financial environment, with sudden and malignant events having a greater magnitude of impact on the tourism industry.

3.2 Changes in the spatial characteristics of inbound tourism flows

In order to explore the changes in the spatial characteristics of inbound tourism flows in Tibet during 2005-2018, this paper takes the major events affecting inbound tourism flows in 2008 and 2012 as the dividing points, constructs a flow data matrix among various tourist attractions, constructs the spatial distribution structure of inbound tourism flows in Tibet, analyzes the changing trends in the spatial distribution of inbound tourism flows in Tibet over the past 14 years, and uses degree centrality, average path length, and average clustering coefficient to explore the changes in the spatial structure of inbound tourism flows in Tibet. The calculation results of some attractions are shown in Table 2, and the Tibet inbound tourism flow network was visualised according to the relative degree centrality values, and the results are shown in Figures 5-7.

Table2. Formatting sections, subsections and subsubsections.

	2005-2007	2008-2011	2012-2018
Network Nodes	78	99	80
Directed edges	151	331	151
Average centrality	1.936	3.357	1.85
Average path length	3.841	3.057	3.398
Average clustering coefficient	0.089	0.132	0.093

The spatial distribution of inbound tourism flows in Tibet from 2005 to 2018 has obvious spatial bias

clustering, with obvious axis-scattering characteristics. As shown in Figure 5, inbound tourism flows in Tibet during 2005-2007 were mainly concentrated in the central area with Lhasa city as the core, including Shannan city, Shigatse city and the surrounding counties, extending southwestward along the G318 national highway to Mount Everest, forming a By the Shishapangma, Everest, Potala Palace, Jokhang, to the Namtso northeast-southwest direction axis. The outer areas are less connected to each other and are mainly connected to the attractions within Lhasa, forming a pattern with Lhasa as the core radiating to the surrounding areas.

Fig 5. Inbound Tourism Flows to Tibet 2005-2007

As shown in Figure 6, during the period 2008-2011, due to the opening of the entire Qinghai-Tibet Railway, there was a significant increase in the connection between attractions along the Qinghai-Tibet Railway north of Lhasa and other attractions, and inbound tourism flows in Tibet radiated in three directions, north, east and south, with Lhasa as the centre along National Highway G318 and National Highway G109, with the core area extending to the cities of Nagqu, Linzhi and Jilong County, which also led to a spatially axial. The spatial distribution of tourism flows is characteristic. The peripheral areas, apart from the famous attractions such as the Lake Manasarovar, Mount Kailash and the Salween as the centre to form a scattering point, are all blank areas, especially the Ali-region in northwest Tibet, which is obviously not attractive to inbound tourists.

Fig 6. Inbound Tourism Flows to Tibet 2008-2011

As shown in Figure 7, during the period 2012-2018, due to the impact of the financial crisis, the inbound tourism market continued to be sluggish, the connection

between inbound tourism flow nodes became weaker, and the heat of tourism flows in the direction of Everest and Shishapangma declined, and the core areas showed signs of contraction, for example, the heat of core attractions such as Namtso, Cona Lake, Tibet Museum and Jokhang declined, but the emerging southern Tibetan regions of attractions such as the ancient city of Old Town of Taizhao, Kading Valley, Lebu Valley and Namkapub Ri did not show a significant decline, especially Lebu Valley and Namkapub Ri, which showed a strong correlation. The correlation between the two attractions is as high as 85.7%, and all tourists departing from Lebu Valley have travelled to Namkapub Ri, indicating that the lack of tourist transport routes and low accessibility in southern Tibet have led to a spatially scattered tourist flow.

Fig 7. Inbound Tourism Flows to Tibet 2012-2018

On the whole, inbound tourism flows in Tibet are constrained by the roads and do not form a closed tourism route, but are more in the mode of "single passage". The opening of roads and railways has had a strong stimulating and gathering effect on tourists, with the majority of them visiting Lhasa, Shigatse, Shannan, Nagqu and Linzhi, and fewer visiting the Ali region in northern Tibet and the city of Chamdo in northeastern Tibet.

4 Conclusions

The inbound tourism flow to Tibet has a clear temporal seasonal variation, with a large difference between the winter and summer months. The peak season is from May to October and the low season from November to April, which is mainly due to the unique plateau climate conditions in Tibet, where the temperature rises from May to October, rainfall increases and the concentration of oxygenates in the air improves, making the climate conditions more suitable for tourists to visit; in terms of year-to-year variation, Tibet The opening of the Qinghai-Tibet Railway in 2007 had a positive impact on the entire Tibetan inbound tourism economy, hence the significant upward trend in inbound tourists in 2007, with an increase of 124.84%, while the political turmoil in 2008 and the contraction of the international tourism market in 2012 had a negative impact on inbound tourism in Tibet, hence the the number of inbound tourists produced a precipitous decline in both 2008 and 2012, indicating the need for a stable political environment for the development of tourism. According to the Tibetan Statistical Yearbook, the

Tibetan tourism market developed steadily in 2018, which is at variance with the Flickr statistics, which may be related to the decline in the use of Flickr in recent years, although Wood et al. proved in 2013 that Flickr data can be used to effectively estimate the rate of visits to tourist attractions, verifying the reliability of Flickr data, but with the recent the explosion of social platforms such as Twitter and Facebook, whether Flickr data can still be used for tourism flow studies beyond 2018 needs to be further explored.

Inbound tourism flows in Tibet have obvious characteristics of axis-scatter structure in spatial distribution, with strong spatial bias clustering. Inbound tourism flows are mainly concentrated in Lhasa and its surrounding areas, radiating along the highway lines to the peripheral areas, while the peripheral areas are obviously not attractive enough for inbound tourists, except for the Mount Everest and Lake Manasarovar, which are slightly attractive, while the rest, such as the Pali Grassland, Old Town of Taizhao and Tsomon Ling, can only attract a small number of tourists, which reflects the extreme imbalance of tourism development in Tibet from another side. As transportation is an important factor influencing tourists' decision making, vigorously developing Tibet's transportation is the key point to promote the further development of Tibet's inbound tourism.

The spatial structure of inbound tourism flows in Tibet is not significantly small-world in character, and the network of attractions is low in density and loose in structure. 14 years on, the average path length in Tibet shows a decreasing trend, and the links between attractions are strengthening, with tourism routes gradually transforming from a linear structure to a mesh structure. The attraction fever of the southern routes of Tibet is declining, while the attraction fever of the northern and eastern routes is increasing, which is closely related to the construction of transport infrastructure in the north and east.

Inbound tourists have a particular preference for 'cold spots', and after 2013 the link between Lebu Valley, an ordinary Tibetan spot with fewer visitors, and Namkapub Ri has increased. The preference of inbound tourists for Lebu Valley and Namkapub Ri is a good indication of the unique appeal of Tibetan towns to foreign tourists, and the possibility of combining culture and tourism to create special towns and provide a new impetus to Tibetan tourism.

References

1. Jinfeng, W. Xuli, P. (2010) Characteristic of Beijing and Shanghai Inbound Tourists Flow Network. *Scientia Geographica Sinica*, 15(3): 370-376.
2. Xiuqiong, C. Fucai, H. (2006) Temporal change of regional disparity in chinese inbound tourism. *Acta Geographica Sinica*, 61(12): 1271-1280.
3. HU, S. (2005) Modeling of tourist flow to huangshan. *Modern Physics Letters B*, 19(28n29): 1691-1694.
4. Lew, A. Mckercher, B. (2006) Modeling tourist movements: a local destination analysis. *Annals of*

Tourism Research, 33(2): 403-423.

Tourism Management, 76:

5. Fabien, G. et al. (2008) Leveraging explicitly disclosed location information to understand tourist dynamics: a case study. *Journal of Location Based Services*, 2(1): 41-56.
6. Oender, I. Koerbitz, W. Hubmann-Haidvogel, A. (2016) Tracing tourists by their digital footprints: the case of Austria. *Journal of Travel Research*, 55(5): 566-573.
7. Ning, C. Xia, P. & Zhou, H. (2016) Popularity analysis of tourist attraction based on geotagged social media big data. *Science of Surveying and Mapping*, 3(12): 67-171
8. Tibet Autonomous Region Bureau of Statistics. (2005) Statistical Bulletin of National Economic and Social Development of Tibet Autonomous Region in 2004. *Tibet Development Forum*, 000(003): 55-56.
9. Lu, G. Tianling, L. & Xiaohui, W. (2013) The empirical study of domestic tourist's perceived Tibet tourism image. *Tourism Science*, (02): 73-82.
10. National Bureau of Statistics. (2021) 2020 China Statistical Yearbook. *Statistical theory and practice*, 01: 2.
11. Sancho, M.T.M. et al. (2018) Using the spatial knowledge of map users to personalize city maps: a case study with tourists in Madrid, Spain. *International Journal of Geo-Information*, 7(8): 1-16.
12. Yongsai, M. et al. (2018) Multilevel Visualization of Travelogue Trajectory Data. *ISPRS International Journal of Geo-Information*, 7(1):
13. Imran, M. et al. (2015) Travel Recommendation Using Geo-tagged Photos in Social Media for Tourist. *Wireless Personal Communications*, 80(4) 1347-1362.
14. Juan, C. Javier, G. & Carmen, M. (2015) Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography*, 63: 408-417.
15. KUO, C. CHAN, T. FAN, I. et al. (2018) Efficient method for POI/ROI discovery using Flickr geotagged photos. *ISPRS International Journal of Geo-Information*, 7(3): 121.
16. ABDUL, M. LING, C. GENCAI, C. et al. (2013) A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science*, 27(4): 662-684.
17. RUIZ, C. SPILIOPOULOU, M. MENASALVAS, E. (2010) Density-based semi-supervised clustering. *Data Mining & Knowledge Discovery*, 21(3): 345-370.
18. CRISTÓBAL, C. ÁNGELES, G. MARÍA-ROSA, G. (2016) Social network analysis in tourism. *Current Issues in Tourism*, 19(12):
19. NAIXIA, M. RONGZHENG, Y. TENGFEI, Y. et al. (2020) Exploring spatio-temporal changes of city inbound tourism flow: The case of Shanghai, China.