Study on Spatial-Temporal Characteristics and Influencing Factors of Network Attention: The Case of Guilin, China

Jiansheng Yao\(^{1,2}\) and Yanling Liu\(^{1,2}\)

\(^{1}\)College of Tourism & Landscaper Architecture, Guilin University of Technology, Guilin, China
\(^{2}\)Institute of Guangxi Tourism Industry, Guilin, China
Email: yjs789@glut.edu.cn

Abstract. To research the spatial-temporal distribution characteristics and influencing factors of tourism network attention is meaningful for understanding the current situation of the development tourism, revealing tourism customer market, and providing guidance for tourism planning, service, management. Took Guilin of China as an example and based on Baidu Index, this paper studied the spatial-temporal characteristics and influencing factors of tourism network attention by using seasonal concentration index, variation index, geographic concentration index, coefficient of variation and regression analysis. The research found that mobile network attention has a major share and has obvious seasonality. The PC network attention occupies a low share, which is not affected by the peak seasons, but fluctuates periodically with working days and rest days. In Guilin, the main factor influencing the time distribution of the network attention is precipitation, and the climate comfort has not reached a significant level. The regional distribution of the network attention is neither centralized nor balanced. The population size, the network development and the distance are important factors affecting the spatial distribution of network attention.

1. Introduction
Exploring the characteristics and influencing factors of tourism network attention is beneficial to the development and marketing of tourism products, adjusting the supply of related products or services, so as to improve economic benefits. Tourism demand, therefore, has become a hot issue in tourism research [1]. In recent years, the Internet penetration rate in China has gradually increased. In the Internet age, more and more people use the Internet to publish, share and obtain information, which is an important basis for behaviour decision-making. Ginsberg J found that the network search data could reflect the user's attention to a certain thing, which was correlated to real social behaviours to some extent [2]. Previous studies have shown that the search information of tourism can also reflect people's attention to and demand for tourism destinations. It, therefore, can reveal the spatial-temporal characteristics of tourism demand of tourism destinations to study the spatial-temporal characteristics of network attention of tourists.

At present, "Google trend" is the main data source to study the network attention [3]. However, "Baidu" is the most important search engine in China. According to StatCounter statistics, "Baidu search" currently occupies nearly 2/3 of the Chinese search engine market share. "Baidu Index" is a data sharing platform provided by Baidu based on the search behaviour of Baidu's massive netizens. Therefore, it has great advantages to study the network attention of Chinese tourist destinations based on "Baidu Index" [4,5]. In recent years, Chinese scholars have conducted related research on China's tourism network attention based on the Baidu Index. Li Shan used the Baidu Index for the first time to analyse the network attention of 53 AAAAA-level scenic spots in China, and concluded that the time
distribution of the network attention of scenic spots showed a "premonitory" effect [6]. Later, researchers carried out a lot of studies on the network attention based on Baidu Index, mainly focused on the following aspects: Tourist flow prediction research based on the relationship between the network attention of tourism and the actual passenger flow; The spatial and temporal distribution characteristics and influencing factors of network attention, etc.

In China, Guilin is a famous international tourism city, and tourism is the pillar industry of Guilin. In this paper, we using Baidu index to research the spatial-temporal distribution characteristics and influencing factors of Guilin tourism network attention. The results are beneficial to understand the current situation of the development of Guilin tourism, reveal the Guilin tourism customer market distribution characteristics and influencing factors, and provide guidance for Guilin tourism planning, service, management and marketing.

2. Case and Data

2.1. Case
Guilin has enjoyed the reputation of "landscapes are the best in the world" for many years. It is rich in tourism resources. It has dozens of scenic spots above AAAA level, such as Li River, Elephant Hill, Longji Terrace, Yulong River and so on. In China, Guilin is the approval of the construction of the international tourism resort, opening up to the international tourism city, national tourism development area in innovation, build a world-class tourist destination demonstration zone. It is the home of the UNWTO /Asia Pacific Tourism Association International Forum on tourism trends and prospects, and one of the best tourist cities in China recommended by the UNWTO. In 2019, Guilin received 138.366 million tourists, with a total tourist expenditure of 187.425 billion RMB. Tourism is the pillar industry of Guilin, so it is great significance to study the tourism demand of Guilin.

2.2. Data
"Baidu Index" can directly reflect the search frequency of a keyword in a period of time. In this paper, the Baidu search index is taken as the index to measure network attention. Based on Python web crawler technology, the daily Baidu search index data of the whole year in 2017 is obtained, including PC-index (using Baidu search via PC computer), mobile-index (using Baidu search via mobile Internet) and all-index (PC-index + mobile-index).

To select keywords, this article fully draws lessons from the existing literature keywords selection principles and takes into account the comprehensiveness, multifaceted nature and particularity of different cases. Finally, the keywords, including Guilin tourism, Guilin scenic spots, Li river and Yangshuo, was selected. Other relevant data, such as regional population size, economic status, network development status, rainfall, temperature, humidity, etc., are from China Statistical Yearbook. The distance to Guilin is computed using the latitude and longitude of Guilin government and other regional government on the Gao DE Map.

3. Spatial-temporal Distribution Characteristics
This paper analyses and visualizes Guilin tourism Baidu index data using Python libraries, including Pandas, matplotlib, pyecharts and statsmodels, etc. To further, we study the spatial-temporal characteristics and influencing factors of network attention to Guilin.

3.1. Time Distribution Characteristics
The time series of Baidu Index was analysed with different time granularity of month and day (Fig. 1 and Fig.2). The results are as follows.

1. The mobile-index accounts for a large proportion (about 75.48%), which is basically consistent with the trend of all-index. The PC-index accounts for a relatively small proportion, which has little impact on the trend of all-index. With the gradual popularization of mobile devices and the great improvement of functions, the advantage of being able to surf the Internet anytime and anywhere makes mobile user occupy the main share. In view of the fact that in most cases, the trend of mobile-
index is consistent with that of all-index, we will only describe the mobile-index and PC-index except when necessary.

2. According to the monthly series diagram of the mobile-index (Fig.1), Guilin tourism has obvious seasonality. April to October is the peak season. January to March and November to December are the off-season. The seasonality of the PC-index is not obvious. To further verify the conclusion, we analyzed the seasonal concentration index \( I \) calculated by Formula 1. Where \( I \) is the seasonal concentration index, \( x_i \) is the Baidu index of the \( i^{th} \) month, and \( \bar{x} \) is the mean of \( x_i \). The results are shown in Table 1. It is found that the seasonal concentration of mobile-index is higher than that of PC-index, while that of PC-index is not obvious.

\[
I = \left( \frac{1}{12} \sum_{i=1}^{12} (x_i - \bar{x})^2 \right)^{1/2}
\]

Table 1. Seasonal concentration index

|                  | all-index | mobile-index | PC-index |
|------------------|-----------|--------------|----------|
| Seasonal concentration index\% | 2.596 | 3.2435 | 0.9682 |

3. According to the daily series curve of mobile-index (Fig.2), Guilin's network attention has four peaks. Combined with the schedule of some holidays in 2017, Guilin tourism is also affected by Holiday factors, forming "four peaks" as follows: Spring Festival is a peak in the off-season, "5.1", "10.1" and the summer vacation in August.

4. According to the monthly variation index of formula (2), we design the daily variation index, such as formula (3). Where \( S_i \) and \( S_{di} \) are the monthly and daily variation index respectively, \( x_i \) is the Baidu index of the \( i^{th} \) month or the \( i^{th} \) day. We draw the monthly and daily variation index chart as Fig.3 and Fig.4.

\[
S_i = x_i \left( \frac{1}{12} \sum_{k=1}^{12} x_k \right)^{-1} \times 100\%
\]

\[
S_{di} = x_i \left( \frac{1}{365} \sum_{k=1}^{365} x_k \right)^{-1} \times 100\%
\]
The results show that: a) the trend of mobile variation index and time series is roughly the same, while PC variation index is not affected by the low and peak seasons. b) the monthly variation index is relatively gentle, but the daily variation index fluctuates frequently and has obvious periodicity in weekly. Because using PC is limited by time and place, it has great volatility and periodicity with working days and rest days.

3.2. Time Influence Factor
Climate comfort can affect the physical and mental state of tourists and promote or inhibit tourism demand. Holidays determine whether and how long people have time to travel. Generally, the longer the holiday time, the stronger the demand for tourism. In Guilin, the most typical tourism resource is the Li River, and the water flow of the Li River is one of the important factors affecting Guilin tourism. The degrees of climate comfort are calculated by \( THI = (1.8t + 32) - 0.55(1 - 0.01f)(1.8t - 26) \). Where THI, t and f are the climate comfort, the temperature and the humidity respectively. The virtual factors are assigned according to literature [7]. The water flow of Li River is simply simulated according to the rule \( W_i = r_i + \frac{1}{2} r_{i-1} + \frac{1}{4} r_{i-2} \). Where \( W_i \) is the water flow of Li River on the \( i \)th month, and \( r_i \) is the precipitation of Guilin on the \( i \)th month. Although \( W_i \) cannot replace the actual water flow, it can approximate the general trend of water flow of the Li River.

Step-by-step multiple linear regression analysis showed that climate comfort had no significant influence on the Guilin's tourism network attention, and the model results only included two factors: holiday virtual factor (\( X_1 \)) and Li River water flow (\( X_2 \)). The standardized coefficients are 0.47 and 0.53, respectively, \( p \) values are 0.017 and 0.006, \( R^2 = 0.812 \), \( \text{adj} R^2 = 0.77 \).

3.3. Time Influence Factor
Taking 34 provinces and municipalities in China as the spatial distribution unit, the geographic concentration index \( G \) and coefficient of variation \( CV \) are calculated according to formulas (4) and (5). Where \( x_i \) is the Baidu index of the \( i \)th spatial unit, \( \overline{x} \) and \( S \) are their mean and sum, respectively.

\[
G = 100 \times \left( \frac{1}{S} \sum_{i=1}^{n} (\frac{1}{S} x_i)^2 \right)^{1/2} \tag{4}
\]

\[
CV = \frac{1}{\overline{x}} \left( \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \right)^{1/2} \tag{5}
\]

The results are shown in Table 2: PC-index and mobile-index is similar in geographical concentration and coefficient of variation. The regional distribution is neither concentrated nor balanced.

| Table 2. Geographic concentration index and coefficient of variation |
|-------------------------|--------------|------------------|
|                         | all-index    | PC-index         | mobile-index    |
| Geographic concentration index | 20.876 1     | 20.815 0         | 21.069 4        |
| Coefficient of variation   | 0.694 1      | 0.687 8          | 0.713 7         |
Taking all the indexes of each region as the data value, the geographical distribution map and bar (Fig.5) are drawn to intuitively analyse the degree of network attention and geographic concentration.

The tourism network attention for Guilin in Guangdong and Guangxi has significantly higher than other provinces. This is because their close geographical locations and similar cultural. In addition to Guangdong and Guangxi, the top 10 followed by Hunan, Jiangsu, Beijing, Zhejiang, Shanghai, Henan, Shandong, Sichuan, Hebei, and Hubei. It can be seen that geographical factors, demographic factors, and economic development affect the network attention. Hunan, Hubei and Sichuan are closer to Guilin geographically. Jiangsu, Beijing, Zhejiang, and Shanghai are relatively developed regions in terms of economy and network. Henan and Shandong are provinces with large populations.

Among the areas with low network attention, although Hong Kong, Macao and Taiwan are geographically close to Guilin, their population is relatively small, so the overall attention is low. Hainan, Yunnan and Guizhou are close to Guilin in geography, but the tourism resources of the three provinces are relatively rich, especially in the Yunnan-Guizhou region, some tourism resources are similar to Guilin. These result in the low network attention of the three provinces to the Guilin tourism. Tibet, Qinghai, Ningxia, Xinjiang, Gansu, Inner Mongolia, Jilin and Heilongjiang are far away from Guilin. In addition, Tibet, Qinghai and Ningxia have relatively small population. It is worth noting that the Northeast region of China (Heilongjiang-Jilin-Liaoning) has the most unfavourable factors in distance, but the network attention is relatively not the lowest. This is due to the great difference in cultural and landscapes between the Northeast and Southwest of China.

3.4. Space Influence Factor

Previous studies have shown that the network attention of different regions can reflect the tourism demand of the region, which is generally affected by the economic development level, population size, network development degree and the distance between the two places [7]. In view of the availability of data, this paper takes 31 provinces and cities in mainland China as the statistical analysis samples.

Since some variables may have linear correlation or cannot explain the dependent variables well, the multiple linear stepwise regression method is used to obtain the optimal regression model. Results show that total population ($X_1$), network development degree ($X_2$) and distance between two places ($X_3$) have significant influences on the spatial distribution characteristics of network attention. The corresponding standardized coefficients were 0.53, 0.40, -0.36, $p$ values were 0.000, 0.001 and 0.005, $R^2 = 0.685$, adj $R^2 = 0.650$. It can be seen that the three regression coefficients of the model have passed the significance test, and the influence factors of regional population size are relatively large.

![Geographical distribution](image-url)
The network attention is positively correlated with the regional population size and network development degree, and negatively correlated with the distance.

4. Conclusion and Discussion
1. Tourism network attention in Guilin is obvious seasonality, and the off-season of Guilin Tourism has large space for development. This is because: a) There are long winter holidays and Spring Festival holidays in the off-season, and the climate are comfort. b) The water flow of Li River is the main factor affecting the seasonal characteristics of Guilin tourism. The government of Guilin, however, has solved the problem of "flood in rainy season and suspension of navigation in dry season" by building three reservoirs. Therefore, tourism demand in the off-season will be greatly improved, and will exceed the current peak season.

2. The mobile network attention accounts for the main share (more than 3/4), with obvious seasonality. The PC network attention fluctuates periodically with working days and rest days. In the era of "Internet plus", we should pay more attention to network marketing, especially mobile Internet marketing strategy. In addition, the Internet is a double-edged sword, which may also spread negative news. Therefore, tourism department should pay special attention to the establishment of good eWOM.

3. The regional distribution of tourism network attention in Guilin is neither centralized nor balanced. The regional population size has a greater impact on the it, followed by the network development and distance. With the gradual improvement of aviation and high-speed rail network, the traffic is more and more convenient, and the factors affecting travel distance will be less.

4. In conclusion, Guilin should develop winter market for Northeast and Northwest of China, that is, "winter tour Guilin". In winter, Northeast and Northwest of China is a bleak season for all things. On the contrary, Guilin's landscape characteristics, southwest minority culture and comfortable climate can bring greater visual, cultural and feel impact to tourists of northeast and northwest. These will enhance the tourist experience and promote the winter market of Guilin. Therefore, northeast and northwest of China have great market development potential. Guilin tourism enterprises and management departments should combine the Internet and new media to do a good job in the activities and publicity of "winter tour of Guilin".

5. Nevertheless, this research has some shortcomings, the data source is from Baidu Index, relatively single, data breadth has limitations. At present, OTAs, social apps and small programs occupy more time online. They greatly reduce the proportion of people using Baidu to search tourism information, but these also provide a richer data source for the study of tourism network attention.

5. Acknowledgments
The research reported in this paper is supported by the Guilin Scientific Research and technology Development Project No. 20180102-2, the Guangxi Young and Middle-aged Teachers Basic Ability Enhancement Project No. 2019KY0292, and the Research Foundation of Guilin University of Technology No. GUTQDJJ2017141.

6. References
[1] Wenqi Ruan, Shuning Zhang, Quanyong Li, Xiangmin Zheng, Spatio-temporal differentiation and influencing factors of chinese's tourism demand to Thailand, Tourism Tribune, 2019, vol. 34, no.5 pp. 76–89.
[2] Ginsberg J, Mohebbi M H, Patel R S, et al, Detecting influenza epidemics using search engine query data, Nature, 2009, vol.457, pp. 1012-1014.
[3] Dergiades T, Mavragani E, Pan B, Google trends and tourists' arrivals: emerging biases and proposed corrections, Tourism Management, 2018, vol.66, pp.108-120.
[4] Li S, Chen T, Wang L, et al, Effective tourist volume forecasting supported by PCA and improved BPNN using baidu index, Tourism Management, 2018, vol.68, pp.116-126.
[5] Sun S, Wei Y, Tsui K, et al, Forecasting tourist arrivals with machine learning and internet search index, Tourism Management, 2019, vol.70, pp.1-10.
[6] Shan L, Rongxu Q, Ling C, Cyberspace attention of tourist attractions based on baidu index: temporal distribution and precursor effect, *Geography and Geo-information Scienc*, 2008, vol.24, no.6, pp. 102-107.

[7] Lijun M, Yun L, Spatiotemporal Characteristics of Residents Tourism Demand for Typical Scenic Spots in Hunan Province Based on Network Attention, *Economic Geography*, 2017, vol.37, no.2, pp. 201-208.