Using Geotagged Social Media Data to Explore Sentiment Changes in Tourist Flow: A Spatiotemporal Analytical Framework

Wei Jiang 1,2,3, Zhengan Xiong 1, Qin Su 1, Yi Long 3,*, Xiaqing Song 3 and Peng Sun 1

1 School of Geography and Tourism, Anhui Normal University, Wuhu 241003, China; jiangweiGIS@whu.edu.cn (W.J.); xza@ahnu.edu.cn (Z.X.); QnSu_ahnu1@ahnu.edu.cn (Q.S.); PengSun_1290s@ahnu.edu.cn (P.S.)
2 Engineering Technology Research Center of Resource Environment and GIS, Anhui Province, Wuhu 241003, China
3 Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Nanjing 210023, China; xqsong@whu.edu.cn
* Correspondence: longyi@njnu.edu.cn; Tel.: +86-27-6877-8969; Fax: +86-25-8589-1742

Abstract: Understanding sentiment changes in tourist flow is critical in designing exciting experiences for tourists and promoting sustainable tourism development. This paper proposes a novel analytical framework to investigate the tourist sentiment changes between different attractions based on geotagged social media data. Our framework mainly focuses on visualizing the detailed sentiment changes of tourists and exploring the valuable spatiotemporal pattern of the sentiment changes in tourist flow. The tourists were first identified from social media users. Then, we accurately evaluated the tourist sentiment by constructing a Chinese sentiment dictionary, grammatical rule, and sentiment score. Based on the location information of social media data, we built and visualized the tourist flow network. Last, to further reveal the impact of attractions on the sentiment of tourist flow, the positive and negative sentiment profiles were generated by mining social media texts. We took Beijing, a famous tourist destination in China, as a case study. Our results revealed the following: (1) the temporal trend of tourist sentiment has seasonal characteristics and is significantly influenced by government control policies against COVID-19; (2) due to the impact of the attraction’s historical background, some tourist flows with highly decreased sentiment strength are linked to attractions; (3) on the long journey to the attraction, the sentiment strength of tourists decreases; and (4) bad traffic conditions can significantly decrease tourist sentiment. This study highlights the methodological implications of visualizing sentiment changes during collective tourist movement and provides comprehensive insight into the spatiotemporal pattern of tourist sentiment.

Keywords: tourist flow; sentiment change; spatiotemporal analysis; geotagged social media data

1. Introduction

Tourism is an important element of many regional economies and accounts for a large amount of human movement within cities. Tourist flow, which refers to the collective movement of tourists from one particular location to another [1,2], is a major concern for tourism researchers and the tourism industry. Visualizing and investigating the multidimensional features of tourist flow are critical in designing leisure and recreation sites, transport, tourism development, and other aspects of the urban structure of cities [3–5].

Traditional data for tourist flow studies have been obtained mainly from questionnaire-based diaries and statistics captured by tourism management departments. The information obtained from questionnaire-based diaries is relatively complete, but the process of collecting diaries is labor intensive, and the data size is limited [6]. Although the statistics can reflect the collective movement of tourists between different cities or countries, these statistics are published mainly on an annual basis and are highly aggregated [7,8]. Due
to the disadvantages of traditional data, it is difficult for most studies based on such data to reveal the spatiotemporal pattern of tourist flow within cities. Other data resources are needed to investigate specific information about tourist movement between different attractions. Location data might provide a solution.

At present, different types of location data, such as mobile phone location data and geotagged social media data, have introduced new opportunities for exploring tourist flow within a city [9]. Compared with traditional data, location data can record more detailed spatial, temporal, demographic, and sentiment information about collective tourist movements. Based on mobile phone location data, Xi et al. [10] investigated the influencing factors of tourist movement from the perspective of party size in Xi’an, China. Using geotagged social media data, Chua et al. [11] proposed a new method of characterizing and visualizing the spatiotemporal pattern of tourist flow among different demographic groups in Cilento, Italy. Tang and Li used Sina Weibo (the Chinese version of Twitter) data to construct a spatial network of tourist flow [12]. Their results revealed the spatial characteristics of the tourist flow network and the function of the tourist node. Location data have been widely applied in tourist flow analysis [13].

However, to the best of our knowledge, the existing studies based on location data have focused mainly on the volume, direction, and demographic features of tourist flow. Few studies concern specific sentiment changes in the process of collective tourist movement. Compared with other features, the sentiment feature of tourist flow may be more important for creating a service stage that offers tourists memorable moments [14,15]. To address this shortfall, a spatiotemporal analytical framework is proposed to address two specific research questions: (1) how can the detailed sentiment changes in tourist flow be effectively visualized, and (2) what is the valuable spatiotemporal pattern of the sentiment changes in tourist flow?

We took Beijing, a famous tourist destination in China, as a case study. Our method first identified tourists from many social media users. Based on social media texts and the sentiment word dictionary, the sentiment of tourists was accurately evaluated. By exploring the trajectory of tourists, we constructed a network of tourist flow and investigated the spatiotemporal dynamics of the sentiment of tourist flow. To further reveal the impact of attractions on the sentiment of tourist flow, the positive and negative sentiment profiles were built by mining social media texts. The results indicate that our method can visualize the detail sentiment change in tourist flow between different attractions in a cost-effective manner. In addition, the results of our study can help managers have a better understanding of the spatiotemporal pattern of the sentiment change in tourist flow, which is essential for providing tourists with optimal tour routes and emotional experiences.

2. Related Works

2.1. Tourist Flow Analysis

In recent years, a large number of scholars have attempted to shed light on tourist flow in certain times and spaces [11,16]. A variety of data types have been applied to tourist flow analysis. Based on the data type, existing studies related to tourist flow can be divided into two groups: those using traditional data and those using location data.

2.1.1. Tourist Flow Analysis Based on Traditional Data

Traditional data refers mainly to statistical data, questionnaire-based diaries, self-administered diaries, and on-site interviews. Most statistical data are collected annually and globally by tourism industry organizations; due to the lack of information about the process of tourist movement, it is difficult to apply these data to investigation of the tourist flow within a city [17]. Travel diaries and interviews can provide detailed spatiotemporal information about tourist movement. Furthermore, diaries and interviews also contain certain information that cannot be obtained from other types of data, such as tourists’ income and travel motivation. Some previous studies using diaries and interviews have focused on revealing the spatial pattern of tourist flow between different attractions [16].
However, there are several shortcomings of diaries and interviews. The spatiotemporal precision of these two data types is lower than that of location data [11]. In addition, the data size is limited, and the process of data collection is time consuming [18]. Due to these disadvantages, the application of traditional data in tourist flow studies is limited.

2.1.2. Tourist Flow Analysis Based on Location Data

Due to the development of location technology, location data have become an important data resource for tourist flow analysis. Location data include mainly GPS, cellular call data records, and geotagged social media data. GPS data are collected by tracking devices and can record detailed spatiotemporal trajectories of tourists. Currently, GPS data show significant potential in studies related to tourist flow, such as tourist movement prediction [19] and tour route planning [20]. The consent of tourists is the premise of GPS data collection, and most tourists are generally not willing to share their movement trajectories in detail. Therefore, the number of volunteers and their active locations are limited [21]. In addition, the cost of purchasing a large number of tracking devices is relatively high [22].

Cellular call data refers to the records of mobile phone activities and is generated by a telecommunication base station. During the process of tourist movement, the telecommunication base station records tourists’ location when they use mobile phones. Exploring the cellular call data of tourists can reveal important and interesting findings related to the patterns of tourist flow [23]. The spatiotemporal precision of cellular call data is significantly lower than that of GPS data because the base station location is used to represent the true location of the tourist. Additionally, cellular call data can record only the locations of the active use of mobile phones, such as making calls, sending messages, and connecting to the Internet [24]. Therefore, the tourist trajectories extracted from cellular call data are incomplete.

In recent years, social media services, such as Twitter, Sina Weibo, Flickr, and Facebook, have been widely used [25]. Each social media user can be treated as a social sensor and can generate a large amount of geotagged social media data. Geotagged social media data can provide information on texts, social networks, and detailed spatial locations down to road level. A growing body of studies has capitalized on these multiple types of information by applying geotagged social media data to the investigation of destination image [26], destination recommendation [27], tourist intention [28], and sentiment [29]. Some previous studies also try to map and analyze tourist flow with geotagged social media data. These studies mainly take each attraction as a node and construct the tourist flow network between different attractions [30–32]. To further explore detailed information about tourist movement, some studies have divided the study area into regular grids and revealed the spatiotemporal characteristics of tourist flow between grid centers [11]. Currently, geotagged social media data is widely used to investigate the volume and direction of tourist flow. However, to the best of our knowledge, few studies have applied geotagged social media data to explore the pattern of sentiment changes in tourist flow.

2.2. Sentiment Analysis

Sentiment analysis refers to utilizing computational linguistics and natural language processing to the investigate people’s sentiment and opinion [33]. Traditionally, sentiment analysis relies heavily on manual coding and data collection through surveys or interviews. Traditional studies suffer from limitations, such as high cost and recall bias [34]. Online content can reflect the publisher’s feeling and sentiment, such as sadness, happiness, anger, and depression. Customer reviews and social media texts, two types of online content, have recently been applied in sentiment analysis related to tourism [6,35].

Based on machine learning, dictionary-based methods, and hybrid methods, most studies have explored the sentiment of online content from two perspectives: the classification of sentiment polarity and the measurement of sentiment strength. From the perspective of polarity classification, a sentiment can be defined as binary or ternary. In the
binary classification, the sentiment is assumed to be either “positive” or “negative” [36]. When “neutral” is added to the polarity between “positive” and “negative”, the binary classification can be extended to a ternary classification [37].

In studies of sentiment measurement, the sentiment strength of customer reviews or social media text is evaluated as a value or level. Compared to polarity classification, sentiment measurement is more suitable for exploring online content that does not reflect a clear sentiment polarity or contains mixed polarities. The dictionary-based method is one of the most widely used methods in the evaluation of sentiment strength; this method measures the strength of text by relying heavily on a sentiment dictionary and predefined rules. Each word in a sentiment dictionary is manually coded as a value or level. In addition, the dictionary-based method requires the researchers to redefine grammatical and syntactical conventions, which can strongly impact the sentiment strength.

The study of tourist sentiment is still in an early stage. Most existing studies have focused on exploring customer feelings about hotel services [38] and investigating spatial patterns of tourist sentiment [29]. Duan et al. evaluated customer experiences in hotels by analyzing customer reviews on the Internet [39]. Based on the evaluated results, they found a strong relationship between customer experiences and satisfaction ratings. Park et al. combined emotion theory and text mining technology to classify tourist sentiment in theme parks into four dimensions [29]. Their results revealed the spatial distribution characteristics of each sentiment dimension. Most existing studies have analyzed and evaluated tourist sentiment within tourism attractions from the perspective of space [39,40]. However, few studies have considered the spatiotemporal pattern of tourist sentiment changes between different attractions.

3. Study Area and Data
3.1. Study Area

Beijing is not only a political and cultural center but is also one of the most famous tourist destinations in China. In 2019, the city received more than 322 million tourists, and tourism revenue amounted to 622.4 billion yuan. The tourism industry plays a very important role in Beijing’s economy. Therefore, we took Beijing as a case study to explore the spatiotemporal pattern of sentiment changes in tourist flow.

There are many historical relics and examples of natural scenery in Beijing. To select the attractions, we investigated the content related to attraction recommendations on Qunar.com (https://www.qunar.com/ (accessed on 1 February 2021)), such as “10 Best Attractions in Beijing” and “12 Things to Do in Beijing”. Qunar.com is one of the most popular travel service platforms in China and is used by millions of users. Through this platform, users can check in at attractions, post their travel diaries, and browse attraction recommendations. Based on the content posted on Qunar.com, 13 popular attractions were selected, as shown in Figure 1. Among these attractions, Wangfujing is the business district and attracts many tourists each day.

3.2. Data Collection and Preprocessing

The geotagged social media data were collected using application programming interfaces. In this study, we applied Sina Weibo data to explore the sentiment of tourist flow. Sina Weibo can be considered the Chinese version of Twitter and is one of the most widely used social media platforms in China [12]. The Sina Weibo company provided an application programming interface named “place/nearby timeline” for searching and collecting geotagged social media data. Based on this application programming interface, we obtained and stored 22,932,987 Sina Weibo microblogs posted between 1 July 2017, and 31 October 2020. The representative samples are shown in Table 1. Some of the attributes of the microblogs are as follows: (1) “ID” and “User_ID” refer to the identification of the microblog and user, respectively; (2) “Created_at” refers to the posting time of the microblog; (3) “Geo” indicates the latitude and longitude of the posting location; and
(4) “Source” refers to the name of the application or phone model that was applied to post the microblogs.

![Figure 1. The distribution of the selected attractions in Beijing.](image)

**Table 1.** Samples of geotagged Sina Weibo microblogs.

| ID | User_ID | Created_at | Text | Geo | POI_Title | Source | Registration Place |
|----|---------|------------|------|-----|-----------|--------|--------------------|
| XX | XX      | 07:45:20 01 May 2018 | 不到长城非好汉，到了长城不遗憾 [Yeah] [Yeah] [Yeah] (["Not a true man unless he comes to the Great Wall"]. It is not regrettable to arrive at the Great Wall [Yeah] [Yeah] [Yeah]). | 116.01387; 40.356033 | 北京八达岭长城 (The Great Wall at Badaling, Beijing) | iPhone 7 | 山东 青岛 (A city in China) |
| XX | XX      | 09:29:51 12 August 2019 | 一年四季的故宫都值得来看看 [微风] (The Forbidden City is worth visiting all the year round [Breeze]). | 116.397316; 39.91814 | 故宫博物院 (The Palace Museum) | iPhone | 陕西 西安 (A city in China) |
| XX | XX      | 14:12:04 24 October 2019 | 观京城美景，练健康体魄! [加油] [微风] (See Beijing’s beauty and exercise the body! [Strong] [Breeze]). | 116.186951; 39.991596 | 香山公园 (Xiangshan Park) | iPhone | 湖北 武汉 (A city in China) |
| XX | XX      | 11:06:53 18 August 2017 | (助我赢取77.77元现金大奖) 助ofo小黄车集齐5种卡，赢77.77元现金大奖 (Help me win a prize of RMB 77.77. Collect 5 kinds of cards by riding shared bikes.) | 116.447613; 39.951815 | Null | PP时光滑 (PP time machine) | 河北 廊坊 (A city in China) |
To create a reliable database for analyzing tourist sentiment, noise needed to be filtered out in the preprocessing phase. The noise among Sina Weibo microblogs refers mainly to reposted microblogs, advertisements, and microblogs posted by bots. For example, the last microblog in Table 1 is noise. Geotagged Sina Weibo microblogs cannot be reposted. Therefore, geotagged microblogs have much less noise than microblogs without location information. Based on the noise filtering method proposed by previous studies [41], we removed the noise and retained 22,911,295 Sina Weibo microblogs for further analysis.

4. Method

In this section, we provide a detailed discussion of our framework for analyzing the spatiotemporal pattern of sentiment changes in tourist flow.

4.1. Tourist Identification

Tourist identification is the premise of exploring the sentiment of tourist flow. Based on the method of previous studies [11], users with two particular characteristics were identified as tourists: (1) posting Sina Weibo microblogs within attractions, and (2) having registered in a place other than Beijing. We first filtered out the users who had not posted any microblogs within attractions. Then, the tourists were identified by analyzing their place of registration.

On the Sina Weibo platform, users are required to select the nation, province, and city of their registration place when they create a new account. Owing to privacy considerations, some users select the “other” option for the registration place, and it is difficult to extract tourists from these users. To fill this gap, we applied a 4-step identification method based on the theory of Chua et al. (2016): (1) we applied the “statuses/user timeline” to collect all geotagged Sina Weibo microblogs; (2) we calculated the average day spent in Beijing by all users, $\tilde{d}$; (3) for each user $i$ with the “others” option, we obtained the probability index $p_i$ by the following equation:

$$p_i = \begin{cases} 
\frac{Bd_i}{Cd_i} & d_i \geq \tilde{d} \\
0 & d_i < \tilde{d}
\end{cases} \quad (1)$$

where $Bd_i$ and $Cd_i$ are the number of days spent by user $i$ in Beijing and China, respectively; and (4) we identified user $i$ as a tourist when $p_i \leq 0.5$ because tourists will not spend most of their time in Beijing. In this study, a total of 761,480 tourists were finally detected for sentiment evaluation.

4.2. Tourist Sentiment Evaluation

Social media texts can reflect a large amount of information related to tourist sentiment. A dictionary-based method was proposed to evaluate tourist sentiment strength by exploring social media texts. As existing studies have noted, sentiment evaluation is domain dependent [42]. Therefore, to improve the accuracy of tourist sentiment evaluation, we first constructed a sentiment dictionary in the tourism domain.

4.2.1. Sentiment Dictionary Construction

A sentiment dictionary includes the part of speech and sentiment polarity of each word. By constructing a sentiment dictionary, we can quantitatively evaluate the positive and negative sentiment strength of social media texts. Our sentiment dictionary was constructed on the basis of the dictionary named “HowNet”. The HowNet dictionary is provided by the China National Knowledge Infrastructure and is one of the commonly used Chinese sentiment dictionaries [43]. Chinese words in this dictionary are classified as different types, such as positive, negative, degree adverb, and negative adverb. The HowNet dictionary is designed without considering the characteristics of tourist sentiment and social media texts. Therefore, the positive and negative words in this dictionary are not sufficient for evaluating tourist sentiment in social media texts with high accuracy.
To expand the HowNet dictionary, the sentiment words and emoji in the tourism domain were identified manually. Ten volunteers were gathered to analyze 20,000 social media texts. For each text, each volunteer first extracted sentiment words and emoji; then, the sentiment polarities of words and emoji were classified. A total of 373 sentiment words and 98 sentiment emoji were finally identified. After comparison with the HowNet dictionary, 204 new words and all emoji were added to the HowNet dictionary to construct a new sentiment dictionary. The constructed dictionary contained 6778 sentiment words and 98 sentiment emoji.

4.2.2. Grammatical Rule Construction

In addition to sentiment words and emoji, degree adverbs, negative adverbs, and adversative conjunctions can significantly influence sentiment strength in social media texts [44]. To consider the impacts of adverbs and conjunctions, we constructed grammatical rules for degree adverbs, privative words, and adversative conjunctions that embody grammatical conventions for emphasizing or weakening sentiment strength.

1) Degree adverbs. In the HowNet dictionary, most Chinese degree adverbs were included and divided into 6 categories based on their intensity. Category 1 to category 6 were assigned multiples of 0.5 to 3, meaning the sentiment strength of a word or emoji is multiplied by 0.5 to 3 when it is used with a degree adverb.

2) Privative words. Privative words, such as “别” (do not), “没有” (none), and “不曾” (have not), play an important role in sentiment strength evaluation. As some studies have pointed out, Chinese privative words can deny the sentiment polarity of words or emoji that appear behind the privative words [44]. For a sentence containing privative words, we focused on the number and position of the privative words. Specifically, the sentiment polarities of the words and emoji behind the privative words are reversed when the number of privative words is odd; otherwise, the polarity remains unchanged.

3) Adversative conjunctions. Based on the theory of D. Zheng, Tian, and Zhang [45], Chinese adversative conjunctions can be divided into two categories, as shown in Table 2. The adversative conjunctions in the first category, such as “虽然” (although) and “不管” (whatever), are mainly used in the subordinate clause and indicate that the sentiment polarity reversed in the principal clause is reversed. “但是” (but) and “只不过” (nothing but) are representative words in the second category; the words in this category are used in principal clauses. In a sentence containing second-category words, the sentiment is expressed mainly in the principal clause. Based on consideration of the impacts of adversative conjunctions, the rules for determining sentiment polarity were constructed, as shown in Table 3.

Table 2. The list of adversative conjunctions.

| Category       | Number | Samples                                                                 |
|----------------|--------|-------------------------------------------------------------------------|
| The first category | 14     | “虽然”, “虽是”, “虽说”, “尽管”, “固然”, “即使”, “纵使”, “即使”, “无论”, “纵然”, “不论”, “不管”, “任凭”, “原本” (The meaning of Chinese words in the first category is similar to “although” or “whatever”) |
| The second category | 14     | “但是”, “可是”, “不过”, “倒是”, “然而”, “然则”, “但”, “却”, “只是”, “只不过”, “才”, “可”, “还是”, “而” (The meaning of Chinese words in the second category is similar to “but”) |
Table 3. The rules for determining the sentiment polarity of texts with adversative conjunctions.

| Sentence Structure | Sentiment Polarity |
|--------------------|--------------------|
| First category + positive words | Negative |
| First category + odd number of privative words + positive words | Positive |
| First category + even number of privative words + positive words | Negative |
| First category + negative words | Positive |
| First category + odd number of privative words + negative words | Negative |
| First category + even number of privative words + negative words | Positive |
| Second category + positive words | Positive |
| Second category + odd number of privative words + positive words | Negative |
| Second category + even number of privative words + positive words | Positive |
| Second category + negative words | Negative |
| Second category + odd number of privative words + negative words | Positive |
| Second category + even number of privative words + negative words | Negative |

4.2.3. Sentiment Score Construction

The sentiment score was constructed to evaluate the comprehensive sentiment strength of social media texts. There are two steps for constructing the sentiment score of text \( t \) which was posted by tourist \( p \). First, we checked the structure of each sentence in social media text \( t \). For the sentences which are consistent with the structures in Table 3, the positive or negative sentiment polarity of these sentences were quantified by applying grammatical rules. Second, the left sentences were segmented into words. By checking each word, we can count the numbers of the positive and negative sentiment words. Based on the results of the two steps, the sentiment score of the text \( t \) was determined as follows:

\[
S_p^t = \left( S_{pos}^t - S_{neg}^t \right) + \left( W_{pos}^t - W_{neg}^t \right)
\]

where \( S_{pos}^t \) and \( S_{neg}^t \) refer to the number of positive and negative sentences which are measured by applying grammatical rules, respectively; \( W_{pos}^t \) and \( W_{neg}^t \) refer to the number of positive and negative words, respectively. The sentiment score was applied to visualize the sentiment of the tourist flow.

4.3. Sentiment Visualization of the Tourist Flow Network

Sentiment is one of the important features of tourist flow. Visualizing the sentiment of tourist flow is the basis of sentiment analysis. In this study, sentiment scores and spatiotemporal information of social media data were combined to construct and visualize sentiment changes in the tourist flow network.

Quantifying the sentiment change of an individual tourist during his or her movement is the premise of visualizing sentiment in tourist flow. Based on the sentiment score of the text of geotagged social media data, we can obtain the sentiment of individual tourists in different locations. To visualize tourist movement between different locations, the trajectory of each tourist was extracted. According to previous study \([11]\), each sequential pair of geotagged social media data whose time interval was less than 6 h can be treated as a pathway and can reflect the movement of a tourist from one location to another. Based on each sequential pair of geotagged microblogs, two steps were needed to extract the trajectory of individual tourist. First, the origin and destination locations of each pathway were restricted to the nearest nodes of the road network. Second, to extract detailed information for each pathway, the shortest network path was obtained by applying the Dijkstra algorithm. Each trajectory was represented by the shortest network path, which is constituted by a sequence of directed links between adjacent network nodes. For example, for individual tourist \( p \), he or she posted one geotagged microblog in node \( i \) and node \( i \), sequentially. His or her trajectory can be extracted and shown in Figure 2.
Based on the directed links between adjacent network nodes, the sentiment change of an individual tourist during his or her movement can be quantified. The directed link refers to the process that a tourist moves from one node to the next adjacent node. For individual tourist $p$, the sentiment changes during the tourist's movement from node $i - 1$ to node $i$ can be calculated as follows:

$$SC_{i-1 \rightarrow i}^p = S_{i}^p - S_{i-1}^p$$  \hspace{1cm} (3)$$

where $S_{i-1}^p$ and $S_{i}^p$ refer to the sentiment score of tourist $p$ when he or she was located in node $i - 1$ and node $i$. Based on previous studies [11,29], the sentiment strength of tourist $p$ can be assumed to be the sentiment strength of a tourist at node 1 before he or she reached node $i$.

The sentiment change in the tourist flow network was visualized by aggregating an individual tourist’s trajectory. To reveal the spatial pattern of the tourist flow network in different time periods, the directed links in each trajectory were categorized as daytime (6:00–17:59) and nighttime (18:00–5:59). By aggregating directed links in different time periods, the sentiment changes in directed tourist flow from node $i - 1$ to node $i$ in two time periods can be calculated as follows:

$$FSC_{i-1 \rightarrow i}^{\text{day}} = \sum_{p=1}^{n} SC_{i-1 \rightarrow i}^p$$ \hspace{1cm} (4)$$

$$FSC_{i-1 \rightarrow i}^{\text{night}} = \sum_{j=1}^{m} SC_{i-1 \rightarrow i}^j$$ \hspace{1cm} (5)$$

where $SC_{i-1 \rightarrow i}^j$ refers to the sentiment changes of tourist $j$ who moved from node $i - 1$ to node $i$; $n$ and $m$ refer to the number of tourists who moved from node $i - 1$ to node $i$ in daytime and nighttime, respectively.

Based on the calculated sentiment changes in directed tourist flow, the sentiment in tourist flow network in daytime and nighttime can be constructed. The number of distributions of geotagged microblogs for each node in daytime and nighttime are shown in Figure 3. For the nodes in daytime, the minimum number of geotagged microblogs
was 87 and the average number was 2810. For the nodes in nighttime, the minimum number of geotagged microblogs was 75 and the average number was 2656. The number of distributions of directed links with changed sentiment strength in each directed flow are shown in Figure 4. In the daytime, 17.8% flows contained no more than 10 links and the average number was 55. In the nighttime, 30.9% flows contained no more than 10 links and the average number was 38. Figure 4 shows that some flows contained very few directed links. To avoid the bias of a small set of links, the flows with no more than 10 links were identified and marked as insignificant.

Figure 3. The number of distributions of geotagged microblogs for each node in daytime (a) and nighttime (b).

Figure 4. The number of distributions of directed links with changed sentiment strength in each directed flow in daytime (a) and nighttime (b).

To optimize the visualization of the sentiment changes of the tourist flow network, the sentiment change was normalized as follows:

\[ MSC_{\text{day}}^{i-1 \rightarrow i} = 2 \times \frac{\text{FSC}_{\text{day}}^{i-1 \rightarrow i} - \text{FSC}_{\text{min}}^{\text{day}}}{\text{FSC}_{\text{max}}^{\text{day}} - \text{FSC}_{\text{min}}^{\text{day}}} - 1 \]  

\[ MSC_{\text{night}}^{i-1 \rightarrow i} = 2 \times \frac{\text{FSC}_{\text{night}}^{i-1 \rightarrow i} - \text{FSC}_{\text{min}}^{\text{night}}}{\text{FSC}_{\text{max}}^{\text{night}} - \text{FSC}_{\text{min}}^{\text{night}}} - 1 \]  

where \( \text{FSC}_{\text{min}}^{\text{day}} \) and \( \text{FSC}_{\text{max}}^{\text{day}} \) refer to the minimum value and maximum value, respectively, of sentiment change in daytime; \( \text{FSC}_{\text{min}}^{\text{night}} \) and \( \text{FSC}_{\text{max}}^{\text{night}} \) refer to the minimum value and maximum value, respectively, of sentiment change in nighttime. Based on the optimized tourist flow network, sentiment changes in the directed flow were finally visualized by drawing arrows from the starting node to the ending node. The color of the directed flow represents the level of the changes in sentiment strength.
4.4. Sentiment Profile Construction

To reveal the reasons for sentiment change in tourist flow, a sentiment profile was constructed to explore the social media texts. A sentiment profile is a network structure that is generated by counting and clustering high-frequency words related to sentiment characteristics. In this study, prepositions were first filtered out to remove the impact of meaningless words. Then, we applied the ROST CM6 software to quantify the co-occurrence relationship between high-frequency words in positive or negative social media texts. Based on the co-occurrence matrix, the semantic relationship between different words was clustered using Gephi. Gephi is popular software used for clustering, visualizing, and network analysis. After clustering, Gephi visualized each high-frequency word as a network node; the size of the node represents the number of lines that link to this node. The nodes that have the same color indicate the same cluster. The line between nodes represents the co-occurrence relationship between words. The width of the line indicates the weight value between words.

5. Results

5.1. Temporal Pattern

The dynamics of tourist sentiment follow certain temporal patterns. The daily sentiment strength from 1 June 2017, to 31 October 2020, was determined, as shown in Figure 5. Figure 5 shows that the trend of sentiment strength has obvious seasonal characteristics. Specifically, there is a breakpoint in the sentiment strength in January or February. The Spring Festival in these two months may lead to this breakpoint. During the Spring Festival holiday, Chinese people tend to gather to celebrate the festival, and some of them then spend their remaining holiday time traveling to Beijing. Therefore, the total sentiment strength in Beijing during January and February first decreases and then increases.

![Figure 5. The dynamics of tourists’ daily sentiment strength.](image)

The temporal trend of tourist sentiment strength is significantly influenced by government control policies against COVID-19. Figure 5 shows that tourist sentiment strength remained at a low level between January 2020 and May 2020. In January 2020, COVID-19 was first proved to be transmitted from person to person. To prevent the transmission of COVID-19, the Chinese government introduced stringent policies to limit people’s activities, especially long-distance travel. Therefore, the number of tourists in Beijing was low and their sentiment tended to be negative. After May 2020, the sentiment strength significantly increased and peaked in October 2020. This is because tourism recovered quickly when the epidemic was under control. In October 2020, a large number of tourists enjoyed their first seven-day holiday in Beijing after the outbreak of COVID-19.
5.2. Spatial Pattern

To reveal the spatial pattern of sentiment changes during tourist movement, the spatial distribution of the sentiment changes in tourist flow was analyzed. In addition, the tourist flow networks in daytime and nighttime were compared.

5.2.1. Daytime

Figure 6 shows the spatial distribution of sentiment change in tourist flow in the daytime. The color of the directed flow represents the level of increased or decreased sentiment strength. The attraction-related node refers to the node around or within attractions. Most increased sentiment flows at a high level (the flows shown in red) are attraction-related flows. Attraction-related flow refers to the tourist flow that is linked to attraction-related nodes. The experiences within attractions may result in a significant increase in tourist sentiment, whereas not all attraction-related flows represent highly increased sentiment flows. Some decreased sentiment flows at a high level (the flows shown in dark blue) were found to be around the Old Summer Palace. In addition, for the Badaling Great Wall in the northwest part of the study area, we find that the sentiment of the tourist flow did not increase until the tourists reached the Great Wall. The Great Wall is far from the city center of Beijing, and it takes tourists considerable time to reach this attraction; thus, the long journey may decrease the tourist sentiment strength. After experiencing the Great Wall, the sentiment of the tourist flow significantly increased.

![Figure 6](image-url)

**Figure 6.** The spatial distribution of sentiment change in tourist flow in the daytime.

To reveal the impact of attractions on tourist sentiment in the daytime, we constructed the sentiment profiles of all attraction-related tourist flows. The positive sentiment profile is
shown in Figure A1 in Appendix A. Based on Figure A1, we identified some co-occurrence relationships: (1) “Tiananmen”–“Forbidden City”–“red wall”–“beautiful”; (2) “South Luogu Lane”–“delicious”; (3) “lovely”–“girls”. These relationships indicate that tourists’ positive sentiment is strongly related to the attractions. Specifically, the beautiful scenery, lovely girls, and delicious food within attractions can significantly increase the tourist sentiment strength.

The negative sentiment profile is shown in A2 in Appendix A. Based on Figure A2, some findings can be listed: (1) “Great Wall”–“on the way”–“tired” indicates that the long journey to the Great Wall can prompt the expression of negative sentiment. This co-occurrence relationship can explain why tourist sentiment strength always decreased on the way to the Great Wall; (2) “mask”–“epidemic situation”–“against” suggests that the COVID-19 epidemic can decrease the tourist sentiment strength. Due to the fear of the COVID-19, tourist express negative sentiment; (3) “sadness”–“unfortunately”–“Old Summer Palace”–“ruin” demonstrates that the ruins of the Old Summer Palace make tourists feel sad. The Old Summer Palace was destroyed, which is considered a shame for Chinese people. The history of the attraction has a significant impact on tourist sentiment. Therefore, dark blue flows can be found to be linked to the Old Summer Palace.

Figure 6 illustrates that decreased sentiment flows at a high level appeared in Area A. To explore the reasons for dark blue flows, the negative profile for these flows within Area A were constructed. As shown in Figure A3 in Appendix A, we find “Worry”–“Traffic jam”–“minute” and “subway”–“hate”–“anger”, indicating that tourist sentiment was decreased by traffic congestions.

5.2.2. Nighttime

As shown in Figure A4 in Appendix A, the spatial distribution of sentiment changes during tourist movement in the nighttime is different from that in the daytime. The highly increased sentiment flows around some attractions disappeared; these attractions, such as the Great Wall, Xiangshan Park, and Summer Palace, were located mainly in the northwest part of the study area. Some highly increased flows still appear around the attractions in the middle part of the Beijing and 798 Art District. Among these attractions, Shichahai and South Luogu Lane are entertainment districts. Wangfujing is the business district and attracts tourists in the nighttime. In addition, highly decreased flows can also be found within Area A in the nighttime.

The positive profile of the attraction-related flows in the nighttime is shown in Figure A5 in Appendix A. Some attractions, such as the Great Wall, Old Summer Palace, and Summer Palace, cannot be found in the positive profile. This is because these attractions were closed and cannot impact tourist sentiment. Compared to the positive profile in the daytime, the nighttime profile contains more words related to food. The negative profile of attraction-related flows in the nighttime is shown in Figure A6 in Appendix A. Due to the impact of alcoholic drinks in bars, tourists tend to express negative sentiment. “Wangfujing”–“go shopping”–“cost performance” indicates that commodities in the Wangfujing have low price performance. The bad shopping experiences decrease the tourist sentiment strength. Based on the results of the spatial analysis of sentiment changes, spatial patterns of tourist sentiment can be summarized as follows:

(1) Most highly increased flows were linked to the attraction-related nodes.
(2) Due to the impact of an attraction’s historical background, highly decreased flows can be found around attractions.
(3) On the long journey to the attraction, the sentiment strength of tourists decreased.
(4) Bad traffic conditions can significantly decrease tourist sentiment.

6. Discussion

Existing studies applying geotagged social media data for tourist flow analysis have focused mainly on the volume changes in tourist flow [16,46]. For instance, some studies have provided insights into the spatial, temporal, and demographic characteristics of col-
lective tourist movement [11]. However, what is the spatiotemporal pattern of sentiment changes in tourist flow? Sentiment is an important feature of tourist flow and is necessary for optimizing the tourist experience [15]. Our analytical framework indicates that geotagged social media data can be a reliable data resource for exploring tourist sentiment. Specifically, our approach attempts to reveal the spatiotemporal pattern of tourist sentiment changes. In comparison to the existing methods, our approach can visualize sentiment changes in tourist flow. Although some previous studies have attempted to explore tourist sentiment, they have been limited to the spatial characteristics of tourist sentiment within attractions [29,40].

Some limitations of this study should be noted. Our approach is driven principally by multidimensional information that social media users generate without a ground truth for verification. Therefore, the discovered spatiotemporal pattern of sentiment changes might be somewhat misleading if the representation of social media users is unaccounted for. Some previous studies have demonstrated that the majority of Sina Weibo users are young adults [41]. Social media users cannot represent the actual tourist population. In addition, the geotagged social media data cannot reflect the complete travel itineraries of individual tourists. From this perspective, it is important to acknowledge that the disadvantages of social media data may impact the accuracy of sentiment analysis results. Despite these limitations, the proposed analytical framework provides valuable and alternative insights into the spatiotemporal pattern of tourist sentiment changes. Our findings can complement the current understanding of tourist flows from the new perspective of tourist sentiment.

7. Conclusions

This paper proposes a new analytical framework for exploring the spatiotemporal pattern of tourist sentiment changes based on geotagged social media data. The framework focuses on investigating the spatial distributions of tourist sentiment changes in different time periods. The results provide a comprehensive insight into sentiment changes during collective tourist movements. Beijing was taken as a case study, and geotagged Sina Weibo data posted from 2017 to 2020 were applied to quantify tourist sentiment. We can draw the following conclusions:

1. The temporal trend of tourist sentiment has seasonal characteristics and is significantly influenced by government control policies against COVID-19.
2. Most highly increased flows were linked to the attraction-related nodes.
3. Due to the impact of an attraction’s historical background, tourist flows with highly decreased sentiment strength can be found around attractions.
4. On the long journey to the attraction, the sentiment strength of tourists decreased.
5. Bad traffic conditions can significantly decrease tourist sentiment.

In summary, the results indicate that our method can provide sophisticated descriptions of sentiment changes in tourist flow in a cost-effective manner. In addition, the framework can reveal valuable spatiotemporal patterns of tourist flow. Based on our study, tourism managers can identify more effective strategies to optimize tour routes and emotional experiences. For example, from our results, managers can identify the roads where tourist sentiment was significantly decreased by bad traffic conditions; by replacing these roads, the emotional experiences can be optimized.

Our study was driven by geotagged social media data. The age structure of social media users is different from that of the real world. Therefore, social media data can only be used as an approximate representation of tourist sentiment in the real world. In the future, our research team will investigate the impact of the representability of social media data on spatiotemporal patterns of tourist sentiment. In addition, the privacy of social media users is also a crucial concern. We will focus on the protection of the privacy of social media users and provide guidance on developing academic ethical standards in the application of social media data.
Author Contributions: Conceptualization, Wei Jiang and Zhengan Xiong; methodology, Zhengan Xiong; software, Zhengan Xiong and Qin Su; validation, Yi Long and Xiaqing Song; formal analysis, Peng Sun; investigation, Qin Su; resources, Wei Jiang; data curation, Zhengan Xiong; writing—original draft preparation, Wei Jiang; writing—review and editing, Wei Jiang; visualization, Xiaqing Song; supervision, Yi Long; project administration, Qin Su; funding acquisition, Wei Jiang. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Nature Science Foundation of Anhui province, China, grant number 1908085QD165; the Open Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, grant number 20104; the National Natural Science Foundation of China, grant number 41571382; the National Natural Science Foundation of China, grant number 42071175.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from SINA and are available https://open.weibo.com/ (accessed on 15 February 2021) with the permission of SINA.

Acknowledgments: The authors would like to thank the anonymous reviewers and the editors for their valuable comments and suggestions on earlier versions of the manuscript.

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

Appendix A

Figure A1. The positive sentiment profile of tourist flows linked to attractions in the daytime.
Figure A2. The negative sentiment profile of tourist flows linked to attractions in the daytime.

Figure A3. The negative sentiment profile of tourist flows within Area A in the daytime.
Figure A4. The spatial distribution of sentiment change in tourist flow in the nighttime.

Figure A5. The positive sentiment profile of flows linked to attractions in the nighttime.
References

1. Jin, X.; Qu, M.; Bao, J. Impact of crisis events on Chinese outbound tourist flow: A framework for post-events growth. *Tour. Manag.* 2019, 74, 334–344. [CrossRef] [PubMed]

2. Huang, Y.; Shi, K.; Zong, H.; Zhou, T.; Shen, J. Exploring Spatial and Temporal Connection Patterns among the Districts in Chongqing Based on Highway Passenger Flow. *Remote Sens.* 2019, 12, 27. [CrossRef]

3. Ali, F.; Kim, W.G.; Li, J.; Jeon, H.-M. Make it delightful: Customers’ experience, satisfaction and loyalty in Malaysian theme parks. *J. Destin. Mark. Manag.* 2018, 7, 1–11. [CrossRef]

4. Basarangil, I. The relationships between the factors affecting perceived service quality, satisfaction and behavioral intentions among theme park visitors. *Tour. Hosp. Res.* 2018, 18, 415–428. [CrossRef]

5. Chen, B.Y.; Li, Q.; Wang, D.; Shaw, S.-L.; Lam, W.H.; Yuan, H.; Fang, Z. Reliable Space–Time Prisms Under Travel Time Uncertainty. *Ann. Assoc. Am. Geogr.* 2013, 103, 1502–1521. [CrossRef]

6. Claster, W.; Pardo, P.; Cooper, M.; Tajeddini, K. Tourism, travel and tweets: Algorithmic text analysis methodologies in tourism. *Middle East. J. Manag.* 2013, 1, 81. [CrossRef]

7. Leask, A. Visitor attraction management: A critical review of research 2009–2014. *Tour. Manag.* 2016, 57, 334–361. [CrossRef]

8. Shi, Y.; Qi, Z.; Liu, X.; Niu, N.; Zhang, H. Urban Land Use and Land Cover Classification Using Multisource Remote Sensing Images and Social Media Data. *Remote Sens.* 2019, 11, 2719. [CrossRef]

9. Zhou, Y.; Fang, Z.; Thill, J.-C.; Li, Q.; Li, Y. Functionally critical locations in an urban transportation network: Identification and space–time analysis using taxi trajectories. *Comput. Environ. Urban Syst.* 2015, 52, 34–47. [CrossRef]

10. Xi, Z.; Xiaoni, L.; Yuanyuan, L.; Jun, L.; Jun, A. Tourist movement patterns understanding from the perspective of travel party size using mobile tracking data: A case study of Xi’an, China. *Tour. Manag.* 2018, 69, 368–383. [CrossRef]

11. Chua, A.; Servillo, L.; Marchegiani, E.; Moere, A.V. Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tour. Manag.* 2016, 57, 295–310. [CrossRef]

12. Tang, J.; Li, J. Spatial network of urban tourist flow in Xi’an based on microblog big data. *J. China Tour. Res.* 2016, 12, 1–19. [CrossRef]

13. Mou, N.; Yuan, R.; Yang, T.; Zhang, H.; Tang, J.; Makonen, T. Exploring spatio-temporal changes of city inbound tourism flow: The case of Shanghai, China. *Tour. Manag.* 2020, 76, 103955. [CrossRef]

14. Philander, K.; Zhong, Y. Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *Int. J. Hosp. Manag.* 2016, 55, 16–24. [CrossRef]

15. Bigné, J.E.; Andreu, L.; Gnoth, J. The theme park experience: An analysis of pleasure, arousal and satisfaction. *Tour. Manag.* 2005, 26, 833–844. [CrossRef]
