A Tourist Behavior Analysis Framework Guided by Geo-Information Tupu Theory and Its Application in Dengfeng City, China

Zhihui Tian, Yi Liu, Yongji Wang * and Lili Wu

Abstract: With the development of tourism and the change in urban functions, the analysis of the spatial pattern of urban tourist flows has become increasingly important. Existing studies have explored and analyzed tourist behavior well, using the appropriate digital footprint data and research methods. However, most studies have ignored internal mechanisms analysis and tourism decision making. This paper proposed a novel framework for tourist behavior analysis inspired by geo-information Tupu, including three modules of the spatiotemporal database, symptom, diagnosis, and implementation. The spatiotemporal database module is mainly used for data acquisition and data cleaning of the digital footprint of tourists. The symptom module is mainly used for revealing the spatial patterns and network structures of tourist flows. The diagnosis and implementation module is mainly used for internal mechanism analysis and tourism decision making under different tourist flow patterns. This paper applied the proposed research framework to Dengfeng City, China, using online travel diaries as the source of digital footprint data, to analyze its tourist behavior. The results were as follows: tourist flows of Dengfeng were unevenly distributed, thus forming an obvious core–periphery structure with intense internal competition and unbalanced power. The difference in tourism resources between its northern and southern areas remains a challenge for future tourism development in Dengfeng.

Keywords: geo-information Tupu; tourist flow; digital footprint; online travel diary; social network analysis; Dengfeng

1. Introduction

Tourist flow refers to the migration of tourists in the tourism space [1], which reflects the flow of tourists and the difference in tourism resources. Moreover, tourist flow reveals the phenomenon of spatial interaction between tourist nodes. The analysis of interactive networks based on tourist flow can be helpful to express the spatial agglomeration and diffusion phenomena of tourists, and analyze the roles, functions and interactions of tourist nodes. Therefore, it is indicative of allocating regional tourism resources and rationally integrating regional tourism space [2].

Tourist flow studies, which aim to reveal tourist behavior from the two perspectives of data sources and research methods, have become a hot topic in tourism research. In terms of data sources, the approaches to acquiring data have been gradually enriched, from early, traditional ways to advanced, diverse ways. Early, traditional methods mainly included questionnaire surveys [3–6] and statistical yearbooks [7–9]. With the deep integration of the Internet and the tourism industry, tourists use various electronic products to record travel time and trajectory, forming a rich digital footprint [10]. Advanced digital footprint data can be obtained in diverse ways through various approaches, such as online travel diaries [11–13], Weibo check-in data [14–16], mobile positioning data [17–19], and geo-tagged photos [20–22]. It has the advantages of low collection cost, wide temporal and
spatial coverage, and sustainable tracking [23]. Therefore, the tourists’ digital footprint data obtained in social media provide strong support for the acquisition of data sources for tourist flow research. In terms of research methods, based on the digital footprint, the approaches of tourist flow study have been gradually enriched, from traditional geographic methods to interdisciplinary and multi-perspective methods. The early methods mainly focus on spatial pattern analysis of tourist flow, including temporal characteristic analysis [24], spatial characteristic analysis [25], and spatiotemporal evolution characteristic analysis [26]. With the development of interdisciplinary integration, some scholars have used social network analysis [27–29] to study the network structure of tourist flow from the perspective of structural relationships. Social network analysis can quantify the roles, functions, and spatial interactions of the nodes [8], which is a more intuitive way to analyze tourist behavior. For example, Bindan Zeng [30] analyzed the network structural characteristics of Japan’s inbound tourism based on the social network analysis method, using tourists’ digital footprint data. Han et al. [31] used a questionnaire survey as the data source and used social network analysis to analyze the tourism networks and purchase trends of Chinese and Japanese tourists visiting Korea.

Existing studies have explored and analyzed tourist behavior (or tourist flow) well, using the appropriate digital footprint data and research methods. However, most studies have ignored internal mechanism analysis and tourism decision making. A complete tourist behavior pattern should include the following elements: data acquisition, tourist flow analysis, internal mechanisms analysis, and tourism decision making [32,33]. It coincides with geo-information Tupu theory [34,35]. Geo-information Tupu is a geographical spatiotemporal analysis method consisting of symptom, diagnosis, and implementation [36]. It is used to discover the spatial and temporal knowledge and laws of geology, and to provide application services for social and economic development [37]. For example, based on geo-information Tupu theory, Du et al. [38] used crop classification data from 2009–2017 to analyze the planting patterns in a black soil area. Then, corresponding agricultural decisions were made, such as making a rotational fallow system of cultivated land, improving soil organic matter content, increasing crop yield, and so on. Overall, geo-information Tupu is the combination of cognition, methods, and maps [39]. Applying it to the study of tourist flow characteristics may provide a new research perspective for tourist behavior analysis.

To achieve a more complete tourist behavior analysis process, this paper proposes a novel framework from the perspective of geo-information Tupu theory for studying tourist behavior, and applies it to Dengfeng City, China. The proposed framework could help managers have a better understanding of the spatiotemporal pattern for tourist flow and then provide managers with a more suitable decision-making reference for tourism planning management. In addition, it further enriches the theory and method of tourist behavior analysis.

2. Materials and Methods

2.1. Case Study and Data Source

Dengfeng City is located in the west–central region of Henan Province, China (Figure 1). There is Songshan Mountain in the north and Jishan Mountain and Daxiongshan Mountain in the south. At present, Dengfeng City has one world cultural heritage and one national 5A tourist attraction. In recent years, tourism in Dengfeng has undergone vigorous development. However, the proportion of the tourism industry in the city’s economy is still low. It has world-level cultural tourism resources, but has not exerted its advantages [40]. To make full use of the advantages of Dengfeng’s cultural tourism resources, this paper analyzed Dengfeng’s tourist behavior and made corresponding tourism decisions.
Online travel diaries are mainly produced by tourists with certain travel experiences. These travel diaries record the time and trajectory of tourists’ travels, with good traceability and detail [41]. After comparing the common travel websites and social media of China, the online travel diary data generated by tourists visiting Dengfeng from Qunar.com (www.qunar.com (accessed on 16 July 2021)), Ctrip.com (www.ctrip.com (accessed on 16 July 2021)), and Mafengwo.com (www.mafengwo.cn (accessed on 16 July 2021)) were selected as the digital footprint data sources.

2.2. Methods

Geo-information Tupu is a series of multidimensional maps that uses geoscience analysis to describe the status quo and establishes a spatiotemporal model to analyze the past and virtual future [42], including the three parts of the spatiotemporal database, symptom, and diagnosis and implementation. Inspired by the hierarchical structure of geo-information Tupu [37], this paper established three modules of the spatiotemporal database, tourist flow analysis (symptom), and tourism decision making (diagnosis and implementation) to study tourist behavior. Among them, the spatiotemporal database module was used for data acquisition and data cleaning of the tourists’ digital footprint. Based on the establishment of the database, the symptom module was used to analyze tourist flow patterns by information extraction models. According to the expression of tourist flow from the symptom tupu module, the diagnosis and implementation module was used to analyze the internal mechanisms and make tourism decisions under different tourist flow patterns. The detailed schematic is shown in Figure 2.

2.2.1. Establishment of Spatiotemporal Database Based on Digital Footprint Data

The spatiotemporal database module is mainly used for data acquisition and data cleaning of the digital footprint of tourists. First, this paper used online travel diaries as the source of the digital footprint data. We used the octopus collector (www.bazhuayu.com (accessed on 16 July 2021)) to collect travel diaries on tourism websites from 1 January 2015 to 31 December 2018. Note that there were some information errors and logical problems in the online travel diaries, such as advertising posts, incomplete travel diaries, duplicate travel diaries, and isolated point data. Therefore, it was necessary to manually clean the data before analyzing the tourist flow. For the cleaned data, we used mathematical statistics to aggregate and generate the attractions database.

Second, the spatiotemporal database was established, which included a database of online travel diaries, a database of attraction coordinates, and a database of attraction visits. In the digital footprint database, each record mainly included the user ID, tourist date,
tourist node, as shown in Table 1. In the database of attraction coordinates, each record mainly included the name of the attraction, along with its longitude and latitude, as shown in Table 2. In the database of attraction visits, each record mainly included the name of the attraction and the visit frequency of tourists in 2015–2018, as shown in Table 3.

Figure 2. Analysis framework of tourist flow network inspired by geo-information Tupu theory.

Table 1. Online travel diaries data.

<table>
<thead>
<tr>
<th>Field</th>
<th>Explanation</th>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Tourists’ nicknames on travel websites</td>
<td>Varchar (100)</td>
<td>Intoxicated Evening Breeze (Qidong)</td>
</tr>
<tr>
<td>DATE</td>
<td>Time of the tour</td>
<td>Varchar (4)</td>
<td>2018</td>
</tr>
<tr>
<td>NODE</td>
<td>Attractions visited</td>
<td>Varchar (100)</td>
<td>San Dengfeng Village; Pagoda Forest; Shaolin Temple</td>
</tr>
</tbody>
</table>

Table 2. Attraction coordinates data of online travel diaries.

<table>
<thead>
<tr>
<th>Field</th>
<th>Explanation</th>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME</td>
<td>Attraction name</td>
<td>Varchar (100)</td>
<td>Shaolin Temple</td>
</tr>
<tr>
<td>LON</td>
<td>Longitude of the attraction</td>
<td>Decimal (18.8)</td>
<td>112.93523381</td>
</tr>
<tr>
<td>LAT</td>
<td>Latitude of the attraction</td>
<td>Decimal (18.8)</td>
<td>34.5082926</td>
</tr>
</tbody>
</table>

Table 3. Attraction visits data of online travel diaries.

<table>
<thead>
<tr>
<th>Field</th>
<th>Explanation</th>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME</td>
<td>Attraction name</td>
<td>Varchar (100)</td>
<td>Shaolin Temple</td>
</tr>
<tr>
<td>2015</td>
<td>The number of tourists who visited the attraction in 2015</td>
<td>Int (4)</td>
<td>63</td>
</tr>
<tr>
<td>2016</td>
<td>The number of tourists who visited the attraction in 2016</td>
<td>Int (4)</td>
<td>88</td>
</tr>
<tr>
<td>2017</td>
<td>The number of tourists who visited the attraction in 2017</td>
<td>Int (4)</td>
<td>91</td>
</tr>
<tr>
<td>2018</td>
<td>The number of tourists who visited the attraction in 2018</td>
<td>Int (4)</td>
<td>98</td>
</tr>
</tbody>
</table>
2.2.2. Spatial Pattern Analysis of Tourist Flow

The symptom tupu module aims to generate multidimensional maps using information extraction models to express the tourist flow. Some models are chosen to analyze tourist behavior from two perspectives: the spatial patterns of tourist flow and the network structure of tourist flow. This paper used the gravity center model and three-dimensional density analysis to depict the phenomenon of the agglomeration and diffusion of tourist flow in tourism space. To analyze the network structure, this paper used social network analysis to reveal the role, function, and interaction of tourist nodes.

The gravity center model [41] is an important tool to study the spatial characteristics of the gravity center of the tourist flow in the process of regional development. The gravity center model of tourist flow takes the attractions as the basic calculation unit, and sets the tourist flow intensities of the attractions as the weights. It calculates the gravity center of different activities in the region by simulating the equilibrium center of traction between tourist nodes with different weights:

\[ X = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \quad Y = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \]  

(1)

where \((X, Y)\) are the coordinates of the gravity center; \(n\) is the total number of attractions; \((x_i, y_i)\) are the geographic coordinates of attraction \(i\), expressed by the latitude and longitude coordinates; and \(w_i\) is the weight of attraction \(i\), expressed by the intensity of the tourist flow, that is, the time-frequency attraction that appears in the online travel diary.

The three-dimensional density analysis can express the evenness of the spatial distribution of tourist flows. It uses the fixed-point symbol method to abstract and symbolize the attractions. Landmark elements of attractions are placed in the three-dimensional virtual geographic environment system based on latitude and longitude. The graphs generated by the symbolization of the footprint density data of the attractions are placed on the landmark elements. The size of the graphical symbols is used to quantitatively describe the differences in tourist density among the attractions.

Social network analysis explores the roles, functions, and connections of attractions based on structural relationships. It mainly involves three elements: nodes, relationships, and connections. [43] Each attraction in the region is equivalent to a point in the social network structure. The mapping relationship between points in the social network structure reflects the connections between attractions. The connection represents the traffic access between attractions.

The evaluation indicators of social network analysis mainly consist of two parts: single node structure and overall network structure [44].

Structural holes and centrality indicators are important tools to measure the role and function of nodes. Using these indicators can quantify the competitive position and core degree of tourist nodes in the tourist flow network.

Structural holes indicate the degree of a node’s advantageous position in the network. The nodes with more structural holes are less affected by surrounding nodes and thus have a strong regional competitive advantage. Generally, the effective size and constraint metrics are used to measure the structural holes.

The higher the effective size value is, the less repetitive the nodes in the network and the greater the possibility of the existence of structural holes. It can be calculated as follows:

\[ ES_i = \sum_{j}^{n} \left( 1 - \sum_{q}^{n} p_{iq} m_{iq} \right), \quad q \neq i, j, \quad \text{where} \quad p_{iq} = \frac{z_{iq} + z_{ij}}{\sum_{j}^{n} (z_{ij} + z_{ji})}, \quad i \neq j, \quad m_{ij} = \frac{z_{ij} + z_{ji}}{\max(z_{jk} + z_{kj})}, \quad j \neq k \]  

(2)

where \(z_{ij}\) is the number of connections from node \(i\) to node \(q\); \(p_{ij}\) is the proportional relationship between node \(i\) and node \(q\), that is, the number of connections between node \(i\) and node \(q\) divided by all the connections of node \(i\); \(m_{ij}\) is the marginal strength between \(i\) and \(q\), which is the number of connections between nodes \(j\) and \(q\) divided by the maximum
number of connections between node \( j \) and other nodes; and \( n \) is the number of nodes in the tourist flow network.

The lower the constraint value is, the less the dependence of the node on other nodes, and the more capable the node owning the structural hole. It can be calculated as follows:

\[
CT_i = \sum_{j=1}^{n} (p_{ij} + \sum_{q \neq j} p_{iq}p_{qj})^2, q \neq i, j
\]

where \( p_{ij} \) is the proportional relationship between node \( i \) and node \( j \); \( p_{iq} \) is the proportional relationship between node \( i \) and node \( q \); \( p_{qj} \) is the proportional relationship between node \( q \) and node \( j \); and \( n \) is the number of nodes in the tourist flow network.

Centrality indicates the degree of a node’s core position in the network. Nodes with higher centrality values have a high level of influence and dominance in the tourist flow network. Generally, the three metrics of degree centrality, closeness centrality, and betweenness centrality are used to measure centrality.

The higher the degree centrality value is, the more connections it has with other nodes, and the more important its position. Degree centrality is divided into in-degree centrality and out-degree centrality, which can be calculated as follows:

\[
C_D^{(\text{in})}(n_i) = \sum_{j=1}^{l} R_{ij^{(\text{in})}}, \quad C_D^{(\text{out})}(n_i) = \sum_{j=1}^{l} R_{ij^{(\text{out})}}
\]

where \( C_D^{(\text{in})}(n_i) \) is the in-degree centrality, \( C_D^{(\text{out})}(n_i) \) is the out-degree centrality, \( l \) is the number of tourist nodes in the network, \( R_{ij^{(\text{in})}} \) indicates that there is a directional connection from node \( j \) to node \( i \), and \( R_{ij^{(\text{out})}} \) indicates that there is a directional connection from node \( i \) to node \( j \).

The higher the closeness centrality value is, the shorter the tourist distance between the node and other nodes, and the better the accessibility of the node. Closeness centrality can be divided into in-closeness centrality and out-closeness centrality, which can be calculated as follows:

\[
C_c(n_i) = \frac{1}{\sum_{j=1}^{l} d(n_i, n_j)}
\]

where \( C_c(n_i) \) is closeness centrality, \( d(n_i, n_j) \) is the shortest path distance between node \( n_i \) and node \( n_j \), and both in-closeness centrality and out-closeness centrality are expressed by this formula.

The higher the betweenness centrality value is, the more obvious the role of the transportation hub, and the stronger the control of the node on other tourist nodes. It can be calculated as follows:

\[
C_B(n_i) = \sum_{j} \sum_{k} \frac{g_{jk}(n_i)}{g_k}, j \neq k \neq i
\]

where \( C_B(n_i) \) is the betweenness centrality of node \( i \), \( g_{jk}(n_i) \) is the number of shortest travel lines from node \( j \) to node \( k \) and through travel node \( i \) in the network, and \( g_k \) is the number of shortest travel lines from node \( j \) to node \( k \).

The core–periphery indicator is an essential way to measure connection among tourist nodes. First, the core-periphery indicator uses the relative density of nodes in the tourist flow network to classify attractions into core and edge areas [45]. Second, it can quantify the cohesiveness within core and edge areas. Third, it can quantify the connections between the core area and edge area. These connections include the driving effect of the core area on the edge area and the driving effect of the edge area on the core area. Higher values indicate stronger interactions. In addition, the core status and edge status are determined
not only by the level of the attraction’s own development but also, more importantly, by the linkage and driving effect between different attractions. It can be calculated as follows [46]:

\[ \rho = \sum_{i,j} a_{ij} \delta_{ij} \]

\[ \delta_{ij} = \begin{cases} 1 & \text{if } c_i = \text{CORE and } c_j = \text{CORE} \\ 0 & \text{if } c_i = \text{PERIPHERY and } c_j = \text{PERIPHERY} \\ \text{otherwise} & \end{cases} \]

where \( \rho \) is essentially an unnormalized Pearson correlation coefficient applied to matrices rather than vectors, \( a_{ij} \) indicates the presence or absence of a tie in the observed data, \( \delta_{ij} \) (subsequently called the pattern matrix) indicates the presence or absence of a tie in the ideal image, \( c_i \) refers to the class (core or periphery) that actor \( i \) is assigned to, and “.” indicates off-diagonal regions of the matrix outside the core and edge blocks. The correlation coefficient \( \rho \) between the observed data and the ideal image is maximized by finding the ideal image with the largest density of core blocks and the smallest density of edge blocks. Then, the core-periphery structure of the network is determined.

2.2.3. Tourism Decision-making Based on Diagnosis Tupu

The diagnosis tupu module is mainly used for internal mechanisms analysis and tourism decision making. In this paper, the diagnosis tupu was generated by combining the results of the gravity center model, three-dimensional density analysis, and social network analysis. First, the location of the gravity center was marked on the diagnosis tupu to depict the phenomenon of the agglomeration of tourist flow in tourism space, which was guided by the gravity center tupu. Second, the popular attractions were marked on the diagnosis tupu to depict the phenomenon of the diffusion of tourist flows, indicated by the three-dimensional density tupu. Third, the attractions and routes were marked with different notes and graphics on the diagnosis tupu, as indicated by the social network tupu. The unbalanced power of the attractions could be visualized to analyze the network structure of the tourist flow. Finally, tourism decisions were made based on the analysis results, and tourism decisions mainly included developing differentiated strategies for cultivating attractions and promoting the tourist flows between attractions.

3. Results

3.1. Establishment of Spatiotemporal Database

The online travel diary data was collected and cleaned, and a total of 404 available data remained, including 1635 visits to 22 attractions. Then, the database of online travel diaries, the database of attraction coordinates, and the database of attraction visits were organized. The database of online travel diaries is shown in Table 4.

<table>
<thead>
<tr>
<th>ID</th>
<th>DATE</th>
<th>NODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intoxicated evening breeze (Qidong)</td>
<td>2018</td>
<td>San Dengfeng Village, Pagoda Forest, Shaolin Temple</td>
</tr>
<tr>
<td>Listen to the spring breeze (Zhejiang)</td>
<td>2018</td>
<td>San Dengfeng Village, Pagoda Forest, Shaolin Temple; Songyang Academy, Junji Peak, Fawang Temple, Songyang Academy</td>
</tr>
<tr>
<td>Tian Xingjian (Zhengzhou)</td>
<td>2018</td>
<td>Fawang Temple, Junji Peak</td>
</tr>
</tbody>
</table>

3.2. Tourist Flow Analysis

3.2.1. Spatial Patterns of Tourist Flow

(1) Spatial agglomeration pattern analysis of tourist flow: gravity center model

The spatial agglomeration phenomenon of the tourist flow was visualized based on the gravity center model, as shown in Figure 3. The gravity center of the tourist flow in
Dengfeng remained relatively stable from 2015 to 2018. It was concentrated in an area of 0.25 miles in the northern area of Dengfeng. Over the years, the gravity center of the tourist flow shifted from the northeast to the southwest in 2015–2016, shifted back to the northeast in 2016–2017, and then shifted back to the northwest from 2017 to 2018. However, the displacement was small, with a cumulative offset distance of only approximately 0.48 miles.

Figure 3. Evolution of the gravity center of tourist flows in Dengfeng from 2015 to 2018: (a) location of the gravity center; (b) evolution of the gravity center.

(2) Spatial diffusion pattern analysis of tourist flow: three-dimensional density analysis

The spatial diffusion phenomenon of the tourist flow was visually analyzed by using three-dimensional density, as shown in Figure 4. From 2015 to 2018, the number of tourists to Dengfeng increased year by year. Shaolin Temple, Pagoda Forest, San Dengfeng Village, Songyang Academy, and Junji Peak were the most popular visiting locations. They were all located in the northern region of Dengfeng. This result indicated that tourists were more likely to visit attractions with strong cultural heritage and regional characteristics, which led to a significant difference in the spatial distribution of the tourist flows.
Figure 4. Frequency of visits to attractions in Dengfeng from 2015 to 2018: (a) distribution of digital footprint in 2015; (b) distribution of digital footprint in 2016; (c) distribution of digital footprint in 2017; (d) distribution of digital footprint in 2018.
3.2.2. Network Structures of Tourist Flow

Based on the perspective of structural relationships, social network analysis was used to reveal the roles, functions, and interactions of tourist nodes, as shown in Figures 5 and 6. The size of the nodes in Figures 5 and 6 represents the level of the nodes, and the thickness of the connections between the nodes indicates the volumes of the tourist flows.

Figure 5. Single node structure of tourist flows in Dengfeng: (a) effective size; (b) constraint; (c) in-degree centrality; (d) out-degree centrality; (e) in-closeness centrality; (f) out-closeness centrality; (g) betweenness centrality.
Figure 6. The overall network structure of tourist flow in Dengfeng based on the core-periphery indicator.

(1) Single node structure analysis: structural holes and centrality indicators

The effective size and constraint were used to measure the structural holes, and the results are shown in Figure 5a,b. Tourist nodes with higher effective size values and lower constraint values have a strong regional competitive advantage. The effective size values of Songyang Academy, Zhongyue Temple, and Shaolin Temple were relatively higher, and their constraint values were lower. These three attractions were less influenced by the tourist flows of their surrounding attractions and had obvious competitive advantages. Songshan Ski Resort and Zhougong Observatory had relatively lower effective size values and higher constraint values. It revealed that both were more dependent on the development of tourist flows from surrounding attractions. Thus, these two attractions were at a disadvantage in the competition. However, the effective scale values and constraint values of The First Patriarch Temple, Zen Shaolin Music Ceremony, and Huishan Temple were relatively higher. This showed that although they had a competitive advantage in terms of tourist flow, they still heavily depended on the surrounding attractions. The possible reason may be that both of the two attractions are famous in Dengfeng but are located in the “hot spot” of tourism. They were surrounded by Shaolin Temple and Songyang Academy, which have a deeper cultural heritage, splitting the tourist flow.

The in-degree centrality and out-degree centrality were used to measure the degree of centrality, and the results are shown in Figure 5c,d. Tourist nodes with higher in-degree centrality values and out-degree centrality values have more connections with other nodes, and thus have a leading role in the tourist flow network. The results showed that Shaolin Temple and Songyang Academy had relatively higher in-degree centrality values and out-degree centrality values, indicating that they were highly connected to other nodes. It revealed that they had core competitiveness and dominant roles in the regional tourism system.

The in-closeness centrality and out-closeness centrality were used to measure the closeness centrality, and the results are shown in Figure 5e,f. Tourist nodes with higher in-closeness centrality values and out-closeness centrality values have more proximity to other nodes, and thus have higher accessibility in the process of tourist flow transfer. The results showed that Shaolin Temple and Songyang Academy had relatively higher in-closeness centrality values and out-closeness centrality values, indicating that they were highly connected to other nodes. It revealed that they had core competitiveness and dominant roles in the regional tourism system.

Betweenness centrality quantified the number of nodes acting as transfer stations in the tourist flow network, and the results are shown in Figure 5g. Tourist nodes with higher betweenness centrality values have higher intermediary capacity in the process of tourist flow transfer. The results showed that Zhongyue Temple, Songyang Academy, and Shaolin Temple had relatively higher betweenness centrality values. This revealed
that these three attractions played a transit role in the tourist flow network, assuming the function of tourism channels.

(2) Overall network structure analysis: core–periphery indicator

The core-periphery indicator classified attractions into core and edge areas and quantified the degree of interaction between areas, as shown in Figure 6. The results showed that the attractions in the core area were distributed in a “plate” and “point axis” pattern, and the core tourist flows formed an unclosed “streamline” and closed “triangle” pattern with multiple nodes in the series. The density within the core area was higher than that within the edge area (0.625 vs. 0.077). This result indicated that tourists were concentrated in the core area and rarely flowed to the edge area. Moreover, although there were connections within the edge area, the degree of interaction was relatively low. In addition, the linkage density of the core area to the edge area was higher than that of the edge area to the core area (0.259 vs. 0.223). The core area exhibited strong internal connectivity and was less driven to the edge area, but still greater than the edge area to the core area. Overall, the tourist flow network presented a structure of significant “core–periphery”.

3.3. Tourism Decision Making

The diagnosis tupu was generated by combing the multi-symptom tupu (gravity center model, three-dimensional density analysis, and social network analysis), as displayed in Figure 7.

The high concentration of tourism in the northern region (Figure 7) is a constraint to the development of tourism in the central and southern regions. Therefore, it is important for the future construction of Dengfeng as a tourism city to consider how to correctly deal with the differences in tourism resources in the northern and southern regions and solve the problem of unbalanced tourism development in the northern and southern attractions. From this perspective, the following recommendations are provided.

Firstly, according to the trajectory of the gravity center of the tourist flow (Figures 3 and 7), the opening of the Daxiongshan Xianren Valley helped the development of tourism in the southern regions, at least to some degree, but the area still lacked strong attractiveness for tourists. Therefore, Dengfeng City should develop special tourism projects in the Daxiongshan Xianren Valley and develop differentiated marketing strategies to balance the tourism development of the northern and southern regions.

Secondly, according to the results of structural holes (Figure 5b), Dengfeng Astronomical Observatory, located in the central region, was less vulnerable to negative impacts from nearby attractions due to its low constraint value. Therefore, simulating the tourism potential of Dengfeng Astronomical Observatory may offer a potential solution to balance the tourism development of the northern and central regions.
Thirdly, according to the results of centrality (Figure 5c,d), Shaolin Temple and Songyang Academy had a high degree of centrality values and were the core attractions in the tourist flow network. Therefore, they occupied a dominant position in Dengfeng tourism. In the future, if tourist flow connections between these core attractions and attractions in the south–central region are strengthened, such as through combined marketing, it will be possible to promote the tourism development of the northern and southern regions.

4. Discussion

In the existing literature, most tourist behavior studies mainly focus on inbound tourist flows or tourist flows of famous and popular cities [47], ignoring those of small and not well-known cities that have tourism development potential. This is likely a weakness for improving the overall tourism competitiveness of the country. Thus, a focus on the tourist behavior of small and not well-known cities with tourism development potential could be necessary. However, there is often a lower popularity of the attractions and smaller tourist numbers in this type of city. This also led to a smaller amount of data collected from online travel diaries.

The representativeness of sample is also an important issue. The representativeness of sample refers to the degree to which a sample can represent the underlying population. Assessing sample representativeness takes two general approaches [48]. The first approach is examining the sample selection process to see whether the sample is obtained through probabilistic sampling procedures. However, online travel diary data are one type of volunteered geographic information (VGI), which shares the commonality of voluntary and non-expert geographic information creation. Thus, this approach is not suitable to evaluate the representativeness of online travel diary data. The second approach in assessing sample representativeness is to compare the sample with the population on comparison variables, rather than on target variables. Comparison variables are those variables believed to be related to the target variables in a certain way, and they should be obtainable for both the sample and the population. Yang et al. [49] evaluated the representativeness of the AmeriFlux network of eddy covariance towers to represent the environments contained within the coterminous United States by comparing environmental similarity between eco regions. Similarly, we compared the correlation between the distributions of tourists among the attractions obtained from different travel websites, and the results are shown in Table 5. The correlations between the distributions of tourists obtained from different travel websites are significant. It could imply that the sample data are valid for analyzing the spatial distribution of the actual number of visitors.

Table 5. Pearson correlation coefficient of the distributions of tourists among the attractions between different travel websites.

<table>
<thead>
<tr>
<th></th>
<th>Mafengwo.com</th>
<th>Qunar.com</th>
<th>Ctrip.com</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mafengwo.com</td>
<td></td>
<td>0.939 **</td>
<td>0.927 **</td>
</tr>
<tr>
<td>Qunar.com</td>
<td>0.939 **</td>
<td></td>
<td>0.973 **</td>
</tr>
<tr>
<td>Ctrip.com</td>
<td>0.927 **</td>
<td>0.973 **</td>
<td></td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level.

In addition, we collected the true number of tourists from the government website of the Dengfeng Tourism Bureau, as shown in Table 6. The increasing trend of the number of online travel diaries is consistent with the true number of tourists to Dengfeng City. It may imply that the sample data are valid for analyzing the changes in the spatial distribution of tourists from 2015 to 2018.
Table 6. Comparison of the true number of tourists and the number of valid online travel diaries.

<table>
<thead>
<tr>
<th>Year</th>
<th>True Number of Tourists (Thousand)</th>
<th>Number of Online Travel Diaries</th>
<th>Number of Valid Online Travel Diaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>11,040</td>
<td>177</td>
<td>62</td>
</tr>
<tr>
<td>2016</td>
<td>12,000</td>
<td>248</td>
<td>98</td>
</tr>
<tr>
<td>2017</td>
<td>13,530</td>
<td>299</td>
<td>111</td>
</tr>
<tr>
<td>2018</td>
<td>15,307</td>
<td>321</td>
<td>133</td>
</tr>
</tbody>
</table>

The existing studies of tourism in small and not well-known cities mainly used qualitative analysis. For example, Li [50] outlined the current situation of tourism development in Dengfeng and analyzed the existing problems. Pulido-Fernandez et al. [51] characterized the olive oil tourism typology and identified its main activities in the Mediterranean basin based on a thorough bibliographical review and an expert panel. Their studies analyzed tourism from different perspectives and had certain practical significance. However, they are subjective due to the lack of quantitative analysis. Therefore, it is necessary to quantitatively analyze the current situation of tourism. There are many ways to quantitatively study tourist behavior. This paper provides a complete framework for the study of tourist behavior analysis, combining geo-information Tupu theory and tourist flow analysis. Note that this paper tends to provide an analysis framework, not the specific model. The framework shows the ability to (1) provide a new research perspective for tourist behavior analysis, (2) generate a more comprehensive understanding of tourist flow patterns, and (3) take different/new input data and analysis models for its future application in other locations. It contributes sufficiently to enrich the body of knowledge on tourist behavior analysis from the geo-information Tupu theory perspective.

The change on the gravity center was small (Figure 3), with a cumulative offset distance of only approximately 0.48 miles. However, it was sufficient to suppose changes in tourist behaviors. The gravity center of the tourist flow shifted from northeast to southwest in 2015–2016. The exposure of the Yongxin Shi scandal in 2015 reduced the tourist enthusiasm for Shaolin Temple, which led to the shift of the gravity center of the tourist flow to the southwest from 2015 to 2016. In 2016–2017, the gravity center of the tourist flow shifted back to the northeast, which may have been due to the opening of the Daxiongshan Scenic Area. Although Songshan Scenic Area (including Shaolin Scenic Area, Songyang Scenic Area, and Zhongyue Scenic Area) is a historical and cultural center, some tourists flowed into Daxiongshan Scenic Area because of the development of new tourist products. This led to a shift in the gravity center of tourist flow. However, in the early stages of the development of the Daxiongshan Scenic Area, problems appeared in tourism management and service provision owing to the rapid growth in the number of tourists. In addition, the attractions of Daxiongshan Scenic Area were mainly related to natural scenery and entertainment facilities. Compared with the Songshan Scenic Area, which combined culture and nature, the Daxiongshan Scenic Area had difficulty maintaining its attractiveness to tourists for a long time. As a result, the gravity center of the tourist flow began to shift to the northwest from 2017 to 2018.

According to the results of three-dimensional density analysis (Figure 4), Shaolin Temple, Pagoda Forest, San Dengfeng Village, Songyang Academy, and Junji Peak were the most popular attractions, and Zhongyue Temple ranked eighth. However, the results of social network analysis showed that Songyang Academy, Shaolin Temple, and Zhongyue Temple had strong competitive advantages. These results seemed to be inconsistent. The possible reason may be that the advantageous degree of an attraction depends not only on the level of development of the attraction itself but also, more importantly, on the relevance and driving effect of the attraction on other attractions. Songyang Academy, Shaolin Temple, and Zhongyue Temple had more interactions of tourist flow with surrounding attractions compared with other attractions. Although these attractions had low tourist flow intensity, they were closely connected with surrounding nodes and less influenced by
other attractions. Therefore, they were nodes with strong competitive advantages in the tourism network.

According to the results of social network analysis (Figure 7), core paths tended to be relatively short compared with important paths and common paths, which meant that tourist flows were affected by distance decay. This could lead to a high concentration of tourists in core areas and a constraint for developing tourism in edge areas. Therefore, the traffic guidance around the core area needs attention. The overall vitality of Dengfeng’s tourism market will be promoted through measures such as optimizing public transportation and shortening tourist routes. In addition, there were fewer strong competitive attractions in the northern region, resulting in a lack of alternative attractions and paths and creating bottlenecks in tourist flows. Therefore, Dengfeng City should actively develop special tourism projects to cultivate advantageous attractions with strong competitiveness, such as Pagoda Forest, San Dengfeng Village, Fawang Temple, and Pagoda at Songyue Temple. While maintaining the competitive advantage of Zhongyue Temple, attention should be given to transforming it into a core attraction. Then, Zhongyue Temple can cluster and radiate other attractions, alleviating internal unhealthy competition.

It is obvious that social media users could not reflect the actual number of tourists, thus online travel diaries could not accurately reflect the complete travel routes of tourists [12]. Furthermore, young and educated travelers are more likely to use these online travel websites [41]. From these perspectives, the digital footprint data captured based on online travel diaries may affect the accuracy of tourist behavior analysis results. In the future, attempts to combine online travel data with official survey data, because the latter is based on a stratified random sample of the total population [52], could significantly improve the precision of the data.

5. Conclusions

This paper proposed a novel research framework for analyzing tourist behavior, inspired by geo-information Tupu theory. Unlike traditional tourist behavior analysis, this framework attempts to systematically elaborate methods of data acquisition, tourist flow analysis, internal mechanism analysis, and tourism decision making, thus providing a more complete tourist behavior analysis process and developing a new research perspective for tourist behavior analysis. To verify the validity of the framework, this paper used Dengfeng City, China, as a case study for tourist behavior analysis. First, we chose online travel diaries from 2015–2018 as the data source and cleaned the data to realize the dataset construction. Then, the traditional quantitative methods of spatial analysis (gravity center model and three-dimensional density analysis) and social network analysis were combined to analyze the spatial pattern and network structure of tourist flows. Finally, we analyzed the internal mechanism of tourist flow and made some tourism decisions. The conclusions can be summarized as follows:

Firstly, the results of the gravity center model showed that Daxiongshan Xianren Valley, as an emerging tourist node, had difficulty maintaining its attractiveness to tourists compared with other tourist nodes. It is necessary to improve its reputation and enhance its publicity. There are two alternative solutions: one is to use joint promotion with popular attractions, another is to optimize public transportation and shorten tourism routes between the core node and Daxiongshan Xianren Valley.

Secondly, according to the results of three-dimensional density analysis, tourists preferred to visit attractions with strong cultural heritage and regional characteristics. Therefore, the joint promotion of the same type of attractions is conducive to achieving collective development. Some thematic tourism routes, such as “Shaolin Temple, Songyang Academy and Zhongyue Temple” can be created to form religious tourism (Buddhism, Confucianism, and Taoism).

Thirdly, as indicated by the results of social network analysis, Shaolin Temple was the core node of the tourist flow network in Dengfeng City. Furthermore, it had an important tourism brand effect in China. Dengfeng City can promote the development of the whole
tourism industry by taking advantage of the natural tourism and humane tourism of Shaolin Temple.

**Author Contributions:** Conceptualization, Zhihui Tian; methodology, Yongji Wang and Lili Wu; validation, Yi Liu and Yongji Wang; formal analysis, Yi Liu and Yongji Wang; investigation, Yi Liu; data curation, Yi Liu; writing—original draft preparation, Yi Liu, Yongji Wang and Lili Wu; writing—review and editing, Yi Liu, Yongji Wang and Lili Wu; visualization, Yi Liu and Yongji Wang; supervision, Zhihui Tian; project administration, Zhihui Tian; funding acquisition, Yongji Wang. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the key scientific research projects of colleges and universities in Henan Province, grant number 22A170021.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available from the author upon reasonable request.

**Acknowledgments:** The authors would like to acknowledge all of the reviewers and editors.

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

**References**


