



International Journal of
Geo-Information

Geo Data Science for Tourism

Edited by

Andrea Marchetti and Angelica Lo Duca

Printed Edition of the Special Issue Published in *IJGI*

Geo Data Science for Tourism

Geo Data Science for Tourism

Editors

Andrea Marchetti

Angelica Lo Duca

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin



Editors

Andrea Marchetti
Institute of Informatics and
Telematics—National
Research Council (IIT-CNR)
Italy

Angelica Lo Duca
Institute of Informatics and
Telematics—National
Research Council (IIT-CNR)
Italy

Editorial Office

MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *ISPRS International Journal of Geo-Information* (ISSN 2220-9964) (available at: https://www.mdpi.com/journal/ijgi/special.issues/GIS_tourism).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. <i>Journal Name</i> Year , <i>Volume Number</i> , Page Range.
--

ISBN 978-3-0365-5029-9 (Hbk)

ISBN 978-3-0365-5030-5 (PDF)

© 2022 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.

The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

Contents

About the Editors	vii
Preface to “Geo Data Science for Tourism”	ix
Lingjin Wang, Xiao Wu and Yan He Nanjing’s Intracity Tourism Flow Network Using Cellular Signaling Data: A Comparative Analysis of Residents and Non-Local Tourists Reprinted from: <i>ISPRS Int. J. Geo-Inf.</i> 2021 , <i>10</i> , 674, doi:10.3390/ijgi10100674	1
Xu Zhang, Chao Song, Chengwu Wang, Yili Yang, Zhoupeng Ren, Mingyu Xie, Zhangying Tang and Hongtu Tang Socioeconomic and Environmental Impacts on Regional Tourism across Chinese Cities: A Spatiotemporal Heterogeneous Perspective Reprinted from: <i>ISPRS Int. J. Geo-Inf.</i> 2021 , <i>10</i> , 410, doi:10.3390/ijgi10060410	27
Yifan Zuo, Huan Chen, Jincheng Pan, Yuqi Si, Rob Law and Mu Zhang Spatial Distribution Pattern and Influencing Factors of Sports Tourism Resources in China Reprinted from: <i>ISPRS Int. J. Geo-Inf.</i> 2021 , <i>10</i> , 428, doi:10.3390/ijgi10070428	45
Peiyuan Qiu, Jialiang Gao and Feng Lu Identifying the Relatedness between Tourism Attractions from Online Reviews with Heterogeneous Information Network Embedding Reprinted from: <i>ISPRS Int. J. Geo-Inf.</i> 2021 , <i>10</i> , 797, doi:10.3390/ijgi10120797	67
Xiao Zhou, Jiangpeng Tian and Mingzhan Su Tour-Route-Recommendation Algorithm Based on the Improved AGNES Spatial Clustering and Space-Time Deduction Model Reprinted from: <i>ISPRS Int. J. Geo-Inf.</i> 2022 , <i>11</i> , 118, doi:10.3390/ijgi11020118	87
Yuanhong Qiu, Jian Yin, Ting Zhang, Yiming Du and Bin Zhang Spatiotemporal Dynamic Analysis of A-Level Scenic Spots in Guizhou Province, China Reprinted from: <i>ISPRS Int. J. Geo-Inf.</i> 2021 , <i>10</i> , 568, doi:10.3390/ijgi10080568	117
Chengkun Huang, Feiyang Lin, Deping Chu, Lanlan Wang, Jiawei Liao and Junqian Wu Spatiotemporal Evolution and Trend Prediction of Tourism Economic Vulnerability in China’s Major Tourist Cities Reprinted from: <i>ISPRS Int. J. Geo-Inf.</i> 2021 , <i>10</i> , 644, doi:10.3390/ijgi10100644	139
Yen-Cheng Chen, Ching-Sung Lee, Ya-Chuan Hsu and Yin-Jui Chen Why Is Green Hotel Certification Unpopular in Taiwan? An Analytic Hierarchy Process (AHP) Approach Reprinted from: <i>ISPRS Int. J. Geo-Inf.</i> 2021 , <i>10</i> , 255, doi:10.3390/ijgi10040255	161

About the Editors

Andrea Marchetti

Andrea Marchetti (Senior Technologist) works at the Institute of Informatics and Telematics of the National Research Council, Italy. His main research activities are data science, data quality, and web application. He applies them to various fields such as cultural heritage, tourism, health, and other topics of national interest. He has participated in several European and national projects including OpeNer, Caper, and GeoMemories. He is currently in charge for the IIT of the interreg “Pitem Pace Far Conoscere” project and the “Osiris-FO” project. Since 2012, he has been teaching courses on Web Design and Data Journalism at the University of Pisa.

Angelica Lo Duca

Angelica Lo Duca (Researcher) works at the Institute of Informatics and Telematics of the National Research Council, Italy. She is also an external professor of Data Journalism at the University of Pisa. Her research interests include Data Science, Data Journalism, and Web Applications. She used to work on Network Security, Semantic Web, Linked Data, and Blockchain. She has published more than 40 scientific papers at national and international conferences and journals. She has participated in different national and international projects and events. She has been a member of the Program Committee at different conferences. She is also part of the Editorial Team of the HighTech and Innovation Journal.

Preface to “Geo Data Science for Tourism”

Tourism is one of the largest and most important industries in the world. It directly employs millions of people and generates billions of dollars in revenue each year. Given its importance, it is not surprising that data science is increasingly being used to understand and optimise the tourism industry. Geodata science, in particular, is playing a key role in this effort. By analysing large data sets, geodata scientists are able to identify patterns and trends that can be used to improve the efficiency and effectiveness of the tourism sector.

Geodata Science for Tourism aims at investigating the recent challenges in tourism seen from the point of view of data science. There are many challenges that tourism businesses face when it comes to data science. One of the biggest is understanding how to use data to drive decision making. With so much data available, it can be difficult to identify which features are most relevant and how to use them to improve operations. Additionally, data science can be used to better understand customer behaviour and preferences, which can help businesses tailor their offerings and better meet customer needs. However, collecting and analysing customer data can be costly and time-consuming, making it a challenge for smaller businesses in particular. Another key challenge is staying ahead of the curve in terms of technology and analytics. As the tourism industry evolves, so too do the tools and techniques that data scientists use to understand it. Businesses need to invest in keeping their data science teams up to date with the latest developments in order to stay competitive.

By understanding customer behaviour and preferences, businesses can make more informed decisions about where to open new locations, what type of amenities to offer, and how to price their services. Additionally, data can help businesses track and understand trends in the industry, such as shifts in customer demand or changes in competitors’ offerings.

Operationally, data can be used to optimise everything from staffing levels to inventory management. By understanding which times of day are busiest or which services are most popular, businesses can staff accordingly and ensure they have the necessary supplies on hand.

In short, data is essential for businesses in the hospitality industry to succeed. Geodata science can help these businesses make better use of data to improve their decision making and operations.

There are many different types of data that can be used to study tourism. This includes data on tourist destinations, travel patterns, and spending. Geodata science is a relatively new field that uses geographical data to study tourism and its impact on the environment. Geospatial data is data that captures the location and shape of an object on the earth’s surface. This type of data can be used to track the movement of people and objects, as well as to identify patterns and trends.

One of the benefits of using geodata science for tourism research is that it can help to identify trends and patterns in tourist behaviour, as you will see in the first four articles of the book. Another benefit of geodata science is that it can help to assess the impact of tourism on the environment. This includes looking at things such as water consumption, energy use, and carbon emissions. By understanding the environmental impact of tourism, we can make more sustainable choices about how we travel. Geodata science for tourism can also be used to predict future trends, as you will read in the last two chapters of the book. This information can be used by policymakers to make decisions about where to allocate resources.

We hope that this book will help you to extend your awareness of the benefits of using geodata science in the tourism industry.

Andrea Marchetti and Angelica Lo Duca
Editors

Article

Nanjing's Intracity Tourism Flow Network Using Cellular Signaling Data: A Comparative Analysis of Residents and Non-Local Tourists

Lingjin Wang, Xiao Wu * and Yan He

School of Architecture, Southeast University, Nanjing 210096, China; lingjin_wang@seu.edu.cn (L.W.); 230189014@seu.edu.cn (Y.H.)

* Correspondence: 101010124@seu.edu.cn; Tel.: +86-138-5178-0536

Abstract: With the rapid development of transportation and modern communication technology, “tourism flow” plays an important role in shaping tourism’s spatial structure. In order to explore the impact of an urban tourism flow network on tourism’s spatial structure, this study summarizes the structural characteristics of the tourism flow networks of 43 scenic spots in Nanjing from three aspects—tourism flow network connection, node centrality, and communities—using cellular signaling data and the social network analysis method. A comparative analysis revealed the tourism flow network structures of residents and non-local tourists. Our findings indicated four points. Firstly, the overall network connectivity was relatively good. Core city nodes displayed high spatial concentration and connection strength. However, suburban nodes delivered poor performance. Secondly, popular nodes were intimately connected, although there were no “bridging” nodes. Lesser-known nodes were marginalized, resulting in severe node polarization. Thirdly, regarding the network community structure, the spatial boundary between communities was relatively clear; the communities within the core city were more closely connected, with some parts encompassing suburban nodes. Most suburban communities were attached to the communities in the core area, with individual nodes existing independently. Fourthly, the primary difference in the tourism flow network structures between residents and non-local tourists was that the nodes for residents manifested a more balanced connection strength and node centrality. Core communities encompassed more nodes with more extensive coverage. Conversely, the nodes for non-local tourists showed wide discrepancies in connection strength and node centrality. Furthermore, core communities were small in scale with clear boundaries.

Keywords: tourism flow; cellular signaling data; social network analysis; network connection; node centrality; communities

Citation: Wang, L.; Wu, X.; He, Y. Nanjing's Intracity Tourism Flow Network Using Cellular Signaling Data: A Comparative Analysis of Residents and Non-Local Tourists. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 674. <https://doi.org/10.3390/ijgi10100674>

Academic Editors: Andrea Marchetti, Angelica Lo Duca and Wolfgang Kainz

Received: 20 July 2021

Accepted: 1 October 2021

Published: 4 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since the 1960s, due to the continued developments in modern science and technology (including computer and network information technology, advanced transportation, modern communications, globalization, and informatization), global networking has become a significant development tendency. Against this background, by integrating the Marxist theory of globalization, information theory, and postmodern space theory, sociologist Manuel Castells proposed a novel social research theory—the space of flows theory—revealing a new perspective on the organizational logic of the modern social system [1]. Under its influence, “space of flows” has become a research hotspot in the geography domain, and has inspired many research topics, including information flow, traffic flow, knowledge flow, culture flow, and technology flow.

Tourism space is the projection of tourist activities in space. Regarding tourism destinations, tourists display the characteristic of mobility [2]. Therefore, the concept of tourism flow appeared when the spatial structures of tourism, at urban and regional levels,

were studied from the perspective of tourist behavior [3]. It has become an important research topic in western tourism geography, which unfolds mainly from the spatial models of tourism flow, its causes and impact mechanisms, and tourists' spatiotemporal behavior. The study of the spatial models of tourism flow started in 1977 when Hill et al. proposed the "core-edge" model of tourism flow [4]. Most of the literature that followed also began from the spatial analysis angle, and successively proposed such laws or models as distance decay, the gravity model, and spatial dimension [5]. The research on the causes and impact mechanisms of tourism flow began in the 1980s, when knowledge from related subjects (including mathematics, geography, and economics) was introduced to explore the causal mechanism of the economic impact of tourism flow on destinations, and the law of its endogenous occurrence and occurrence pattern simulations. Since the beginning of the 21st century, such research has gradually matured into a system. There is quantitative research focusing on the related political factors and economic constraints [6], and the impacts of various supply-side resources [7]. The research on tourists' spatiotemporal behavior only began in the late 1990s. It focused on the prediction of the direction and quantity of tourism flow, and tourism consumption [8,9]. By comparison, tourism flow-related research in China started late, commencing in the 1980s. It focused mainly on spatiotemporal distribution and the law of evolution. The relevant explorations centered around the laws of the spatiotemporal evolution of tourism flow [10], and its node transfer mechanism on a countrywide, provincewide, and citywide scale [11,12], summarized in the spatial structural model of tourism flow [13,14]. In a preliminary exploration of the driving-force mechanisms of tourism flow, researchers usually relied on annual and seasonal change indices. They based their studies on statistical indicators, such as in-degree and out-degree nodes in social networks [15] and the skewness index [16]. They performed their research using time-series and cluster analyses, while simultaneously drawing on the push-pull theory, origin-destination (OD) distribution theory, and driving-force theory [17,18]. Whether it is information, traffic, or tourism flow, in this study, they are all significant channels through which we may glimpse the structure of urban social and spatial networks.

Tourism flow shaped the structure of the tourism network, which is of great significance to the spatial study of tourism. Tourism networks are fundamentally social networks constituted by tourist behavior. In recent years, scholars have brought the social network analysis from the sociological domain into the analysis of tourism flow network structures [19,20]. A social network is a structure made of mutually connected behavioral objects, which is considered to be a structure that is constituted by social relations. It is used extensively in studying social media networks, information communication networks, friend circles, business networks, kinship networks, and disease transmission networks [21]. Social network analysis is a research method that is based on social network theory and is applied to the social interaction between individuals in complex relationships [22]. This method has established rich model parameters for studying node objects, edge objects, and the constitutive features of networks themselves (including centrality analysis, community analysis, network correlation, and core-edge structure). These areas became the research directions for multiple-domain applications in sociology, mathematics, and computer and communication science. With the rapid development of information technology, the large-scale statistics and a vast amount of data regarding tourists' spatiotemporal behavior are readily available. This data enables the study of tourism flow network structures, based on the social network analysis method, to move further toward refinement and quantification. Currently, the related big data studies concentrate on national and regional scales [23]. Scholars study tourists' spatial distribution and construct tourism flow networks using big data [24] (such as the GPS location information of pictures [25], independent itineraries [26], network travel notes, and tour routes in online bookings [27]), and further investigate the influencing evolutionary factors or network formation. Some scholars use location-based social networks to understand human mobility and people's behavior by mining check-in patterns, studying the influence of hidden structural patterns in social network nodes, and

the changes in external environment, on user check-in patterns [28]. In addition, due to the current global spread of COVID-19, many scholars have combined the geo-tagged data of social media with epidemic prevention needs to establish a reference model to predict the infection risk of social interaction and travel between residents and tourists [29,30].

Considering the lack of existing research, this study highlights three aspects of innovation. Firstly, it highlights the combination of the application of big data and research on tourism flows within cities at the medium and micro scales. Due to the limited precision of big data, the demand for washing and screening users' phone signaling data within a city is higher; therefore, the existing research is mainly from a regional perspective, focusing on the tourist flow between cities and the relationship between the cities' tourism resources [31]. This study carried out data processing on a variety of filter condition experiments, and the processing results, multiple times, meaning the relatively accurate basic data could be used to analyze the spatial structure of tourism flow within the city. Secondly, it highlights the combination of the social network analysis method and traditional space theory. Traditional space theory emphasizes static material space expression; however, this study used the social network analysis method to connect human activities with spatial structures, which can reflect the functional connection in urban space more accurately. Thirdly, it analyses the differences between local residents and tourists in a tourism network structure. The differences in tourism flow network structures, caused by different sources of tourists (local and non-local, age structure differences, consumption level differences, etc.) have been ignored in the existing research. This study attempts to conduct a comparative analysis of the tourism flow network that is formed by local residents and tourists, and to excavate the differences between them.

Therefore, in order to further explore the impact of an urban tourism flow network on tourism's spatial structure, this study uses the social network analysis method and cellular signaling data. From the perspective of tourists' spatiotemporal behavior, and using a summary of the overall characteristics of the tourism flow network structure of Nanjing city proper, this paper differentiates between resident and non-local tourism flow network structures. In addition to addressing the deficiency in the existing research, regarding the application of big data and the absence of research scale, it provides a scientific basis for the differentiated organization of tourism space and tourism routes, urban infrastructure, transportation planning, and tourism social management.

2. Materials and Methods

2.1. Research Districts

Nanjing city has a long history and culture, and unique natural landscapes, making it a major scenic tourist destination at the national level. In 2015, the number of domestic tourists reached 99.9266 million, ranking it among the top tourist cities in the country. According to the 2018 monthly "Report on the Platform for Operation Monitoring of Smart Tourism Big Data in Nanjing", issued by the Nanjing Municipal Bureau of Culture and Tourism, the number of tourists visiting scenic spots with grades of 2A or above reached 91.8% (referring to China's "Classification and Evaluation of Tourist Scenic Spot Quality Grade", the scenic spot classification includes five grades; 5A is the highest grade for scenic spots). This study took Nanjing city proper as its spatial scope, selecting all three A-grade or above scenic spots in the city as the specific research districts, including 43 popular scenic spots (see Figure 1), for tabulating the statistics of tourist spatial behaviors.

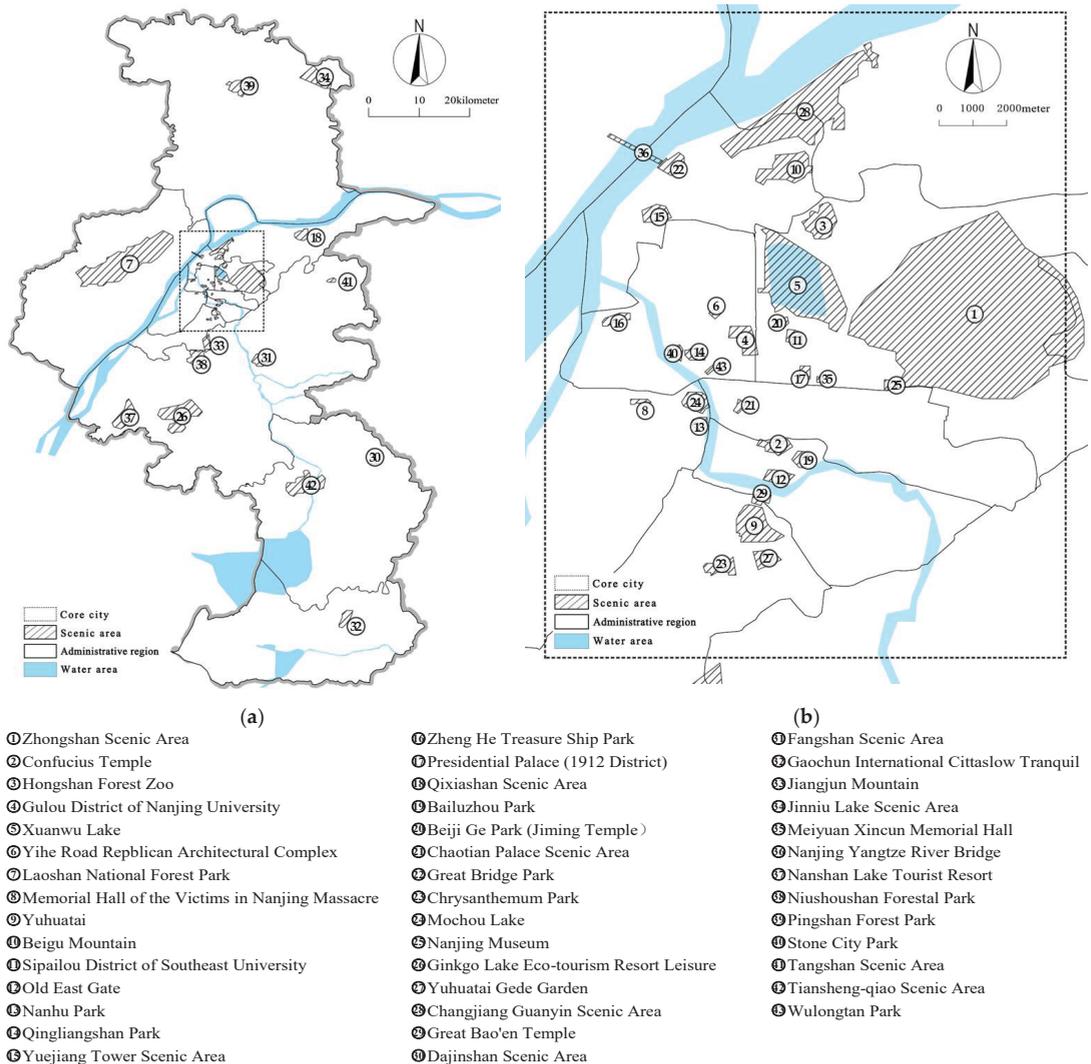


Figure 1. Research scope and selected scenic spots: (a) entire city of Nanjing; (b) core city of Nanjing.

2.2. Data Sources

This study performed two statistical analyses on different temporal and spatial scales, using cellular signaling data, to examine the phenomenon of tourism flow resulting from the spatial displacement of tourist crowds. This method overcame the predicament of the insufficient temporal and spatial accuracy of tourist behavior, encountered in previous research adopting the traditional analytical method of regional perspective. The data in this study came mainly from the cellular signaling data (user ID, number attribution, geographical position of triggering base station, and triggering moment) provided by Nanjing Mobile. The concrete selection steps are outlined as follows. First, data collection appearing in Nanjing city proper on four statutory holidays in 2015, viz., November 14, 15, 21, and 27. Second, tourist monitoring at related scenic spots in Nanjing city, referencing average length of visit for each scenic spot in the city. Data were selected from at least

two scenic spots within the research scope (see the above 44 scenic spots) that tourists visited for over an hour. These two scenic spots were treated as the spatial departure and destination points. Third, on this basis, per number attribution, the data of local and non-local users were identified. Based on the length of evening visits (staying for more than 3 consecutive hours between 24:00 and 07:00) on working days, versus day visits during working hours (staying for more than 4 consecutive hours between 07:00 and 19:00), the residences and workplaces of local users were identified. Based on the time spent on handling business in the daytime (staying for more than 2 consecutive hours between 08:00 and 19:00), the destinations of non-local users on business travel were identified, which were later excluded. In the end, from the 2.365 million active users (who appeared at least twice on the four statutory holidays), the data of 1.212 million local users and 108,000 non-local users were identified and used as the foundational data for this comparative study on tourism flow structures of local and non-local users.

Of the sampled users, Nanjing Mobile's market share was 64.7%, meaning the above identification results were tantamount to a large sample. Based on the permanent resident population of approximately 8 million in Nanjing's 6th population census, the sample of identified local users equaled 23.4% of the total population. Considering the average daily reception of 270,000 non-local tourists in Nanjing, the amount of non-local users in the sample was 10%. Such samples are far larger than those in traffic and manual questionnaire surveys. No statistical data were available to examine the tourist spots identification results in terms of identification accuracy. However, on the four holidays, the correlation coefficients between the daily distribution of any two recreation areas all reached over 0.8, indicating that the recreation areas' identification results were stable and should also be credible, thereby bearing overall representativeness.

2.3. Research Methods

The transfer and diffusion of tourists between tourism destinations produces a certain connection between each destination, giving rise to a dynamic evolutionary system. The social network analysis method can precisely depict the various relationships within a system, from the perspective of macro-structural relations. For this reason, has been used reliably in recent tourism studies. Therefore, this study, with the aid of the social network analysis software UCINET [32], analyzed three aspects—namely, tourism flow network connection, node centrality, and network communities. In addition, the overall characteristics of Nanjing's tourism flow network structure were summarized visually using the ArcGIS digital technology platform. On this basis, this study further compared the differences in the characteristics of tourism flow network structures, between local residents and non-local tourists. The concrete research methods are shown in Figure 2.

2.3.1. Methods for Analyzing Tourism Flow Network Connections

Based on social network analysis, the spatial departure statistics and tourist destination points in users' tourist activities were collected at each scenic spot, per their spatial positions. The scenic spots were set as network nodes, and visiting tourists were sequentially linked to form a network. By collecting the flow volume statistics of each node, we obtained the spatial concentration of each node in the entire tourism flow network to represent the out-degree centrality. Next, an asymmetric adjacency matrix was constructed, the multi-valued matrix was converted to a binary matrix, and the ArcGIS software analyzed the strength of network connection between nodes.

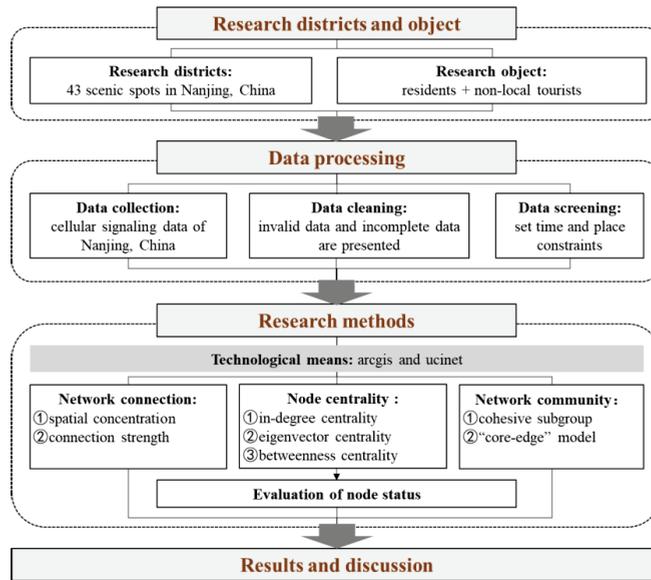


Figure 2. Research methodology.

2.3.2. Methods for Analyzing Node Centrality in Tourism Flow Networks

The centrality of nodes is a significant indicator in social network analysis. Through the depiction of the centrality of different standards, the value and status of a node's existence can be reflected in tourism flow network structures. Three forms of centrality have been selected for analysis in this study:

- In-degree centrality (popularity and attractiveness)

In-degree centrality measures the popularity of a certain node, representing the extent of a scenic node's clustering ability. This study selected the in-degree centrality and standardized it to the [0,1] interval. The calculating formula of the standard in-degree centrality C_{RD_i} of node i is as equation:

$$C_{RD_i} = \frac{\sum_j a_{ji}}{\max(\sum a)} \tag{1}$$

- Eigenvector centrality (latent attractiveness)

This depends on the direct relation of a node with its adjacent point and represents the connectivity level of tourism nodes. Regarding adjacent matrix A , the score X_i of the relative centrality of node i should exist:

$$X_i = \frac{1}{\lambda} \sum_j a_{ji} X_j \tag{2}$$

Rewriting the expression gives the eigenvector calculating equation:

$$Ax = \lambda x \tag{3}$$

Solving and standardizing the matrix gives the calculating equation of the standard eigenvector centrality $C_{R\beta}$:

$$C_{R\beta} = \alpha(I - \lambda A)^{-1} AI \tag{4}$$

In the equation, α is the standardized constant, λ is the eigenvalue corresponding to the first eigenvector, which determines the importance of the adjacent point to centrality, and I is the identity matrix.

- Betweenness centrality (controlling ability of mediation)

This measures the ability of a node to control the movement of tourists between node pairs in the tourism flow network, manifesting the controlling ability of a node, or its network mediating and moderating effects.

Assuming that the number for the shortest path between node j and node k is g_{jk} , and the number for the shortest path between node j and node k that must pass through i is $g_{jk(i)}$, then the probability of i situating in the shortest path between j and k can indicate the betweenness centrality of node i . The calculating formula of the standard betweenness centrality is as calculating equation:

$$C_{RB_i} = \frac{2 \sum_{j \neq k} \frac{g_{jk(i)}}{g_{jk}}}{(n-1)(n-2)} \quad (5)$$

Based on the above standard centrality, centrality was divided into three classes (high, middle, and low) through natural breaks, in order to ensure that the internal difference within the same class was minimized, while the differences between different classes was maximized. This paper combined the characteristics of node centrality (in-degree centrality, eigenvector centrality, and betweenness centrality) (see Table 1), and constructed an evaluation system for the node status in the tourism network through the (matrix) model.

Table 1. Combination relationship of three types of node centrality in the tourism network.

	Low In-Degree Centrality	Low Eigenvector Centrality	Low Betweenness Centrality
High in-degree centrality	—	Nodes in more popular areas, although in single tourism routes.	Nodes themselves are popular, although they are usually treated as the departure and destination points on tourism routes.
High eigenvector centrality	Exist in the areas around nodes of high popularity.	—	Nodes and their adjacent nodes are located in areas of highly intensive activity, and possess interactive diversity.
High betweenness centrality	Although less known, such nodes are located where visitors are highly likely to pass on their way to access other nodes.	Rare type. Nodes in a monopoly position when connecting with marginal nodes, and are areas of highly intensive activity.	—

2.3.3. Methods for Analyzing the Communities in the Tourism Flow Network

Community analysis is another important focus of research in social network analysis. The communities that are demarcated through nodes, and their connections, can reflect the extent of independence and popularity change of network tourism routes. This paper used the CONCOR method of the UCINET analysis software to perform a cohesive subgroup analysis, measuring the strength of node connections in flow volume. The entire network was divided into several sub-networks with powerful internal connections. Through multiple iterations, a correlation coefficient matrix was created. The higher the numerical value of the density matrix, the closer the connection between subgroups. It has guiding significance regarding tourists in their selection of combinations of tourism nodes, and designation of tourism routes. Furthermore, through the “core-edge” model, we further determined the status of tourism nodes in the overall network, and summarized and analyzed the structural model of Nanjing’s tourism flow network.

3. Results

3.1. Analysis of Tourism Network Connections

3.1.1. Spatial Concentration of the Tourism Flow Network

According to the research methods stated in Section 2.3.1, of the 43 scenic spots comprising the spatial scope, we selected 10,252 pairs of nodes with vectors. We collected the statistics of each node's visit volume. Then, the nodes were divided into five classes, according to natural breaks. The larger the nodes, the higher their spatial concentration. The spatial concentration of the overall–local–non-local networks (see Figure 3) was thus derived, representing their out-degree centrality. The results showed the following.

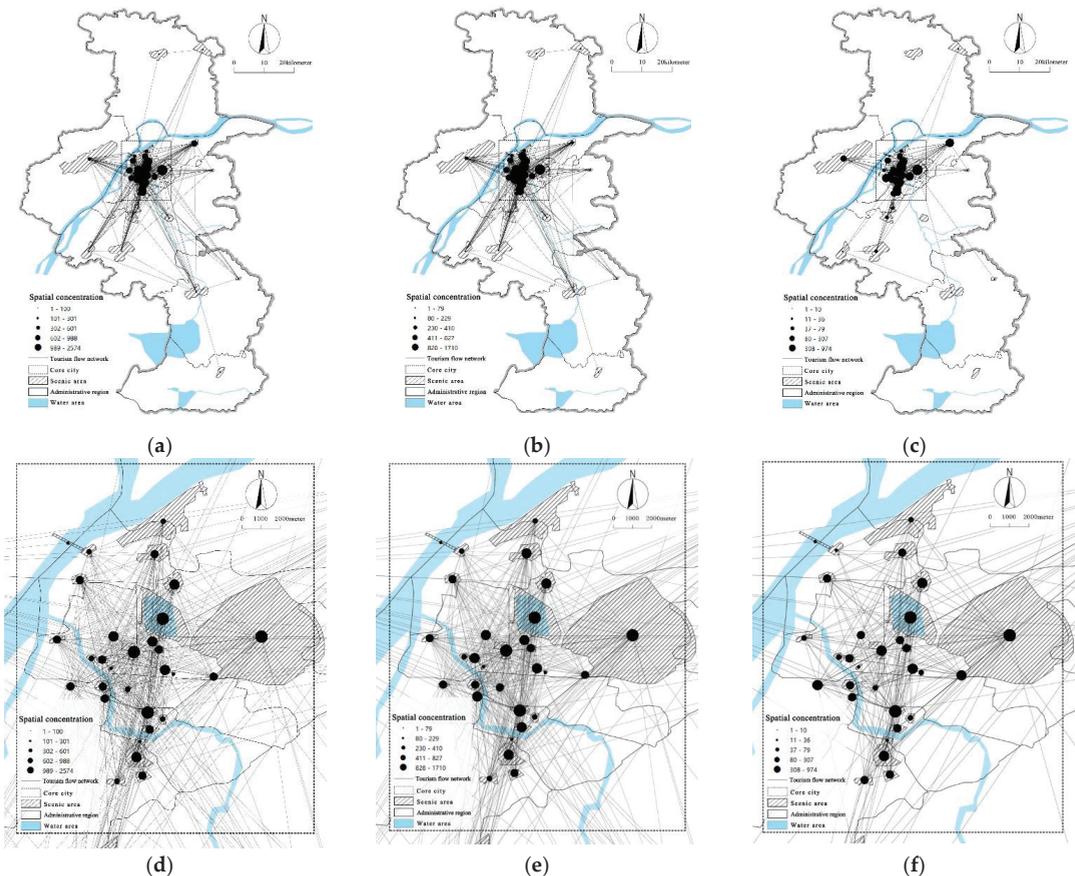


Figure 3. Spatial concentration of the tourism flow network under three types of network context (overall, resident, and non-local tourist): (a) overall tourism flow network of entire city; (b) resident tourism flow network of entire city; (c) non-local tourism flow network of entire city; (d) overall tourism flow network of core city; (e) resident tourism flow network of core city; (f) non-local tourism flow network of core city.

As a whole, the overall connectivity of Nanjing's tourism flow network was relatively good. No nodes were completely isolated. From the entire city proper, the network density was not high. However, the nodes in the core city were highly popular, with flows exhibiting the clustering feature. Nodes whose spatial concentration was in the first class (Zhongshan Scenic Area, Confucius Temple, Xuanwu Lake, Presidential Place (1912 District), Nanjing University Gulou District, and Hongshan Forest Zoo) were all

located in the core city. Their spatial concentrations were far higher than those in the second class, thus demonstrating their absolute core status. Additionally, all of the nodes whose spatial concentration ranked in the first three classes were in the core city. Most of the 14 nodes with the lowest class spatial concentration were in the suburbs outside the core city (including Gaochun International Cittaslow Tranquil, Pingshan Forest Park, Jinniu Lake Scenic Area, Fangshan Scenic Area, Dajinshan Scenic Area, Tiansheng Bridge Scenic Area, etc.), except for the Nanjing Yangtze River Bridge and Meixian Xincun Memorial Hall, which were in the core city. Generally, the spatial concentration of Nanjing's tourism flow showed a network characteristic of "exceptionally high spatial concentration for core city nodes, low spatial concentration for suburban nodes, and outstanding performance of core nodes".

By comparing the nodes' spatial concentrations formed by the tourism behavior of residents and non-local tourists, we discovered that the local network visits to scenic spots manifested higher densities than non-local tourists, although both displayed the characteristic of high spatial concentration for the core city network, and low spatial concentration for the suburban network. The spatial concentration of residents and non-local tourists related to core city nodes was generally similar. Residents displayed higher spatial concentration than non-local tourists, in terms of humanistic scenic spots (e.g., the Nanjing Museum, Nanjing University Gulou District, and Yihe Road Republican Architecture Complex). In comparison with non-local tourists, residents showcased higher and more comprehensive spatial concentration regarding suburban scenic spots. For non-local tourists, the nodes of Pingshan Forest Park, Fangshan Scenic Area, and Gaochun International Cittaslow Tranquil were isolated. In sum, both groups, residents and non-local tourists, displayed a characteristic of "relatively consistent overall spatial concentration, with slight differences existing in individual nodes".

3.1.2. Analysis of the Connection Strength of the Tourism Flow Network

Based on the above conclusion, this study further analyzed the strength of the network connection between nodes. The lines between tourism nodes represent the volume of tourism flows, with the thickness of the lines indicating the volume size, and the arrows signifying the direction of the flows. The connection strength was obtained from the overall-local-non-local networks (see Figure 4). Our findings showed the following.

Generally, over 85% of Nanjing's node connections were concentrated in the core city area. Fundamentally, the node connections gradually weakened as the distance from the core city area increased. Most of the suburban nodes showed a unidirectional inflow. Such connections resulted from the spillover from the core city nodes. The top five node pairs in the network connections (Zhongshan Scenic Area–Confucius Temple, Zhongshan Scenic Area–Xuanwu Lake, Beiji Ge Park (Jiming Temple)–Xuanwu Lake, Hongshan Forest Zoo–Xuanwu Lake, and Xuanwu Lake–Hongshan Forest Park) were all bidirectionally connected. Among them, Zhongshan Scenic Area, Confucius Temple, and Xuanwu Lake formed strong connections with two or more nodes and, thus, served as the absolute core in the network connection. Most of the second class node pairs formed from the first class node connection with other nodes, and so on. Thus a "tree" network structure manifests in Nanjing city's tourism flow network, where nodes with high connection strength spill over progressively to those with low connection strength. On the whole, the connection strength of Nanjing's tourism flow showed a "high-strength connection between nodes in the core city area, with a progressive decrease towards the periphery spatially".

Comparing the node connection strength that was formed by residents and non-local tourist behaviors, we found that both exhibited high connection strength in the core city, and low connection strength in the suburbs. However, the nodes that residents visited showed a higher overall connection strength than those of non-local tourists. The connection strength of the nodes visited by residents displayed a relatively balanced spatial distribution, and more diverse types. Comparatively, the connection strength of the nodes visited by non-local tourists showed a more concentrated spatial distribution, with blank spaces appearing

in the network connection of some suburb nodes. Additionally, the spatial concentration tended to fall on higher-rated and better-known node connections. Resident-visited nodes that fell in the first grade of connection strength included the following four pair nodes: Xuanwu Lake–Zhongshan Scenic Area, Xuanwu Lake–Beijing Ge Park (Jiming Temple), Xuanwu Lake–Hongshan Forest Zoo, and Old East Gate–Confucius Temple; non-local resident-visited nodes with first grade connection strengths include the following two node pairs: Xuanwu Lake–Zhongshan Scenic Area and Confucius Temple–Zhongshan Scenic Area. This shows that marked differences existed between these two groups. In terms of tourism route selection, residents attached greater priority to proximity in spatial location, whereas non-local tourists tended to choose well-known scenic spots. Generally, concerning node connection strength, “the nodes residents visited displayed a spatial balance in connection strength and diversification of types, while the nodes visited by non-local tourists exhibited a spatial concentration and inclination toward highly-rated nodes. The two groups showed distinct differences in tourism route selection”.

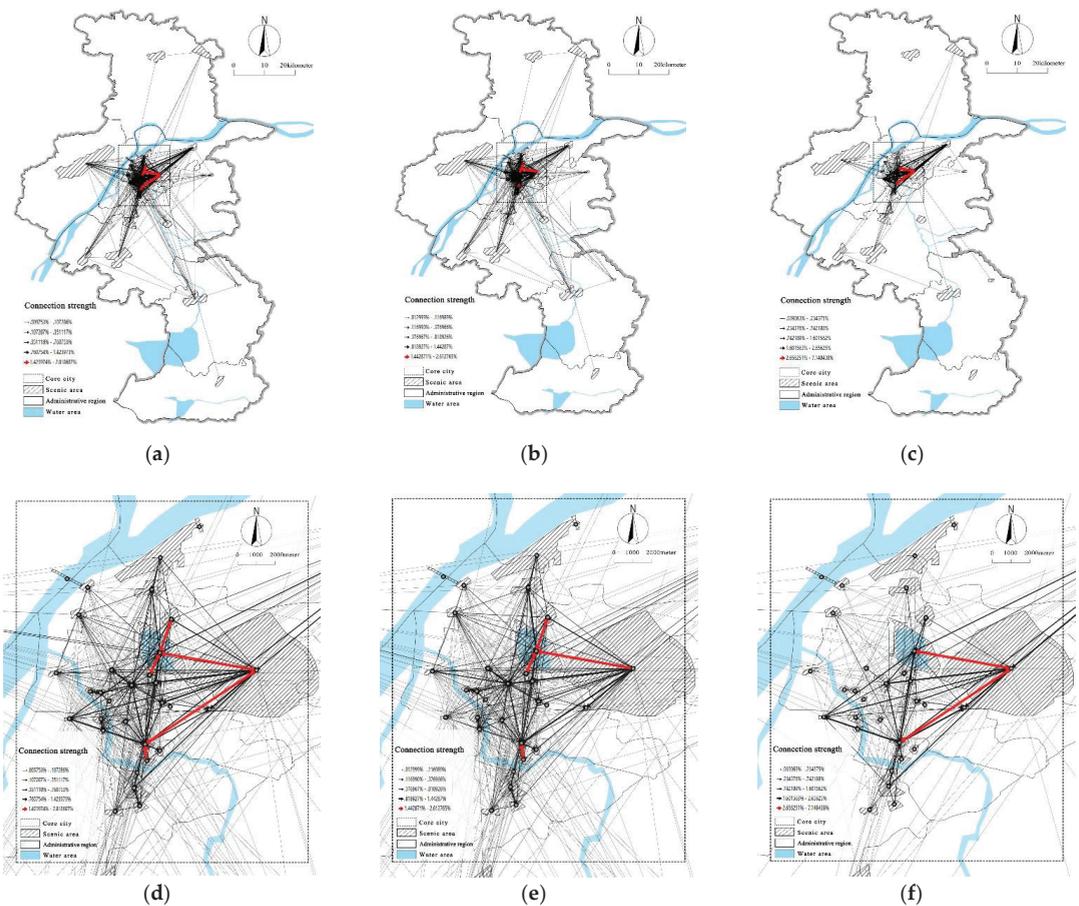


Figure 4. Connection strength of tourism flow networks under three types of network context (overall, resident, and non-local tourist): (a) overall tourism flow strength of entire city; (b) resident tourism flow strength of entire city; (c) non-local tourism flow strength of entire city; (d) overall tourism flow strength of core city; (e) resident tourism flow strength of core city; (f) non-local tourism flow strength of core city.

3.2. Analysis of Node Centrality in the Tourism Flow Network

3.2.1. Node Centrality Analysis

Centrality is a quantitative statistic of node power, as seen in the analytical methods for tourism flow networks in Section 2.3.2. Centrality can depict the values of nodes in tourism networks. The calculations in this paper produced the differentiated measure values of the three types of centrality under different tourism flow network contexts—namely, overall, local, and non-local (see Table 2). The natural break classification categorized them into three levels. The internal difference within the same level was minimized, while the differences between different classes was maximized. A spatial visualization was also performed (see Figures 5–7).

Table 2. Tabulation of three types of node centrality in the tourism flow network.

	In-Degree Centrality			Eigenvector Centrality			Betweenness Centrality		
	Overall	Resident	Non-Local Tourist	Overall	Resident	Non-Local Tourist	Overall	Resident	Non-Local Tourist
Zhongshan Scenic Area	0.929	0.929	0.667	0.977	0.913	1	0.109	0.101	0.148
Confucius Temple	0.833	0.833	0.667	0.933	0.894	0.913	0.049	0.051	0.138
Gulou District of Nanjing University	0.833	0.81	0.595	0.933	0.933	0.875	0.043	0.052	0.091
Xuanwu Lake	0.786	0.714	0.667	0.857	0.857	0.824	0.029	0.025	0.084
Yihe Road Republican Architectural Complex	0.738	0.714	0.214	0.84	0.84	0.618	0.018	0.023	0.001
Hongshan Forest Zoo	0.738	0.738	0.381	0.824	0.824	0.712	0.027	0.034	0.014
Qingliangshan Park	0.738	0.714	0.31	0.808	0.764	0.667	0.018	0.018	0.015
Memorial Hall of the Victims in Nanjing Massacre	0.69	0.667	0.357	0.792	0.737	0.724	0.031	0.031	0.014
Yuhuatai	0.69	0.69	0.31	0.792	0.778	0.7	0.026	0.028	0.007
Old East Gate	0.69	0.667	0.214	0.764	0.75	0.689	0.015	0.015	0.004
Presidential Palace (1912 District)	0.667	0.667	0.762	0.792	0.792	0.689	0.012	0.014	0.153
Beigu Mountain	0.667	0.667	0.31	0.792	0.778	0.677	0.01	0.012	0.013
Sipailou District of Southeast University	0.667	0.667	0.286	0.712	0.712	0.646	0.007	0.008	0.003
Nanhu Park	0.643	0.643	0.143	0.778	0.764	0.636	0.016	0.016	0.004
Zheng He Treasure Ship Park	0.619	0.595	0.262	0.792	0.792	0.667	0.007	0.009	0.002
Laoshan National Forest Park	0.619	0.548	0.238	0.778	0.778	0.646	0.056	0.035	0.023

Table 2. Cont.

	In-Degree Centrality			Eigenvector Centrality			Betweenness Centrality		
	Overall	Resident	Non-Local Tourist	Overall	Resident	Non-Local Tourist	Overall	Resident	Non-Local Tourist
Yuejiang Tower Scenic Area	0.619	0.595	0.286	0.737	0.724	0.636	0.01	0.012	0.002
Qixiashan Scenic Area	0.571	0.524	0.31	0.857	0.824	0.808	0.014	0.013	0.028
Chaotian Palace Scenic Area	0.571	0.524	0.286	0.7	0.7	0.609	0.004	0.004	0
Yuhuatai Gede Garden	0.548	0.524	0.262	0.792	0.764	0.677	0.008	0.008	0.007

* Note: the in-degree centrality is ranked in descending order, and only the first 15 are shown.

• Popularity and Attraction of Network Nodes (In-degree Centrality)

As shown in Figure 5, the popularity and attraction of the core city nodes were far higher than the suburban nodes. A total of 90% of the core city nodes were within the high and medium levels of popularity and attraction, with a relatively small range. As for the suburban nodes, their popularity and attractiveness were low, with only a small number at the high and medium levels of popularity and attractiveness.

A comparison of the popularity and attractiveness of the nodes visited by residents and non-local tourists showed that, in both groups, core city nodes were more popular than suburban nodes. The most popular nodes included Xuanwu Lake, Zhongshan Scenic Area, Confucius Temple, and Nanjing University Gulou District. The nodes that attracted residents were more balanced in their spatial distribution (particularly in the core city), with many of the nodes falling within the high and middle classes. By comparison, the nodes that attracted non-local tourists were more concentrated, and the number of popular nodes was far lower than that of the popular nodes visited by residents.

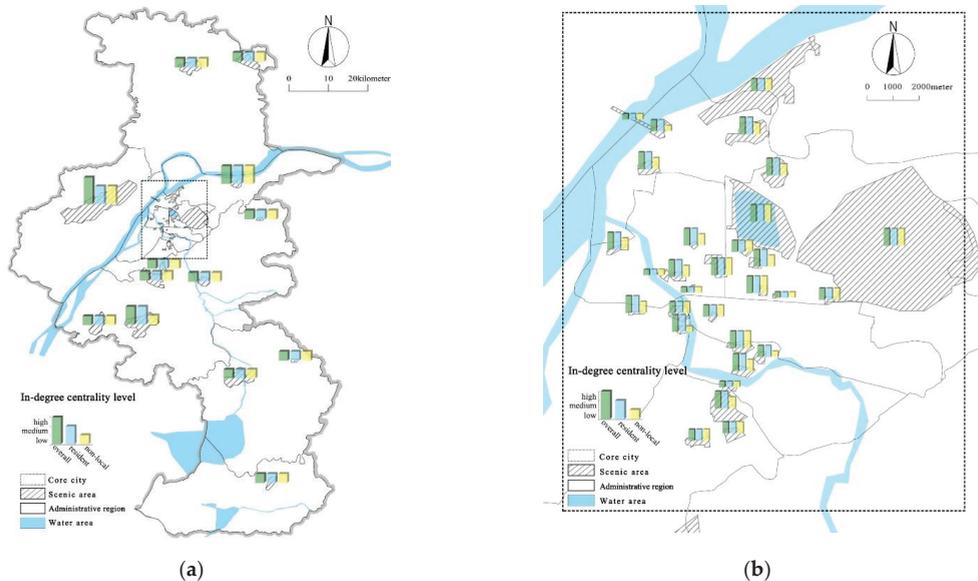


Figure 5. Classification map of the in-degree centrality of nodes under three types of network context (overall, resident, and non-local tourist): (a) entire city of Nanjing; (b) core city of Nanjing.

- Latent Attraction of Network Nodes (Eigenvector Centrality)

As indicated in Figure 6, the intensity of the activities around the core city nodes was far higher than in the suburban nodes. Six nodes were located in centers of intense activity (Zhongshan Scenic Area, Xuanwu Lake, Hongshan Forest Zoo, Confucius Temple, Nanjing University Gulou District, and Yihe Road Republican Architectural Complex). Qixiashan Scenic Area, in the suburbs, had a high-level latent attraction, which resulted from the high intensity in the core city nodes.

Comparing the latent attraction of the nodes that were visited by residents and non-local tourists, our results showed that both bore similarities in the core city area. The nodes that were visited by residents achieved a higher rate of latent attraction than those that were visited by non-local tourists. In the suburbs, the nodes in the non-local tourist network displayed a higher level of latent attraction than those that were visited by residents. The findings show that residents had more flexible and varied selections of tourism nodes, while non-local tourists selected popular nodes as often as possible.

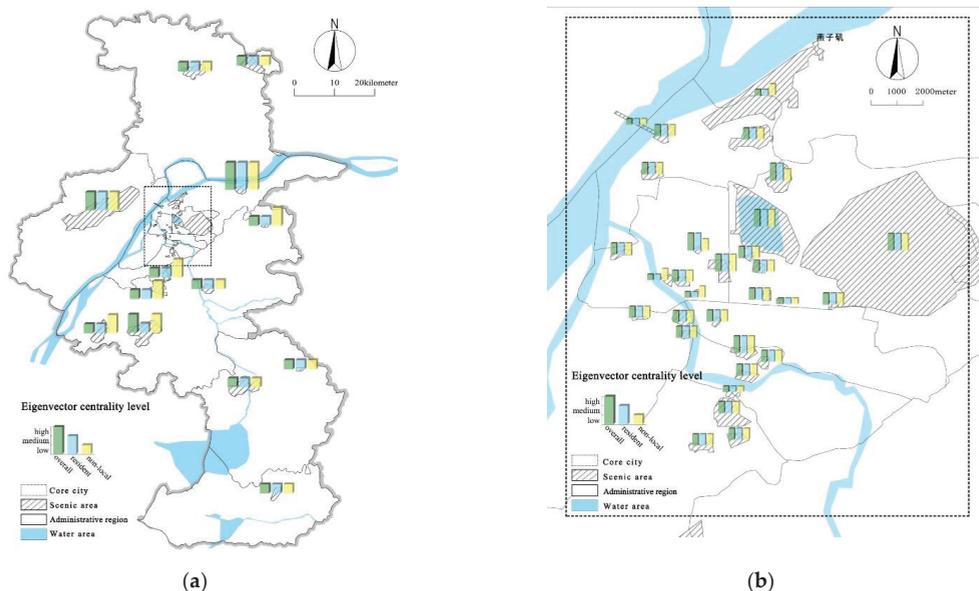


Figure 6. Classification map of the eigenvector centrality of nodes under three types of network context (overall, resident, and non-local tourist): (a) entire city of Nanjing; (b) core city of Nanjing.

- Mediating and Controlling Power of Network Nodes (Betweenness centrality)

As shown in Figure 7, there was an unclear relationship between the mediating and controlling power of the nodes in the city proper and their spatial locations. Almost all of the nodes were at medium and low levels, with Zhongshan Scenic Area being the only essential node. This indicates that all of Nanjing city's nodes have weak mediating and controlling power, meaning each node was relatively independent, or mostly combinations of nodes, with no nodes being necessary. Thus, the network structure was rather loose, with no significantly popular tourism routes (links whose node number was greater than two).

Similarities were found in the comparison of the mediating and controlling power of the nodes visited by residents and non-local tourists. The nodes that non-local tourists selected had slightly stronger mediating and controlling power than those that were selected by residents. This shows that the combinations of nodes in the non-local tourism flow network were relatively uniform and stable, while those in the resident networks were more flexible and varied.

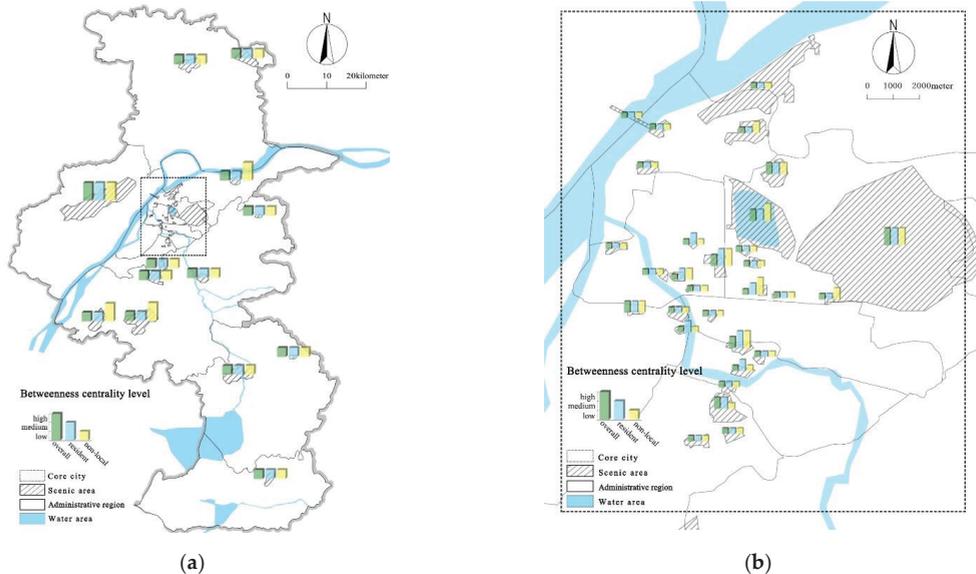


Figure 7. Classification map of the betweenness centrality of nodes under three types of network context (overall, resident, and non-local tourist): (a) entire city of Nanjing; (b) core city of Nanjing.

3.2.2. Evaluation of Node Status Based on Node Centrality

A correlation exists between all of the types of tourism node centrality. Therefore, these three types of node centrality were arranged and combined to create a more comprehensive status evaluation system. Figure 8 depicts a coordinate system of differences in centrality combinations. In this system, axes X, Y, and Z represent in-degree centrality, eigenvector centrality, and betweenness centrality, respectively, and each divides into high, medium, and low levels. If a node had a combination of high and medium levels, or medium and low levels, the range was one. However, if it consisted of all three levels, or only high and low levels, the range was two. When all of the three centrality type levels were identical, there was no gradation. When all of the three node centrality types were low, there was always an edge node in the tourism network. When all of the three node centrality types were at the medium level, this meant that the node was still in balanced development. When the three node centrality types were high from beginning to end, it was a popular core node. This coordinate system arranged and combined the three different centrality type levels (high, medium, and low) to highlight tourism significance.

All of the nodes populated the coordinate system according to the three centrality combination types depicted in the node status evaluation system (see Figure 9). Three nodes with low levels of eigenvalue centrality were excluded from the diagram to improve readability. In the overall table, the more common combinations were: (1) middle in-degree, middle eigenvalue, and low betweenness—nodes had a certain popularity, as did the adjacent points, although it is very likely that they were the tourist's starting and destination points; (2) high in-degree, medium eigenvalue, and low betweenness—nodes were popular, and the adjacent points also had a certain level of popularity, being situated in the core position of non-core communities; (3) high in-degree, high eigenvalue, and medium betweenness—nodes and adjacent points were popular, situated at the “bridging” position between edge nodes and the core district, or between small communities.

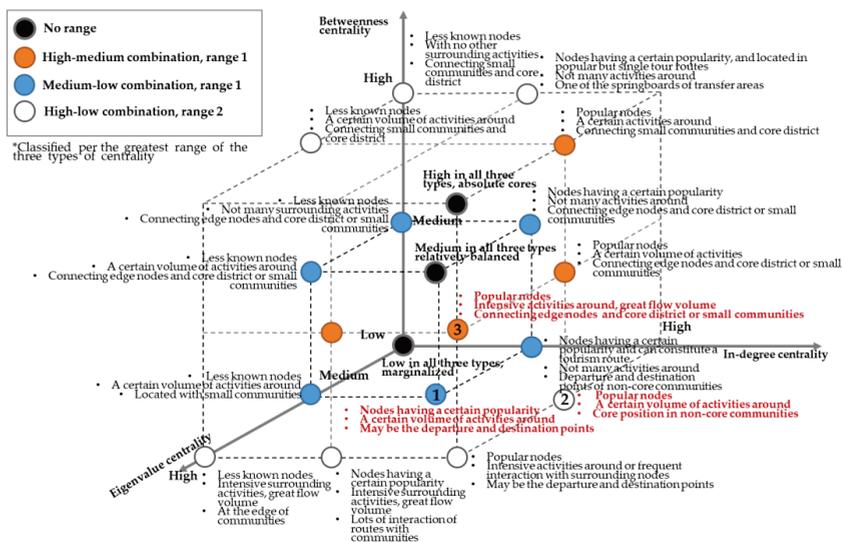


Figure 8. Node status evaluation system for tourism flow networks.

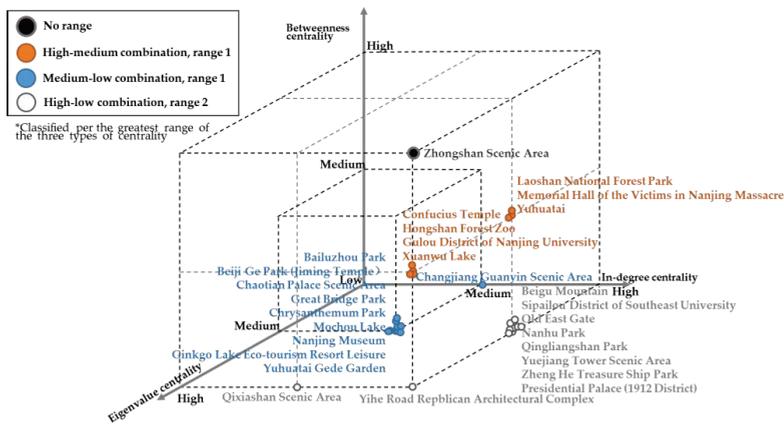


Figure 9. Evaluation of the node status in the overall tourism flow network.

This study further compared the resident and non-local tourist networks (see Figure 10). We revealed and ranked the resident network’s first three combinations: (1) medium in-degree, medium eigenvalue, and low betweenness; (2) medium in-degree, high eigenvalue, and medium betweenness; (3) high in-degree, medium eigenvalue, and low betweenness. This showed “high in-degree centrality, generally moderate eigenvalue centrality, and seriously polarized betweenness centrality”. The results indicate that the popular nodes that were approved by residents were more diversified. Some of the popular nodes showed close connections and balanced development, while others existed independently. The first three combinations in the ranking of non-local tourist networks were: (1) medium in-degree, medium eigenvalue, and low betweenness; (2) low in-degree, medium eigenvalue, and low betweenness; (3) medium in-degree, medium eigenvalue, and medium betweenness—manifesting the characteristic of “overall low in-degree centrality, higher eigenvalue centrality, and seriously polarized betweenness centrality”. Compared to those visited by residents, the popular nodes that were approved by non-local tourists

were relatively concentrated and stable, and mostly in a state of moderate popularity or marginalization.

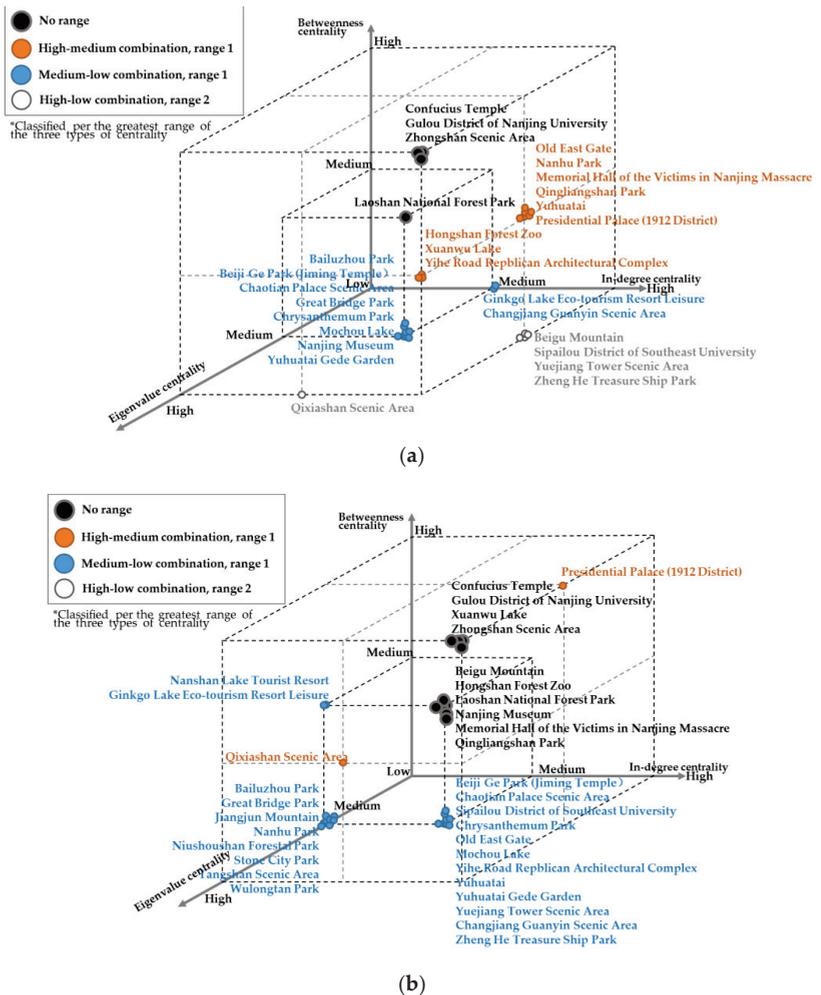


Figure 10. Comparison of the node status evaluation in resident networks and non-local tourist networks: (a) resident networks; (b) non-local tourist networks.

3.3. Analysis of Tourism Flow Network Communities

3.3.1. Cohesive Subgroup Analysis of the Tourism Flow Network

The CONCOR method in the UNICET software was used to perform a cohesive subgroup analysis. We calculated the coefficients of each row (or column) in the matrix, with the final results shown in Table 3. The higher the numerical value of the density matrix, the closer the subgroup connection.

The results of the cohesive subgroup analysis revealed the substructures within the tourism flows, and show more tourist route combinations. In the cohesive subgroup density matrix of the overall tourism flow network in Nanjing city (see Table 3), eight substructures formed, each with varying closeness of flow connections between the subgroups. Subgroup 5, 6, and 8 showed the most frequent interactions between internal node members. Subgroup 7 showed a higher frequency in the internal interaction between its subgroup nodes

and their external connections. In this light, combination marketing, or combined tourism tickets, may appeal to these two subgroups. The nodes in subgroup 1 and 3 were mainly dependent on their connections with other subgroups. These node types may be treated as additional products to the above subgroups. Subgroup 2 and 4 were relatively independent, with comparatively weak internal and external connections, and can be branded separately as tourism products with unique features.

Table 3. Density matrix of the cohesive subgroups in the tourism flow network.

Overall	1	2	3	4	5	6	7	8
1	0.000	0.241	1.000	1.000	0.065	0.039	0.463	0.000
2	0.004	0.000	0.000	0.049	0.029	0.183	0.048	0.002
3	1.000	0.328	0.460	0.000	0.162	0.124	0.362	0.043
4	0.277	0.188	0.017	0.074	0.000	0.000	0.000	0.000
5	0.228	0.553	0.324	0.162	1.000	0.541	0.885	0.048
6	0.096	1.000	0.198	0.000	0.479	1.000	0.374	0.049
7	0.448	0.706	0.441	0.239	0.461	0.340	1.000	0.047
8	0.067	0.281	0.349	0.000	0.101	0.177	0.329	1.000
AVG1 (degree of internal connection)	0.000	0.000	0.230	0.037	0.500	0.500	0.500	0.500
AVG2 (degree of external connection)	0.352	0.258	0.311	0.138	0.288	0.257	0.367	0.107

* Note: $R^2 = 0.159$, $AVG = 0.298$. 1—Bailuzhou Park, Nanshan Lake Tourist Resort, Jiangjun Mountain, Pingshan Forest Park; 2—Wulongtan Park, Great Bao'en Tower (Porcelain Tower), Meiyuan Xincun Memorial Hall, Laoshan National Forest Park, Tangshan Scenic Area, Fangshan Scenic Area, Jinniu Lake Scenic Area, Stone City Park, Gaochun International Citta slow Tranquil; 3—Yuhuatai Gongde Park, Memorial Hall of the Victims in Nanjing Massacre, Yuhuatai, Chrysanthemum Park; 4—Tianshen-qiao Scenic Area, Niushoushan Forestal Park, Nanjing Yangtze River Bridge, Ginkgo Lake Eco-tourism Resort Leisure; 5—Changjiang Guanyin Scenic Area, Hongshan Forest Zoo, Great Bridge Park, Beigushan, Beiji Ge Park (Jiming Temple), Yuejiang Tower Scenic Area, Xuanwu Lake, Zhongshan Scenic Spot; 6—Sipailou Campus of Southeast University, Yihe Road Republican Architectural Complex, Zheng He Treasure Ship Park, Nanjing University Gulou District, Qinglinagshan Park; 7—Dajinshan Scenic Area, Confucius Temple, Chaotian Palace Scenic Area, Qixiashan Scenic Area, Old East Gate, Nanjing Museum, Presidential Palace (1912 District); 8—Mochou Lake, Nanhu Lake Park.

Based on the analysis of the eight cohesive subgroups in the overall tourism flow network, these subgroups fall into four classifications: (1) **endogenous agglomeration type**—AVG1 (degree of internal connection $> AVG$, and AVG2 (degree of external connection) $\leq AVG$, manifesting as the subgroups converging inward, powerful connections between internal individual members, and weak connections outside of the subgroups; (2) **internal-external balance type**—AVG1 (degree of internal connection) $> AVG$, and AVG2 (degree of external connection) $> AVG$, manifesting as the subgroups having strong internal and external connections; (3) **externally attached type**—AVG1 (degree of internal connection) $\leq AVG$, and AVG2 (degree of external connection) $> AVG$, manifesting as weak connections between members and internal subgroups, although forming a strong connection with one or several external subgroups; (4) **individual independence type**—AVG1 (degree of internal connection) $\leq AVG$, and AVG2 (degree of external connection) $\leq AVG$, with members within subgroups having relatively weak internal and external connections, and failing to form obvious network connections with other subgroups and individuals (or forming a connection with individual subgroups).

Likewise, the local and non-local tourism flow networks could be classified according to this criterion, which produced our results for the analysis of the tourism flow networks of the three types of subgroup (overall, local, non-local) (see Table 4, Figure 11).

Overall, in Nanjing city, all of the tourism nodes with a high visit volume (tourism brand image) were within the endogenous agglomeration and internal-external balance types, thereby becoming the absolute cores of these two kinds of subgroups. Spatially, they connected to several adjacent nodes, with well-defined edges to the subgroups, and all were within the core city of Nanjing. Furthermore, some nodes that were on the periphery of the core city existed by attaching to these two kinds of subgroups. There were also some relatively independent, small-scale node groups, with only a single flow path. These results align with the study on the all-for-one tourism policy of cities.

Table 4. Cohesive subgroups of tourism flow networks under three types of network context.

Type	Overall ($R^2 = 0.159$)
Endogenous agglomeration type	<ol style="list-style-type: none"> 1. Changjiang Guanyin Scenic Area, Hongshan Forest Zoo, Great Bridge Park, Beigu Mountain, Beiji Ge Park (Jiming Temple), Yuejiang Tower Scenic Area, Xuanwu Lake, Zhongshan Scenic Area 2. Sipailou District of Southeast University, Yihe Road Republican Architectural Complex, Zheng He Treasure Ship Park, Gulou District of Nanjing University, Qingliangshan Park <ol style="list-style-type: none"> 3. Mochou Lake, Nanhu Park
Internal-external balance type	<ol style="list-style-type: none"> 1. Dajinshan Scenic Area, Confucius Temple, Chaotian Palace Scenic Area, Qixiashan Scenic Area, Old East Gate, Nanjing Museum, Presidential Palace (1912 District)
Externally attached type	<ol style="list-style-type: none"> 1. Bailuzhou Park, Nanshan Lake Tourist Resort, Jiangjun Mountain, Pingshan Forest Park 2. Yuhuatai, Yuhuatai Gede Garden, Memorial Hall of the Victims in Nanjing Massacre, Chrysanthemum Park
Individual independence type	<ol style="list-style-type: none"> 1. Wulongtan Park, Great Bao'en Temple, Meiyuan Xincun Memorial Hall, Laoshan National Forest Park, Tangshan Scenic Area, Fangshan Scenic Area, Jimniu Lake Scenic Area, Stone City Park, Gaochun International Cittaslow Tranquil 2. Tiansheng-qiao Scenic Area, Niushoushan Forestal Park, Nanjing Yangtze River Bridge, Ginkgo Lake Eco-tourism Resort Leisure
Type	Resident ($R^2 = 0.155$)
Endogenous agglomeration type	<ol style="list-style-type: none"> 1. Stone City Park, Qingliangshan Park, Meiyuan Xincun Memorial Hall 2. Great Bridge Park, Laoshan National Forest Park, Yihe Road Republican Architectural Complex, Changjiang Guanyin Scenic Area, Beiji Ge Park (Jiming Temple), Zheng He Treasure Ship Park, Beigu Mountain, Xuanwu Lake, Sipailou District of Southeast University, Zhongshan Scenic Area, Yuejiang Tower Scenic Area, Hongshan Forest Zoo, Gulou District of Nanjing University 3. Chrysanthemum Park, Great Bao'en Temple, Yuhuatai, Yuhuatai Gede Garden 4. Jiangjun Mountain, Niushoushan Forestal Park
Internal-external balance type	<ol style="list-style-type: none"> 1. Bailuzhou Park, Wulongtan Park, Presidential Palace (1912 District), Chaotian Palace Scenic Area, Ginkgo Lake Eco-tourism Resort Leisure, Old East Gate, Memorial Hall of the Victims in Nanjing Massacre, Confucius Temple, Nanjing Museum, Tangshan Scenic Area, Nanshan Lake Tourist Resort, Nanhu Park
Externally attached type	<ol style="list-style-type: none"> 1. Gaochun International Cittaslow Tranquil, Jimniu Lake Scenic Area, Tiansheng-qiao Scenic Area
Individual independence type	<ol style="list-style-type: none"> 1. Mochou Lake, Fangshan Scenic Area 2. Nanjing Yangtze River Bridge, Dajinshan Scenic Area, Pingshan Forest Park

Table 4. Cont.

Type	Non-local tourist ($R^2 = 0.089$)
Internal-external balance type	1. Xuanwu Lake, Tangshan Scenic Area, Niushoushan Forestal Park, Qixiashan Scenic Area, Presidential Palace (1912 District), Confucius Temple, Gulou District of Nanjing University, Tiansheng-qiao Scenic Area, Yuejiang Tower Scenic Area
	2. Old East Gate, Nanjing Museum, Chrysanthemum Park, Chaotian Palace Scenic Area, Nanhu Park, Yuhuatai, Yuhuatai Gede Garden, Memorial Hall of the Victims in Nanjing Massacre, Zhongshan Scenic Area
Externally attached type	1. Wulongtan Park, Ginkgo Lake Eco-tourism Resort Leisure, Dajinshan Scenic Area, Great Bridge Park, Meiyuan Xincun Memorial Hall, Fangshan Scenic Area, Zheng He Treasure Ship Park, Gaochun International Citaslow Tranquil, Laoshan National Forest Park, Pingshan Forest Park, Changjiang Guanyin Scenic Area
Individual independence type	1. Nanjing Yangtze River Bridge
	2. Jimiu Lake Scenic Area, Bailuzhou Park
	3. Hongshan Forest Zoo, Sipailou District of Southeast University, Yihe Road Republican Architectural Complex, Stone City Park, Nanshan Lake Tourist Resort, Beigu Mountain, Qingliangshan Park, Beiji Ge Park (Jiming Temple)
	4. Jiangjun Mountain, Mochou Lake, Great Bao'en Temple

From the residents' point of view, more nodes were classified as endogenous agglomeration and internal-external balance types. These form an even greater agglomeration network in the tourism flow, breaking down the spatial boundary of the core city, and absorbing more suburban nodes. Furthermore, the subgroups of externally attached and individual independence types had relatively few nodes, and showed the tendency of gradual absorption by the subgroups with a high agglomerative nature.

From the perspective of non-local tourists, no subgroup emerged from the endogenous agglomeration type. Additionally, almost all of the subgroups of the internal-external balance type were within the core city. Such subgroups were formed by core nodes and a few adjacent nodes, and were, therefore, subject to spatial limitations. More than half of the nodes were situated within the subgroups of the externally attached and individual independence types, indicating that only the core nodes in tourism flow network structures performed prominently, whereas the overall connection was insufficient and fragmented.

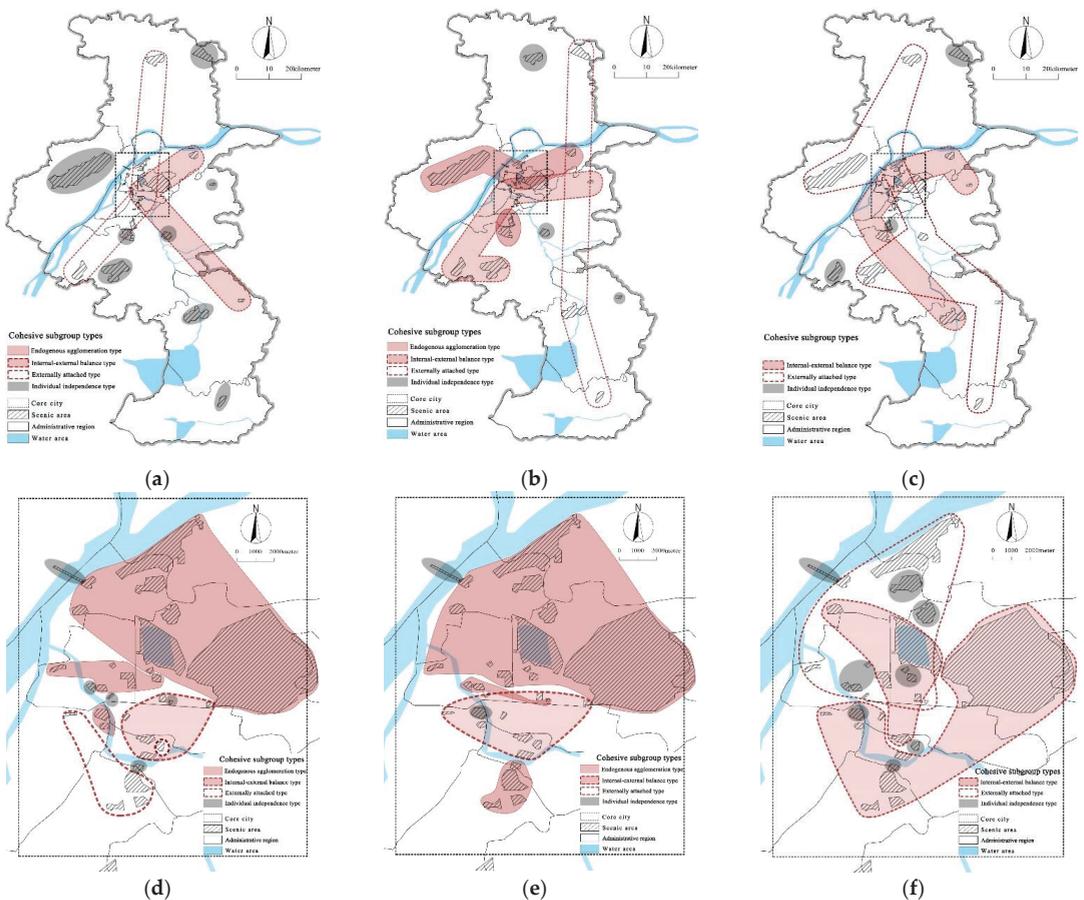


Figure 11. Cohesive subgroups' spatial distribution of tourism flow networks under three types of network context (overall, resident, and non-local tourist): (a) overall cohesive subgroup of entire city; (b) resident cohesive subgroup of entire city; (c) non-local cohesive subgroup of entire city; (d) overall cohesive subgroup of core city; (e) resident cohesive subgroup of core city; (f) non-local cohesive subgroup of core city.

3.3.2. “Core-Edge” Model

Based on the binarization results of the communities in the tourism flow networks that were extracted using the “core-edge” model of UNICET, the results of the overall tourism flow in Nanjing indicate the following (see Table 5, Figure 12). The core district members included four nodes—namely, Confucius Temple, Presidential Palace (1912 District), Xuanwu Lake, and Zhongshan Scenic Area. The remaining 39 nodes were at the district’s edge. Regarding the correlation degree, core district members reached 0.78, while district edge members were only 0.08, indicating obvious structural stratification in the tourism network in Nanjing. Furthermore, the correlation degree between core members and edge members reached 0.29, indicating that a connection between core and district edge was also relatively close. Regarding spatial distribution, the core district nodes were all located within the core city. Spatially, a diminished connection density was observed between the district edge and the core district nodes, from the core to the periphery. Therefore, aside from actively developing the core tourism districts, it is necessary to simultaneously enhance the overall tourism competitiveness of Nanjing city through positive cultivation, systematic and active expansion, and linkage to edge tourism districts.

By comparing the “core-edge” structures that were formed by resident and non-local tourist tourism flow networks (see Table 5), this study found that the respective core district members had slightly different compositions. The core district that was formed by resident tourism flow included Zhongshan Scenic Area, Xuanwu Lake, Beiji Ge Park (Jiming Temple), Confucius Temple, and Hongshan Forest Zoo. The non-local tourist district included Zhongshan Scenic Area, Confucius Temple, Xuanwu Lake, and Presidential Palace (1912 District). These findings show that the approval rate of Beiji Ge Park (Jiming Temple) and Hongshan Forest Zoo was only high within Nanjing, while the Presidential Place (1912 District) was generally a check-in spot for non-local tourists. In terms of correlation degree, the core districts that were formed by the non-local tourist tourism flow had a higher degree of internal correlation than those that were formed by residents, although residents had a higher degree of internal correlation for the edge districts, and between the core members and edge members. This shows that the node combinations for non-local tourists were relatively stable, and their polarization was more serious. On the other hand, resident tourism routes have begun to develop in a diversified and individualized direction, with some niche scenic spots beginning to enter their horizons. In terms of spatial distribution, the core districts that were formed by non-local tourists in the tourism flow network showed a concentrated and contiguous distribution. Meanwhile, residents have just begun to break through the spatial contiguity, manifesting a spatial form of separation between the two groups. Therefore, the differences between the tourist groups should be considered when a tourism development strategy is formulated for differentiation, individualization, and diversification.

Table 5. “Core-edge” model under three types of network context (overall, resident, and non-local tourist).

	Overall	Resident	Non-Local Tourist
Core	Confucius Temple, Presidential Palace (1912 District), Xuanwu Lake, Zhongshan Scenic Area (a total of 4, correlation = 0.782)	Zhongshan Scenic Area, Xuanwu Lake, Beiji Ge Park (Jiming Temple), Confucius Temple, Hongshan Forest Zoo (a total of 5, correlation = 0.710)	Zhongshan Scenic Area, Confucius Temple, Xuanwu Lake, Presidential Palace (1912 District) (a total of 4, correlation = 0.842)

Table 5. Cont.

	Overall	Resident	Non-Local Tourist
Edge	<p>Hongshan Forest Zoo, Gulou District of Nanjing University, Beiji Ge Park (Jiming Temple), Yuhuatai, Yihe Road Republican Architectural Complex, Memorial Hall of the Victims in Nanjing Massacre, Qixiashan Scenic Area, Nanhu Park, Old East Gate, Qingliangshan Park, Nanjing Museum, Beigu Mountain, Sipailou District of Southeast University, Yuhuatai Gede Garden, Mochou Lake ...</p> <p>(a total of 39, only the first 15 are listed, correlation = 0.081)</p>	<p>Yihe Road Republican Architectural Complex, Gulou District of Nanjing University, Presidential Palace (1912 District), Nanhu Park, Qingliangshan Park, Old East Gate, Yuhuatai, Beigu Mountain, Sipailou District of Southeast University, Yuhuatai Gede Garden, Mochou Lake, Zheng He Treasure Ship Park, Yuejiang Tower Scenic Area, Memorial Hall of the Victims in Nanjing Massacre, Qixiashan Scenic Area ...</p> <p>(a total of 38, only the first 15 are listed, correlation = 0.090)</p>	<p>Memorial Hall of the Victims in Nanjing Massacre, Gulou District of Nanjing University, Qixiashan Scenic Area, Nanjing Museum, Yuhuatai, Zheng He Treasure Ship Park, Hongshan Forest Zoo, Old East Gate, Laoshan National Forest Park, Nanhu Park, Sipailou District of Southeast University, Beigu Mountain, Beiji Ge Park (Jiming Temple), Yuhuatai, Qingliangshan Park ...</p> <p>(a total of 39, only the first 15 are listed, correlation = 0.078)</p>

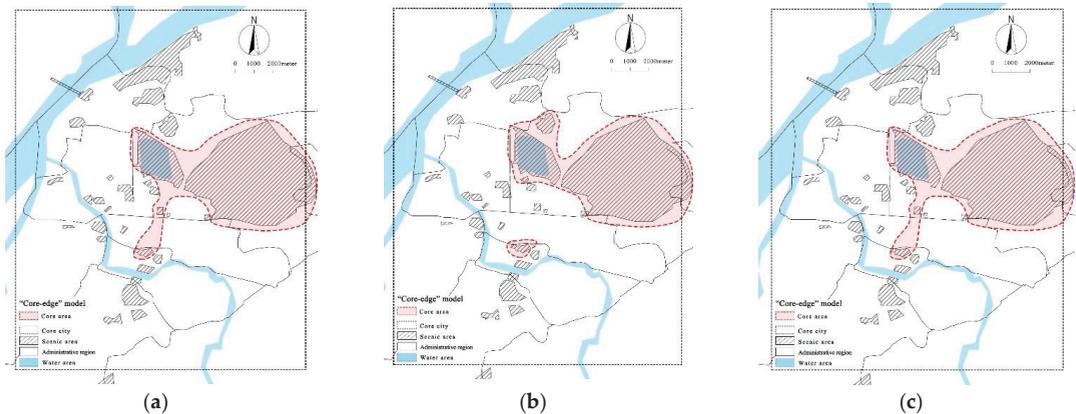


Figure 12. “Core-edge” model spatial distribution under three types of network context (overall, resident and, non-local tourist): (a) overall “core-edge” model of core city; (b) resident “core-edge” model of core city; (c) non-local “core-edge” model of core city.

4. Discussion

This study summarized the structural characteristics of the tourism flow network of 43 scenic spots in Nanjing city from three aspects: tourism flow network connection, node centrality, and communities. A comparative analysis revealed the tourism flow net structures that were formed by resident and non-local tourist behavior. The results are presented in Table 6.

Table 6. Summary of tourism flow network structures under three context types (overall, resident and, non-local tourist).

Analytical Perspective	Overall	Resident	Non-Local Tourist
Network connection	Relatively good overall connectivity, and high spatial concentration and connection strength of core city nodes. Poor performance of suburban nodes.	High spatial concentration and connection strength of core city nodes. Relatively balanced on the whole, with emergent tendency of a gradually increasing connection between core city and suburban nodes.	Spatial concentration and connection strength of core city nodes far higher than suburbs. Close connection between nodes with high spatial concentration and insufficient external connection.
Node centrality	Close mutual connection between popular nodes, no “bridging” nodes, less known nodes were marginalized, and relatively serious phenomenon of node polarization.	Connection between popular nodes relatively strong, with “bridging” nodes existing between them and marginal nodes. Influence of some popular nodes beginning to radiate outward to lesser-known nodes, and in a state of balanced development.	Concentration of popular nodes with very close internal connection, showing endogenous characteristics, weak external influence, and marginal nodes isolation.
Communities	Relatively clear spatial boundary between communities. The communities within the core city more closely connected, with some parts of the core absorbing suburban nodes. Most suburban communities attached to the communities of the core city, while individual nodes existed independently.	Obscure spatial boundary between communities. Community organization beginning to break down the spatial boundary of core districts, expanding to the whole region in a wedge shape. Attachment-type and independence-type communities gradually reduced.	Clear spatial boundary between communities. Communities in core city closely connected. Suburban communities marginalized and existing independently.

The tourism network structure in Nanjing city exhibits the characteristics of single center (core city), core-edge differentiation, and spatial agglomeration and diffusion. Furthermore, differences exist between the resident and non-local tourist characteristics. What are the reasons for the formation of such a tourism flow network structure in Nanjing? Generally speaking, at the urban scale, traffic flow is the main carrier of tourism flow. Therefore, traffic accessibility and travel mode must be the most directly influencing factors. Nanjing's road and rail traffic, for example, is radiated outward, from a central city layout; the density inside the core city road network is much higher than in other areas, meaning that the good traffic accessibility attracts tourists from all parts of the city. This directly caused the single center tourist flow network structure, which is located in the core city center. At the same time, due to the low accessibility of public transportation in the outer suburbs' scenic spots, road trips have become the main means of arrival. Additionally, local residents are more qualified to use this means of travel, which results in the tourism flow network of local residents virtually covering the whole city, while non-local tourists are mostly concentrated in the inner city of the central city.

Moreover, these differences are due to the combined influence of scenic spot characteristics and tourists forming the tourism flow network structure. Tourist spot characteristics (in addition to traffic factors) concretely involve the locations of scenic spots, grade, type differences (historical-cultural, natural landscape, indoors versus outdoors), and popularity (Internet heat). Furthermore, tourist characteristics include tourist source locations (local and non-local), residence, travel time (travel season, length of commute), travel motives, knowledge of and affection for destinations, information sources, revisit rate, and travel modes.

In light of the aforementioned conclusion, this study suggests that we should grasp the characteristics of the overall structure and nodes of the Nanjing city tourism flow network, and develop the entire region in unison. Efforts could include the following: (1) bringing nodes to maturity to form core tourism subdistricts; (2) energizing developing nodes; (3) devising specialized and themed tourism routes; (4) developing individualized ways of traveling for relatively independent nodes; (5) focusing on group differences among tourists and precisely positioning the audience groups for scenic spots; (6) executing accurate publicity.

The result of this study is the network structure of tourist flow represented by people flow. Therefore, it cannot reflect all of the characteristics of spatial structure, and different conclusions may be drawn by using other data, such as information flow, logistics and traffic flow. Secondly, the research uses mobile phone signaling data as the basic data, which is less authoritative than traditional official statistics; however, the characteristics of tourism flow networks are difficult to be realized using traditional data analysis. Although the data itself, and the processing process, will produce errors, the identification results were tested to ensure the randomness of the sampling, and can reflect the overall characteristics and rules. In addition, due to the length limit, this study focused on grasping the overall structural characteristics of the tourism flow network in Nanjing city and does not elaborate on the strategies used. Future studies should discuss in-depth methods to optimize tourism flow networks.

Author Contributions: Conceptualization, Lingjin Wang; methodology, Lingjin Wang; software, Lingjin Wang and Yan He; validation, Lingjin Wang and Xiao Wu; formal analysis, Lingjin Wang; investigation, Lingjin Wang; resources, Lingjin Wang; data curation, Yan He; writing—original draft preparation, Lingjin Wang; writing—review and editing, Lingjin Wang; visualization, Lingjin Wang; supervision, Xiao Wu; project administration, Xiao Wu; funding acquisition, Xiao Wu. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China(51878142), National Key Research and Development Program of China(2019YFD1100800), Postgraduate Research&Practice Innovation Program of Jiangsu Province(KYCX20_0143) and the Fundamental Research Funds for the Central Universities(3207032101D).

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from China Mobile Communications Group Co., Ltd., and are available from Lingjin Wang, with the permission of China Mobile Communications Group Co., Ltd.

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

References

1. Castells, M. *Rise of the Network Society*; Blackwell Publishers: Oxford, UK, 1996.
2. Feng, Z. Tourism and transport: Modes, networks and flows. *Tour. Geogr.* **2010**, *12*, 579–582. [[CrossRef](#)]
3. Zhang, J.; Jensen, C. Comparative advantage—Explaining tourism flows. *Ann. Tour. Res.* **2007**, *34*, 223–243. [[CrossRef](#)]
4. Hills, L.; Lundgren, J. The impact of tourism in the Caribbean: A methodological study. *Ann. Tour. Res.* **1977**, *5*, 248–267. [[CrossRef](#)]
5. Lundgren, J. Geographic concepts and the development of tourism research in Canada. *Geo J.* **1984**, *9*, 17–25. [[CrossRef](#)]
6. Prideaux, B. Factors affecting bilateral tourism flows. *Ann. Tour. Res.* **2005**, *32*, 780–801. [[CrossRef](#)]
7. Jensen, C.; Zhang, J. Trade in tourism services: Explaining tourism trade and the impact of the general agreement on trade in services on the gains from trade. *J. Int. Trade Econ. Dev.* **2013**, *22*, 398–429. [[CrossRef](#)]
8. Santeramo, F.G.; Morelli, M. Modelling tourism flows through gravity models: A quantile regression approach. *Curr. Issues Tour.* **2015**, *19*, 1077–1083. [[CrossRef](#)]
9. Sharipov, S.; Demirkol, H.G. Analysing the flow pattern in the tourism city of eskisehir. *Int. J. Tour. Cities* **2018**, *4*, 342–354. [[CrossRef](#)]
10. Ding, Z. A Research on the Space-Time Evolution of the Domestic Tourist Flow in Nanjing. *Tour. Trib.* **2004**, *2*, 37–40.
11. Zhang, Y.; Ma, Y.; Ma, H.; Gu, J. Research on the Transition of the Equilibrium Point of Inbound Agglomeration and Diffusion Tourism Flow in Beijing. *Tour. Trib.* **2009**, *24*, 31–35.
12. Hong, T.; Ma, T.; Huan, T.-C. Network behavior as driving forces for tourism flows. *J. Bus. Res.* **2015**, *68*, 146–156. [[CrossRef](#)]
13. Qiu, Y.; Ge, P.; Ren, P. An Empirical Study on Cluster Distributaries of the Initial State of Jiuzhaigou. *Resour. Sci.* **2010**, *32*, 118–123.
14. Jin, Z. An Analysis of Optimization Effects of the Changes in Product Spatial Structure of Tourist Flow—A Case Study on Hangzhou. *Tour. Trib.* **2006**, *7*, 42–47.
15. Jin, C.; Cheng, J.; Xu, J. Using User-Generated Content to Explore the Temporal Heterogeneity in Tourist Mobility. *J. Travel Res.* **2018**, *57*, 779–791. [[CrossRef](#)]
16. Liu, Z.; Li, H.; Shi, C.; Wang, X.; Zhang, H. Response of Temporal Distribution of Short-term Tourist Flows to Regional Tourism Spatial Structure—A Case Study of Golden Week Tourist Flows in Yunnan Province. *Acta Geogr. Sin.* **2010**, *65*, 1624–1632. [[CrossRef](#)]
17. Liu, F.; Zhang, J.; Zhang, J.; Chen, D.; Liu, Z.; Lu, S. Roles and functions of tourism destinations in tourism region of south anhui: A tourist flow network perspective. *Chin. Geogr. Sci.* **2012**, *22*, 755–764. [[CrossRef](#)]
18. Wang, Z.; Liu, Q.; Xu, J.; Fujiki, Y. Evolution characteristics of the spatial network structure of tourism efficiency in china: A province-level analysis. *J. Destin. Mark. Manag.* **2020**, *18*, 100509. [[CrossRef](#)]
19. Li, A.; Mou, N.; Zhang, L.; Yang, T.; Liu, W.; Liu, F. Tourism Flow Between Major Cities during China’s National Day Holiday: A Social Network Analysis Using Weibo Check-in Data. *IEEE Access* **2020**, *8*, 225675–225691. [[CrossRef](#)]
20. Sun, Y.; Shao, Y.; Chan, E. Co-visitation network in tourism-driven peri-urban area based on social media analytics: A case study in Shenzhen, China. *Landsc. Urban Plan.* **2020**, *204*, 103934. [[CrossRef](#)]
21. Chen, Y.; Chen, R.; Hou, J.; Hou, M.; Xie, X. Research on users’ participation mechanisms in virtual tourism communities by bayesian network. *Knowl.-Based Syst.* **2021**, *226*, 107161. [[CrossRef](#)]
22. Heinz, E. *Social Structures: A Network Approach*; Cambridge University Press: New York, NY, USA, 1989.
23. Hu, M.; Qiu, R.T.R.; Wu, D.C.; Song, H. Hierarchical pattern recognition for tourism demand forecasting. *Tour. Manag.* **2021**, *84*, 104263. [[CrossRef](#)]
24. Chen, S.; Wang, X.; Zhang, H.; Wang, J.; Peng, J. Customer purchase forecasting for online tourism: A data-driven method with multiplex behavior data. *Tour. Manag.* **2021**, *87*, 104357. [[CrossRef](#)]
25. Nistor, M.M.; Nicula, A.-S.; Dezsai, S.; Petrea, D.; Kamarajugedda, S.A.; Carebia, I.A. Gis-based kernel analysis for tourism flow mapping. *J. Settl. Spat. Plan.* **2020**, *11*, 137–145. [[CrossRef](#)]
26. Li, T.; Li, Q.; Liu, J. The spatial mobility of rural tourism workforce: A case study from the micro analytical perspective. *Habitat Int.* **2021**, *110*, 102322. [[CrossRef](#)]
27. Oender, I.; Koerbitz, W.; Hubmann-Haidvogel, A. Tracing tourists by their digital footprints: The case of Austria. *J. Travel Res.* **2016**, *55*, 566–573. [[CrossRef](#)]
28. Comito, C. NextT: A framework for next-place prediction on location based social networks. *Knowl.-Based Syst.* **2020**, *204*, 106205. [[CrossRef](#)]
29. Shakeri, E.; Far, B.H. Exploring the requirements of pandemic awareness systems: A case study of COVID-19 using social media data. In Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering Workshops, Association for Computing Machinery, Virtual Event, Australia, 21–25 December 2020; pp. 33–40.

30. Nurbakova, D.; Ermakova, L.; Ovchinnikova, I. In Understanding the Personality of Contributors to Information Cascades in Social Media in Response to the COVID-19 Pandemic. In Proceedings of the 2020 International Conference on Data Mining Workshops (ICDMW), Sorrento, Italy, 17–20 November 2020; pp. 45–52.
31. Qian, C.; Li, W.; Duan, Z.; Yang, D.; Ran, B. Using mobile phone data to determine spatial correlations between tourism facilities. *J. Transp. Geogr.* **2021**, *92*, 103018. [[CrossRef](#)]
32. Hanneman, R.; Riddle, M. *Introduction to Social Network Methods*; University of California: Los Angeles, CA, USA, 2005.

Article

Socioeconomic and Environmental Impacts on Regional Tourism across Chinese Cities: A Spatiotemporal Heterogeneous Perspective

Xu Zhang ^{1,†}, Chao Song ^{2,3,4,†}, Chengwu Wang ^{1,*}, Yili Yang ⁴, Zhoupeng Ren ², Mingyu Xie ⁴, Zhangying Tang ¹ and Honghu Tang ¹

- ¹ State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, School of Geoscience and Technology, Southwest Petroleum University, Chengdu 610500, China; zhangxu.gis@foxmail.com (X.Z.); tzycd@163.com (Z.T.); swputanghh@163.com (H.T.)
 - ² State Key Laboratory of Resources and Environmental Information System (LREIS), Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; chaosong@scu.edu.cn (C.S.); renzp@lreis.ac.cn (Z.R.)
 - ³ HEOA Group, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu 610041, China
 - ⁴ Institute for Healthy Cities and West China Research Center for Rural Health Development, Sichuan University, Chengdu 610041, China; yy1115@hotmail.com (Y.Y.); xie.mingyu@outlook.com (M.X.)
- * Correspondence: chengwuwanggis@163.com
† These authors contributed equally to this paper.

Citation: Zhang, X.; Song, C.; Wang, C.; Yang, Y.; Ren, Z.; Xie, M.; Tang, Z.; Tang, H. Socioeconomic and Environmental Impacts on Regional Tourism across Chinese Cities: A Spatiotemporal Heterogeneous Perspective. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 410. <https://doi.org/10.3390/ijgi10060410>

Academic Editors: Andrea Marchetti and Angelica Lo Duca

Received: 29 April 2021
Accepted: 10 June 2021
Published: 14 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Understanding geospatial impacts of multi-sourced drivers on the tourism industry is of great significance for formulating tourism development policies tailored to regional-specific needs. To date, no research in China has explored the combined impacts of socioeconomic and environmental drivers on city-level tourism from a spatiotemporal heterogeneous perspective. We collected the total tourism revenue indicator and 30 potential influencing factors from 343 cities across China during 2008–2017. Three mainstream regressions and an emerging local spatiotemporal regression named the Bayesian spatiotemporally varying coefficients (Bayesian STVC) model were constructed to investigate the global-scale stationary and local-scale spatiotemporal nonstationary relationships between city-level tourism and various vital drivers. The Bayesian STVC model achieved the best model performance. Globally, eight socioeconomic and environmental factors, average wage (coefficient: 0.47, 95% credible intervals: 0.43–0.51), employed population (−0.14, −0.17–−0.11), GDP per capita (0.47, 0.42–0.52), population density (0.21, 0.16–0.27), night-time light index (−0.01, −0.08–0.05), slope (0.10, 0.06–0.14), vegetation index (0.66, 0.63–0.70), and road network density (0.34, 0.29–0.38), were identified to have nonlinear effects on tourism. Temporally, the main drivers might have gradually changed from the local macro-economic level, population density, and natural environment conditions to the individual economic level over the last decade. Spatially, city-specific dynamic maps of tourism development and geographically clustered influencing maps for eight drivers were produced. In 2017, China formed four significant city-level tourism industry clusters (hot spots, 90% confidence), the locations of which coincide with China's top four urban agglomerations. Our local spatiotemporal analysis framework for geographical tourism data is expected to provide insights into adjusting regional measures to local conditions and temporal variations in broader social and natural sciences.

Keywords: Chinese regional tourism; socioeconomic and environmental drivers; spatiotemporal influencing factors; spatiotemporal estimation mapping; Bayesian STVC model; spatiotemporal nonstationary regression; geographical data modeling analysis

1. Introduction

Tourism is an underlying industry that promotes the development of the global economy [1]. According to the World Travel & Tourism Council (WTTC), tourism contributed 10.3% (8.9 trillion US dollars) of global GDP and provided one-tenth of the total number of jobs (330 million positions) in 2019 before the pandemic [2]. Through developing the tourism industry, local governments can markedly improve the level of infrastructure construction, increase employment opportunities, improve people's living conditions, and promote urban economic growth [3–5]. In addition, tourism development is a fundamental part of a sustainable development strategy, which is recognized as a green industry by the world due to its low energy consumption and light pollution characteristics in the development process [6].

Despite being one essential force promoting regional economy, regional tourism itself is greatly influenced by socioeconomic status [7,8], including GDP [8], employment status [9], personal income [10], health and hygiene [7], industrial production index [11] and social media [12]. Besides the socioeconomic condition, research also identified the notable role of the environment in affecting regional tourism [13–15], especially climatic conditions, such as temperature [16], precipitation [17], sunshine [18], and relative humidity [19]. Road infrastructure was also a critical environmental driver enhancing the tourism industry [20,21]. However, all these previous studies only adopted a limited number of factors. It is necessary to consider the comprehensive impacts on tourism by combining socioeconomic conditions with environmental conditions.

When investigating relationships between regional tourism and potential explanatory factors, an unrealistic assumption persistently embedded in previous literature was that the variables' relationships were homogeneous, which had been defined as stationarity. For instance, non-spatial tourism studies using qualitative analysis [22], feasible generalized least square (FGLS) regression [19], linear and quantile regression [23], or logit regression [24] are regarded as global-scale analyses and also ignore the existence of spatial effects. Likewise, some geospatial tourism studies using the spatial regression models, such as the exploratory spatial data analysis (ESDA) [25] or spatial econometric models [26], are capable of incorporating spatial effects for intercept or residual but are still unable to estimate a set of space-scale coefficients to characterize the varying region-specific relationships between variables. Hence, a more reasonable assumption in the real world highlights the heterogeneous or varying impacts of explanatory variables on tourism development due to region-specific situations, especially for studies conducted across large domains at finer geospatial scales. Such spatially heterogeneous variables relationships are called spatial nonstationarity in the field of statistics. At present, the geographically weighted regression (GWR) [27] is frequently used in tourism research, aiming at exploring such spatial nonstationarity between tourism and various influencing factors [28,29]. However, to the best of our knowledge, no study has been conducted from the spatiotemporal integrated nonstationary perspective, to fully explore both socioeconomic and environmental drivers on regional tourism development.

In China, as the area of interest in this study, there has long been an issue of regional tourism development disparities [30], which obstructed regional tourism sustainability to some extent [31]. Although these geospatial disparities have been extensively discussed at a provincial-level scale [32] or city group scale [33], seldom have studies explored the city-specific disparities of regional tourism, especially over mainland China. Based on tourism connotations and tourism elements, Chinese scholars have established a comprehensive indicator framework of influencing the urban tourism industry from multiple dimensions. Socioeconomic and environmental aspects are also considered indispensable indicators reflective of regional tourism industry development [34]. However, no existing studies ever investigated the joint impacts of socioeconomic and environmental conditions on China's city-level tourism from a spatiotemporal heterogeneous perspective, to provide evidence-based implications for assisting the formulation of tourism-related policies at governmental levels in a timely and effective manner.

In an attempt to find effective factors affecting regional tourism outcomes to provide tourism strategies tailored for specific local spatial conditions and changing temporal circumstances, we constructed an explanatory variable framework composed of 30 variables, including socioeconomic and environmental conditions. We explored spatiotemporal heterogeneous relationships between the regional tourism development and the multi-source explanatory factors from 2008 to 2017 across Chinese cities by employing the Bayesian spatiotemporally varying coefficients (STVC) model [35,36]. The establishment of such an explanatory variable framework in our study also served as a contributor to the current literature in this field in terms of improving the comprehensiveness of the existing research index system as well as adding novel perspectives into this field based on the consideration of both spatial and temporal heterogeneity.

2. Materials and Methods

2.1. Study Area and Data

Considering the unbalanced development speed and regional differences in China's tourism industry during the last decade, in this study, 343 prefecture-level areas were selected as the underlying research units (excluding Hong Kong, Macao, and Taiwan). Total tourism revenue was employed as a proxy variable to describe the regional tourism development level from 2008 to 2017 [30]. Figure 1 illustrates the original geographical distribution of city-level total tourism revenue across China in 2017.

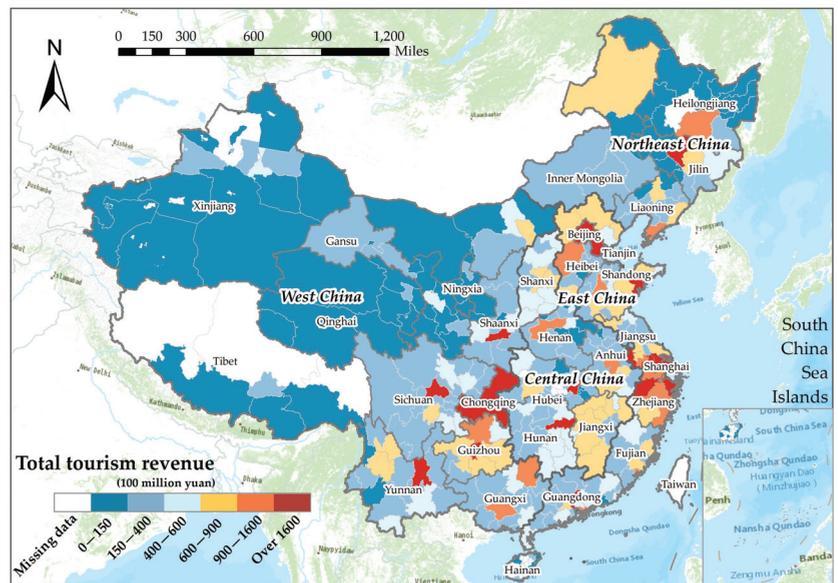


Figure 1. Geographical distribution of the original city-level total tourism revenue across China in 2017.

Correspondingly, we collected a relatively comprehensive system of 30 explanatory variables at the city level, including 21 socioeconomic factors and nine environmental variables (summarized in Table 1), to detect their impacts on total tourism revenue in China. The total tourism revenue and socioeconomic data were retrieved from the China City Statistical Yearbook and Statistical Bulletin. The climate data (EV1-EV4) were collected from the National Meteorological Information Center (<http://data.cma.cn/>, accessed on 28 April 2021). The other environmental factors (EV5-EV8) were downloaded from the Resource and Environment Science and Data Center (<http://www.resdc.cn/>, accessed on 28 April 2021). As a list of environmental variables, including elevation, road network density, slope, and nighttime light index, were not temporally continuous data, these

variables were only added as a part of the local-scaled modeling for spatial nonstationary analysis. Other socioeconomic and environmental factors had spatiotemporal variation characteristics, satisfying the hypothesis of spatiotemporal nonstationarity.

Table 1. City-level potential explanatory variables of regional tourism in China: SV1-21 denote twenty-one socioeconomic factors, and EV1-9 denote nine environmental factors.

Identifier	Socioeconomic Variables	Identifier	Environmental Variables
SV1	Gross domestic product (GDP) per capita (yuan)	EV1	Precipitation (0.1 mm)
SV2	Population density (person/km ²)	EV2	Temperature (centigrade)
SV3	Employment density of the first industry (person/km ²)	EV3	Air pressure (1 N/m ²)
SV4	Employment density of the Second industry (person/km ²)	EV4	Humidity (hPa)
SV5	Employment density of the tertiary industry (person/km ²)	EV5	NDVI (/)
SV6	Mobile phone penetration rate (subscriber/person)	EV6	Road network density (km/km ²)
SV7	Internet broadband penetration rate (subscriber/person)	EV7	Elevation (meter)
SV8	Local general budget revenue per capita (yuan)	EV8	Slope (°)
SV9	Local government budgetary expenditures per capita (yuan)	EV9	Nighttime light index (/)
SV10	Employees population density (person/km ²)		
SV11	Savings deposits of per capita residents (yuan)		
SV12	Loans of financial institutions per capita (yuan)		
SV13	Industrial enterprises density (number/km ²)		
SV14	Social fixed asset investment per capita (yuan)		
SV15	Social consumable total retail sales per capita (yuan)		
SV16	Student's density of ordinary middle school (person/km ²)		
SV17	Student's density of primary school (person/km ²)		
SV18	Hospital density (number/km ²)		
SV19	Hospital beds per capita (number/person)		
SV20	Employment density of urban units (person/km ²)		
SV21	Average wage of employed persons in urban units (yuan)		

2.2. Statistical Methods

2.2.1. Variable Selection

Two widely adopted approaches, namely multicollinearity assessment and random forest [37], were employed in a progressive manner as a screening step for identifying the most representative influencing factors on the tourism industry from 30 candidate variables. Precisely, the indicator variance inflation factor (VIF) was first adopted to measure the multicollinearity effect, referring to a correlation between explanatory factors [38]. Commonly, $VIF < 10$, representing mild and negligible multicollinearity, is adopted as the threshold to screen variables [39]. Here, given the adequacy of candidate variables involved in this analysis, a stricter standard was adopted, indicating that a candidate variable with $VIF > 5$ was removed. Following the VIF step, random forest, an integrated machine learning approach relying on the decision tree, was adopted for further screening the explanatory variables according to the calculation of an indicator named mean decrease impurity (MDI), which has been commonly used for reflecting the ranking of a factor's relative importance [40]. For a candidate variable, a higher value of MDI is associated with the increased importance of the variable. This random forest step is typically empirical and data-driven, as MDI is not a relative measure [41].

2.2.2. Bayesian STVC Model

The Bayesian spatiotemporally varying coefficients (STVC) model is a recently burgeoning local spatiotemporal regression developed under the Bayesian hierarchical modeling (BHM) framework. It is mainly designed to quantitatively characterize structured and heterogeneous spatiotemporal impacts (expressed as local-scale coefficients) of different covariates on the outcomes of the variable of interest, that is, to explore the spatiotemporal nonstationarity inherent in geospatial research phenomena [35,36].

For China's tourism case, Y_{it} denotes the space-time monitoring data of the total tourism revenue indicator, in which $i = 1, \dots, I$ ($I = 343$) are the administrative geographical units of the cities. For each city, data are available for a ten-year period from 2008 to 2017, labeled as $t = 1, \dots, T$ ($T = 10$). Then, the structured additive predictor $\zeta_{it} = g(Y_{it})$ within a reduced Bayesian STVC model is formulated in Equations (1)–(3), i.e.,

$$\zeta_{it} = g(Y_{it}) = \eta + \sum_{k=1}^K f_{space}(\omega_{ik} SX_{itk}) + \sum_{m=1}^M f_{time}(\varphi_{tm} TX_{itm}), \quad (1)$$

$$\omega_i | \omega_{-i}, \tau_\omega, W \sim N\left(\frac{\sum_{j=1}^I w_{ij} \omega_j}{\sum_{j=1}^I w_{ij}}, \frac{1}{\tau_\omega \sum_{j=1}^I w_{ij}}\right), i = 1, \dots, I, \quad (2)$$

$$\varphi_{t+1} - \varphi_t | \tau_\varphi \sim N\left(0, \frac{1}{\tau_\varphi}\right), t = 1, \dots, T - 1, \text{ or}$$

$$\varphi_t - 2\varphi_{t+1} + \varphi_{t+2} | \tau_\varphi \sim N\left(0, \frac{1}{\tau_\varphi}\right), t = 1, \dots, T - 2, \quad (3)$$

In Equation (1), $g(\cdot)$ denotes a log-Gaussian likelihood function for this case to link Y_{it} and ζ_{it} . η denotes the intercept with fixed effect. SX signifies K main covariates with the spatial nonstationary assumption. TX represents M main covariates that are assumed to be temporally nonstationary. The parameter ω_{ik} is named as space-coefficients (SCs) and φ_{tm} is named time-coefficients (TCs), which are two fundamental outputs of the STVC model. $f_{space}(\cdot)$ and $f_{time}(\cdot)$ signify the spatial and temporal latent Gaussian models (LGMs) that are used for fitting the random effects of spatial and temporal nonstationarity to estimate local parameters SCs and TCs [42,43].

In Equation (2), on account of the spatial LGM $f_{space}(\cdot)$, the prior intrinsic conditional autoregressive (iCAR) model is adopted for fitting the spatial autocorrelation characteristics that are also called the spatial structured random effects within a BHM [44], where ω_{-i} denotes every spatial unit in ω apart from the i -th spatial unit, $W = (w_{ij})$ represents the spatial relation matrix in which $w_{ij} = 1$ if spatial units i and j are neighbors, e.g., spatial adjacency relations here, and $w_{ij} = 0$ otherwise, as well as τ_ω further indicates the precision parameter [45].

In Equation (3), the prior random walk (RW) model is used as the temporal LGM $f_{time}(\cdot)$ to estimate the temporal autocorrelation characteristics of TCs, where the structured temporal random effect of covariates φ can be a random walk of order one or two, with τ_φ being the precision parameter [46]. The prior RW model of order two is more suitable for the research object with a clear linear time trend, compared with the prior RW model of order one.

2.2.3. Model Implementation and Comparison

To explore both the global homogeneous and local heterogeneous impacts of socioeconomic and environmental factors on city-specific outcomes of China's tourism, we implemented four types of Bayesian regressions, the multiple linear regression (MLR, model 1), the ordinary generalized additive model (GAM, model 2), the global spatiotemporal regression (model 3), and the Bayesian STVC model (model 4), which belongs to the local spatiotemporal regression family. We chose these models based on the following considerations. First, model 1 and model 2 were traditional mainstream models. We used them to fit the overall linear and nonlinear impacts of covariates on tourism [47]. Then, we used a widely applied spatiotemporal regression (model 3), which mainly served as a spatiotemporal descriptive tool, to depict the original smoothed spatial variations and temporal trends of tourism in China [42]. However, models 1–3 are regarded as the global-based type of regression, meaning that covariate impacts (coefficients) were homogeneous across space and over time [35]. Given this underlying limitation of stationarity,

model 4 was finally employed to explore the structured heterogeneous (varying) impacts of covariates at both space and time scales [36].

To be specific, the equation of an MLR (model 1) is given by

$$\zeta_{it} = g(Y_{it}) = \eta + \sum_{k=1}^K \chi_k X_{itk}, \quad (4)$$

where χ_k denotes the overall coefficient of the k -th covariate X , which qualifies the linear numerical impacts of explanatory factors on Y_{it} .

An ordinary GAM (model 2) is formulated as [48]

$$\zeta_{it} = g(Y_{it}) = \eta + \sum_{k=1}^K f_{GAM}(\delta_{hk} X_{itk}), \quad (5)$$

where $f_{GAM}(\cdot)$ denotes the nonparametric smooth function for fitting a set of coefficients δ_{hk} with h groups, representing the numerical nonlinear impacts of the k -th covariate. Unlike model 1, model 2 is useful in identifying response–covariate numerical nonlinear relationships. However, both model 1 and model 2 cannot consider the spatiotemporal effects essential for geospatial analysis.

A global spatiotemporal regression (model 3) can be modeled with [42,46,49]

$$\zeta_{it} = g(Y_{it}) = \eta + \sum_{k=1}^K \chi_k X_{itk} + f_{space}(\mu_i) + f_{time}(\lambda_t), \quad (6)$$

where μ_i signifies the space-intercepts (SIs) representing the structured spatial distribution of Y_{it} , λ_t signifies the time-intercepts (TIs) representing the structured temporal trend of Y_{it} , LGMs $f_{space}(\cdot)$ and $f_{time}(\cdot)$ are the same as in Equation (1).

Model 4, as fully introduced in Equations (1)–(3), has been as a reduced Bayesian STVC regression by removing the spatiotemporal random effects of intercepts to ensure noticeable variations of both spatial and temporal nonstationary impacts of different explanatory factors on the target response variable [36,50].

Finally, the optimal model from the above four regressions with the best model fitness and predictability was further utilized to estimate the complete spatiotemporal maps of Y_{it} .

2.2.4. Model Inference and Evaluation

Alternative regression models were established using the Bayesian statistics based on the advanced hierarchical modeling strategy, that is, a BHM framework. Non-informative priors were assigned to parameters within the BHM to embody the idea of data-driven modeling [47]. The integrated nested Laplace approximation (INLA) algorithm, an approximate Bayesian inference technique, was adopted to estimate these regression models using the R-INLA package in the R environment [51] due to its advantage of producing reliable estimated results with a relatively short computation time [52]. The performances of these alternative regressions are evaluated in terms of three aspects, including the degree of model fitting, model complexity, and predictive ability [46]. Specifically, the deviation information criterion (DIC) [53] and the Watanabe–Akaike information criterion (WAIC) are used for reflecting the degree of fitting of the Bayesian regression, for which a smaller value indicates a better model fit. Likewise, the complexity of the Bayesian regression is evaluated with two indices (P_{DIC} and P_{WAIC}) that can be simultaneously obtained via the adoption of both the DIC and WAIC methods, for which smaller values are also reflective of better model performances. In terms of the model predictive power, a logarithmic score (LS) retrieved from the conditional predictive ordinates under a leave-one-out cross-validation is used, with smaller values associated with better predictive capacities [54].

3. Results

3.1. Selected Drivers for Modeling

As indicated in Figure 2, through setting $VIF < 5$ as the inclusion criteria, potential explanatory variables with higher rankings of MDI were selected from the screening outcomes and were added into the regression modeling. To be specific, first, we removed factors with higher multicollinearity based on the exclusion threshold of 5 for VIF, as shown in Figure 2a. This step left 13 socioeconomic factors (i.e., SV1, SV2, SV3, SV6, SV7, SV9, SV10, SV14, SV17, SV18, SV19, SV20, and SV21) and four environmental factors (i.e., EV5, EV6, EV8, and EV9). Further, using Figure 2b, we selected the top eight factors (i.e., EV9, SV21, EV8, EV5, EV6, SV20, SV1, and SV2), which had relative higher importance (contribution) to the response variable. Because the selection of MDI is generally empirical, here, the main reason for our choice of MDI is that there was an apparent bluff trend between the two factors of SV2 and SV7. The top eight factors covered four socioeconomic factors and four environmental factors, which was ideal for exploring the combined impacts of the above two critical aspects on tourism development. Hence, based on the perspectives above, the screening threshold applicable to this case is $MDI > 200$. Summing up the above, a core variables system particularly applicable to China's tourism case was created, which contained a total of eight critical factors (renamed as X1–X8 in Table 3), and was further incorporated into the next-step regression analysis.

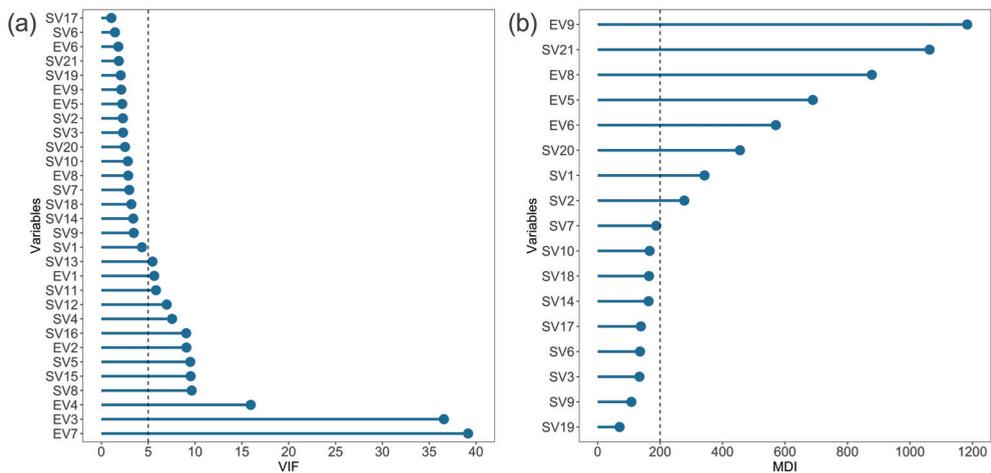


Figure 2. Two-step variables screening procedure: (a) remove variables with higher multicollinearity ($VIF > 5$); (b) select variables with the higher relative importance ($MDI > 200$).

3.2. Model Assessment and Comparison

We assessed the four types of comparative Bayesian regression models' performances by jointly considering model fitness, complexity, and predictive power, for which a total of five representative evaluation indicators are summarized in Table 2. Model 4 (STVC) showed the best performance with the minimum assessment indicators DIC, WAIC, and LS. However, for P_{DIC} and P_{WAIC} , model 4 demonstrated a notable deficiency and it presented a much higher complexity than the other three mainstream benchmark models. The complexity (P_{DIC}) of model 4 was found to be about 99 times higher than that of a multiple linear regression (model 1), which was 2.6 times higher than that of a global spatiotemporal regression (model 3). Two possible reasons were considered for explaining the increased complexity of the STVC model. Specifically, the STVC model demonstrated both superior model fitness and predictive capacity compared with all the other regressions. Moreover, it should be pointed out that the STVC model was the only one that had the

capacity for synchronously detecting both temporal and spatial heterogeneous associations between variables to be further interpreted at a space-time scale. Therefore, model 4 (STVC) was selected as the final regression to explore the spatiotemporal heterogeneous relationships between tourism and eight selected explanatory variables, which was also used for producing a series of estimated spatiotemporal distribution maps reflective of the city-level tourism revenue in China.

Table 2. Bayesian modeling evaluations of the alternative regressions for China’s tourism case account for model fitness, complexity, and predictive power.

Index	DIC	P_{DIC}	LS	WAIC	P_{WAIC}
Model 1	40,238.39	9.02	5.97	40,262.83	30.05
Model 2	39,153.72	64.89	5.81	39,157.01	66.83
Model 3	31,825.74	348.95	4.72	31,848.15	338.66
Model 4	30,307.19	899.21	4.52	30,345.55	761.76

Model 1–4: multiple linear regression, generalized additive model, global spatiotemporal regression, and local spatiotemporal regression STVC model; DIC: deviance information criterion; WAIC: Watanabe–Akaike information criterion; P_{DIC} : effective number of parameters from DIC; P_{WAIC} : effective number of parameters from WAIC; LS: logarithmic score.

3.3. Global-Scale Impacts of Drivers

Two kinds of overall impacts of socioeconomic and environmental variables on tourism were estimated: one was the global-scale linear numerical effects based on model 1; the other one was the global-scale nonlinear numerical effects obtained from model 2. We summarized the critical parameters of model 1 in Table 3, including the overall coefficients representing the stationary relationship among variables, standard deviation (SD), and the 2.5% and 97.5% quantiles of Bayesian credible intervals (CIs). In terms of the four socioeconomic variables, X1 and X2 reflected the income level of individual residents, X3 represented the regional macroeconomic development conditions, and X4 represented the population condition. For the other four environmental variables, X5 and X7 represented the city-specific urbanization process and vegetation coverage based on satellite remote sensing data, respectively. X6 and X8 reflected the general geographical situations characterized by topography and transportation, respectively. Except for X2 and X5, the overall coefficients of the other six factors were found to be greater than zero. This finding indicated that most core variables served as positive stimulants for tourism development from a global-scale perspective. Notably, the NDVI (X7), the average wage of employees in urban units (X1), GDP per capita (X3), and road network density (X8) demonstrated more significant impacts on tourism among the eight factors.

Table 3. Linear numerical impacts of main drivers on China’s city-level tourism industry.

Variables	Socioeconomic and Environmental Aspects	Mean	SD	Q 0.025	Q 0.975
X1	Average wage of employed persons in urban units	0.4694	0.0212	0.4278	0.5110
X2	Employment density of urban units	−0.1428	0.0170	−0.1761	−0.1095
X3	GDP per capita	0.4660	0.0258	0.4154	0.5165
X4	Population density	0.2136	0.0282	0.1582	0.2689
X5	Nighttime light index	−0.0141	0.0310	−0.0750	0.0467
X6	Slope	0.1013	0.0202	0.0615	0.1410
X7	Normalized Difference Vegetation Index (NDVI)	0.6630	0.0187	0.6263	0.6996
X8	Road network density	0.3382	0.0226	0.2937	0.3826

Furthermore, the exponent-scale nonlinear numerical effects of the eight selected drivers were illustrated in Figure 3. We noticed that all the variables’ numerical influencing curves had a similar upward trend. At the same time, we identified the varying impacts of each variable across their development process. It is worth mentioning that X2 and X5 were negatively linearly correlated with the tourism industry, which could not be explained

directly. While by further analyzing the nonlinear results, only X2 and X5 appeared to have a significant downward trend, leading to the overall negative linear association in Table 3. This finding proved that model 2 had a superior interpretation capacity over model 1 in fitting global-scale numerical impacts. However, both linear and nonlinear numerical modeling results were produced based on a stationary assumption. As a result, these global-scale outputs might smooth or hide the local-scale heterogeneous impacts of different variables on the tourism industry over the entire study area and time frame, particularly for a fine-scale space–time dataset.

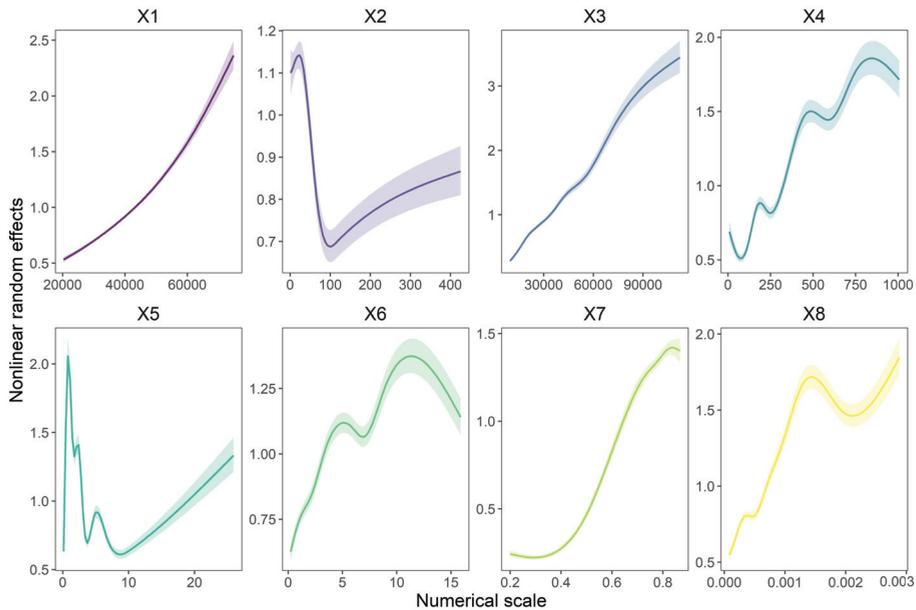


Figure 3. Global-scale nonlinear numerical effects of main drivers on China’s city-level tourism industry: X1—average wage of employed persons in urban units, X2—employment density of urban units, X3—GDP per capita, X4—population density, X5—nighttime light index, X6—slope, X7—NDVI, and X8—road network density.

3.4. Temporally Varying Impacts of Drivers

In Figure 4, we presented a TIs graph and five TCs graphs with 95% Bayesian CIs, to exhibit the crude temporal dynamic trend of tourism and the temporally heterogeneous impacts of main drivers on tourism in China, as well as the uncertainties of these estimated parameters. According to Figure 4a, China’s tourism development level demonstrated a continuously increasing trend from 2008 to 2017, meaning that China’s tourism industry maintained a high development speed spanning ten years. Furthermore, it can be seen from Figure 4b that the temporal tourism–covariates relationships varied non-linearly over periods was an essential feature of the Bayesian STVC model that cannot be facilitated via the adoption of global-scaled coefficients. Generally speaking, X3 (GDP per capita), X4 (population density), and X7 (NDVI) showed a downward trend from 2008 to 2017, which indicated a strong to weak impact of these variables on tourism development over time. While X1 (average wage of employed persons in urban units) and X2 (employment density of urban units) presented an initial downtrend, followed by an upward tendency starting from 2013, suggesting that their roles in promoting tourism development gradually especially after 2013. These findings also meant that groups with high quality of living might be a potentially vital force to promote tourism. In addition, we noticed that the TCs

of X1 and X2 had relatively higher uncertainties (CIs) due to the fluctuation ranges of TCs of X1 (-0.05 – 0.05) and X2 (-0.02 – 0.02) being much narrower compared with those of the other indicators, e.g., X4 (-0.4 – 0.2). In fact, when we plotted all the factors within a single graph using the same vertical coordinate, the uncertainties (CIs) of X1 and X2 turned out to be very small; however, the time trends of X1 and X2 could be smoothed out. From this perspective, the uncertainties of all indicators were within an acceptable range.

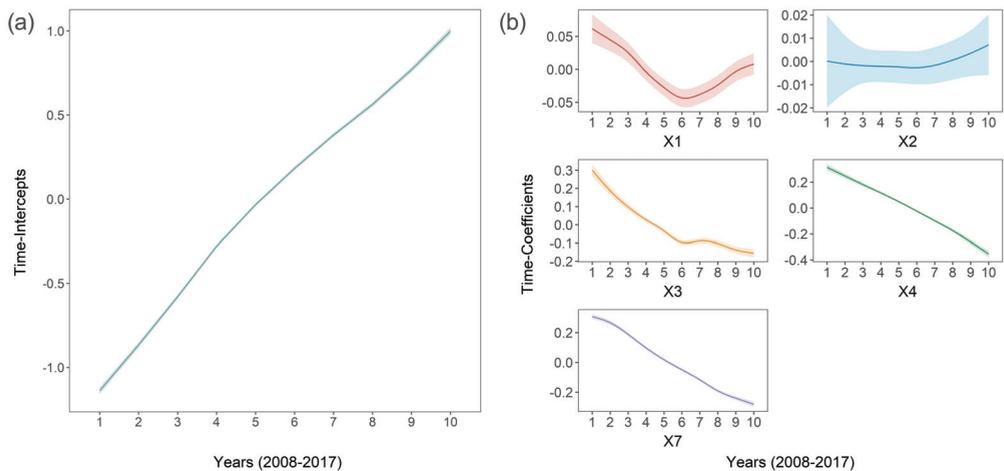


Figure 4. (a) Time-intercepts (TIs) graph: the temporal trend of China’s tourism industry from 2008 to 2017, and (b) time-coefficients (TCs) graphs (covariates’ temporal nonstationarity): the impacts of drivers (X1–X4 and X7) on tourism are varying over ten years. Covariates with time-scale variations: X1—average wage of employed persons in urban units, X2—employment density of urban units, X3—GDP per capita, X4—population density, and X7—NDVI. Shadow areas in each facet are the 95% Bayesian CIs to describe the uncertainties of time-scale parameters.

3.5. Spatially Varying Impacts of Drivers

Spatially, we retrieved the parameters of SIs from model 3 to geographically map the ten-year average tourism revenue distribution across China, as presented in Appendix A Figure A1. In addition, utilizing the SCs parameters from model 4, the variables’ spatial nonstationary maps were depicted in Figure 5a. The cluster maps for parameter SCs were also produced to highlight those significant (>90% confidence) hot spots and cold spots at the city level, as shown in Figure 5b.

From Figure A1, the spatial distribution characteristics of China’s tourism revenue demonstrated a gradual increase from West China to East China. Simultaneously, we detected diverse geospatial tourism–covariates relationships at the city level from Figure 5a and an apparent spatial agglomeration effect of SC maps in Figure 5b. In fact, for every single factor of interest, city-specific areas with higher sensitivity to this particular covariate could be visually identified in terms of achieving regional tourism development, based on direct analysis of the corresponding SC map produced by that covariate. Furthermore, within each city area, a series of urban policies could be proposed to facilitate its tourism development based on the relative impacts of the eight-core variables. The relative effect of each covariate within each city could be assessed by vertically integrating the local-scale information from all the SC maps together [50].

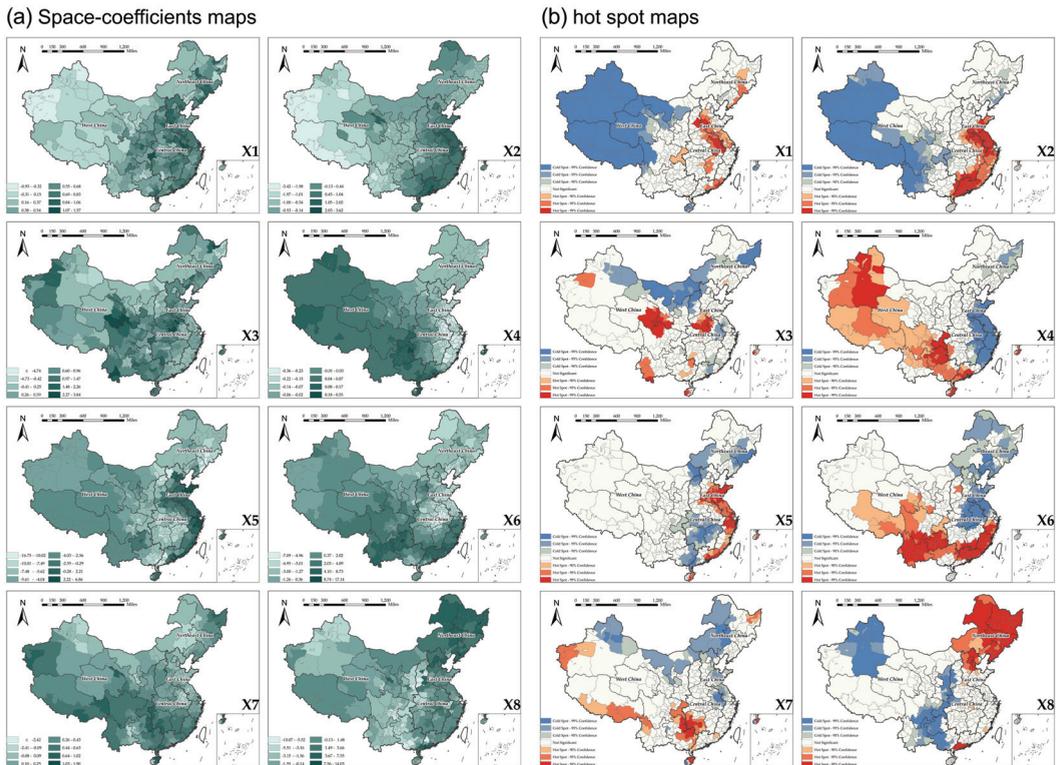


Figure 5. (a) Space-coefficients (SCs) maps (covariates' spatial nonstationarity): spatially varying impacts of main drivers (X1–X8) on total tourism revenue at city level across China, and (b) hot spot analysis for SCs maps: X1—average wage of employed persons in urban units, X2—employment density of urban units, X3—GDP per capita, X4—population density, X5—nighttime light index, X6—slope, X7—NDVI, and X8—road network density.

Looking at the macroscopic regional scale using the hot spot maps in Figure 5b, we may discover that: in Northeast China, X8 may serve as an essential factor for promoting the development of the local tourism industry, while X5 and X6 may have no impacts, and the other factors may also have an individual city-specific impact yet without generating geographic hotspot regions in the past. Likewise, the high-level tourism development in China's eastern region may be mainly promoted by X2 and X5; and X4 are not entirely essential. In Western China, with low-level tourism development, X4 may be a primary determinant to improve its tourism conditions. Meanwhile, X6 and X7 also present spatially positive clustered effects in some areas, such as Yunnan and Sichuan. In the regions of Central China, tourism development seems to be dominated by socioeconomic factors, including X1, X2, and X3.

3.6. Spatiotemporal Estimated Maps of China's City-Level Tourism Revenue

A complete series of spatiotemporal distribution maps of China's city-specific tourism development level from 2008 to 2017 was produced by adopting the optimal Bayesian STVC model (model 4), as shown in Figure 6. The newly model-estimated tourism maps highlighted hidden areas (e.g., cities with missing values) and provided more intuitive information (e.g., smooth the city-level extreme outliers), which were expected to assist in making policies about the sustainable development of tourism. Generally, the overall growth in the city-level tourism industry was identified over the past decade in China, during which time diverse improvement intensities were found among regions at a local

city scale. In Central China, since 2008, about 77% of blue-colored cities with weak tourism industries gradually shifted to yellow/red colors with relatively strong tourism industries. In contrast, such a shifting proportion of East China was about 55%, suggesting that Central China's tourism industry grew faster than that of eastern cities. In terms of West China and Northeast China, the shifting proportions were about 51% and 41%, respectively, which were relatively lower than the other two divisions. In 2017, the low-tourism-level cities with a blue color were mainly distributed in the provinces of Heilongjiang, Gansu, Ningxia, Xinjiang, Qinghai, and Tibet.

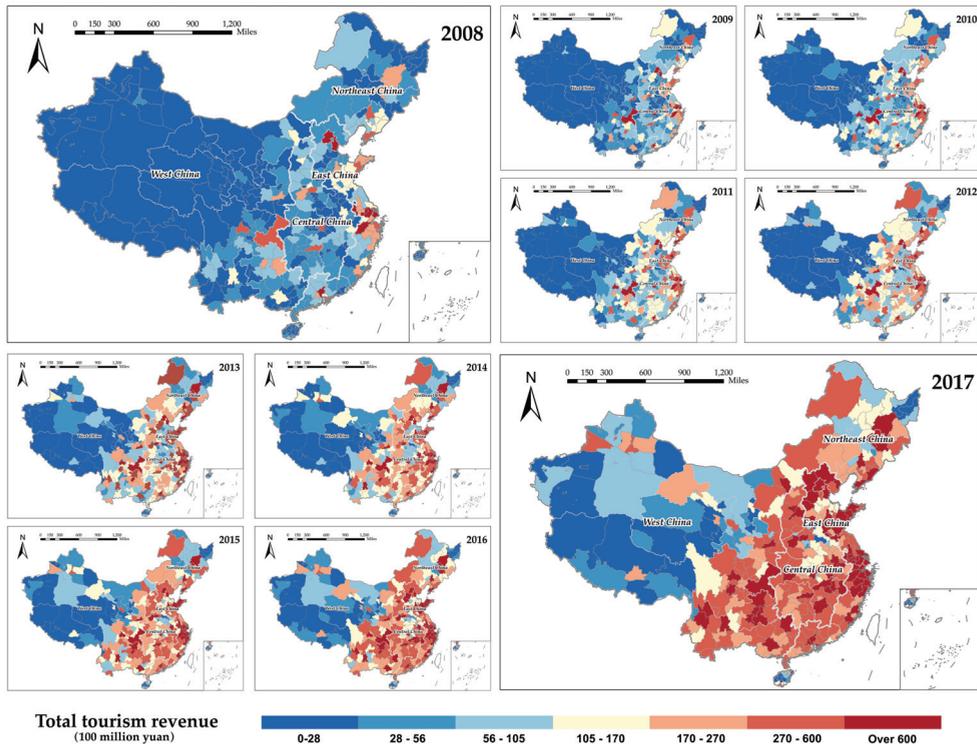


Figure 6. Estimated spatiotemporal maps for showing dynamic variations of city-level tourism development across China from 2008 to 2017.

Lastly, we performed a hot spot analysis for the newly estimated complete tourism maps in 2007 and 2018, respectively, to detect those significant city clusters (>90% confidence) of the tourism industry, as shown in Figure 7. In 2017, we found four significant tourism industry clusters (hot spots) and one less-developed tourism region (cold spot) at the city level, compared with 2008 with two clearly formed hot spots of the tourism industry. These four identified high-level tourism city clusters in 2017 were demonstrated to be consistent with China's top four major urban agglomerations, namely, Beijing–Tianjin–Hebei, the Yangtze River Delta, the Pearl River Delta, and the Sichuan–Chongqing Region. This might reveal that current tourism agglomeration development is closely related to the urbanization degree. Meanwhile, a cold spot was detected in West China, indicating that the tourism development of western cities was relatively slow.

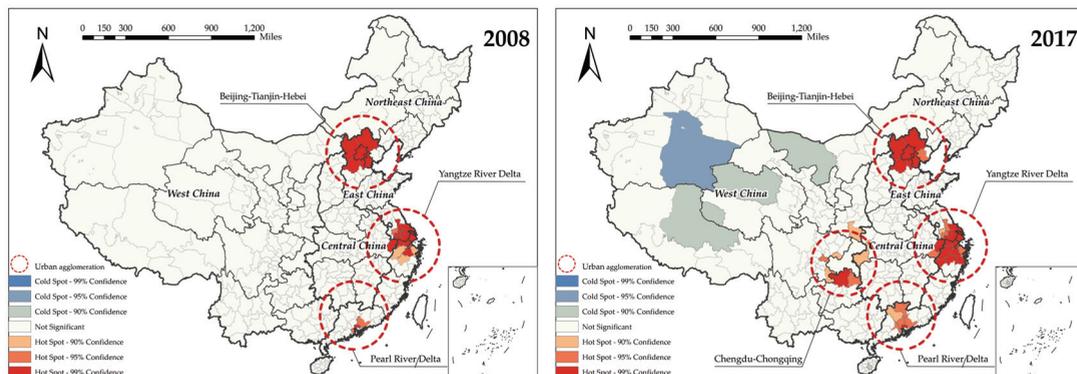


Figure 7. The urban agglomeration of the tourism industry across Chinese cities in 2008 and 2017: hot spot mapping for the areas' total tourism revenue, estimated by the optimal Bayesian STVC model.

4. Discussion

In this study, the multidimensional impacts of socioeconomic and environmental variables, including linear and nonlinear numerical effects and spatiotemporal heterogeneous effects, on regional tourism were comprehensively investigated across Chinese cities along with the first production of a set of spatiotemporal maps depicting China's total tourism revenue. These findings may add innovative insights about the mechanisms of how multi-source geospatial factors have affected the regional tourism industry, and is expected to provide a brand-new viewpoint for policymakers. According to different scales, we have some conclusions, as follows.

Globally, significant effects of both socioeconomic and environmental variables were identified [28,55–57], which highlighted the necessity of taking a wide range of factors into accounts throughout the procedure of tourism policies formulation. Tourism is a comprehensive industry composed of multiple elements, including food, shelter, transportation, travel, entertainment, and purchase. However, the importance of some of these elements embedded in the tourism industry, such as food, shelter, and transportation, is always ignored for the reason that they are simply regarded as the basic service facilities of a city. Therefore, the positive effects of the socioeconomic and environmental factors on tourism are supposed to be focused on the industrial level, which suggests that the idea of developing industries should always be adopted as the guideline for developing the tourism industry regardless of regional or national levels. At present, the “Travel +” strategy being implemented by the Chinese government is exactly based on this idea [58].

Temporally, the development of China's tourism has mainly benefited from comprehensive time-scale impacts of multiple factors. Based on temporal nonstationarity, the predominant stimulants for tourism development were demonstrated to have gradually switched from the regional economy, populational size, and tourism resource attractiveness to personal economic status. These results implied that China's current tourism industry demonstrated a new feature that a transition from sightseeing tourism to leisure and holiday tourism is very much likely to occur. Meanwhile, residents' affluence has been highlighted as an indispensable contributor to nationwide tourism development [59]. Under such a changing background of the tourism industry in China, it is highly suggested that improving personal income, as well as safeguarding the rights and interests of employees, should be adopted as an essential strategy for facilitating the nationwide tourism industry development, which might be achieved via the implementation of multiple tourism-related policies at governmental levels, such as approving paid-leave policies for employees, encouraging enhanced flexibilities of work schedules to be tailored for vocational leaves, as well as encouraging off-peak vocational arrangements.

Spatially, the development of China's tourism could be characterized as "strong in the east and weak in the west" [30], which was affected by various factors. Cities of West China were mainly affected by population size and tourism resources, while personal income, employment and urbanization had more contributions to cities in the east region [60]. The city-level spatial nonstationarity found in this study could serve as an acceptable reference in the procedure of making more targeted policies by governments at all levels. For example, the western region may put forward corresponding talent introduction policies while promoting economic development. In addition, the local government can develop sightseeing and holiday tourism through developing natural landscapes. Cities of East China need to focus on optimizing the protection system of workers' rights and interests and developing characteristic tourism products to provide tourists with high-end, comfortable, and personalized services for stimulating tourism. Northeast China may focus on infrastructure and strengthen the planning and laying of the road networks to enhance regional tourism accessibility. Furthermore, city-level local authorities could utilize local resources rationally and determine the direction of tourism strategies by using the critical drivers' local spatial influencing maps to support ecotourism, sightseeing tourism, vacation tourism, geological tourism, and urban tourism. In addition, the first series of maps displaying China's tourism revenue's spatiotemporal distributions at an administrative city level from 2008 to 2017 was produced, which was further analyzed to provide urbanization-related insights into empirically optimizing the unbalanced development of the tourism industry [61,62].

To sum up, from the multidimensional spatiotemporal heterogeneous perspective, the government should formulate various tourism policies based on region-specific conditions, as well as pursue the development concept of "applying proper measurements in line with local conditions and temporal variations". At present, tourism industry development in areas with relatively high urbanization levels has demonstrated a change from sightseeing tourism to leisure tourism. As a result, socioeconomic status should be continuously considered as a significant factor throughout tourism-related policy-making procedures in these regions. In contrast, regarding cities with low-level urbanization distributed in West China, environmental factors or sightseeing resources, instead of other factors, should be addressed as predominant issues to be considered throughout the formulation of tourism-related policies [60]. Therefore, making city-specific strategies that take city-specific factors into account is expected to improve the accuracy of policy formulation, as well as the effectiveness of strategic implementation, which would further mitigate both "invalid policy" and "weak policy" produced by the "one-size-fits-all" policy.

Finally, we would like to underline the importance of the local spatiotemporal regression approach, namely, the Bayesian STVC model we have selected. As discussed above, introducing a spatiotemporal heterogeneous perspective to regional tourism management could avoid the one-size-fits-all issue via providing multidimensional spatiotemporal information. In the spatial statistics field, local regressions that can deal with such spatiotemporal heterogeneity among variables relationships (spatiotemporal nonstationarity) are relatively rare, which can be generally classified into the frequentist-type model [63–65] and the Bayesian-type model [35,36,66], as they were proposed independently under different statistical traditions. The main reasons we chose the Bayesian STVC model as the applied local spatiotemporal regression lie in the following considerations. First, only the Bayesian-based local spatial or spatiotemporal model is an actual "full-map" modeling technique; thus, the results are more reliable [67,68]. Second, the Bayesian STVC model follows a space–time independent nonstationary assumption, dramatically reducing the computational burden and weakening the overfitting problem. Last but not least, due to its separately fitting of space-coefficients (SCs) and time-coefficients (TCs), the Bayesian STVC model is more user-friendly: stakeholders can directly separately obtain the spatial and temporal autocorrelated regularities [36,50]. Beyond these benefits, the Bayesian STVC model still needs further improvement to solve more complex space–time interaction issues in natural and social sciences.

5. Conclusions

This study verifies that socioeconomic and environmental factors simultaneously affect tourism development over China, globally and locally, supported by the up-to-date space-time data of city-level tourism statistics and a series of advanced Bayesian regressions. Remarkably, the local impacts of socioeconomic and environmental conditions vary heterogeneously at the city level in both time and space dimensions across China, and was demonstrated by adopting the cutting-edge Bayesian STVC model, which was also used for estimating the first series of spatiotemporal maps of city-level tourism development. These fruitful findings provide novel insights into policy-making procedures at multiple levels. Here, the Bayesian STVC model was successfully applied to mine the spatial and temporal autocorrelated nonstationarity inherent in tourism–covariates relationships over China and could serve as an emerging tool to offer new insights on spatiotemporal-oriented influencing factor analysis and high-precision prediction in broader GIScience-related fields of social and natural sciences.

Apart from all these achievements, several concerns should be better addressed in future lines of research. First, the seasonal effect is the main factor affecting tourists' behavior [69], which emphasizes collecting and using quarterly tourism data in tourism research. However, this study is limited because national urban tourism data sources only have annual scale records. Second, other underlying tourism-related factors such as tourism resources were not fully considered in this study [34]. Future studies might focus on a relatively small area with seasonal heterogeneity by using multi-source tourism data to construct more scientific indicators [70] and developing more sophisticated spatiotemporal statistical models for outputting more informative results for regional tourism research.

Author Contributions: Conceptualization, Chengwu Wang and Chao Song; data curation, Xu Zhang, Zhangying Tang and Honghu Tang; formal analysis, Xu Zhang, Chao Song, and Chengwu Wang; funding acquisition, Chao Song, Chengwu Wang and Zhangying Tang; investigation, Chengwu Wang and Xu Zhang; methodology, Chao Song and Zhoupeng Ren; project administration, Chao Song; software, Chao Song and Mingyu Xie; visualization, Xu Zhang, Chao Song, Mingyu Xie and Honghu Tang; writing—original draft, Xu Zhang; writing—review and editing, Chao Song, Chengwu Wang and Yili Yang. All authors have read and agreed to the published version of the manuscript.

Funding: This work was collaboratively supported by the National Natural Science Foundation of China (Grant No. 42071379 and 41701448), the State Key Laboratory of Resources and Environmental Information System (Grant No. 201811), the Applied Basic Research Funded Project of Sichuan Science and Technology Department (Grant No. 2020YJ0117), the Humanities and Social Sciences Research Fund Project of Southwest Petroleum University (Grant No. 2019RW021), the Chengdu Federation of Social Science Association (Grant No. 2021ZC003), and the Fund for Introducing Talents of Sichuan University (Grant No. YJ202157). The funders had no conflicts in study design, data collection and analysis, decision to publish, or manuscript preparation.

Data Availability Statement: Publicly available datasets were analyzed in this study. The tourism and socioeconomic data can be found in China City Statistical Yearbook and Statistical Bulletin. The environmental data can be found here: <http://data.cma.cn/> (accessed on 28 April 2021) and <http://www.resdc.cn/> (accessed on 28 April 2021).

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

Appendix A

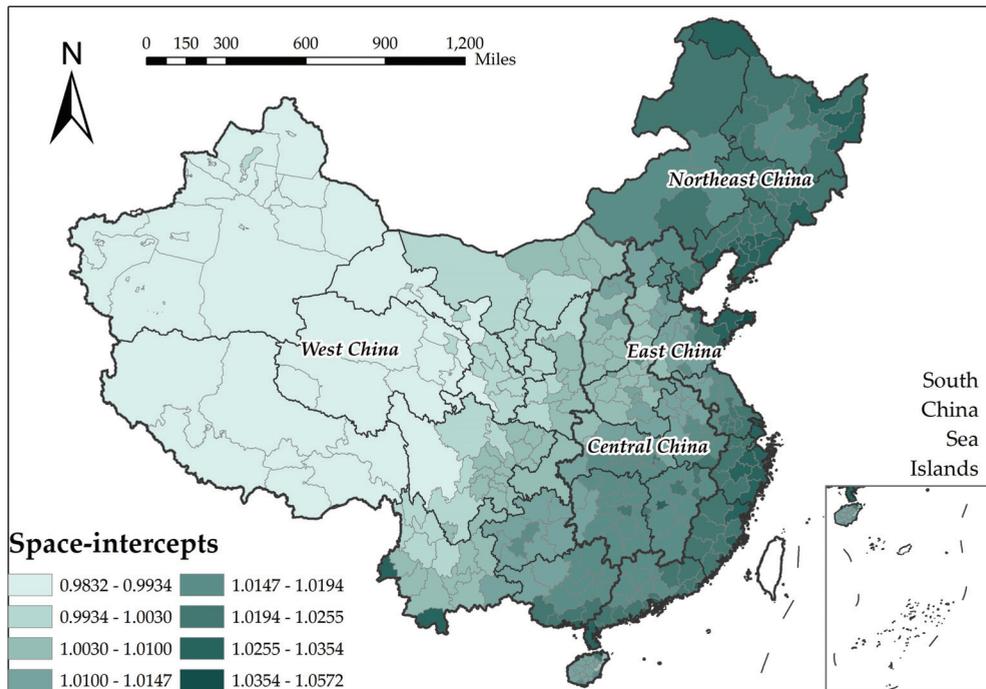


Figure A1. Space-Intercepts (SIs) map: the model-estimated geographical variations of China's ten-year average total tourism revenue at the city level during 2008–2017.

References

- Fayissa, B.; Nsiah, C.; Tadasse, B. Impact of tourism on economic growth and development in Africa. *Tour. Econ.* **2008**, *14*, 807–818. [[CrossRef](#)]
- Travel & Tourism Global Economic Impact & Issues 2017. Available online: <https://wtcc.org/Portals/0/Documents/Reports/2020/Global%20Economic%20Impact%20Trends%202020.pdf?ver=2021-02-25-183118-360> (accessed on 6 June 2021).
- Modeste, N.C. The impact of growth in the tourism sector on economic development: The experience of selected Caribbean countries. *Econ. Internazionale/Int. Econ.* **1995**, *48*, 375–385.
- Proença, S.; Soukiazis, E. Tourism as an economic growth factor: A case study for Southern European countries. *Tour. Econ.* **2008**, *14*, 791–806. [[CrossRef](#)]
- Ma, T.; Hong, T.; Zhang, H. Tourism spatial spillover effects and urban economic growth. *J. Bus. Res.* **2015**, *68*, 74–80. [[CrossRef](#)]
- Coccosis, H. Tourism and sustainability: Perspectives and implications. In *Sustainable Tourism? European Experiences*; CAB INTERNATIONAL: Wallingford, UK, 1996; pp. 1–21.
- Joshi, O.; Poudyal, N.C.; Larson, L.R. The influence of sociopolitical, natural, and cultural factors on international tourism growth: A cross-country panel analysis. *Environ. Dev. Sustain.* **2017**, *19*, 825–838. [[CrossRef](#)]
- Khan, A.; Bibi, S.; Ardito, L.; Lyu, J.; Hayat, H.; Arif, A.M. Revisiting the Dynamics of Tourism, Economic Growth, and Environmental Pollutants in the Emerging Economies—Sustainable Tourism Policy Implications. *Sustainability* **2020**, *12*, 2533. [[CrossRef](#)]
- Tuckova, Z.; Sverak, P. Impact of the Regional Macroeconomics Indicators on Tourism Entities in Plzen and Zlin Regions. *Procedia Econ. Financ.* **2016**, *39*, 313–318. [[CrossRef](#)]
- Jarvis, D.; Stoeckl, N.; Liu, H.-B. The impact of economic, social and environmental factors on trip satisfaction and the likelihood of visitors returning. *Tour. Manag.* **2016**, *52*, 1–18. [[CrossRef](#)]
- Jayaraman, K.; Lin, S.K.; Haron, H.; Ong, W.L. Macroeconomic factors influencing Malaysian tourism revenue, 2002–2008. *Tour. Econ.* **2011**, *17*, 1347–1363. [[CrossRef](#)]

12. Alonso-Almeida, M.d.M.; Borrajo-Millán, F.; Yi, L. Are social media data pushing overtourism? The case of Barcelona and Chinese tourists. *Sustainability* **2019**, *11*, 3356. [[CrossRef](#)]
13. Field, C.B.; Barros, V.R. *Climate Change 2014-Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects: Volume 1, Global and Sectoral Aspects: Working Group II Contribution to the IPCC Fifth Assessment Report*; Cambridge University Press: Cambridge, UK, 2014.
14. Hewer, M.J.; Gough, W.A. Thirty years of assessing the impacts of climate change on outdoor recreation and tourism in Canada. *Tour. Manag. Perspect.* **2018**, *26*, 179–192. [[CrossRef](#)]
15. Martín, M.B.G. Weather, climate and tourism a geographical perspective. *Ann. Tour. Res.* **2005**, *32*, 571–591. [[CrossRef](#)]
16. Day, J.; Chin, N.; Sydnor, S.; Cherkauer, K. Weather, climate, and tourism performance: A quantitative analysis. *Tour. Manag. Perspect.* **2013**, *5*, 51–56. [[CrossRef](#)]
17. Smith, T.; Fitchett, J.M. Drought challenges for nature tourism in the Sabi Sands Game Reserve in the eastern region of South Africa. *Afr. J. Range Forage Sci.* **2020**, *37*, 107–117. [[CrossRef](#)]
18. Atzori, R.; Fyall, A.; Miller, G. Tourist responses to climate change: Potential impacts and adaptation in Florida's coastal destinations. *Tour. Manag.* **2018**, *69*, 12–22. [[CrossRef](#)]
19. Susanto, J.; Zheng, X.; Liu, Y.; Wang, C. The impacts of climate variables and climate-related extreme events on island country's tourism: Evidence from Indonesia. *J. Clean. Prod.* **2020**, *276*, 124204. [[CrossRef](#)]
20. Kanwal, S.; Rasheed, M.I.; Pitafi, A.H.; Pitafi, A.; Ren, M. Road and transport infrastructure development and community support for tourism: The role of perceived benefits, and community satisfaction. *Tour. Manag.* **2020**, *77*, 104014. [[CrossRef](#)]
21. Li, L.S.; Yang, F.X.; Cui, C. High-speed rail and tourism in China: An urban agglomeration perspective. *Int. J. Tour. Res.* **2019**, *21*, 45–60. [[CrossRef](#)]
22. Barkauskas, V.; Barkauskienė, K.; Jasinskas, E. Analysis of macro environmental factors influencing the development of rural tourism: Lithuanian case. *Procedia Soc. Behav. Sci.* **2015**, *213*, 167–172. [[CrossRef](#)]
23. Marrocu, E.; Paci, R.; Zara, A. Micro-economic determinants of tourist expenditure: A quantile regression approach. *Tour. Manag.* **2015**, *50*, 13–30. [[CrossRef](#)]
24. Li, M.; Fang, L.; Huang, X.; Goh, C. A spatial-temporal analysis of hotels in urban tourism destination. *Int. J. Hosp. Manag.* **2015**, *45*, 34–43. [[CrossRef](#)] [[PubMed](#)]
25. Gutiérrez, J.; García-Palomares, J.C.; Romanillos, G.; Salas-Olmedo, M.H. The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. *Tour. Manag.* **2017**, *62*, 278–291. [[CrossRef](#)]
26. Xu, D.; Huang, Z.; Hou, G.; Zhang, C. The spatial spillover effects of haze pollution on inbound tourism: Evidence from mid-eastern China. *Tour. Geogr.* **2020**, *22*, 83–104. [[CrossRef](#)]
27. Brunson, C.; Fotheringham, S.; Charlton, M. Geographically weighted regression. *J. R. Stat. Soc. Ser. D* **1998**, *47*, 431–443. [[CrossRef](#)]
28. Yang, Y.; Fik, T. Spatial effects in regional tourism growth. *Ann. Tour. Res.* **2014**, *46*, 144–162. [[CrossRef](#)]
29. Shabrina, Z.; Buyuklieva, B.; Ng, M.K.M. Short-Term Rental Platform in the Urban Tourism Context: A Geographically Weighted Regression (GWR) and a Multiscale GWR (MGWR) Approaches. *Geogr. Anal.* **2020**. [[CrossRef](#)]
30. Lu, L.; Yu, F. A study on the spatial characteristic of provincial difference of tourism economy. *Econ. Geogr.* **2005**, *3*, 406–410.
31. Li, H.; Chen, J.L.; Li, G.; Goh, C. Tourism and regional income inequality: Evidence from China. *Ann. Tour. Res.* **2016**, *58*, 81–99. [[CrossRef](#)]
32. Yang, Y. Agglomeration density and tourism development in China: An empirical research based on dynamic panel data model. *Tour. Manag.* **2012**, *33*, 1347–1359. [[CrossRef](#)]
33. Gan, C.; Voda, M.; Wang, K.; Chen, L.; Ye, J. Spatial network structure of the tourism economy in urban agglomeration: A social network analysis. *J. Hosp. Tour. Manag.* **2021**, *47*, 124–133. [[CrossRef](#)]
34. Wang, D.; Ap, J. Factors affecting tourism policy implementation: A conceptual framework and a case study in China. *Tour. Manag.* **2013**, *36*, 221–233. [[CrossRef](#)]
35. Song, C.; Shi, X.; Bo, Y.; Wang, J.; Wang, Y.; Huang, D. Exploring spatiotemporal nonstationary effects of climate factors on hand, foot, and mouth disease using Bayesian Spatiotemporally Varying Coefficients (STVC) model in Sichuan, China. *Sci. Total Environ.* **2019**, *648*, 550–560. [[CrossRef](#)] [[PubMed](#)]
36. Song, C.; Shi, X.; Wang, J. Spatiotemporally Varying Coefficients (STVC) model: A Bayesian local regression to detect spatial and temporal nonstationarity in variables relationships. *Ann. GIS* **2020**, *1–15*. [[CrossRef](#)]
37. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
38. Vatcheva, K.P.; Lee, M.; McCormick, J.B.; Rahbar, M.H. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology* **2016**, *6*, 227. [[CrossRef](#)] [[PubMed](#)]
39. Song, C.; He, Y.; Bo, Y.; Wang, J.; Ren, Z.; Guo, J.; Yang, H. Disease relative risk downscaling model to localize spatial epidemiologic indicators for mapping hand, foot, and mouth disease over China. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 1815–1833. [[CrossRef](#)]
40. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
41. Genuer, R.; Poggi, J.-M.; Tuleau-Malot, C. Variable selection using random forests. *Pattern Recognit. Lett.* **2010**, *31*, 2225–2236. [[CrossRef](#)]
42. Blangiardo, M.; Cameletti, M. *Spatial and Spatio-Temporal Bayesian Models with R-INLA*; John Wiley & Sons: Hoboken, NJ, USA, 2015.

43. Martínez-Bello, D.A.; López-Quílez, A.; Torres Prieto, A. Spatio-temporal modeling of Zika and dengue infections within Colombia. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1376. [[CrossRef](#)]
44. Lindgren, F.; Rue, H. Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* **2015**, *63*, 1–25. [[CrossRef](#)]
45. Besag, J. Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. Ser. B* **1974**, *36*, 192–225. [[CrossRef](#)]
46. Blangiardo, M.; Cameletti, M.; Baio, G.; Rue, H. Spatial and spatio-temporal models with R-INLA. *Spat. Spatio Temporal Epidemiol.* **2013**, *4*, 33–49. [[CrossRef](#)]
47. Wang, X.; Yue, Y.; Faraway, J.J. *Bayesian Regression Modeling with INLA*; Chapman and Hall/CRC: London, UK, 2018.
48. Hastie, T.J.; Tibshirani, R.J. *Generalized Additive Models*; CRC Press: Boca Raton, FL, USA, 1990; Volume 43.
49. Ugarte, M.D.; Adin, A.; Goicoa, T.; Militino, A.F. On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Stat. Methods Med Res.* **2014**, *23*, 507–530. [[CrossRef](#)]
50. Song, C.; Wang, Y.; Yang, X.; Yang, Y.; Tang, Z.; Wang, X.; Pan, J. Spatial and Temporal Impacts of Socioeconomic and Environmental Factors on Healthcare Resources: A County-Level Bayesian Local Spatiotemporal Regression Modeling Study of Hospital Beds in Southwest China. *Int. J. Environ. Res. Public Health* **2020**, *17*, 5890. [[CrossRef](#)]
51. Rue, H.; Martino, S.; Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B* **2009**, *71*, 319–392. [[CrossRef](#)]
52. Rue, H.; Riebler, A.; Sørbye, S.H.; Illian, J.B.; Simpson, D.P.; Lindgren, F.K. Bayesian computing with INLA: A review. *Annu. Rev. Stat. Its Appl.* **2017**, *4*, 395–421. [[CrossRef](#)]
53. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B* **2002**, *64*, 583–639. [[CrossRef](#)]
54. Held, L.; Schrödle, B.; Rue, H. Posterior and cross-validators predictive checks: A comparison of MCMC and INLA. In *Statistical Modelling and Regression Structures*; Springer: Cham, Switzerland, 2010; pp. 91–110.
55. Wang, J.; Huang, X.; Gong, Z.; Cao, K. Dynamic assessment of tourism carrying capacity and its impacts on tourism economic growth in urban tourism destinations in China. *J. Destin. Mark. Manag.* **2020**, *15*, 100383. [[CrossRef](#)]
56. Gu, M.; Wong, P.P. Residents' perception of tourism impacts: A case study of homestay operators in Dachangshan Dao, North-East China. *Tour. Geogr.* **2006**, *8*, 253–273. [[CrossRef](#)]
57. Yin, P.; Lin, Z.; Prideaux, B. The impact of high-speed railway on tourism spatial structures between two adjoining metropolitan cities in China: Beijing and Tianjin. *J. Transp. Geogr.* **2019**, *80*, 102495. [[CrossRef](#)]
58. Li, J. Research on the Development of Coastal Rural Tourism in China from the Perspective of Industrial Integration. *J. Coast. Res.* **2020**, *115*, 271–273. [[CrossRef](#)]
59. Lin, V.S.; Qin, Y.; Li, G.; Wu, J. Determinants of Chinese households' tourism consumption: Evidence from China Family Panel Studies. *Int. J. Tour. Res.* **2020**. [[CrossRef](#)]
60. Pratt, S. Potential economic contribution of regional tourism development in China: A comparative analysis. *Int. J. Tour. Res.* **2015**, *17*, 303–312. [[CrossRef](#)]
61. Fang, L. Empirical Research on the Impact of Urbanization on Regional Tourism Economy in China Basing on Panel Data. In Proceedings of the 2018 9th International Conference on E-Business, Management and Economics, Waterloo, ON, Canada, 2–4 August 2018; pp. 101–105.
62. Luo, J.M.; Qiu, H.; Goh, C.; Wang, D. An analysis of tourism development in China from urbanization perspective. *J. Qual. Assur. Hosp. Tour.* **2016**, *17*, 24–44. [[CrossRef](#)]
63. Fotheringham, A.S.; Crespo, R.; Yao, J. Geographical and temporal weighted regression (GTWR). *Geogr. Anal.* **2015**, *47*, 431–452. [[CrossRef](#)]
64. Huang, B.; Wu, B.; Barry, M. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 383–401. [[CrossRef](#)]
65. Wu, C.; Ren, F.; Hu, W.; Du, Q. Multiscale geographically and temporally weighted regression: Exploring the spatiotemporal determinants of housing prices. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 489–511. [[CrossRef](#)]
66. Jiang, H.; Hu, H.; Li, B.; Zhang, Z.; Wang, S.; Lin, T. Understanding the non-stationary relationships between corn yields and meteorology via a spatiotemporally varying coefficient model. *Agric. For. Meteorol.* **2021**, *301*, 108340. [[CrossRef](#)]
67. Wolf, L.J.; Oshan, T.M.; Fotheringham, A.S. Single and multiscale models of process spatial heterogeneity. *Geogr. Anal.* **2018**, *50*, 223–246. [[CrossRef](#)]
68. Finley, A.O. Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods Ecol. Evol.* **2011**, *2*, 143–154. [[CrossRef](#)]
69. E Silva, F.B.; Herrera, M.A.M.; Rosina, K.; Barranco, R.R.; Freire, S.; Schiavina, M. Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tour. Manag.* **2018**, *68*, 101–115. [[CrossRef](#)]
70. Duca, A.L.; Marchetti, A. Open data for tourism: The case of Tourpedia. *J. Hosp. Tour. Technol.* **2019**, *10*, 351–368.

Article

Spatial Distribution Pattern and Influencing Factors of Sports Tourism Resources in China

Yifan Zuo ^{1,2,†}, Huan Chen ^{2,†}, Jincheng Pan ³, Yuqi Si ², Rob Law ⁴ and Mu Zhang ^{2,*}

¹ School of Management, Jinan University, Guangzhou 510632, China; yifanzuo@stu2019.jnu.edu.cn

² Shenzhen Tourism College, Jinan University, Shenzhen 518053, China; huanchen@stu2017.jnu.edu.cn (H.C.); s202036041024@stu2020.jnu.edu.cn (Y.S.)

³ Department of Physical Education, Guizhou University of Finance and Economics, Guiyang 550000, China; 1711621003@student.sus.edu.cn

⁴ School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hong Kong, China; rob.law@polyu.edu.hk

* Correspondence: zhangmu@jnu.edu.cn; Tel.: +86-755-2693-1865

† The authors contribute equally to this article.

Citation: Zuo, Y.; Chen, H.; Pan, J.; Si, Y.; Law, R.; Zhang, M. Spatial Distribution Pattern and Influencing Factors of Sports Tourism Resources in China. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 428. <https://doi.org/10.3390/ijgi10070428>

Academic Editors: Wolfgang Kainz, Andrea Marchetti and Angelica Lo Duca

Received: 22 May 2021

Accepted: 21 June 2021

Published: 23 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Sports tourism is an emerging tourism product. In the sports and tourism industry, resource mining is the foundation that provides positive significance for theoretical support. This study takes China's sports tourism boutique projects as the study object, exploring its spatial distribution pattern through the average nearest neighbor index, kernel density, and spatial autocorrelation. On the strength of the wuli–shili–renli system approach, the entropy value method and geographic detector probe model are used to identify the driving factors affecting the spatial distribution pattern. Findings reveal the following: (1) From 2013 to 2014, the sports tourism resources in China present a distribution pattern with the Yangtze River Delta urban agglomeration as the high-density core area and the Guizhou–Guangxi border area and the western Hubei ecological circle as the sub-density core areas. (2) From 2014 to 2018, China's sports tourism boutique projects increased by 381, and the regional differences among various provinces tended to converge. The high-density core area remained unchanged. The sub-density cores are now the Yunqian border area of the Karst Plateau, the Qinglong border area of the Qilian Mountains, and the Jinji border area of the Taihang Mountains, shaping the distribution trends of “depending on the city, near the scenery” and “large concentration, small dispersion”. (3) The proportion of provincial sports tourism development classified as being in the coordinated stage is 61.29%. (4) The explanatory power of the factors affecting the spatial layout in descending order is natural resource endowment, sports resource endowment, transportation capacity, industrial support and guidance, market cultivation and development, people's living standards, software and hardware services, and economic benefit effects. The explanatory power of the interaction of two different factors is higher than that of the single factor.

Keywords: sports tourism; spatial distribution; geographic detector; influencing factors; China

1. Introduction

Sports tourism is defined as “the use of sports as a vehicle for tourism endeavor” [1]. In recent years, with the increase in public leisure time, continuous enhancement of fitness awareness, and rapid expansion of tourism consumption, the Chinese government has vigorously promoted the development of sports tourism to satisfy the people's yearning for a better life. The government has called for the creation of sports tourism demonstration zones and encourages the construction of relevant boutique projects. However, in the course of its rapid development in various places, problems such as unreasonable layout, inadequate resource utilization, insufficient capital investment, lack of stadiums, and poor matching of resources with sports characteristics have frequently occurred. As an emerging spatial and regional unit, effective analysis of sports tourism patterns and influ-

encing factors has important practical significance for its layout optimization as well as the sustainable and moderate development.

Current research on sports tourism mainly focuses on the impact on society, economy, and culture. In addition, the characteristics, needs, behaviors, and markets of sports tourists are the exploration hotspots, along with the sports tourism destination planning, product development, and safety management. Cooper and Alderman discussed the influence of canceling sports events on the economy, society, culture, and the environment; in response to the COVID-19, relevant sports tourism alternatives are necessary to promote sustainability [2]. Nishio et al. developed a motivational scale for sports fans (social, achievement, relaxation, and games) and a general tourist motivation scale (escape, nature, shopping, and food) for sports tourism [3]. Jin et al. proposed that the event quality affects the respondents' perceived value, destination image, and behavioral intentions. A structural equation model for related tests ultimately showed that the quality of the event and its perceived value have a significant effect on behavioral intentions [4]. Page et al. compared the safety experience of adventure travelers in New Zealand and Scotland, then commented on the adventure tourism accident compensation legislation and jurisprudence. In addition, the study discussed the injury experience and safety management of adventure travel customers in Queensland and analyzed the adventure travel accidents of inbound tourists from 1982 to 1996 in New Zealand [5].

At present, few studies focus on the spatial structure or distribution of sports tourism resources. Fugao and Li construed the ideal evolution of the spatial structure of sports tourism in the entire region, using spatial structure theories such as growth pole, point-axis, patch-corridor-matrix, and network structure to explain the generation and evolution of the spatial structure of "point-line-surface-domain" [6]. Zuo et al. took the lead in exploring the spatial distribution characteristics of Chinese marathon events. The size of urban populations; living standards; and the overall quality of urban residents (including the concept of sports and leisure), the social environment, and other social mechanism factors affect the spatial distribution of marathon events [7]. Additionally, the spatial distribution characteristics of Chinese marathon events are investigated based on the perspective of natural resources [8]. Geneletti took advantage of geographic information system (GIS) technology with biology, physics, landscape, and other indicators to determine the environmental effect assessment of ski tourism destinations [9]. Thus far, only a few studies have analyzed the spatial distribution pattern of sports tourism resources in China [10,11].

In summary, academic circles have not sufficiently probed the issues of sports tourism and rather focused mainly on account of psychology, management, or behavior. Although a few studies involve spatial structure, the scale is mostly limited to regions, provinces, and cities and therefore lacks a macro-level analysis on the systematic review of the spatial distribution pattern of national sports tourism resources. In view of this, this article systematically sorts out China's sports tourism boutique projects from 2013 to 2018, selecting the nodes in 2014 and 2018. Moreover, the spatial distribution and distribution characteristics of China's sports tourism resources are described through average nearest neighbors, nuclear density analysis, and spatial autocorrelation. In addition, the study uses the entropy weighting method and geographic detector model to identify the driving factors affecting the spatial distribution, making theoretical contributions to the study of the spatial distribution pattern of sports tourism resources. The result builds a systematic index system of influencing factors. It is expected to provide countermeasures for the optimization and healthy development of China's future sports tourism spatial layout, and to provide a reference for the reasonable layout and appropriate development of future sports tourism resources in other countries or regions.

2. Materials and Methods

2.1. Methods

2.1.1. Average Nearest Neighbor

The nearest neighbor distance measures the mutual proximity of sports tourism resources in the spatial distribution. The nearest neighbor index reflects the spatial aggregation characteristics of sports tourism resources, that is, the ratio of the actual to the theoretical nearest neighbor distances [12]. The nearest neighbor index is calculated as

$$ANNI = \frac{ANNO}{ANNE} = 2\sqrt{D} \times ANNO \quad (1)$$

In the above formula, ANN represents the nearest neighbor index, $ANNO$ represents the average nearest neighbor distance, $ANNE$ represents the theoretical nearest neighbor distance, and D represents the nearest neighbor density [12], where

$$ANNE = \frac{1}{2\sqrt{n/A}} = \frac{1}{2\sqrt{D}} \quad (2)$$

In Formula (2), A represents the area of the province and n represents the number of sports tourism resources. When $ANNI = 1$ and $ANNO = ANNE$, the sports tourism resources are randomly distributed; when $ANNI < 1$ and $ANNO < ANNE$, the sports tourism resources are in an agglomerated distribution; when $ANNI > 1$ and $ANNO > ANNE$, the sports tourism resources are uniformly distributed. The smaller the $ANNE$, the higher the concentration of sports tourism resources. Both Zuo et al. [7] and Wang et al. [13] used average nearest neighbor to determine the distribution state of the studied elements; quantify the spatial relationship; and judge whether the elements are clustered, random, or dispersed.

2.1.2. Kernel Density

Kernel density analysis is a quantitative estimation of the density of dot-like objects using a moving cell. The assumption is that geographic events can occur at any location in space, but with different probabilities at different locations. The probability of event occurrence is high in areas where dot-like objects are dense and low in areas where dot-like objects are sparse is low [7]. The analytical formula for kernel density is

$$\tilde{\lambda}(s) = \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{s-s_i}{\tau}\right) \quad (3)$$

In the above formula, $k(\cdot)$ represents the kernel function, τ ($\tau > 0$) represents the bandwidth, n represents the number of sample points, and $(s - s_i)$ represents the distance between the dot-like object s and the estimated point s_i [7]. This formula has been tested many times, and the data selection search bandwidth is 333.6 km to more intuitively reflect the spatial distribution of sports tourism resources. Yoo et al. [14] and Allen et al. [15] made use of kernel density in order to determine the center position of a specific element. The density is the highest at the center position, and it decays with distance. The density is zero at the limit distance in the end.

2.1.3. Spatial Autocorrelation (Global Moran's I)

Spatial autocorrelation reflects the degree of correlation between a certain geographic phenomenon or attribute value on a regional unit and the same phenomenon or attribute value on adjacent regional units [16]. This study uses Moran's I index, which is

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij} (X_i - \bar{X})(X_j - \bar{X})}{S^2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij}} \quad (4)$$

In the above formula, ω represents the spatial weight between areas i and j ; n represents the number of regions; and X_i and X_j represents the observation values of locations i and j , respectively. The value range of Moran's I is $[-1, 1]$: Moran's $I > 0$ indicates a positive spatial correlation phenomenon, Moran's $I < 0$ indicates a negative correlation phenomenon, and Moran's $I = 0$ indicates an independent random distribution [16]. Zuo et al. [7] and Zhang et al. [17] used global Moran's I to calculate the Moran's I value of the research elements on a continuous spatial scale to explore the strength of the spatial correlation of the research elements and their changes with the spatial scale.

2.1.4. Entropy Method

Compared with the analytic hierarchy process, the entropy method is more objective. The weight is determined mainly based on the information provided by index data, and not by whether the data are linear or not. This method can effectively avoid the interference of human factors and has a higher credibility [18]. Zhang et al. [19] and Li et al. [20] used the entropy method to weight the indicators according to the connection degree of each indicator or the amount of information provided, effectively avoiding the subjective factors of the indicator system results. When constructing the indicator system, this study uses the entropy method to measure the natural resource endowment, sports resource endowment, hardware and software services, transportation capacity, people's living standards, industrial support and guidance, economic benefit effects, and market development in various provinces and municipalities in China to accurately analyze various influencing factors. Thus, this study provides the premise and foundation for the influence of the spatial layout of sports tourism resources.

First, the range standardization is performed on the original data of different magnitudes and dimensions. The formula is

$$q_{ij} = \begin{cases} x_{ij} - \min(x_{ij}) / \max(x_{ij}) - \min(x_{ij}) & q_{ij} \text{ is the positive index} \\ \max(x_{ij}) - x_{ij} / \max(x_{ij}) - \min(x_{ij}) & q_{ij} \text{ is a negative index} \end{cases} \quad (5)$$

In the above formula, q_{ij} represents the data value after standardized processing; x_{ij} represents the original data value, where i ($i = 1, 2, 3, \dots, m$) is the sequence number of the evaluation index; j ($j = 1, 2, 3, \dots, n$) is the number of points; and $\max(x_{ij})$ and $\min(x_{ij})$ are the maximum and minimum values of the corresponding index of the order parameters at the critical point of system stability, respectively.

The weight of the i th index of a data set containing m indexes and n samples is calculated as

$$W_i = \frac{1 + \frac{1}{\ln n} \sum_{j=1}^n \left(\frac{Q_{ij}}{\sum_{j=1}^n Q_{ij}} \ln \frac{Q_{ij}}{\sum_{j=1}^n Q_{ij}} \right)}{m + \sum_{i=1}^m \left(\frac{1}{\ln n} \sum_{j=1}^n \frac{Q_{ij}}{\sum_{j=1}^n Q_{ij}} \ln \frac{Q_{ij}}{\sum_{j=1}^n Q_{ij}} \right)} \quad (6)$$

In the above formula, W_{ij} represents the weight of the i th index; Q_{ij} represents the standardized data value, and each index is summed; and U_{ij} represents the comprehensive

evaluation value of the factors affecting the spatial layout of sports tourism resources. The formula is

$$U_{ij} = \sum_{i=1}^m W_i \times Q_{ij} \quad (7)$$

2.1.5. Geodetector

Geodetector is a tool used to analyze and detect spatial differentiation by identifying the extent to which a certain factor explains the spatial differentiation of the result variable, thereby revealing the source of its spatial difference [21]. The formula is

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} \quad (8)$$

In the above formula, L represents the variable stratification, that is, classification or partition; N_h and N represent the number of units in layer h and the entire area, respectively; σ_h^2 and σ^2 represent the variance of the result variable in layer h and the entire area, respectively; and q represents a certain front. The magnitude of influence of the dependent variable on the outcome variable is in the range of $[0, 1]$. The closer q is to 1, the greater the explanatory strength of the pre-dependent variable on the outcome variable. Conversely, the closer q is to 0, the smaller the explanatory strength. This study uses the geographic detector method to identify the factors affecting the spatial distribution of sports tourism resources in China.

The purpose of interaction detection is to assess whether the explanatory power of the spatial differentiation of China's sports tourism resources increases or decreases when two factors are working together. The evaluation method is to judge the direction and method of interaction between factors by comparing the values of the single and double factors of q , which can generally be divided into five categories [21]: (1) nonlinear weakening $Q < \text{Min}(q(X1), q(X2))$; (2) single-factor nonlinear weakening $\text{Min}(q(X1), q(X2)) < Q < \text{Max}(q(X1), q(X2))$; (3) two-factor enhancement $Q > \text{Max}(q(X1), q(X2))$; (4) independent $Q = X$; and (5) nonlinear enhancement $Q > X$, where $Q = q(X1) \cap q(X2)$, $X = q(X1) + q(X2)$. Among them, $Q = q(X1) \cap q(X2)$, $X = q(X1) + q(X2)$, where $q(X1)$ and $q(X2)$ are the influencing factors of the spatial differentiation of sports tourism resources in China. Both Chi et al. [22] and Zhang et al. [17] used Geodetector to study the similarity between the independent variable and the dependent variable in the spatial distribution to understand whether different influencing factors have an interactive effect on the spatial distribution.

2.2. Index Selection

This study integrates the particularity of China's sports tourism resources in its development and follows the relevant principles of scientific, representativeness, operability, reliability, and availability in the selection of indicators. The structure of factors affecting the spatial distribution of sports tourism resources is described in view of the wuli–shili–renli (WSR) methodology, a system theory with Eastern philosophy. The basic core of its philosophy and concept is to consider not only the aspects of objects, but also their better applications to material aspects when dealing with complex issues [23]. Given that sports tourism contains many complex components of people and things, involving their composition and relationships, we learn from previous studies and apply the WSR to multidimensional analysis [24]. WSR methodology was proposed by Gu and Zhu. It is not only a methodology, but also a framework tool for solving complex problems. The connection between wuli, shili, and renli is the coordination of the relationship between intention, goal, reality, strategy, plan, and conception, which can coordinate the relationship between input, output, and outcome of system practice [25].

The quantity of sports tourism resources is taken as the dependent variable. At the same time, we construct a model that can explore the main influencing factors of the spatial distribution of sports tourism resources in China. Variables are selected from natural resource endowment, sports resource endowment, hardware and software services,

transportation capacity, people's living standards, industrial support and guidance, economic benefit effects, and market cultivation and development, as shown in Figure 1. The four factors of natural resource endowment, sports resource endowment, software and hardware services, and transportation capacity provide conditions for the occurrence of sports tourism activities and also restrict the scale and efficiency of internal operations. They are the internal motivation of sports tourism activities and are at the core, which is in line with the physical dimension of the understanding of the objective world, belonging to the physical dimension. People's living standards, industrial support, and guidance are external influencing factors that provide power for the demand market of the sports tourism industry. They are the prerequisite and foundation for the smooth operation of the industry, which is consistent with shili dimension's response to events. Therefore, the two factors belong to shili dimension. The two factors of economic benefit effects and market cultivation and development are internal factors. They act on the people's sports tourism practice and open up the sports tourism market, which can promote the improvement and effect of external influencing factors. In line with renli dimension's understanding of the actual effects of the incident, they belong to the renli dimension.

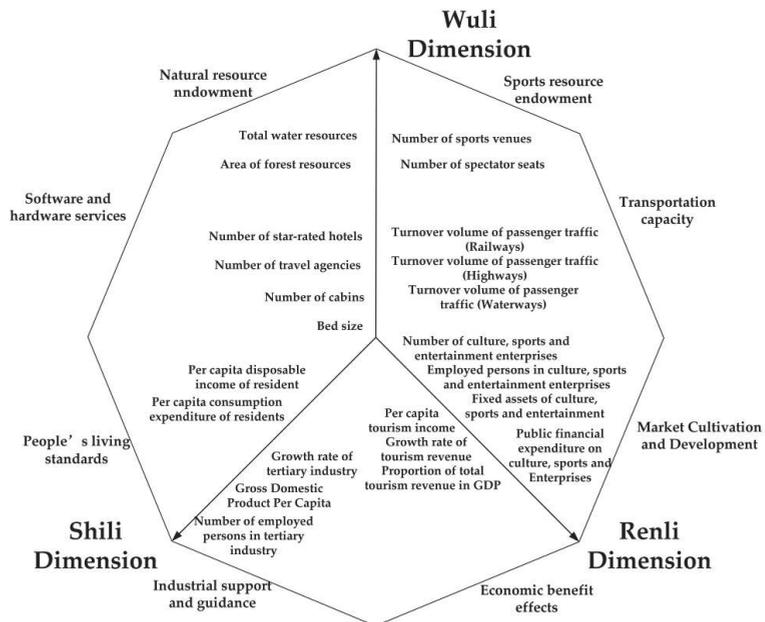


Figure 1. Impact analysis framework of the spatial distribution of sports tourism resources in China.

As shown in Table 1, the choice of variables is built on the following assumptions:

- (1) Natural resource endowment is the major source of demand for the development of sports tourism resources, which primarily include those of water and forests. The richer the natural tourism resources, the better the development of sports tourism [26].
- (2) Sports resource endowment can effectively increase the number of sports events, which has a positive promotion for sports tourism development [7].
- (3) The construction and improvement of software and hardware services are a very critical foundation link for the progress and utilization of sports resources. Improving and supporting software and hardware services can create a good tourism service environment, thereby effectively enhancing the development of sports tourism [26].

- (4) Transportation capacity is one of the important factors for tourists when choosing a destination, and thus the convenience and accessibility of transportation also serve as significant references for the development of tourism resources [27].
- (5) People's consumption has shifted from "subsistence" to "well-off" type. The consumption structure is continuously optimized and upgraded. Meanwhile, sports consumption is moving towards the "participatory and entertainment" type, with sports and recreative tourism as precisely its representative. Therefore, the higher the consumption of residents, the richer the corresponding sports tourism resources [28].
- (6) Sports tourism belongs to the tertiary industry, which has a faster growth rate. As the apparent effect of industrial support and guidance increases, the integrated development of sports, culture, tourism, and other industries increases in quality [29].
- (7) The economic benefit effect can raise employee wages, stimulating greater resources to invest in the tourism industry. Thus, this factor can positively promote the development of sports tourism [2].
- (8) Market cultivation and development can facilitate the convergence of sports and tourism. Thus, expanding the share of sports investment and increasing the level of sports consumption becomes possible. On this basis, market cultivation and development can provide basic support for the development of sports tourism [30].

Table 1. Index selection of factors affecting sports tourism resources in China.

System Framework	First-Level Index	Second-Level Index	Unit of Account	Weighting Target
Wuli Dimension	Natural resource endowment	Total water resources	billion m ³	0.055
		Area of forest resources	10,000 hectares	0.055
	Sports resource endowment	Number of sports venues	Pcs	0.041
		Number of spectator seats	unit	0.071
		Number of star-rated hotels	unit	0.039
	Software and hardware services	Number of travel agencies	unit	0.042
		Number of cabins	room	0.037
		Bed size	bed	0.040
		Turnover volume of passenger traffic (railways)	10,000 persons	0.044
	Transportation capacity	Turnover volume of passenger traffic (highways)	10,000 persons	0.042
Turnover volume of passenger traffic (waterways)		10,000 persons	0.073	
Per capita disposable income of residents		CNY	0.038	
Shili Dimension	People's living standards	Per capita consumption expenditure of residents	CNY	0.035
		Growth rate of tertiary industry	%	0.043
	Industrial support and guidance	Gross domestic product per capita	CNY	0.030
		Number of employed persons in tertiary industry	CNY 10,000/person	0.043
Renli Dimension	Economic benefit effects	Per capita tourism income	CNY 10,000/person	0.035
		Growth rate of tourism revenue	%	0.032
		Proportion of total tourism revenue in GDP	%	0.037
	Market cultivation and development	Number of culture, sports, and entertainment enterprises	unit	0.043
		Employed persons in culture, sports, and entertainment enterprises	10,000 persons	0.041
		Fixed assets of culture, sports, and entertainment	CNY 100 million	0.047
	Public financial expenditure on culture, sports, and Enterprises	CNY 100 million	0.037	

2.3. Data Sources

The data of sports tourism resources came from the recommended list of “China Sports Tourism Boutique Projects” (only the finalists) announced by the General Administration of Sports. In order to promote reasonable regional planning of sports tourism and accelerate the efficiency of sports tourism in releasing new economic kinetic energy, the Chinese government began to cultivate sports tourism boutique projects in 2013. Sports tourism boutique projects are operational tourist attractions, scenic spots, routes, events, festivals, and other projects that are reported by provinces, municipalities, autonomous regions, and municipalities directly under the Central Government and selected by expert appraisal teams. They are based on the market and centered on the sports needs of tourists. Besides, they are supposed to provide tourists with a certain degree of participation and viewing value. Statistics from 2013 to 2018 show a total of 755 sports tourism boutique projects. Supplementary data were found in the China Statistics Bureau, provincial (cities, districts) tourism development statistical bulletins, local tourism industry bulletins, and “China Sports Tourism Boutique Project Development Report”. Repeated declarations were screened. The number of sports tourism boutique projects was 209 in 2014 and 590 in 2018. Sampling was based on the venue for boutique events, coordinates of the visitor center for boutique scenic spots, government location for boutique destinations, and starting point locations for the boutique route.

Considering the consistency of the statistical caliber of relevant indicators involved in sports tourism, data from 31 provinces in China (excluding Hong Kong, Macao, and Taiwan) in 2018 were selected for analysis. The data were mostly derived from the 2018 China Statistical Yearbook, China Tourism Statistical Yearbook, and China Mass Sports Development Report. Several indicators were supplemented by data from local statistical yearbooks, statistical bulletins of the local sports bureaus, and the official website of the Ministry of Finance of China. In addition, the maps of China were all obtained from the Resource and Environmental Science Data Center of the Chinese Academy of Sciences (<http://www.resdc.cn/Default.aspx>, accessed on 12 May 2021).

3. Results

3.1. Pattern of Sports Tourism Resources in 2014

Figure 2 shows that the distribution level of sports tourism resources of the 31 provincial research units in China can be classified as tentative (cumulative ratio = 0%), low (cumulative ratio 0–6%), medium (cumulative ratio 6–44%), or high (cumulative ratio 44–100%). Specifically, Beijing, Hunan, Liaoning, Ningxia, Sichuan, Tianjin, Yunnan, Chongqing, Jilin, and other places do not have shortlisted sports tourism boutique projects in 2013 and 2014 and are considered blank areas. Jiangxi, Tibet, Guangdong, Zhejiang, Shanghai, and Xinjiang account for 6% of the national sports tourism resources, representing areas with low development levels. Guangxi, Heilongjiang, Hainan, Shaanxi, Fujian, Gansu, Henan, Shanxi, Qinghai, Hebei, and other provinces account for 38% of the national sports tourism resources, belonging to the middle-level development area. Shandong, Jiangsu, Inner Mongolia, Hubei, Guizhou, and Anhui account for 56% of the national sports tourism resources, belonging to the high-level development area.

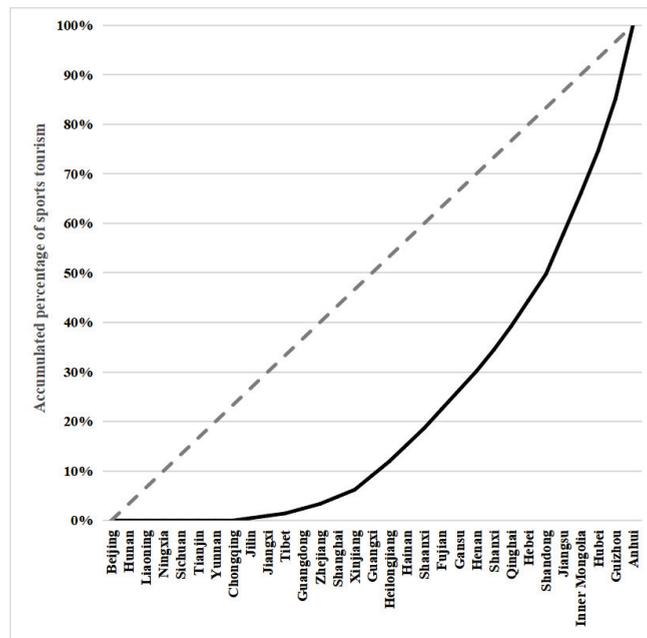


Figure 2. Lorenz curve of the distribution of China's sports tourism resources in 2014.

This study used ArcGIS 10.2 software (ESRI, Inc., Redlands, CA, USA) and average nearest neighbor to analyze China's sports tourism resources in 2014. The results are as follows: average observation distance is 62,725.6701 m, expected average distance is 133,750.1520 m, nearest neighbor ratio R is 0.468976, Z is -14.686483 , and the significance level is $p < 0.001$, indicating that China's sports tourism resources in 2014 showed a clear agglomeration distribution in space.

The spatial agglomeration characteristics of sports tourism resources are discussed through nuclear density mapping. Figure 3 shows that before 2014, China's sports tourism resources displayed the Yangtze River Delta city cluster as the high-density core area, while the Guizhou–Guangxi border and the western Hubei ecological circle are the secondary density core area. Moreover, China's sports tourism resources have a distribution trend of “depending on the city, near the scenery”, which means forming a central city based on the surrounding scenery. Central diffusion gradually forms an axial zone, which superimposes with the central radiation and coexists to form a network surface [6]. Moreover, the spatial characteristics reveal large concentration and small dispersion, that is, mainly concentrated in urban agglomerations and areas with high natural resource endowments, and a small amount scattered in areas with less traffic access but with unique natural resources. The possible reasons are, on the one hand, the sports tourism industry in the Yangtze River Delta urban agglomeration has a good foundation and clear location advantages. On the other hand, the levels of per capita disposable income and per capita consumption expenditure are much higher than those of other regions in the country. As a result, consumption demand is continuously driven and the supply side structure is continuously optimized. The market of sports tourism is vast, leading the country in terms of development [31]. The Guizhou–Guangxi border area and western Hubei ecological circle have superior natural environmental conditions, both of which are karst geomorphic regions. The natural scenery is the development feature. The precious resource of sports tourism is an ideal place to carry out exciting and entertaining activities such as rock climbing and bungee jumping [32]. Urban agglomerations and areas with high natural resource endowments have attracted a large number of sports tourism resources due to their

superior geographical location, sound infrastructure, convenient transportation routes, and industrial policy support and guidance. Zones with poor transportation access but with unique natural resources because of their different tourist experiences can still attract the layout of sports tourism.

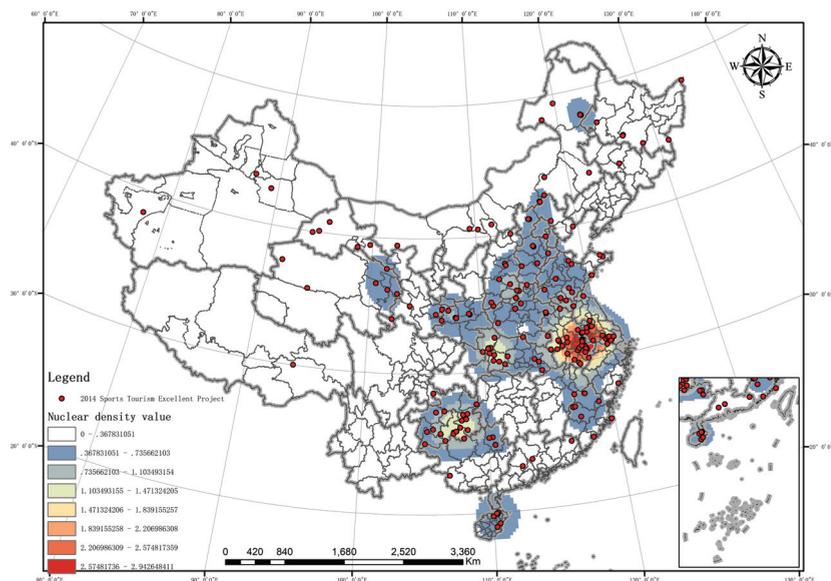


Figure 3. Distribution of core density of China’s sports tourism resources in 2014. The color in the figure changes from gray to red: the redder the color, the more clustered the sports tourism projects. See the nuclear density index as shown in the legend for details.

By using global spatial autocorrelation technology, the spatial characteristics and aggregation effects of China’s sports tourism resources were further explored. Adopting ArcGIS 10.2, the global Moran’s I index was calculated to obtain the global autocorrelation of China’s sports tourism resources. Global Moran’s I is -0.0084 , $Z(I)$ is 0.2503 , and $P(I)$ is 0.376 , illustrating that the national provincial sports tourism resources do not have significant spatial agglomeration trends. The amount of sports tourism resources in each province is not related to those in surrounding provinces. Above all, the effect of “neighboring dependence” has not been formed, which is not conducive to sports tourism development in the entire region.

3.2. Pattern of Sports Tourism Resources in 2018

From 2014 to 2018, China’s sports tourism boutique projects increased by 381, with an average annual increase of 127. Most provinces have a certain degree of growth. Jiangsu, Qinghai, Yunnan, Gansu, Shanxi, Anhui, Guizhou, and Hubei all have over 30 new projects, with a rapid growth trend. Comparing the Lorentz curves of the distributions of sports tourism resources in each province in the two years (Figures 2 and 4), it can be seen that the regional differences tend to converge. The division of the resource distribution in 31 provincial research units across the country also changed into tentative (cumulative ratio = 0%), low (cumulative ratio 0–17%), medium (cumulative ratio 17–46%), and high (cumulative ratio 46–100%). In particular, from 2015 to 2018, no sports tourism boutique projects were shortlisted in Beijing, Hunan, Chongqing, and other places, which are temporarily depicted as blank areas. Sichuan, Tibet, Guangdong, Jilin, Hainan, Ningxia, Tianjin, Shanxi, Shanghai, Liaoning, Jiangxi, Zhejiang, and other provinces account for 17% of the national sports tourism resources and are areas with low development levels. Henan, Shandong,

Heilongjiang, Guangxi, Xinjiang, Hebei, Fujian, and Inner Mongolia account for 17% of the national sports tourism resources. Regions with medium development levels account for 29%. Yunnan, Hubei, Gansu, Shanxi, Jiangsu, Anhui, Guizhou, and Qinghai account for 54% of the national sports tourism resources and are regions with high development levels. Yunnan, Gansu, Shanxi, and Qinghai have risen from the previous blank, medium, and low to high development levels. Xinjiang has risen from a previous low development level to a medium development level, and Liaoning, Ningxia, Sichuan, Tianjin, and Jilin have risen from the previous blank development level to a low level of development. However, Shandong, Hainan, Shaanxi, and Inner Mongolia have been downgraded.

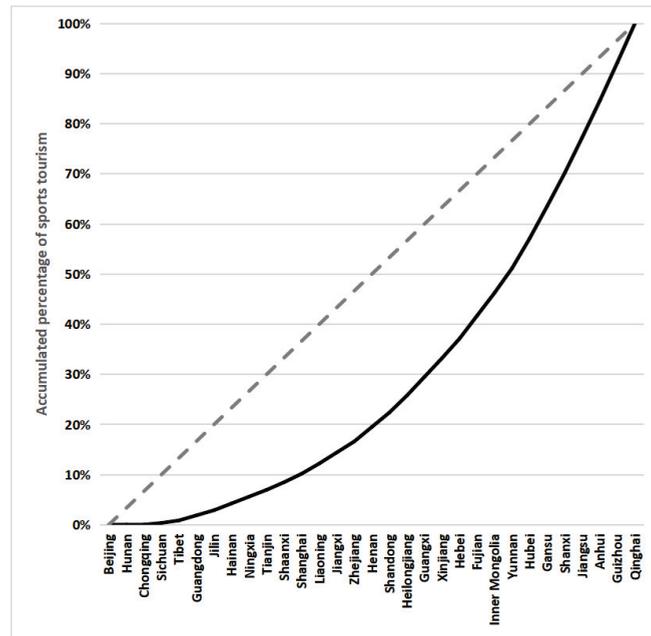


Figure 4. Lorenz curve of the distribution of China's sports tourism resources by province in 2018.

ArcGIS 10.2 (ESRI, Inc., Redlands, CA, USA) was applied to analyze the average nearest neighbors of China's sports tourism resources in 2018. Results show that the average observation distance is 33,333.9247 m, expected average distance is 82,961.2671 m, the nearest neighbor ratio R is 0.401801, Z score is -27.797295 , and the significance level is $p < 0.001$. Thus, China's sports tourism resources in 2018 show a clear agglomeration distribution in space.

As shown in Figure 5. Compared with the results of 2014, while the Yangtze River Delta urban agglomeration remains a high-density core area, the secondary density core areas are now the Yunqian border area of the Karst Plateau, Qinglong border area of the Qilian Mountains, and the Jinji border area of the Taihang Mountains. Nonetheless, the distribution trait is also "depending on the city, near the scenery" and "large concentration, small dispersion". By contrast, the location of the core area of the Yangtze River Delta urban agglomeration has no significant change in 2018, but the sub-density core area extends to the northwest toward the Qinglong border area of the Qilian Mountains and to the southwest towards the Yunqiangui border area of the Karst Plateau. The core area of the western Hubei ecosphere is transformed into the Taihangshan Jinji border area. The possible reason is mainly that the Yangtze River Delta urban agglomeration has the highest level of economic development and the highest residents' living standards in China. The sports tourism market and industrial chain in this area are more mature than in

other regions. Moreover, the development and investment prospects of the sports tourism market are broad. Numerous ethnic minorities reside near the Yunnan–Guizhou Plateau, and their habitats, geomorphology, and climatic conditions are special. The surrounding environment is beautiful and scenic, thus ushering in the explosive period of sports tourism development [33]. Qilian Mountain Qinglongbian District is situated in the golden section of the Silk Road Economic Belt. Rich in geography, water, biological resources, cultural relics, folk customs, sports competitions, and other resources, the area is suitable for the development of sports tourism projects. The relevant resources in the core area have gradually changed from “dispersed” to “intensive” [34]. Natural beauty, historical civilization, and revolutionary historical sites together constitute the unique sports tourism resources Jinji border area of the Taihang Mountains. Relying on the complex and changeable geology, geomorphology, hydrology, and meteorology, as well as a long history and a heavy and ancient sports culture, this area can provide the foundation and guarantee for the development of sports tourism in the Shanxi–Hebei border area of the Taihang Mountains [35].

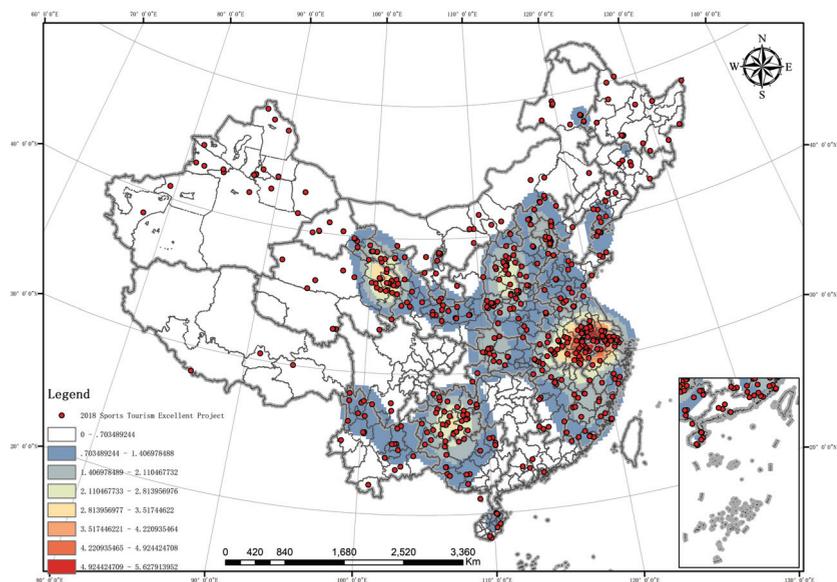


Figure 5. Distribution of core density of China’s sports tourism resources in 2018. The color in the figure changes from gray to red: the redder the color, the more clustered the sports tourism projects. See the nuclear density index as shown in the legend for details.

Using ArcGIS 10.2 to calculate the global Moran’s I index, global Moran’s I is -0.0897 , $Z(I)$ is -0.4782 , and $P(I)$ is 0.325 , demonstrating that no remarkable spatial aggregation trend occurred for sports tourism resources at the provincial level across China in 2018.

As a further exploration, the amount of sports tourism resources in each province was superimposed with the growth rate. These two factors were divided into six development stages using the coupling and coordination model [36]. Figure 6 shows that the provinces coordinating the development of sports tourism account for 61.29%. Fewer provinces exhibit extreme incoordination, namely Chongqing, Hunan, Beijing, and Sichuan. Several provinces are between basic incoordination (Shandong, Ningxia, Jilin Province, Zhejiang) and primary coordination (Hainan, Shanghai, Liaoning, Hebei). The main reason for the above situation is that the development of sports tourism has attracted much attention in recent years. Both the sports and tourism industries are strongly advocating sports tourism, and China has promulgated various policies that are conducive to this develop-

ment. In addition, marathon events and sports hardware facilities in scenic spots have been implemented as its foundation. As such, the coordination stage of sports tourism development in the eastern region is far ahead, which is evidently higher than the national average. The northeast, northwest, and parts of the southwest are at the same level as the national average, which is stable and gradually becoming more coordinated. Sichuan, Chongqing, Hunan, Guangdong, Shaanxi, and other places still have a large room for improvement.

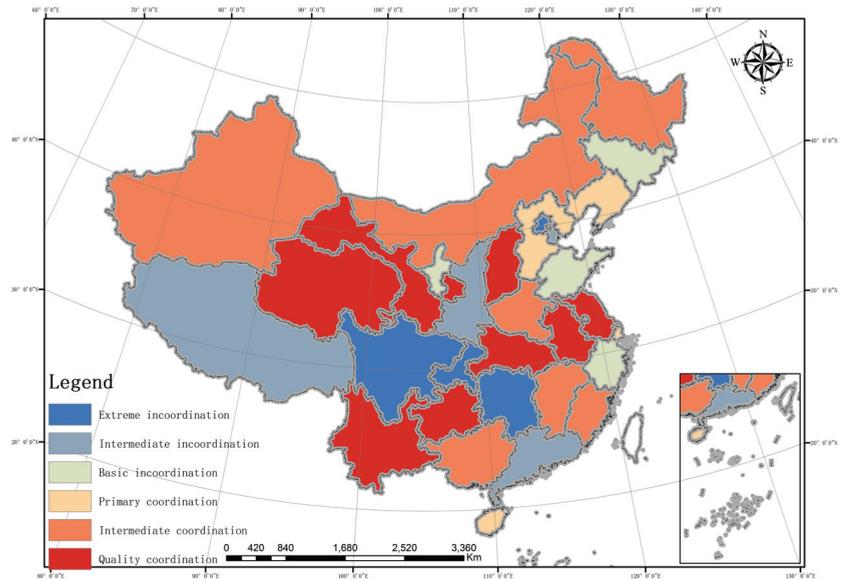


Figure 6. Level of coordination of sports tourism in various provinces in China. The color in the picture changes from blue to red: the redder the color, the more coordinated the development of sports tourism. See the legend for details.

3.3. Factors Influencing the Spatial Distribution of Sports Tourism Resources

3.3.1. Influencing Factors of the Spatial Distribution of Sports Tourism Resources

The geographic detector model was used to explore the essential mechanism of the differences in the spatial distribution of sports tourism resources in China to seek a more scientifically specific optimization path for regional sports tourism development planning. The rapid cluster analysis method in SPSS 24.0 (SPSS Inc., Chicago, IL, USA, 2019) was used to classify the driving factors, such as natural resource endowments, sports resource endowments, software and hardware services, transportation capacity, people's living standards, industry support and guidance, economic benefit effect, and market cultivation and development into five categories from high to low. Then, the geographic detector analysis was carried out to calculate the q value of each driving factor on the spatial distribution of China's sports tourism resources. Table 2 shows the results.

In Table 2, the q value means the extent to which the detection factor explains and affects the spatial distribution of China's sports tourism resources. The larger the q value, the greater the impact of the factor [22]. In general, among the identified eight driving factors, the order of descending impact on the spatial distribution of China's sports tourism resources is as follows: natural resource endowment > sports resource endowment > transportation capacity > industrial support and guidance > market cultivation and development > people's living standards > software and hardware services > economic benefit effect. The principal factors are natural resource endowment, sports resource en-

dowment, transportation capacity, industry support and guidance, and market cultivation and development.

Table 2. Detection results of factors affecting the spatial distribution of sports tourism resources in China.

Driving Factor	q-Statistic	Detection Index	q-Statistic
Natural	0.308	Total water resources	0.300
		Area of forest resources	0.122
Sport	0.219	Number of sports venues	0.116
		Number of spectator seats	0.096
		Number of star-rated hotels	0.225
Service	0.094	Number of travel agencies	0.084
		Number of cabins	0.137
		Bed size	0.278
Transportation	0.172	Turnover volume of passenger traffic (railways)	0.120
		Turnover volume of passenger traffic (highways)	0.309
		Turnover volume of passenger traffic (waterways)	0.078
Living	0.103	Per capita disposable income of residents	0.177
		Per capita consumption expenditure of residents	0.175
Industrial	0.170	Growth rate of tertiary industry	0.206
		Gross domestic product (GDP) per capita	0.200
		Number of employees in the tertiary industry	0.383
Economic	0.079	Per capita tourism income	0.122
		Growth rate of tourism revenue	0.185
		Proportion of total tourism revenue in GDP	0.146
Market	0.136	Number of culture, sports, and entertainment enterprises	0.392
		Employees in culture, sports, and entertainment enterprises	0.167
		Fixed assets of culture, sports, and entertainment	0.233
		Public financial expenditure on culture, sports, and enterprises	0.282

Significance level, $p < 0.05$.

The details are as follows:

- (1) Natural resource endowment, with its explanatory power of 0.308, has a huge impact on the spatial distribution of China's sports tourism resources. In practice, the development of sports tourism resources is conducted in consideration of local conditions, several of which are related to natural resources. Others are based on the characteristics of geo-sports, such as the use of climbing, bungee jumping, surfing, and other activities close to the development of natural resources [26]. Given the abundance of natural resources, the requirements for other essential resources are relatively low.
- (2) The explanatory power of the endowment of sports resources is in the second position with a value of 0.3379. Sports tourism often takes sports fitness and leisure projects, sports events, large-scale stadiums, and other relevant resources as the core attractions to provide conditions for tourist services. These include sports fitness venues and well-known landmark sports buildings (museums) [7].
- (3) The explanatory power of transportation capacity is 0.172, ranking third. Given the scattered sports tourism resources in several areas, the accessibility and convenience of transportation affect the interest of tourists. Often, destinations with better traffic conditions naturally have more tourists. In recent years, national road transportation has shown continuous growth. Counting on the opportunities of global sports tourism, the construction pattern of national boutique sports tourism routes is gradually taking shape, effectively enhancing the spatial spillover effect of sports tourism resources [2].
- (4) Industry support guidance has an explanatory power of 0.170, ranking fourth. Sports tourism is directly related to the tertiary industry in the region and affects the regional GDP. As a labor-intensive and service industry, sports tourism needs to satisfy tourists' food, housing, transportation, travel, shopping, and entertainment needs. These various demands have driven the evolution of related industries and provided more job opportunities [29].

- (5) The explanatory power of market cultivation and development is 0.136, ranking fifth. Government and institutions guide large strategic investors to adjust the investment structure, fully mobilize the enthusiasm of market entities, strengthen talent training, and promote capital flow, which contributes to the sustainable development of sports tourism [2].
- (6) In contrast to the above influencing factors, people's living standards, software and hardware services, and economic benefit effects are weaker in explaining the differences in the spatial distribution of sports tourism resources in China. These factors have an apparent negative impact but an unclear positive effect.

3.3.2. Analysis of Detection Factor Interaction Results

Interaction was used to reveal whether an interactive relationship exists among the abovementioned influencing factors. In Table 3, the results show that the explanatory power of different two-factor interactions is higher than that of single-factor interactions. Meanwhile, the interaction types presented are nonlinear and two-factor enhancements. Specifically, the explanatory powers after Living∩Market, Living∩Industrial, Natural∩Sport, Natural∩Living, and Service∩Living are in the top five of all interaction factors. The biggest differences in explanatory power before and after the interaction are those of Living∩Market, Service∩Living, and Living∩Industrial. The reasons are clear. First, the improvement of people's living standards has greatly satisfied their material needs, allowing people to place greater emphasis on the pursuit of spiritual life. Sports tourism can relax tourists and achieve the purpose of physical and mental pleasure, which can satisfy people's pursuit of spiritual life. Second, the sports tourism industry has gradually increased in importance in the development of China's National Economic Law, which can effectively promote progress in related industries. The sustainable development of sports tourism is inseparable from natural resources, sports resources, and software and hardware services. Consequently, the comprehensive interaction of the above factors can significantly affect the spatial distribution of sports tourism resources. For example, located in the core area of tourism, Anhui has certain advantages in natural resources, transportation, economic benefits, and market cultivation and development. With the relative balance of all influencing factors, sports tourism development in Anhui is comparatively high. In similar situations are Guizhou, Qinghai, Gansu, Shanxi, Inner Mongolia, Yunnan, and Fujian provinces.

Nevertheless, Tibet, Ningxia, and Liaoning, which are the cold spots of sports tourism, are relatively backward in natural resources, sports resources, hardware and software services, transportation capacity, people's living standards, economic benefits, and market cultivation and development. In addition, the levels of these influencing factors show spatial unevenness. However, the above provinces have great potential for improvement, which illustrates that the future development of sports tourism should give full play to the advantages of natural and sports resources and carry out good planning and layout. Making full use of financial support and counting on the tourism public service platform gradually improves the sports tourism service system. Reinforcing the connection and extension of expressways and ordinary roads in remote areas is encouraged to provide better transportation services for tourism. Subsequently, accelerating the development of the tertiary industry to optimize the province's tourism market structure is also a fantastic way to create a distinctive sports tourism brand, which can lead to economic gains and promote related industries through internal penetration, extension, and expansion of the industry.

Table 3. Interaction results and types of detection factors.

	Q	X	Interaction Type	Q	X	Interaction Type
<i>Natural</i> ∩ <i>Sport</i>	0.650	0.527	non-E	0.606	0.197	non-E
<i>Natural</i> ∩ <i>Service</i>	0.545	0.402	non-E	0.306	0.264	bi-E
<i>Natural</i> ∩ <i>Transportation</i>	0.605	0.480	non-E	0.481	0.173	non-E
<i>Natural</i> ∩ <i>Living</i>	0.606	0.411	non-E	0.185	0.230	bi-E
<i>Natural</i> ∩ <i>Industrial</i>	0.519	0.478	bi-E	0.442	0.274	non-E
<i>Natural</i> ∩ <i>Economic</i>	0.448	0.387	bi-E	0.574	0.342	non-E
<i>Natural</i> ∩ <i>Market</i>	0.601	0.444	non-E	0.441	0.251	non-E
<i>Sport</i> ∩ <i>Service</i>	0.350	0.313	bi-E	0.545	0.308	non-E
<i>Sport</i> ∩ <i>Transportation</i>	0.381	0.391	bi-E	0.667	0.273	non-E
<i>Sport</i> ∩ <i>Living</i>	0.449	0.322	non-E	0.333	0.182	non-E
<i>Sport</i> ∩ <i>Industrial</i>	0.383	0.389	bi-E	0.736	0.239	non-E
<i>Sport</i> ∩ <i>Economic</i>	0.436	0.298	non-E	0.506	0.249	non-E
<i>Sport</i> ∩ <i>Market</i>	0.344	0.355	bi-E	0.325	0.306	bi-E
<i>Service</i> ∩ <i>Transportation</i>	0.420	0.266	non-E	0.458	0.215	non-E
			<i>Service</i> ∩ <i>Living</i>			
			<i>Service</i> ∩ <i>Industrial</i>			
			<i>Service</i> ∩ <i>Economic</i>			
			<i>Service</i> ∩ <i>Market</i>			
			<i>Transportation</i> ∩ <i>Living</i>			
			<i>Transportation</i> ∩ <i>Industrial</i>			
			<i>Transportation</i> ∩ <i>Economic</i>			
			<i>Transportation</i> ∩ <i>Market</i>			
			<i>Living</i> ∩ <i>Industrial</i>			
			<i>Living</i> ∩ <i>Economic</i>			
			<i>Living</i> ∩ <i>Market</i>			
			<i>Industrial</i> ∩ <i>Economic</i>			
			<i>Industrial</i> ∩ <i>Market</i>			
			<i>Economic</i> ∩ <i>Market</i>			

Significance level, $p < 0.05$; non-E denotes nonlinear enhanced while bi-E denotes bivariate enhanced.

4. Conclusions

This study comprehensively implemented a combination of the average nearest neighbor, kernel density, and spatial autocorrelation to explore the spatial distribution of 209 and 590 sports tourism boutique projects in 2014 and 2018, respectively. Their influencing factors were determined by combining the entropy method and the geographic probe model. The major takeaways from this study are as follows:

- (1) Before 2014, China's sports tourism resources show that the Yangtze River Delta urban agglomeration is the high-density core area. The Guizhou–Guangxi border area and the western Hubei ecological circle are the sub-density core areas, where the spatial distribution shows obvious agglomeration. The effect of “proximity dependence” between them has not been formed. The distribution of sports tourism resources in the 31 provincial research units across the country can be divided into four stages.
- (2) From 2014 to 2018, the number of China's sports tourism boutique projects increased by 381, with an average annual increase of 127. The regional differences in sports tourism resources in each province tended to converge. In 2018, the urban agglomerations in the Yangtze River Delta are still high-density core areas. However, different from 2014, the sub-density core areas are now the Yunqian border area of the Karst Plateau, the Qinglong border area of the Qilian Mountains, and the Jinji border area of the Taihang Mountains. These areas form the shape of “depending on the city, near the scenery” and “large concentration, small dispersion” and have a clear agglomeration distribution. The effect of “neighboring dependence” has not been formed and is only starting among the provinces.
- (3) Through the coupling and coordination model, the quantity and growth rate of provincial sports tourism resources can be divided into 10 stages. In terms of provincial sports tourism development, 61.29% of provinces are in the coordinated stage. The coordination level of the eastern region is far ahead and significantly higher than the national average. The coordination levels of the northeast, northwest, and parts of the southwest are equal to the national average along with stable development and tendency for coordination. Meanwhile, Sichuan, Chongqing, Hunan, Guangdong, Shaanxi, and other places still have considerable space for progress.
- (4) The influencing factors on the spatial distribution of sports tourism resources show significant variations. The descending order of influence is natural resource endowment > sports resource endowment > transportation capacity > industry support and guidance > market cultivation and development > people's living standards > software and hardware service supporting > economic benefit effect. Moreover, the explanatory power of different two-factor interactions is higher than that of single-factor interactions. The interaction types presented are nonlinear and two-factor enhancements.

5. Discussion

The analysis above shows that China's sports tourism resources present an obvious agglomeration distribution. In their research, Fugao and Li proposed that the development of regional sports tourism has gone through four stages: point symbiosis at the core node of sports tourism, intermittent symbiosis of sports tourism short chain, continuous symbiosis of sports tourism industry chain, and integrated symbiosis of sports tourism industry network. The stages embody the generation and evolution process of the spatial structure of “point–line–surface–domain” [6]. As seen from the results of this study, the spatial distribution of China's sports tourism resources can indeed reflect the spatial structure of “point–line–surface–domain”. The polarization of “spots” is the embryonic stage of sports tourism. Research has found that sports tourism tends to gather in the “spots” that are more developed. The development form is mainly spontaneous, and the spatial form is mainly scattered. The findings of this article are similar to those of Zuo et al. Both the development of sports tourism and the marathon events have emerged in economically developed areas and spread to surrounding areas. For example, they originated from Beijing, Shanghai, and so on, and then spread to Beijing–Tianjin–Hebei region and the Yangtze River Delta.

The analysis also indicates that natural resource endowments and sports resource endowments are the most important factors affecting the spatial distribution of sports tourism in China. It is worth noting that this viewpoint is supported by Kurtzman et al. and Zuo et al. They believe that natural resource endowment is the main source of demand for the development of sports tourism resources. The richer the natural tourism resources, the better the development of sports tourism [26]. The holding of sports events often also promotes the development of the sports tourism industry [7].

As mountain biking, skiing, rock climbing, sailing, swimming, and other sports activities often occur in natural settings such as rivers, mountains, skiing, forests, lakes, seashores, hot springs, grasslands, and other places, the natural resources of tourist destinations are the key attraction for tourists [11]. Besides, our findings are similar to those of Kurtzman and Zauhar, who found that sports events can attract a large number of spectators to participate in tourism [37]. In particular, the sports boutique events attract a large number of sports enthusiasts every year. Among them, the more influential and representative events include the upcoming Beijing Winter Olympics and the Paralympics, as well as a series of international marathons [7]. At present, a large number of studies have shown that transportation capacity restricts the development of tourism. According to our research results, the development of sports tourism is also affected by transportation. The completeness of transportation corridors and other carriers can effectively promote the development of sports tourism activities. Yang et al. believe that the pivotal factor for the sustainable development of sports tourism lies in mass transit planning. Traffic problems often have a great negative impact on sports tourism, while sustainable mass transit planning can reduce the risk of traffic problems [38].

In allusion to the assessment of the spatial distribution pattern and influencing factors of sports tourism resources in China, the following proposals are put forward:

First, the development of sports tourism resources is concerned with local conditions. It is necessary to consider the advantages of local natural resources in combination with the selection of a sports ontology resource as the core attraction. With the help of highlighting unique characteristics, it becomes possible to create fine products and lift the matching degree of local tourism resources and sports tourism. Notably, emphasis on the importance of resources in a region is vital to the development of its sports tourism. Thus, the so-called unique sports events must rely on natural or sports resources with local characteristics. To develop these unique projects, localities must first clarify their own resource advantages. However, the current actual situation is that such natural resources are insufficient without a deep understanding that the core of the sports tourism industry is composed of natural and sports resources. Although the existing abundant sports resources can support the tourism industry for a period of time, creating a unique brand is not sustainable and even difficult. Hence, establishing a concept of “creating a solid basic resource environment” is indispensable. At this time, nature and sports resources may be coordinated to promote the development of the sports tourism industry. A single type of resource should not be overexploited to further develop surrounding sports projects and corresponding industries.

Second, with the opportunity of region-wide sports tourism, the connection and extension of expressways and ordinary roads in remote areas can be improved. This will not only strengthen the construction of tourist passages in border areas, but also elevate the traffic capacity inside and in neighboring provinces. Simultaneously, accelerating the connection of sports characteristic towns and sports tourism may be safely carried out. The construction of high-quality projects and the national sports tourism demonstration base for dedicated roads can expand the radiation range of popular scenic spots and provide better transportation services. In the core areas, consideration of dependence on excellent tourist cities and 5A-level scenic spots can unite sports tourism resources. For evacuation areas, greater attention is needed on the link function of roads, and emphasis should be placed on integrating sports tourism resources in eco-tourism to adapt to the demands of self-driving groups.

Finally, stronger guidance is necessary for local industries to meet the needs of tourists for food, housing, transportation, travel, shopping, entertainment, and other aspects of sports tourism to promote the development of related industries. As a result, more employment opportunities become available for more people, which generates a virtuous circle. According to the local living standards, everything can be carried out within capabilities and invested appropriately. To strengthen cooperation and sharing with the surrounding sports industry market, the government has to increase public financial expenditures for sports and introduce large-scale strategic investors. The enthusiasm of market players should also be fully mobilized. In the meantime, aggrandizement for personnel training is necessary. Routes for mutual promotion, mutual delivery of tourists, information interchange, and complementary features with other types of tourism activities can be formed. Regarding other factors, supporting amenities and the overall service quality of sports tourism should be advanced, which can increase the construction of public service facilities such as sports service and consulting centers. However, certain literature recommends clarifying tourism's own positioning before building supporting restaurants, hotels, or restaurants. Cultivating a new driving force for economic development by accelerating the development of sports tourism is a better option. The awareness of high-quality development of sports tourism can even enhance the quality of sports tourism resources and the ability to develop sustainably.

Through the analysis of the spatial distribution characteristics of sports tourism resources, this study reveals the main driving factors that affect the distribution of sports tourism resources on a national scale. Certain practical significance is provided for the scientific and reasonable layout and the appropriate and sustainable development of sports tourism resources. However, this study still has the following limitations: First, being limited by the availability of selected indicators, tourism boutique projects in recent years could not be selected as the research object, which restricts the topic pertinence and timeliness. Second, as a new mode of integration and development of two industries, sports tourism has been the subject of relatively few quantitative explorations of the coordination and coupling in different regions and the local natural environment, politics, economy, population, and other related factors. Therefore, deepening research on these issues in the future is necessary to enhance the pertinence and effectiveness of sports tourism development measures. For instance, the new pattern of sports tourism resources after 2018 or the factors such as topography, climatic conditions, hydrological conditions, and population economy can be examined. In turn, the effects of reasonable and effective layout, construction according to local conditions, and coordination of the matching resources and sports characteristics can be identified.

Author Contributions: Conceptualization, Yifan Zuo and Yuqi Si; methodology, Yifan Zuo and Yuqi Si; software, Yifan Zuo; validation, Jincheng Pan; investigation, Jincheng Pan; data curation, Yifan Zuo and Huan Chen; writing—original draft preparation, Yifan Zuo, Huan Chen and Rob Law; writing—review and editing, Yifan Zuo, Huan Chen, Rob Law and Mu Zhang; visualization, Yifan Zuo. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Social Science Fund of China (grant No. 19BTY066).

Data Availability Statement: The raw data supporting the conclusions of this manuscript can be made available by the authors to qualified researchers.

Acknowledgments: We would like to thank the three anonymous reviewers and the editors for their valuable comments and suggestions.

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

References

1. Kurtzman, J. Sports tourism categories. *J. Sport Tour.* **2005**, *10*, 15–20. [[CrossRef](#)]
2. Cooper, J.A.; Alderman, D.H. Cancelling March Madness exposes opportunities for a more sustainable sports tourism economy. *Tour. Geogr.* **2020**, *22*, 525–535. [[CrossRef](#)]

3. Nishio, T.; Larke, R.; van Heerde, H.; Melnyk, V. Analysing the motivations of Japanese international sports-fan tourists. *Eur. Sport Manag. Q* **2016**, *16*, 487–501. [[CrossRef](#)]
4. Jin, N.; Lee, H.; Lee, S. Event quality, perceived value, destination image, and behavioral intention of sports events: The case of the IAAF World Championship, Daegu, 2011. *Asia Pac. J. Tour. Res.* **2013**, *18*, 849–864. [[CrossRef](#)]
5. Page, S.J.; Bentley, T.; Walker, L. Tourist safety in New Zealand and Scotland. *Ann. Tour. Res.* **2005**, *32*, 150–166. [[CrossRef](#)]
6. Fugao, J.; Li, C. All-for-One Sports Tourism: Connotation Characteristics, Spatial Structures and Development Patterns. *J. Shanghai Univ. Sport* **2020**, *44*, 12–23.
7. Zuo, Y.; Zou, L.; Zhang, M.; Smith, L.; Yang, L.; Loprinzi, P.D.; Ren, Z. The Temporal and Spatial Evolution of Marathons in China from 2010 to 2018. *Int. J. Environ. Res. Public Health* **2019**, *16*, 5046. [[CrossRef](#)] [[PubMed](#)]
8. Ren, Z.; Zuo, Y.; Ma, Y.; Zhang, M.; Smith, L.; Yang, L.; Loprinzi, P.D.; Yu, Q.; Zou, L. The Natural Environmental Factors Influencing the Spatial Distribution of Marathon Event: A Case Study from China. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2238. [[CrossRef](#)] [[PubMed](#)]
9. Geneletti, D. Impact assessment of proposed ski areas: A GIS approach integrating biological, physical and landscape indicators. *Environ. Impact Assess.* **2008**, *28*, 116–130. [[CrossRef](#)]
10. Zhang, Z.; Ding, Z. The Systematic Spatial Characteristics of Sports Tourism Destinations in the Core Area of Marine Economy. *J. Coastal Res.* **2020**, *112*, 109–111. [[CrossRef](#)]
11. Wu, L. Classification and Spatial Optimization of Sports Tourism Resources in Guangdong Province. *J. Landscape Res.* **2019**, *11*, 110–114.
12. Mansour, S. Spatial analysis of public health facilities in Riyadh Governorate, Saudi Arabia: A GIS-based study to assess geographic variations of service provision and accessibility. *Geo-spat. Inf. Sci.* **2016**, *19*, 26–38. [[CrossRef](#)]
13. Wang, X.; Zou, Z. Open Data Based Urban For-Profit Music Venues Spatial Layout Pattern Discovery. *Sustainability* **2021**, *13*, 6226. [[CrossRef](#)]
14. Yoo, D.; Chun, B.; Min, K.; Lim, J.; Moon, O.; Lee, K. Elucidating the Local Transmission Dynamics of Highly Pathogenic Avian Influenza H5N6 in the Republic of Korea by Integrating Phylogenetic Information. *Pathogens* **2021**, *10*, 691. [[CrossRef](#)]
15. Allen, M.J.; Allen, T.R.; Davis, C.; McLeod, G. Exploring Spatial Patterns of Virginia Tornadoes Using Kernel Density and Space-Time Cube Analysis (1960–2019). *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 310. [[CrossRef](#)]
16. Sokal, R.R.; Oden, N.L. Spatial autocorrelation in biology: 1. Methodology. *Biol. J. Linn. Soc.* **1978**, *10*, 199–228. [[CrossRef](#)]
17. Zhang, X.; Lin, Y.; Cheng, C.; Li, J. Determinant Powers of Socioeconomic Factors and Their Interactive Impacts on Particulate Matter Pollution in North China. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6261. [[CrossRef](#)]
18. Zou, Z.; Yi, Y.; Sun, J. Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment. *J. Environ. Sci. China* **2006**, *18*, 1020–1023. [[CrossRef](#)]
19. Zhang, S.; Xie, X.; Zhao, M. Asset Specificity on the Intention of Farmers to Continue Land Recuperation: Based on the Perspective of Farmer Differentiation. *Land* **2021**, *10*, 603. [[CrossRef](#)]
20. Li, J.; Dou, K.; Wen, S.; Li, Q. Monitoring Index System for Sectors' Digital Transformation and Its Application in China. *Electronics* **2021**, *10*, 1301. [[CrossRef](#)]
21. Wang, J.; Xu, C. Geodetector: Principle and prospective. *Acta Geogr. Sin.* **2017**, *72*, 116–134.
22. Chi, Y.; Qian, T.; Sheng, C.; Xi, C.; Wang, J. Analysis of Differences in the Spatial Distribution among Terrestrial Mammals Using Geodetector—A Case Study of China. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 21. [[CrossRef](#)]
23. Zhu, Z. WSR: A systems approach for information systems development. *Syst. Res. Behav. Sci. Off. J. Int. Fed. Syst. Res.* **2000**, *17*, 183–203. [[CrossRef](#)]
24. Jie, S.; Qingjie, S.; Hongjian, L. Research on the Influencing Factors of Sports Tourism System with WSR Analysis Framework. *China Sport Sci. Technol.* **2010**, *46*, 139–145.
25. Gu, J.; Zhu, Z. Knowing Wuli, sensing Shili, caring for Renli: Methodology of the WSR approach. *Syst. Pract. Act. Res.* **2000**, *13*, 11–20. [[CrossRef](#)]
26. Kurtzman, J.; Zauhar, J. Sports tourism consumer motivation. *J. Sport Tour.* **2005**, *10*, 21–31. [[CrossRef](#)]
27. Perić, M. Estimating the perceived socio-economic impacts of hosting large-scale sport tourism events. *Soc. Sci.* **2018**, *7*, 176. [[CrossRef](#)]
28. Bulatović, J.; Rajović, G. Some aspects of the sport tourism with the review of Montenegro: Overview. *Tour. Educ. Stud. Pract.* **2017**, *4*, 18–28.
29. Yang, S.; Xu, J.; Yang, R. Research on coordination and driving factors of sports industry and regional sustainable development—Empirical research based on panel data of provinces and cities in eastern China. *Sustainability* **2020**, *12*, 813. [[CrossRef](#)]
30. Taleghani, G.R.; Ghafary, A. Providing a management model for the development of sports tourism. *Procedia-Soc. Behav. Sci.* **2014**, *120*, 289–298. [[CrossRef](#)]
31. Jiao, G.; Lu, L.; Chen, G.; Huang, Z.; Cirella, G.T.; Yang, X. Spatiotemporal Characteristics and Influencing Factors of Tourism Revenue in the Yangtze River Delta Urban Agglomeration Region during 2001–2019. *Sustainability* **2021**, *13*, 3658. [[CrossRef](#)]
32. Ruban, D.A. Karst as important resource for geopark-based tourism: Current state and biases. *Resources* **2018**, *7*, 82. [[CrossRef](#)]
33. Donaldson, J.A. Tourism, development and poverty reduction in Guizhou and Yunnan. *China Q.* **2007**, *2007*, 333. [[CrossRef](#)]
34. Wang, Y.; Wu, C.; Wang, F.; Sun, Q.; Wang, X.; Guo, S. Comprehensive evaluation and prediction of tourism ecological security in droughty area national parks—A case study of Qilian Mountain of Zhangye section, China. *Environ. Sci. Pollut. R* **2021**, *28*, 16816–16829. [[CrossRef](#)]

35. Cao, X. Challenges and potential improvements in the policy and regulatory framework for sustainable tourism planning in China: The case of Shanxi Province. *J. Sustain. Tour.* **2015**, *23*, 455–476. [[CrossRef](#)]
36. Shi, T.; Yang, S.; Zhang, W.; Zhou, Q. Coupling coordination degree measurement and spatiotemporal heterogeneity between economic development and ecological environment—Empirical evidence from tropical and subtropical regions of China. *J. Clean. Prod.* **2020**, *244*, 118739. [[CrossRef](#)]
37. Kurtzman, J.; Zauhar, J. A wave in time-The sports tourism phenomena. *J. Sport Tour.* **2003**, *8*, 35–47. [[CrossRef](#)]
38. Yang, J.; Chuang, Y.; Lo, H.; Lee, T. A Two-Stage MCDM Model for Exploring the Influential Relationships of Sustainable Sports Tourism Criteria in Taichung City. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2319. [[CrossRef](#)] [[PubMed](#)]

Article

Identifying the Relatedness between Tourism Attractions from Online Reviews with Heterogeneous Information Network Embedding

Peiyuan Qiu ^{1,2}, Jialiang Gao ^{2,3} and Feng Lu ^{2,3,4,5,*}

¹ School of Surveying and Geo-Informatics, Shandong Jianzhu University, Jinan 250101, China; qiupeiyuan20@sdjzu.edu.cn

² State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; gaojl@lreis.ac.cn

³ University of Chinese Academy of Sciences, Beijing 100049, China

⁴ The Academy of Digital China, Fuzhou University, Fuzhou 350002, China

⁵ Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

* Correspondence: luf@lreis.ac.cn; Tel.: +86-10-6488-8966

Abstract: The relatedness between tourism attractions can be used in a variety of tourism applications, such as destination collaboration, commercial marketing, travel recommendations, and so on. Existing studies have identified the relatedness between attractions through measuring their co-occurrence—these attractions are mentioned in a text at the same time—extracted from online tourism reviews. However, the implicit semantic information in these reviews, which definitely contributes to modelling the relatedness from a more comprehensive perspective, is ignored due to the difficulty of quantifying the importance of different dimensions of information and fusing them. In this study, we considered both the co-occurrence and images of attractions and introduce a heterogeneous information network (HIN) to reorganize the online reviews representing this information, and then used HIN embedding to comprehensively identify the relatedness between attractions. First, an online review-oriented HIN was designed to form the different types of elements in the reviews. Second, a topic model was employed to extract the nodes of the HIN from the review texts. Third, an HIN embedding model was used to capture the semantics in the HIN, which comprehensively represents the attractions with low-dimensional vectors. Finally, the relatedness between attractions was identified by calculating the similarity of their vectors. The method was validated with mass tourism reviews from the popular online platform MaFengWo. It is argued that the proposed HIN effectively expresses the semantics of attraction co-occurrences and attraction images in reviews, and the HIN embedding captures the differences in these semantics, which facilitates the identification of the relatedness between attractions.

Keywords: relatedness between attractions; online tourism reviews; heterogeneous information network; embedding; attraction image; topic extraction

Citation: Qiu, P.; Gao, J.; Lu, F. Identifying the Relatedness between Tourism Attractions from Online Reviews with Heterogeneous Information Network Embedding. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 797. <https://doi.org/10.3390/ijgi10120797>

Academic Editors: Andrea Marchetti, Angelica Lo Duca and Wolfgang Kainz

Received: 28 August 2021

Accepted: 26 November 2021

Published: 29 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The relatedness between geographic objects captures a broad relation between objects that can be close or far apart in location, can be linked by interaction, or may simply share a common property [1]. Identifying the relatedness between tourism attractions can be used in a variety of tourism applications, such as (1) destination collaboration, e.g., evaluating the connection between attractions and find the core attractions in a tourist destination [2]; (2) commercial marketing, e.g., testing how changes in links between destinations influence market equilibrium [3]; (3) travel recommendation, e.g., recognizing the popular tourist areas for tourism route recommendation based on the interactions between attractions [4].

In recent years, with the development of ICT (information and communications technology), big data, such as UGC (user-generated content) data, device data, and transaction data, has made great contributions to improving tourism research [5]. In particular, massive travel reviews of tourists are becoming easily accessible through social networks, such as Yelp, TripAdvisor, Booking, and so on. These reviews support the different types of information about visited attractions, visited times, travel notes and basic profiles of tourists, labels, ranks, review texts, and basic attributes of attractions. Intuitively, the relatedness between attractions can be identified by measuring the co-occurrence of attractions from the above information: the higher the frequency of co-occurrence of attractions (namely, the attractions are mentioned more in the information at the same time), the stronger the relatedness between them. On the one hand, the co-occurrence of attractions is reflected in the lists of tourists' visited attractions, which can be used to construct an attraction flow network. Then, the relatedness between attractions can be identified with network analytics. The results of identified relatedness are helpful to cognize the tourism movement patterns [6,7], evaluate the market position of different attractions [7,8], and reveal the factors affecting the network structure of the tourist flows [9,10]. On the other hand, the co-occurrence of attractions is expressed in review texts or travel note texts. For example, Haris et al. extracted the semantic relationships between tourist places from travel notes through the natural language processing (NLP) technique, then constructed a points of interest (POIs) graph to find the popular attractions and popular trip patterns which consist of the related attractions [11]. Yuan et al. implemented the frequent pattern mining method to identify the city's popular locations by their sequenced co-occurrences from travel blogs, then develop a max-confidence-based method to detect travel routes from the popular location network [12].

In addition to the co-occurrence of attractions, the implicit semantic information in tourism online reviews definitely contributes to modelling the relatedness from a more comprehensive perspective. The attraction image is one of these information types, which is the impression attractions on tourists, and it has different topics, such as the attractions to be seen (e.g., sand and beach), the environment to be perceived (e.g., weather, public hygiene), and experiences to remember (e.g., surfing, swimming) [13]. Thus, if two attractions have more similar images, they will have a stronger relatedness. Due to the attraction image being described in review texts and travel note texts, a topic model can be used to "understand" and extract the attraction image topics from these texts and divide the images into different semantic dimensions. The topic model is a probabilistic model for uncovering the underlying the semantic structure of a document collection based on a hierarchical Bayesian analysis of the original texts [14]. In tourism research, the topic model is used to discover the abstract "topics" in texts [15,16]. Then, the attraction images by tourists in different dimensions are obtained by fusing the topics related to this attraction, and the relatedness between attractions can be measured. The extracted attraction images facilitate the tourism destination analysis [13,17] or tourism personalized recommendation [18–21].

The key to using multi-dimensional semantic information to comprehensively identify relatedness is to quantify the importance of different dimensions of information and fuse them. That is, if two attractions have a higher frequency of co-occurrence, or more similar images, or both, they should have stronger relatedness. Determining the importance of these from massive online travel reviews manually is difficult. Thus, in this paper, we introduce a heterogeneous information network (HIN) to represent the tourism online reviews to characterise the co-occurrence and images of attractions, then comprehensively identify the relatedness between attractions through the HIN embedding technique automatically.

In the HIN, the type of nodes (or objects) or edges (or relations, links) is greater than one [22]. Therefore, the HIN can better model the real interacting system existing in multiple types of relationships. For example, a bibliographic information network can be organized as a HIN, which expresses many facts "one or more authors written a paper", "a paper has been published in a venue", and "a paper cited one or more papers" [23]. In this HIN, the types of nodes are "author", "paper", and "venue", and the types of edges are

“written” (links author and paper), “published in” (links paper and venue), and “citing” (links paper and paper). Then, the relationships between authors can be characterized with the semantics of research area topic from this HIN compared with the homogeneous information network. Moreover, the social network [24–26] and bioinformatic network [27–29] have been modelled as HINs. HINs have been applied to massive tasks as clustering, classification, link prediction, ranking, recommendation, information fusion, influence propagation, and so on [30]. The HIN embedding technique characterizes the nodes of HIN with low-dimensional vectors, i.e., embeddings [24]. Then, the semantic information is embedded in the low-dimensional vector space and the relationships between nodes can be calculated by vector operation.

Taking the HIN’s advantage in expressing the different types of semantics between nodes, we utilize it to represent the tourism online reviews and use HIN embedding to comprehensively identify the relatedness between attractions. First, an online review-oriented HIN is designed to form the different types of elements in the reviews. Second, a topic model is employed to extract the nodes of the HIN from the review texts. Third, an HIN embedding model is used to capture the semantics in the HIN and comprehensively represent the attractions with low-dimensional vectors. Finally, we conduct several experiments to verify the effectiveness of the proposed method.

The remainder of this paper is structured as follows. Section 2 proposes the structure of an online review HIN, the construction method, and the embedding method of this HIN. Section 3 conducts a case study using online tourist review data. Section 4 is devoted to discussions, and Section 5 concludes this work.

2. Materials and Methods

The procedure of identifying the relatedness between tourism attractions from online reviews with HIN embedding is shown as Figure 1. Firstly, a structure of HIN is designed to represent the tourism online reviews. Next, the original online reviews are transformed into the form of the proposed HIN through direct extraction and image topic extraction. Then, the attractions in HIN are embedded into the n-dimensional vectors by HIN embedding technology. Finally, the relatedness between attractions is calculated based on the vector similarity.

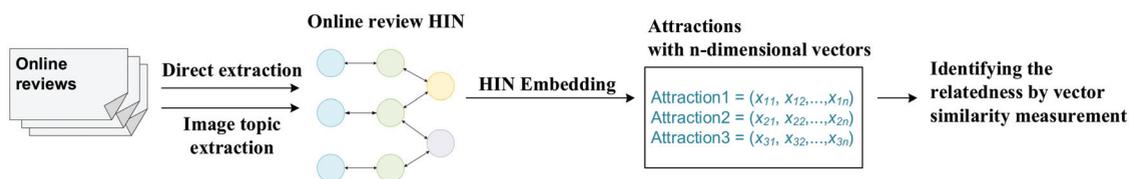


Figure 1. Flowchart for identifying the relatedness between tourism attractions with HIN embedding.

2.1. Online Review HIN Structure

In this research, we built an HIN for representing tourists’ online reviews. Online reviews support which attractions are visited and the attraction images of tourists. Specifically, the attraction image in review is expressed around one or more topics, such as cost, dining, feature of attraction, traffic, and so on. So, the types of nodes in the proposed online review HIN are “attraction”, “tourist”, “review”, and “topic”. The types of edges between these nodes are “havingreview” (an attraction has a review), “reviewof” (a review of an attraction), “writing” (a tourist writes a review), “writtenby” (a review is written by a tourist), “hastopic” (a review has a topic) and “topicof” (an image topic of a review). Figure 2 illustrates an example of the online review HIN from four reviews about two tourists, three attractions and two topics.

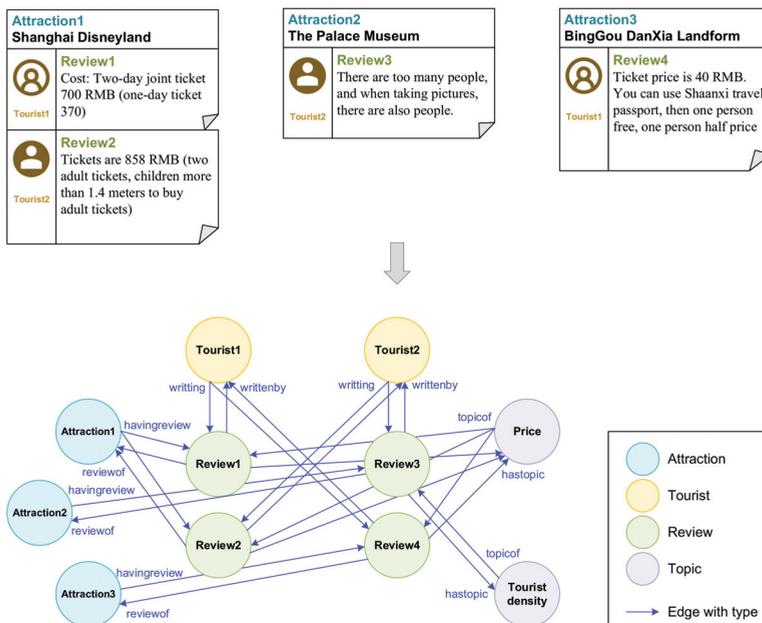


Figure 2. Example of an online review HIN.

In this online review HIN, the node path “attraction” → “review” → “tourist” → “review” → “attraction”, that is, the edge path “havingreview” → “writtenby” → “writing” → “reviewof”, holds the co-occurrence of attractions visited by the same tourists, and the node path “attraction” → “review” → “topic” → “review” → “attraction”, that is, the edge path “havingreview” → “hastopic” → “topicof” → “reviewof”, holds the relationship between attractions by the same topics of attraction images. Thus, this online review HIN expresses the co-occurrence of attractions and attraction images through the above long hop paths.

2.2. Topic Extraction and HIN Construction

The key task of constructing the presented online review HIN is extracting the nodes and edges from the reviews. The nodes “attraction” and “review” and their edge “havingreview”/“reviewof” can be directly extracted from the review list of the attraction. For the nodes “tourist” and “review”, their edge “writing”/“writtenby” can be directly parsed from the basic information of the review, which contains tourist name, score given to an attraction, time of posting the review, etc. However, the image topic is not provided as basic information by the online review, so the node “topic” and the edge “hastopic”/“topicof” between “review” and “topic” are not directly extracted from the online review. Meanwhile, the image topic can be represented by certain words which make up the review text, so the image topic can be acquired from the review text through topic extraction.

Topic models are widely used for extracting abstract “topics” and hidden semantic structures from vast textual documents. Topic models as unsupervised machine learning models can automatically analyse the documents in the corpus and extract potential topics according to the co-occurrence of words in documents. For example, particular words such as “train”, “subway” and “taxi” would co-occur more frequently in a document about the topic “traffic”. In this study, we use the Latent Dirichlet Allocation (LDA) model [31], which is the most popular topic model, to extract the topics of review from the review text. Inputting several documents, the two main outputs of the LDA model are the probabilities that each document belongs to the different topics and the high-frequency keywords

of each topic. Then, the meaning of each topic can be summed up manually from its high-frequency keywords.

However, the original LDA model experiences large performance degradation over short texts due to the lack of word co-occurrence information in each short text [32]. Meanwhile, most of the tourism online review texts are short texts, and the word count in these texts is less than 100. Thus, we introduce the word embedding technique to extend the context of online tourism review texts to meet the word count requirement for the original LDA. For word embedding, the words in the corpus are encoded into a continuous low-dimensional semantic vector space, where each word is represented by a fixed dimensional real-valued vector [33,34]. For instance, the words “France” and “U.S.A” are represented by the 200-dimensional real-valued vectors, respectively, through word embedding; then, their distance can be calculated in the vector space. If the distance between two words is close, these words have similar semantics or related semantics [35]. For example, the distance between “France” and “U.S.A” (or “France” and “French”) is less than the distance between “France” and “Mountain” in the vector space. Thus, words with similar semantics to the original words in a review text can be obtained through a similarity calculation.

The detailed procedure of acquiring the edges “hastopic”/“topicof” between “review” and “topic” from the online reviews through topic extraction is shown in Figure 3.

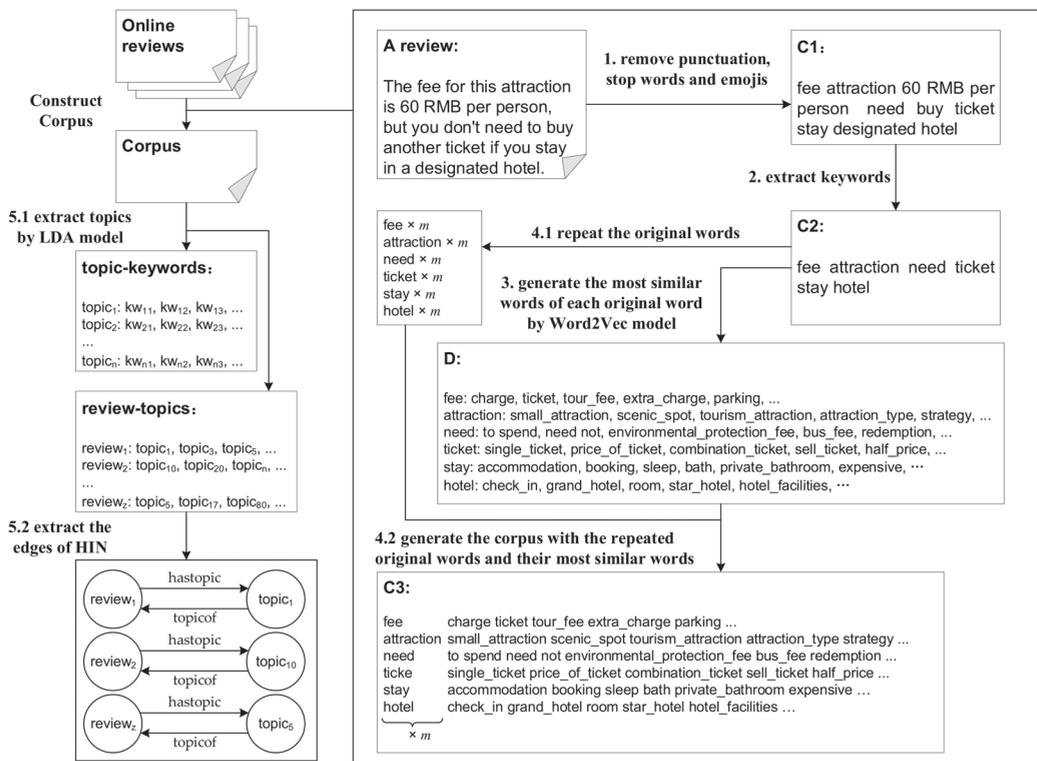


Figure 3. Flowchart for acquiring the edges “hastopic”/“topicof” between nodes “review” and “topic” through topic extraction.

Firstly, the punctuation, stop words, and emojis are removed from the original review text to reduce the interference of this meaningless information on the subsequent processing. The processed texts form a corpus “C1”.

Secondly, a TextRank [36] algorithm is conducted to extract the keywords of each review in the corpus “C1” for highlighting the key information in review. The extracted k keywords of each review represent this review and form a new corpus “C2”.

Thirdly, we use the word embedding model Word2Vec to obtain the low-dimensional semantic vectors of each word in the corpus “C2”. Then, the semantic similarity between words can be measured by the cosine similarity as follows:

$$\text{CosSim}(x, y) = \cos(\theta) = \frac{x \cdot y}{\|x\| \|y\|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (1)$$

where x and y are the vectors of two words. x_i and y_i are components of vector x and y , respectively.

For one word, the semantic similarities between this word and each other word can be measured by Equation (1) and ranked in ascending order. Then, a dictionary records the top l most similar words of each word built. We can use this dictionary “D” to quickly obtain a similar word set of an input word.

Fourthly, each word of the review in the corpus “C2” has l semantic similar words as its extended words from the dictionary “D”. The original words and their extended words in the corpus “C2” consist of a new corpus, e.g., extending the context of reviews. To avoid the importance of original words being diluted by their extended words, the original words can be repeated m times, respectively, in the new corpus. The final corpus is named “C3”.

Fifthly, the number n topics, with their high-frequency keywords and the probabilities that each review belongs to the different topics, were obtained through using the corpus “C3” to train the LDA model. The meaning of each topic can be summed up manually from its high-frequency keywords. The topic with the highest probability is the image topic of a review. Then, the edges “hastopic”/“topicof” between reviews and topics are constructed from the reviews and their topics.

2.3. HIN Embedding and Identifying the Relatedness between Attractions

In order to achieve good performance in such tasks as clustering, classification, link prediction, recommendation, etc., the HIN embedding technique is proposed to embed the nodes of HIN into low-dimensional vectors, and then the embedded nodes can be input into the advanced machine learning models. In recent years, many HIN embedding models have been proposed, such as Metapath2Vec [37], HIN2Vec [38], HAN [39], HERec [24], and so on. While these models have been used to represent the nodes in HINs of a bibliography (e.g., from DBLP, AMiner), social media platforms (e.g., from an online blog, Flickr, Yelp, Douban), bioinformatics (e.g., from HMDD, aBiofilm), etc., they have not been applied to the HIN of tourism information before.

In this research, we select the HIN embedding model HIN2Vec to embed the online review HIN. The HIN2Vec model captures the semantic information contained in meta-paths (namely the node path or edge path mentioned in Section 2.1) and the whole network structure. Then, the relevant nodes which have semantic relationships are close to each other in the low-dimensional vector space. Compared with other HIN embedding models, the HIN2Vec model automatically constructs meta-paths with a given path length and captures the semantic information in these meta-paths instead of the limited short hop (one-hop or two-hop) meta-paths in other models. Thus, HIN2Vec can capture the semantic information in the long hop meta-paths of the online review HIN mentioned in Section 2.1: “havingreview” → “writtenby” → “writing” → “reviewof” and “havingreview” → “hastopic” → “topicof” → “reviewof”.

Specially, the HIN2Vec model is a neural network model which learns the low-dimensional vectors of nodes and edges in HIN by a prediction task: input nodes x , nodes y and edges r to the model to predict whether r exists between x and y . The structure of the HIN2Vec model is shown in Figure 4. The input layer accepts the one-hot vectors \vec{x} , \vec{y} and \vec{r} of x , y and r . The latent layer transforms \vec{x} , \vec{y} and \vec{r} into latent vectors $W'_x \vec{x}$,

$W'_Y \vec{y}$ and $f_{01}(W'_R \vec{r})$ in the d-dimensional vector space. Then, a Hadamard function is used to aggregate these latent vectors and an Identity function is applied for activation. Finally, the output layer uses the Summation as the input function and the Sigmoid function for activation to finish the prediction. The goal of the HIN2Vec model is to learn the optimal vectors $W'_X \vec{x}$, $W'_Y \vec{y}$ and $f_{01}(W'_R \vec{r})$ of x , y and r to ensure that the predicting result is true if r exists between x and y in the real HIN, and false if r does not exist between x and y in the real HIN.

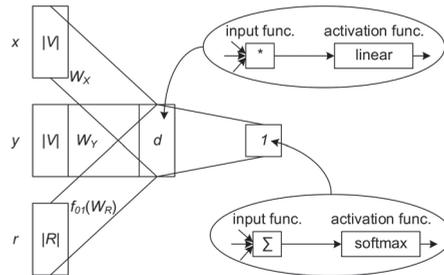


Figure 4. The structure of HIN2Vec model [38].

The process of identifying the relatedness between attractions through the HIN embedding model HIN2Vec is shown in Figure 5. Each edge in the online review HIN is re-represented by the tuple form $\langle node_i, node_j, edge_k \rangle$ for meeting the input of the model’s prediction task and used to train a HIN2Vec model, where $node_i$ and $node_j$ are the head node and tail node in the edge $edge_k$. Then, the vectors of “attraction” nodes are extracted from the trained HIN2Vec model. Finally, the relatedness between two attractions can be identified by a variety of vector similarity measurements such as Euclidean distance, Manhattan distance, cosine similarity, and so on, according to applications.

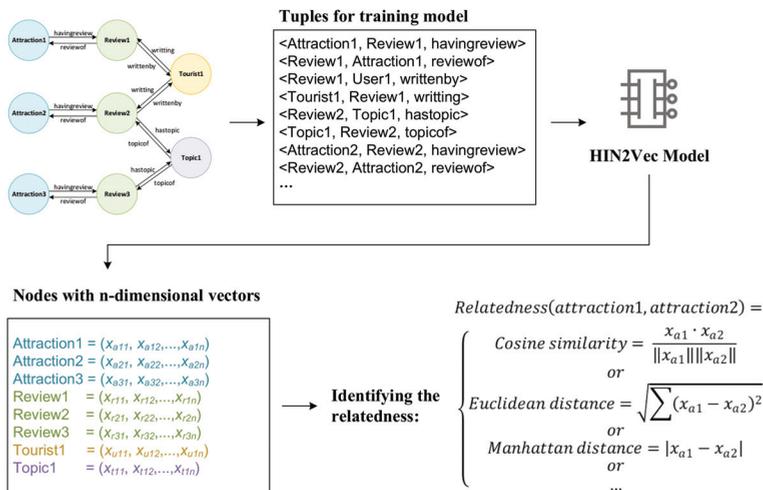


Figure 5. The process of identifying the relatedness between attractions through HIN2Vec model.

3. Case Study

In this section, we verify the performance of the proposed method with the mass tourism reviews. Firstly, the tourism review data and the constructed online review HIN are described. Then, three experiments are conducted: (1) visualization of the HIN embedding result, (2) top related attractions finding, and (3) attractions clustering.

3.1. Review Data

The tourism online review data were collected from the popular tourist-oriented information sharing platform MaFengWo (www.mafengwo.cn/). We selected attractions with reviews from within China (except Hong Kong, Macao, and Taiwan), and the period of reviews was from 2014 to 2018. Moreover, to ensure that each attraction had enough reviews for extracting tourists and topics to build paths to other attractions, attractions with fewer than 20 reviews were filtered out. The final review data to conduct the experiments contained 11,122 attractions, 202,777 tourists, and 1,087,438 reviews. The spatial distribution of attractions is shown in Figure 6.



Figure 6. Spatial distribution of attractions in the review data. (Base map is obtained from Map World: <http://lbs.tianditu.gov.cn/server/MapService.html>).

3.2. Online Review HIN Construction

3.2.1. Image Topic Extraction

In the attraction image topic extraction, some model parameters are set considering the amount of data and efficiency. Primarily, the original review text was segmented into word sequences using the Chinese word segmentation tool because Chinese texts do not use space or another symbol to indicate different words. We used the HanLP2 tool (www.hanlp.com/) to segment the tourist reviews. Then, for TextRank which is also implemented in the HanLP2 tool, the maximum number k of keywords extracted was 50. Next, we used the gensim tool (radimrehurek.com/gensim/) to train the Word2Vec and

LDA models. For the Word2Vec model, the dimension of the vector is 250, the window is 5, the minimum word frequency is 5, and the skip-gram model [34] was selected. The number l of words used to extend the context of reviews is 25, and the repeat times m of original words to avoid the importance of original words being diluted by their extended words is 12. Then, the average word number of reviews in the corpora “C1”, “C2”, and “C3” are 21.98, 12.21, and 440.78, respectively. So, the length of the reviews in corpus “C3” with extended context is suitable for the LDA model to extract image topics. The topic number n of the LDA model was set as 200 to fully distinguish the semantic differences between potential image topics. All training was conducted on a computer equipped with two 2.20 GHz Intel Xeon CPUs and 128 GB RAM.

The probabilities that each document belongs to the different topics and the high-frequency keywords of each topic are the two main outputs of LDA. The extracted 200 image topics can be further divided into 13 categories and 155 sub-categories through manually interpreting the high-frequency keywords of each image topic. The categories and sub-categories are shown in Table 1. Table 2 shows 20 image topics of the above 155 sub-categories and their top 10 high-frequency keywords. These results indicate that the proposed image topic extraction method can capture the semantic difference in the reviews, which will facilitate the relatedness identification between attractions. Finally, the topic with the highest probability is the image topic of a review. It needs to be emphasized that the interpreted categories were used for understanding the meanings of images and to verify the validity of the topic extraction results. So, we kept all the extracted 200 image topics in the next HIN construction instead of merging the topics that belonged to the same category for the HIN2Vec model, which captures the slight semantic differences between the topics.

Table 1. Image topic category and sub-categories (the number of extracted topics belong to the category is in the brackets).

Category	Sub-Categories
dining (5)	no sub-category (2); seafood (1); farmer meal (1); flavor (1)
cost (6)	no sub-category (1); ticket (4); discount (1)
environment (25)	no sub-category (1); quietness (1); safety (1); furnishings (1); panorama (1); scenery (7); air (1); facility (2); beauty (1); tourist density (2); color (2); hygiene (2); atmosphere (1); sunshine (1); vegetation (1)
advice (9)	no sub-category (6); book ticket (1); tour guide (1); scheduling (1)
traffic (11)	no sub-category (3); walk (2); subway (1); bus (1); riding (1); parking (1); road trip (1); rental car (1)
experience (26)	no sub-category (24); climbing mountain (2)
feature of attraction (58)	ice & snow (1); museum (2); urban core (1); Hmong village (1); village (1); panda (1); landmark (1); sculpture (1); animal (2); cave (1); high-rise building (1); college (1); park (1); historic site (1); piazza (1); beach (1); aquarium (1); waterfall (1); lake (1); flower (3); building (1); river (1); alley & street (1); attraction (2); Kaifeng (1); national custom (2); bridge (1); place to film (1); forest (1); mountain (2); business (1); stone carving (1); water town (1); temple (1); theme park (1); railroad (1); hot spring (1); protected historic site (2); cultural heritage (1); canyon (1); modern city (1); town (1); recreation (1); art (1); ginkgo (1); playground (1); garden (1); arboretum (1); natural landform (1); natural scenery (1)
introduction (22)	no sub-category (3); preservation condition (1); locals (1); allusion (1); style (1); custom (1); scale (1); building structure (1); history (3); area (1); entrance (2); picture (1); location (2); development (1); culture (1); religion (1)
evaluate (15)	no sub-category (2); negative (4); positive (9)
time (5)	peak and slack season (1); queuing time (1); opening time (2); duration (1)
activity (16)	no sub-category (1); boating (1); taking cable car (1); temple fair (1); rafting (1); family activity (1); shopping (2); photography (3); show (3); exhibition (1); worship (1)
accommodation (1)	no sub-category (1)
MaFengWo self (1)	no sub-category (1)

Table 2. Examples of image topics and top 10 high-frequency keywords.

Topic ID	Category	Top 10 High-Frequency Keywords
0	activity: show	role; music; image; wax; vivid (2); lifelike; dance; song and dance; wonderful
16	feature of attraction: landmark	different; landmark (3); seem; county; mark; road; landmark building; difference
21	time: opening time	aspect; open (2); crowds; crowd; close; museum close; off duty; closed on Monday
26	experience	pass by; relax; excellent; just pass by; highlight; destination; cannot miss; by pure chance; specially; mind
28	cost: ticket	free of charge; free to visit; watch; open for free; Admission with ID; get ticket; exchange; verify ID (2); Admission with ticket
34	evaluate: positive	be worth; tourism; be very worth; small and beautiful; nice nice nice; nice; with somebody; play; be worth to come
40	dining: seafood	seafood; characteristic; fresh; bathroom; mantis shrimp; seafood market; Musculus senhousei; taste; food stall; corner
41	traffic: parking	park; parking (2); parking fee; hold; park at; park at will; no bathroom; toll collector; park to the side of the road
54	introduction: history	a part; founded in (2); built in; divided into; AD; access; formal; period; old name
77	activity: photography	take photo with (3); take photo (3); Fairy Lake; be tempted to; population; area
85	environment: air	air; fun; breathe (2); interesting; natural oxygen bar; fresh air; facility; site; anion
97	feature of attraction: panda	lovely; panda; cute (2); panda kindergarten; very cute; charmingly nave; cute critter; so lazy; see panda
109	introduction: location	located (5); adjoin; to the north; take road as boundary; west of; to the west
126	environment: tourist density	tourist; protect; worshipper; foreign tourist; endless stream of tourists; surge; vociferously; popular; large party; too popular
140	evaluate: negative	problem; owner; manage; on file; warn; attitude; service attitude; very poor; chaos; servant
169	accommodation	hotel (2); romantic; see; big bedroom; check in; suite; starred hotel; booking; standard room
171	experience: climbing mountain	up; add; flower perfume; up to; climbing mountain; climb up; quite steep (2); wear knee; tired
183	feature of attraction: beach	sand beach; seawater; beach; comfortable; sand; white sand; soft; sandiness; swimming; fine sand beach
191	advice: booking ticket	in advance; put in; luggage; tourist center; information; book on; taobao; book; online shopping; buy ticket
195	traffic: subway	subway; scene; subway station; ordinary people; seek; Line 1; station; Xintiandi Station; Xincheng Station; Line 10

Chinese words translated into the same English word are merged and the number of these Chinese words is shown in brackets.

3.2.2. Online Review HIN

The final online review HIN was constructed based on the nodes and edges acquired from the above original review data and image topic extraction result. The online review HIN contains 1,301,537 nodes and 7,017,522 edges. The number of different types of nodes and edges is shown in Table 3.

Table 3. Statistics of nodes and edges in the online review HIN.

Node			
#Attraction	#Tourist	#Review	#Topic
11,122	202,777	1,087,438	200
Edge			
#Having Review/Review of	#Writing/Written by	#Hastopic/Topic of	
2,176,454	2,176,454	2,664,614	

3.3. Online Review HIN Embedding

The HIN2Vec model was implemented by the open code of the author (<https://github.com/csiesheep/hin2vec>). For training a HIN2Vec model, some important param-

eters were set considering the amount of data and efficiency: the dimension of vectors is 150, the number of negative samples is 5, and the length of random walks is 1000. The length of meta-paths was set to 4 to capture the semantics in the two edge paths “havingreview” → “writtenby” → “writing” → “reviewof” and “havingreview” → “hastopic” → “topicof” → “reviewof” presented in Section 2.1.

Inspired by the work of Liu et al. [40], we used t-SNE to reduce the HIN2Vec embedding result with 150 dimensions to two dimensions for visualization on a two-dimensional plane. The results are shown in Figure 7: (a) is the visualization of all nodes, and (b) is that of the nodes except for the “review” nodes. It illustrates that all nodes are mixed in the visualization result, but the nodes of “attraction”, “topic”, and “tourist” are grouped. The possible reason is the “review” nodes are connected with all the other kinds of nodes in the online review HIN, so the HIN2Vec model cannot discriminate the difference of semantics between “review” nodes and other kinds of nodes in the embedding process. Consequently, the HIN2Vec model captures the semantic differences between attractions, topics, and tourists, which ensures the effectiveness of the relatedness identification between attractions.

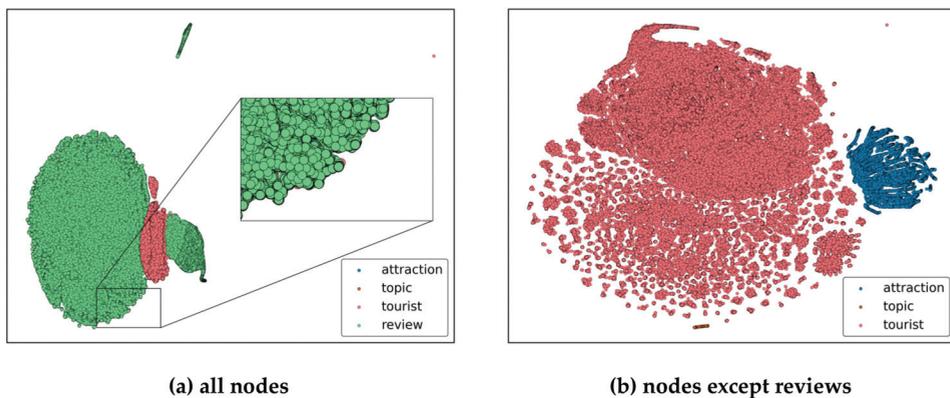


Figure 7. Visualization of the embedding result.

3.4. Top Related Attractions Finding

This experiment was conducted to find the top related attractions of a given attraction to verify the presented method. The relatedness $rel_{hin}(a_i, a_j)$ between attraction a_i and a_j based on online review HIN embedding was identified through measuring the cosine similarity between the vectors of attractions, which is a common metric of measuring the similarity between high-dimensional vectors in machine learning.

3.4.1. Comparative Relatedness Identification Methods

We used two comparison relatedness identifying methods based on homogeneous co-occurrence attraction network embedding and image topic distribution as the contrasts of the proposed relatedness identification.

(1) Relatedness Identification Based on Homogeneous Network Embedding

We built a homogeneous co-occurrence attraction network from the assumption “a tourist written a review text to a tourism attraction” means “this tourist has visited this tourism attraction”. Thus, if a tourist wrote different reviews of different tourism attractions, this tourist has visited all these tourism attractions. That is, these tourism attractions co-occur, which can be used to identify the relatedness between attractions, as mentioned in Section 1. Specifically, the node in the homogeneous co-occurrence attraction network is attraction. The edge represents that its two nodes (attractions) have been visited by the same tourists. Moreover, the edge has a weight to indicate the number of the same tourists.

A higher weight of the edge means that the nodes (attractions) of this edge are visited together more frequently.

Then, the homogeneous network embedding model LINE (Large-scale Information Network Embedding) was used to characterize the nodes with low-dimensional vectors. The LINE model is suitable for undirected, directed, and/or weighted networks containing millions of nodes [41]. This model (1) captures the first-order proximity between the nodes of the observed links in the network, and (2) explores the second-order proximity between the nodes, which is not measured through the observed links but through the shared neighborhood structures of the nodes. Thus, the LINE model can solve the problem of sparse edges in the large real homogeneous network, which leads to poor performance of node embedding.

The LINE model was implemented by the open code of the author (<https://github.com/tangjianpku/LINE>). For training a LINE model, some important parameters were set: the vector dimension was 128, the number of negative samples was 5, the total number of training samples is 10,000, the edge is undirected, and the *first-order* and second-order proximity were both used. Similarly, to the result of the HIN2Vec model, the relatedness $rel_line(a_i, a_j)$ between a_i and a_j was also identified by calculating their cosine similarity in the vector space embedded by the LINE model.

(2) Relatedness Identification Based on Image Topic Distribution

An attraction has many different reviews, and a review has an image topic, so an attraction has different image topics, namely image topic distribution of this attraction. The image topic distribution of attraction can form a vector of this attraction: the vector dimension is the number of all image topics, and the dimension value is the reviews' number that belongs to the corresponding topic. Thus, the relatedness between two attractions was identified by these vectors: the high relatedness means these two attractions have similar image topic semantics. Specifically, the numbers of an attraction's reviews belonging to each image topic are counted from the result of topic extraction and as the dimension values. Therefore, the vector dimension was 200, consistent with the parameter of the LDA model. After normalizing each vector of the attraction, the relatedness $rel_topic(a_i, a_j)$ between a_i and a_j was measured by the cosine similarity.

3.4.2. Results

Each attraction can obtain its top 1000 related attractions by identifying and sorting the rel_line , rel_topic , and rel_hin , which reflects the perspectives of attraction co-occurrence, image topic semantics, and HIN, respectively. Figure 8 shows the spatial distribution of the top 1000 related attractions of five attractions sorting by rel_line , rel_topic , and rel_hin (abbreviated as SD_line , SD_topic , and SD_hin , respectively, for brevity): the Palace Museum, Shanghai Disneyland, Qingdao Trestle, Mount Siguniang, and Potala Palace. To observe the difference in spatial distribution more clearly, the Kernel Density Estimation (KDE) surface generating from the top related attractions overlays each map. This figure illustrates that, compared with the attractions in SD_hin and SD_topic , the attractions in SD_line were closer to the given attractions. This phenomenon is consistent with the notion that frequent pairwise occurrences of points of interest (POIs) indicate their geographic proximity [11] because the SD_line is conducted from the co-occurrence attraction network. Meanwhile, compared with the attractions in SD_hin and SD_line , the attractions in SD_topic were more scattered in China (e.g., the high-density surfaces are greater). The reason is that the geographic proximity of attractions' image topics is not significant. For some image topics relating to certain types of natural terrain, the spatial distributions of these topics may present some rules. For instance, Qingdao Trestle is a wharf that stretches into the sea at Qingdao, so most of its attractions in SD_topic are located on the coastline of China. Overall, the SD_hin is situated between the SD_line and SD_topic , showing that the proposed method identifies the relatedness between attractions from the perspectives of attraction co-occurrence and attraction image topic comprehensively.

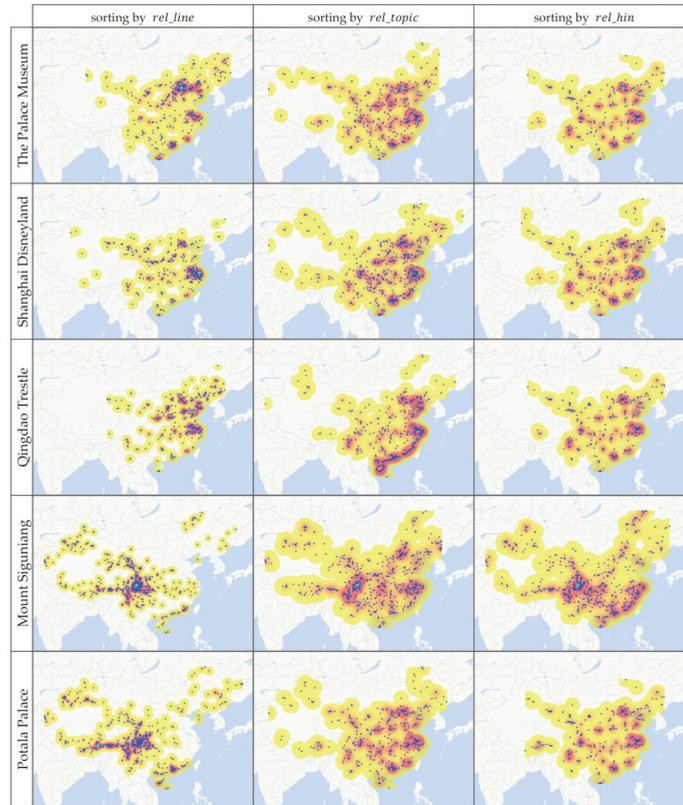


Figure 8. Spatial distribution of the top 1000 related attractions of the given attractions sorting by different relatedness identification (the cyan points are the given attractions; the blue points are the top related attractions; the “yellow-red” surfaces are the KDE surfaces generating from the top related attractions: “yellow” indicates a low density of attractions, and “red” indicates a high density of attractions).

3.4.3. Efficiency Analysis

We calculated the rel_line and rel_topic between each attraction and its top 1000 related attractions which were sorted by rel_hin . The statistical indicators’ average, median, first quartile, and third quartile of the rel_line , rel_topic , and rel_hin on each sorting position are shown in Figure 9. This figure illustrates that the tendencies of rel_line and rel_topic both decreased when the rel_hin decreased. This result indicates that the HIN2Vec model is most efficient in terms of fusing information of attraction co-occurrence and the image topic semantics to comprehensively identify the relatedness between attractions.

Furthermore, we calculated the distances between each attraction and its top 1000 related attractions which were sorted by rel_line , rel_topic , and rel_hin , respectively. The statistical indicators’ average, median, first quartile, and third quartile of the distances on each sorting position are shown in Figure 10. It can be seen that the distances between attractions and their top 1000 related attractions sorting by rel_topic are large, and the differences between the distances on different sorting positions are slight. It illustrates that the geographic proximity of attraction image topics is again not significant. Furthermore, the distances between attractions and their top 1000 related attractions sorting by rel_line and rel_hin increased as the relatedness decreased. Specifically, the distances based on rel_hin increased faster than the distances based on rel_line , e.g., the median distance

of each attraction and its 200th related attraction sorted by *rel_line* is 391.04 km, but the median distance of that sorted by *rel_hin* is 907.71 km. These show that the HIN2Vec model can capture the image topic similarity based on geographic proximity. That is, the HIN embedding listed not only the near co-occurrence attractions as the related attractions of an attraction, but also the attractions far away but with similar image topics.

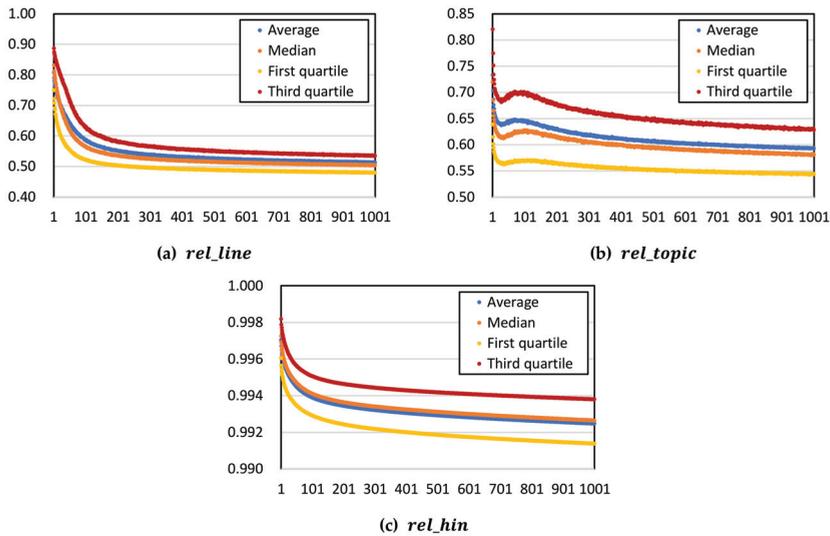


Figure 9. Statistical indicators of *rel_hin*, *rel_line*, and *rel_topic* between each attraction and its each top 1000 related attraction sorted by *rel_hin* (the *rel_hin* is calculated by cosine similarity).

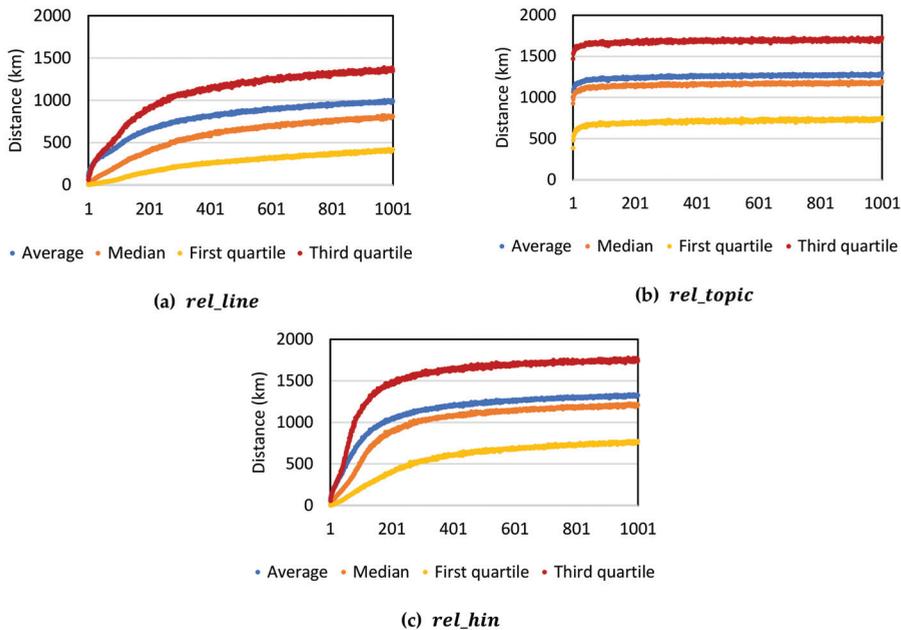


Figure 10. Statistical indicators of distances between each attraction and its top 1000 related attractions sorted by *rel_hin*, *rel_line*, and *rel_topic* (the *rel_hin* is calculated by cosine similarity).

3.5. Attractions Clustering

The attractions can be grouped using a clustering algorithm based on the vectors from HIN embedding. In this case study, the Affinity Propagation (AP) clustering algorithm was selected to group the attractions. AP clustering views each data point as a node in a network, then recursively transmits real-valued messages along the edges of the network until a good set of exemplars and corresponding clusters emerges [42]. Specially, the real-valued messages are divided into responsibility and availability. The former is the message sent from data point i to candidate clustering centre point j , reflecting the suitability that point j is the clustering centre of point i . The latter is the message sent from candidate clustering centre point j to data point i , reflecting the appropriateness that point i selects point j as its clustering centre. AP clustering determines the clustering centre of all data points by the iterative calculation of these two real-valued messages, then finishes the clustering. Thus, the number of clusters of the Affinity Propagation clustering algorithm was not prespecified, which is consistent with the lack of prior knowledge to determine the optimal number of clusters of attractions. Finally, 11,122 attractions were clustered into 467 clusters.

Then, we calculated the average of the relatedness based on the online review HIN (ave_rel), the average of distances (ave_dis) and the standard deviation of distances (std_dis) between all attractions in each cluster. The larger ave_rel of cluster indicated that the attractions in this cluster have stronger relatedness. The larger ave_dis of cluster indicated that the attractions in this cluster were distributed in a larger space range. The larger std_dis of cluster indicated that the attractions in this cluster were distributed more unevenly in space. Because the similarity between data points in AP clustering is measured by the negative Euclidean distance between vectors, we also used negative Euclidean distance to identify the relatedness between attractions in this experiment:

$$relatedness(x, y) = -dist(x, y) = -\sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

where x and y are the vectors of two attractions. x_i and y_i are components of vector x and y , respectively.

Figure 11 indicates the overall trend of ave_dis and std_dis decreasing with ave_rel increasing, while it is not strictly decreasing. It illustrates that the attractions which are spatially close and uniformly distributed have a higher probability of being clustered. That is, the HIN2vec model decides that attraction co-occurrence is a factor that may be more important than image topic in determining the semantic relationship between attractions from the proposed online review HIN. However, the HIN2vec model embeds the attractions from the structure of the online review HIN rather than simply combining co-occurrence relatedness and image topic relatedness between attractions directly. This process may take advantage of additional potential semantic relationships, so the trend is not strictly decreasing.

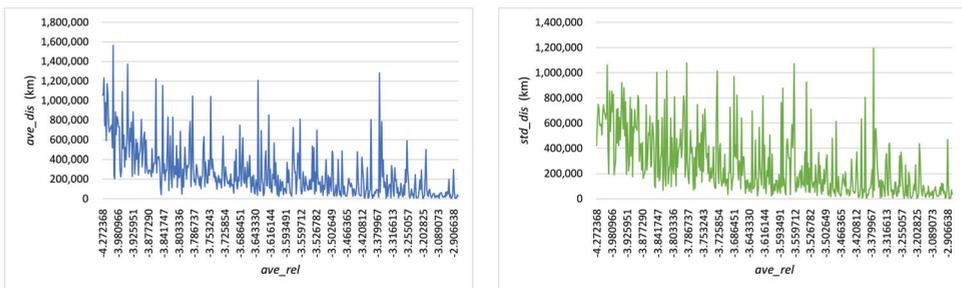


Figure 11. Trend of ave_dis and std_dis between attractions in each cluster as ave_rel increases (the ave_rel is calculated by negative Euclidean distance).

We used Jenks natural breaks classification method to further divide the above 467 clusters into five groups based on the *ave_rel* of these clusters. Then, one cluster for each group was chosen for exploring the validity of the clustering results. The spatial distributions of the five groups and the attractions in the five sample clusters are shown in Figure 12. It indicates that the attractions in each cluster were spatially concentrated as *ave_rel* increased. Most attractions in cluster #6 and all attractions in cluster #20 were concentrated in a city (Harbin and Wuhan). Besides, even the attractions of a cluster are distributed in a large space range, these attractions may have similar image topics, e.g., cluster #441 is about “museum”, cluster #307 is about “beach”, and cluster #6 is about “historic towns”. Meanwhile, the attractions in cluster #20 and cluster #161 are clustered because if these attractions are distributed in a small space range, then they have a higher probability of being co-visited by tourists, resulting in a stronger co-occurrence relatedness between these attractions than image topic relatedness between them. Overall, the attractions in different clusters present co-occurrence relatedness or image topic relatedness, which demonstrates that the HIN embedding automatically adjusts the importance of attraction co-occurrence and attraction image in final relatedness from the characteristics of real data. The clustering result helps one to further discover the attraction communities, of which the attractions can establish close cooperation.

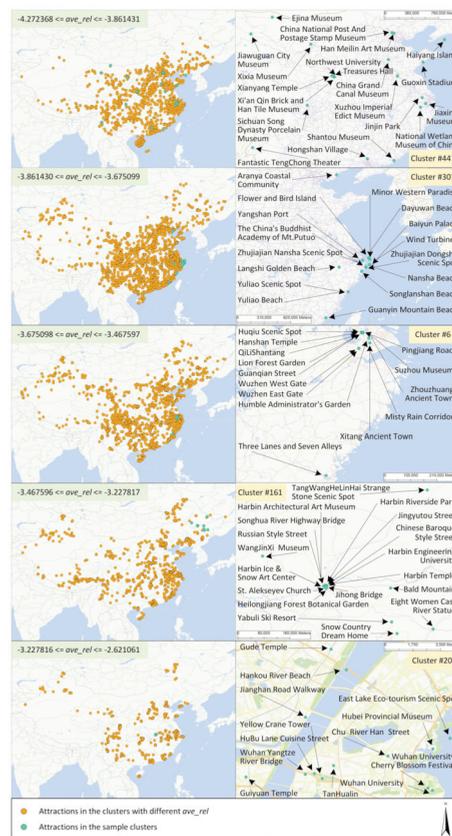


Figure 12. Spatial distributions of the attractions in the clusters with different *ave_rel*. The attractions in the left column are the attractions in all clusters with given range of *ave_rel*. The attractions in the right column are the attractions in the clusters sampled from the corresponding left clusters (the *ave_rel* is calculated by negative Euclidean distance).

4. Discussion

In this study, the HIN2Vec model was used to embed the online review HIN into low-dimensional vector space, whereas there are many other HIN embedding models, as mentioned in Section 2.3, such as Metapath2Vec, HAN, and HERec. These models also show a good performance in representing the nodes in HIN. The reasons we selected the HIN2Vec model in this research are: (1) as presented in Section 2.3, the HIN2Vec model can construct meta-paths automatically and avoid meta-path design. Although the two edge paths “havingreview” → “writtenby” → “writing” → “reviewof” and “havingreview” → “hastopic” → “topicof” → “reviewof” express the semantics of attraction co-occurrence and attraction image, as explained in Section 2.1, we think the other edge paths can still give clues for the HIN2Vec model to mine the semantic relationships between nodes, which may not have significant meanings for people to understand. (2) No model has demonstrated undisputed performance on HIN embedding, because the above models are verified in different tasks and evaluation metrics with different pre-processing [43]. Overall, the emphasis of this research illustrates that the HIN can retain the difference between different relationship semantics when the online reviews are reorganized into a network structure, and the HIN embedding model can capture and fuse these different relationship semantics, which facilitate identifying the relatedness between attractions from a comprehensive perspective.

The proposed relatedness identification between attractions based on online review HIN is a data-driven approach. The HIN2Vec model can automatically capture and fuse heterogeneous semantic information in the online review HIN and give the attraction vectors through fusing all information, without the need to manually set the weights of attraction co-occurrence and attraction image topic. Specifically, the strength of attraction co-occurrence is reflected by the heterogeneous network structure, rather than the weight of edges, as in the traditional network analytics. That is, if two attractions have more reviews written by the same tourists, the HIN2Vec model will ensure these attractions are closer to each other in the embedding vector space, i.e., these attractions have stronger relatedness. Moreover, the HIN2Vec model generates the training data from HIN based on random walk and negative sampling, which overcomes the data-sparsity problem and outputs the effective embedding vectors of attractions that have a few co-occurrences with other attractions or attraction image topics.

While the number of node types in the proposed online review HIN was four and the number of edge types was six, more information in the tourism online reviews should be introduced into the online review HIN in future to better identify the relatedness between attractions, such as the type of attraction, the level of attraction, the residence of the tourist, and so on. Nevertheless, the quality and reliability of the information needs to be noticed to avoid introducing noise into the HIN. For instance, the attraction level “National AAAAA level tourism attraction” is labelled as “National 5A level tourism attraction”, “AAAAA attraction”, “5A level attraction”, etc. in Chinese on MaFengWo. The reason is that the information in social networks lacks strict inspection and revision. Thus, the model will determine these labels as having different semantics if these labels are not uniformed. Furthermore, the data size affects embedding efficiency. The training time of the HIN2Vec model exceeded 15 hours based on the constructed online review HIN. If the length of the meta-paths was set to 5, the HIN2Vec model would not have completed training for five days. Consequently, while HIN embedding showed good performance in identifying the relatedness between attractions, the HIN structure, data size, data quality, and HIN embedding model need to be carefully selected to ensure the training efficiency.

The related attractions of an attraction can be used as the recommendation information when a user browses this attraction online. Meanwhile, the attraction manager can regard the tourists who visited these related attractions as potential customers and take measures to attract these tourists. In addition, the HIN embedding model embeds not only the attractions, but also tourists and image topics in the online review HIN. Thus, the relatedness between tourists can also be identified, which helps to extract tourist profiles,

cluster tourists, and recommend related tourists based on fusing the multiple relationship semantics. Furthermore, the attractions that may be of interest to a tourist can be obtained based on the relatedness between tourists and attractions by the operation of their vectors.

5. Conclusions

Most studies identify the relatedness between attractions through measuring their co-occurrence extracted from online tourism reviews. However, the implicit semantic information in these reviews, which definitely contributes to modelling the relatedness from a more comprehensive perspective, is ignored due to the difficulty of quantifying the importance of different dimensions of information and fusing them. Thus, this paper introduces HIN to reorganize the tourism online reviews for representing the co-occurrence and images of attractions, and then uses HIN embedding to comprehensively identify the relatedness between attractions. First, an online review-oriented HIN was designed to form the different types of elements in the reviews. Second, a topic model was employed to extract the nodes of the HIN from the review texts. Third, an HIN embedding model was used to capture the semantics in the HIN and comprehensively represent the attractions with low-dimensional vectors. The effectiveness of the presented method was validated by three tasks based on the tourist review data from MaFengWo: (1) the visualization illustrates the HIN2Vec model accurately discriminates the attraction, topic, and attraction image types of elements in an online review HIN; (2) the top 1000 related attraction findings show that the presented method comprehensively identifies the relatedness between attractions from the perspectives of both attraction co-occurrence and attraction image; (3) the result of attraction clustering demonstrates the HIN embedding can automatically adjust the importance of attraction co-occurrence and attraction image in final relatedness based on the characteristics of real data. These results indicate that the online review HIN can correctly express the semantics of attraction co-occurrences and attraction images in reviews, and the HIN embedding can capture the differences in these semantics, which facilitates identifying the relatedness between attractions from a comprehensive perspective.

Limitations also exist in this study. Firstly, the structure of the proposed online review HIN only contained four node types and six edge types. Meanwhile, the tourism online reviews provided more types of information, such as the type of attraction, the level of attraction, the residence of the tourist, etc., which helped to identify the relatedness through integrating more semantics. Secondly, we only used the HIN2Vec model to verify the effectiveness of the proposed online review HIN, not to compare the effects of different HIN embedding models. Moreover, while the HIN2Vec model can capture the semantic information in the long hop edge paths, its training time increased significantly with the increase in data size. Therefore, in future work, we would like to (1) extend the online review HIN with more types of information; (2) improve the training efficiency in terms of model selection, model optimization, and HIN structure optimization; and (3) apply the proposed relatedness identification to tourism recommendation and tourism analytics.

Author Contributions: Conceptualization, Peiyuan Qiu and Feng Lu; methodology, Peiyuan Qiu; validation, Peiyuan Qiu and Jialiang Gao; formal analysis, Peiyuan Qiu; investigation, Peiyuan Qiu and Jialiang Gao; resources, Peiyuan Qiu and Jialiang Gao; data curation, Jialiang Gao and Peiyuan Qiu; writing—original draft preparation, Peiyuan Qiu; writing—review and editing, Feng Lu; supervision, Feng Lu; project administration, Feng Lu; funding acquisition, Feng Lu and Peiyuan Qiu. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant No. 41631177, Grant No. 42001341); Doctoral Research Fund of Shandong Jianzhu University, grant number X20084Z; and a grant from State Key Laboratory of Resources and Environmental Information System.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to thank the four anonymous reviewers for their valuable suggestions, which significantly improve the quality of this article.

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

References

- Liu, Y.; Wang, F.; Kang, C.; Gao, Y.; Lu, Y. Analyzing Relatedness by Toponym Co-Occurrences on Web Pages. *Trans. GIS* **2014**, *18*, 89–107. [\[CrossRef\]](#)
- Gu, Z.; Zhang, Y.; Chen, Y.; Chang, X. Analysis of Attraction Features of Tourism Destinations in a Mega-City Based on Check-in Data Mining—A Case Study of Shenzhen, China. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 210. [\[CrossRef\]](#)
- Hong, T.; Ma, T.; Huan, T.-C. Network behavior as driving forces for tourism flows. *J. Bus. Res.* **2015**, *68*, 146–156. [\[CrossRef\]](#)
- Du, S.; Zhang, H.; Xu, H.; Yang, J.; Tu, O. To make the travel healthier: A new tourism personalized route recommendation algorithm. *J. Ambient Intell. Humaniz. Comput.* **2019**, *10*, 3551–3562. [\[CrossRef\]](#)
- Li, J.; Xu, L.; Tang, L.; Wang, S.; Li, L. Big data in tourism research: A literature review. *Tour. Manag.* **2018**, *68*, 301–323. [\[CrossRef\]](#)
- Han, H.; Kim, S.; Otoo, F.E. Spatial movement patterns among intra-destinations using social network analysis. *Asia Pac. J. Tour. Res.* **2018**, *23*, 806–822. [\[CrossRef\]](#)
- Kirilenko, A.P.; Stepchenkova, S.O.; Hernandez, J.M. Comparative clustering of destination attractions for different origin markets with network and spatial analyses of online reviews. *Tour. Manag.* **2019**, *72*, 400–410. [\[CrossRef\]](#)
- Sugimoto, K.; Ota, K.; Suzuki, S. Visitor Mobility and Spatial Structure in a Local Urban Tourism Destination: GPS Tracking and Network analysis. *Sustainability* **2019**, *11*, 919. [\[CrossRef\]](#)
- Liu, B.; Huang, S.; Fu, H. An application of network analysis on tourist attractions: The case of Xinjiang, China. *Tour. Manag.* **2017**, *58*, 132–141. [\[CrossRef\]](#)
- Mou, N.; Zheng, Y.; Makkonen, T.; Yang, T.; Tang, J.; Song, Y. Tourists' digital footprint: The spatial patterns of tourist flows in Qingdao, China. *Tour. Manag.* **2020**, *81*, 104151. [\[CrossRef\]](#)
- Haris, E.; Gan, K.H.; Tan, T.-P. Spatial information extraction from travel narratives: Analysing the notion of co-occurrence indicating closeness of tourist places. *J. Inf. Sci.* **2020**, *46*, 581–599. [\[CrossRef\]](#)
- Yuan, H.; Xu, H.; Qian, Y.; Li, Y. Make your travel smarter: Summarizing urban tourism information from massive blog data. *Int. J. Inf. Manag.* **2016**, *36*, 1306–1319. [\[CrossRef\]](#)
- Lin, M.S.; Liang, Y.; Xue, J.X.; Pan, B.; Schroeder, A. Destination image through social media analytics and survey method. *Int. J. Contemp. Hosp. Manag.* **2021**. Epub ahead of printing. [\[CrossRef\]](#)
- Blei, D.M.; John, D.L. Topic models. In *Text Mining: Classification, Clustering, and Applications*; Taylor and Francis: London, UK, 2009.
- Rossetti, M.; Stella, F.; Zanker, M. Analyzing user reviews in tourism with topic models. *Inf. Technol. Tour.* **2016**, *16*, 5–21. [\[CrossRef\]](#)
- Guo, Y.; Barnes, S.J.; Jia, Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tour. Manag.* **2017**, *59*, 467–483. [\[CrossRef\]](#)
- Wang, J.; Li, Y.; Wu, B.; Wang, Y. Tourism destination image based on tourism user generated content on internet. *Tour. Rev.* **2020**, *76*, 125–137. [\[CrossRef\]](#)
- Kurashima, T.; Iwata, T.; Hoshida, T.; Takaya, N.; Fujimura, K. Geo topic model: Joint modeling of user's activity area and interests for location recommendation. In Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM' 13), Association for Computing Machinery, Rome, Italy, 4–8 February 2013; pp. 375–384.
- Zheng, X.; Luo, Y.; Sun, L.; Zhang, J.; Chen, F. A tourism destination recommender system using users' sentiment and temporal dynamics. *J. Intell. Inf. Syst.* **2018**, *51*, 557–578. [\[CrossRef\]](#)
- An, H.; Moon, N. Design of recommendation system for tourist spot using sentiment analysis based on CNN-LSTM. *J. Ambient Intell. Humaniz. Comput.* **2019**, 1–11. [\[CrossRef\]](#)
- Shafqat, W.; Byun, Y.-C. A Recommendation Mechanism for Under-Emphasized Tourist Spots Using Topic Modeling and Sentiment Analysis. *Sustainability* **2020**, *12*, 320. [\[CrossRef\]](#)
- Shi, C.; Philip, S.Y. *Heterogeneous Information Network Analysis and Applications*; Springer: Berlin/Heidelberg, Germany, 2017.
- Sun, Y.; Han, J. Mining heterogeneous information networks: A structural analysis approach. *ACM SIGKDD Explor. Newsl.* **2013**, *14*, 20–28. [\[CrossRef\]](#)
- Shi, C.; Hu, B.; Zhao, W.X.; Yu, P.S. Heterogeneous Information Network Embedding for Recommendation. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 357–370. [\[CrossRef\]](#)
- Wang, C.; Raina, R.; Fong, D.; Zhou, D.; Han, J.; Badros, G. Learning relevance from heterogeneous social network and its application in online targeting. In Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR' 11), Association for Computing Machinery, Beijing, China, 24–28 July 2011; pp. 655–664.
- Yu, J.; Gao, M.; Li, J.; Yin, H.; Liu, H. Adaptive Implicit Friends Identification over Heterogeneous Network for Social Recommendation. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM' 18), Torino, Italy, 22–26 October 2018; pp. 357–366.

27. Hosseini, A.; Chen, T.; Wu, W.; Sun, Y.; Sarrafzadeh, M. HeteroMed: Heterogeneous Information Network for Medical Diagnosis. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM' 18), Torino, Italy, 22–26 October 2018; pp. 763–772.
28. Chen, X.; Yin, J.; Qu, J.; Huang, L. MDHGI: Matrix Decomposition and Heterogeneous Graph Inference for miRNA-disease association prediction. *PLOS Comput. Biol.* **2018**, *14*, e1006418. [[CrossRef](#)]
29. Zhang, F.; Wang, M.; Xi, J.; Yang, J.; Li, A. A novel heterogeneous network-based method for drug response prediction in cancer cell lines. *Sci. Rep.* **2018**, *8*, 3355. [[CrossRef](#)] [[PubMed](#)]
30. Shi, C.; Li, Y.; Zhang, J.; Sun, Y.; Yu, P.S. A Survey of Heterogeneous Information Network Analysis. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 17–37. [[CrossRef](#)]
31. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
32. Qiang, J.; Qian, Z.; Li, Y.; Yuan, Y.; Wu, X. Short Text Topic Modeling Techniques, Applications, and Performance: A Survey. *IEEE Trans. Knowl. Data Eng.* **2020**, 1–19. [[CrossRef](#)]
33. Bengio, Y.; Ducharme, R.; Vincent, P.; Janvin, C. A Neural Probabilistic Language Model. *J. Mach. Learn. Res.* **2003**, *3*, 1137–1155.
34. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of the Workshop at International Conference on Learning Representations 2013, Scottsdale, AZ, USA, 2–4 May 2013; pp. 1–12.
35. Liu, K.; Gao, S.; Qiu, P.; Liu, X.; Yan, B.; Lu, F. Road2Vec: Measuring traffic interactions in urban road system from massive travel routes. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 321. [[CrossRef](#)]
36. Mihalcea, R.; Tarau, P. TextRank: Bringing Order into Text. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), Association for Computational Linguistics, Barcelona, Spain, 25–26 July 2004; pp. 404–411.
37. Dong, Y.; Chawla, N.V.; Swami, A. Metapath2Vec: Scalable Representation Learning for Heterogeneous Networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD' 17), Halifax, NS, Canada, 13–17 August 2017; pp. 135–144.
38. Fu, T.; Lee, W.-C.; Lei, Z. HIN2Vec: Explore Meta-paths in Heterogeneous Information Networks for Representation Learning. In Proceedings of the 2017 ACM Conference on Information and Knowledge Management (CIKM' 17), Singapore, Singapore, 6–10 November 2017; pp. 1797–1806.
39. Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; Yu, P.S. Heterogeneous Graph Attention Network. In Proceedings of the 2019 World Wide Web Conference (WWW' 19), San Francisco, CA, USA, 13–17 May 2019; pp. 2022–2032.
40. Liu, N.; Huang, X.; Li, J.; Hu, X. On Interpretation of Network Embedding via Taxonomy Induction. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Association for Computing Machinery (KDD' 18), London, UK, 19–23 August 2018; pp. 1812–1820.
41. Tang, J.; Qu, M.; Wang, M.; Zhang, M.; Yan, J.; Mei, Q. LINE: Large-scale Information Network Embedding. In Proceedings of the 24th International Conference on World Wide Web (WWW' 15), International World Wide Web Conferences Steering Committee, Florence, Italy, 18–22 May 2015; pp. 1067–1077.
42. Frey, B.J.; Dueck, D. Clustering by Passing Messages between Data Points. *Science* **2007**, *315*, 972–976. [[CrossRef](#)] [[PubMed](#)]
43. Dong, Y.; Hu, Z.; Tang, J.; Sun, Y.; Wang, K. Heterogeneous Network Representation Learning. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20), Online, 7–15 January 2021; Volume 5, pp. 4861–4867.

Article

Tour-Route-Recommendation Algorithm Based on the Improved AGNES Spatial Clustering and Space-Time Deduction Model

Xiao Zhou ¹, Jiangpeng Tian ^{2,*} and Mingzhan Su ²

¹ College of Computer Science, Sichuan University, Chengdu 610065, China; zhouxiao@infu.ac.cn

² Institute of Geospatial Information, PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, China; smzh625@infu.ac.cn

* Correspondence: tjp_study@infu.ac.cn; Tel.: +86-135-9241-9753

Abstract: This study designed a tour-route-planning and recommendation algorithm that was based on an improved AGNES spatial clustering and space-time deduction model. First, the improved AGNES tourist attraction spatial clustering algorithm was created. Based on the features and spatial attributes, city tourist attraction clusters were formed, in which the tourist attractions with a high degree of correlation among attributes were gathered into the same cluster. It formed the precondition for searching tourist attractions that would match tourist interests. Using tourist attraction clusters, this study also developed a tourist attraction reachability model that was based on tourist-interest data and geospatial relationships to confirm each tourist attraction's degree of correlation to tourist interests. A dynamic space-time deduction algorithm that was based on travel time and cost allowances was designed in which the transportation mode, time, and costs were set as the key factors. To verify the proposed algorithm, two control algorithms were chosen and tested against the proposed algorithm. Our results showed that the proposed algorithm had better results for tour-route planning under different transportation modes as compared to the controls. The proposed algorithm not only considered time and cost allowances, but it also considered the shortest traveling distance between tourist attractions. Therefore, the tourist attractions and tour routes that were suggested not only met tourist interests, but they also conformed to the constraint conditions and lowered the overall total costs.

Keywords: AGNES clustering; tourist attraction clustering; tourist attraction reachability space model; space-time deduction; tour route searching

Citation: Zhou, X.; Tian, J.; Su, M. Tour-Route-Recommendation Algorithm Based on the Improved AGNES Spatial Clustering and Space-Time Deduction Model. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 118. <https://doi.org/10.3390/ijgi11020118>

Academic Editors: Andrea Marchetti, Angelica Lo Duca and Wolfgang Kainz

Received: 27 December 2021

Accepted: 2 February 2022

Published: 7 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tourists are the core of tourism activity. A key issue of smart tourism is how to improve tourist satisfaction and provide the best experience. A complete tourism activity cycle includes pre-travel, traveling, and post-travel activities. The pre-travel experience includes itinerary, planning, and tour-route searching, etc. The travel process itself includes visiting tourist attractions and the travel between locations, etc. The post-travel activity includes the evaluation and feedback on the tourism experience as a whole. In the whole tourism activity cycle, the pre-travel experience is the most important factor to influence tourists' satisfaction and, therefore, their subjective evaluations regarding the quality of their experiences. Tourists will have spent a certain amount of time and cost on their experiences. Therefore, devising and suggesting tour routes according to tourists' needs and desires while realizing the minimum time and cost as well as the maximum benefit is key to optimal tour planning.

In a tour, the tourism objects are tourist attractions. Searching the very tourist attractions that accurately match the tourists' needs is the critical step for the planning and recommending of tour routes. Tourists' needs have large discrepancies, while the tourist

attractions distributed in a city also have different feature attributes and spatial attributes, for which reason, each tourist attraction has relatively large different capacities on meeting tourists' needs. The diversity of tourists' needs and tourist attractions' attributes makes the searching process complex. Thus, rapidly confirming the interested tourist attraction groups according to the tourists' needs can improve the searching algorithm's accuracy and efficiency, so that applying the effective method to generate tourist attraction groups is the key step to search the accurate tourist attractions. In data mining technology, a clustering algorithm can group spatial dots. It absorbs the spatial dots which have the similar attributes into the same group and divides the large scale data mining into a smaller scale one in the group, which can improve the algorithm efficiency. This paper uses the clustering algorithm to generate city tourist attraction clusters in accordance with tourist attraction attributes, and it provides the algorithm basis for searching accurate tourist attractions. There are many kinds of clustering algorithms. This paper uses AGNES as the basic algorithm to set up the clustering model. The agglomerative nesting (AGNES) algorithm is a hierarchical clustering method that operates from bottom to top. It sets the elements as the bottom layer in the spatial distribution and gathers them from bottom to top according to a defined criterion. The AGNES algorithm is a single-link method where each cluster is represented by its arbitrary elements. Therefore, the degree of correlation of two clusters is determined by the two values with the highest degree of correlation in each cluster. The clustering process begins at the discrete distributed bottom layer and gathers each dot within the clusters and ends with the preset number of clusters. A traditional AGNES algorithm is operated with the same spatial distance as the criterion. The reasons for this paper to use AGNES are as follows. First, AGNES is simple and it is easy to implement. AGNES is a naive clustering algorithm, which has a concise principle and process. Its starting and ending conditions are definite and the selecting of the starting seed point is simple. In the clustering process, it only needs to judge the dispersion between the seed point and the non-seed point. Compared with other clustering algorithms, it is more accessible and easier to implement. Second, AGNES has relatively low spatial complexity and time complexity. It has a faster operating rate and consumes less computer memory. Third, AGNES is very suitable for the clustering on small scale dataset. In this paper, the research objects are city tourist attractions; it forms a typical small scale dataset, thus the AGNES is feasible. Fourth, AGNES is more flexible and can realize the multiple layer clustering structures on different granularity by setting different parameters. It has no strict requirements on the samples inputting sequence and can realize the synchronous clustering from different dots to reduce the convergence time.

In tourism clustering research, [1] provided a general introduction on the clustering method, including the AGNES clustering algorithm. In [2], the researchers applied the hierarchical cluster analysis to a set of Indonesian tourism sites in and around Malang City, Malang Regency, and Batu City using the AGNES algorithm to optimize a search engine that could assist tourists when choosing tourist attractions under certain constraints. In another study [3], the AGNES algorithm was applied to the data from the online platform, Airbnb. The collaborative economy of tourism hosts based on their geographic distribution was studied. The city of Guanajuato, Mexico, was selected as the subject city for convenience purposes, and the main touristic attractions were used as parameters to conduct the analysis. According to [4], an ontology-based clustering method was used to analyze the qualitative factors from a semantic perspective to define tourist segments and understand why tourists travelled to a particular destination in the Catalonia region of Spain, and the researchers reported better results using this method as compared to classic clustering algorithm methods. In the literature, the proposed ontology-based clustering method was derived from an extension of the AGNES clustering algorithm. Researchers in [5] designed an original approach to characterize the daily behaviors of tourists by analyzing the sequences of places that were visited by tourists per day, in which the geolocation information of tourists on photo-sharing websites was used as the data, from which the AGNES clustering

algorithm formed clusters and carried out the experiment. The study in [6] proposed a point-of-interest (POI) recommendation method to plan tourism routes.

Different clustering methods have been developed in the design and implementation of tourist route-information recommendation systems based on user POI indices, including AGNES clustering. In [7], the AGNES clustering algorithm was used to identify residents' dependence on public transport. It provided a potential method for choosing the transportation mode for tourists. The researchers in [8] applied semantic clustering to extract tourist preferences. It compared the semantics of tourist preferences with tourist attraction attributes and provided tourist attraction suggestions. The researchers in [9] used the partitioning clustering method to find the nearest tourism destination according to the extracted geotagged photograph-location data. The researchers in [10] studied the cluster-mapping procedures for tourism regions based on the fuzzy-clustering method. This method proved to increase the identification accuracy of the tourism clusters. The researchers in [11] developed a Bali tourism information system by using web-scraping and clustering methods. The clustering algorithm was used to process the word-text data and output word clusters, and then performed clustering on the website. The researchers in [12] proposed a tourist-preference clustering method that was based on tourist facial and background information that were extracted from photographs. The clustering method was used to generate tourist classifications. The researchers in [13] used spatial clustering methods to mine tourist destinations and preferences, in which the regions of tourist attractions for each tourism category were derived by the clustering algorithm. The researchers in [14] used a density-based spatial clustering algorithm to study tourist behavior, and by extracting the tourist behaviors, the tourism hot-spots were extracted as they related to tourist behavior. In [15], the clustering algorithm was used to generate tourist-attraction clusters via network and geographic information system (GIS) analyses, and three tourist-attraction clusters were extracted.

For tour-route algorithms, the researchers in [16] proposed a tour-route-recommendation method using the multiple-criteria tensor model fusing time-space information. The researchers in [17] combined factors of time and space and used the tourist-attraction photographs that were posted on a website by previous tourists to set up a tour-route-recommendation model. The researchers in [18] applied a heuristic method for tour-route recommendation based on urban traffic monitoring. The researchers in [19] employed social-network analysis combined with deep-learning theory to develop a tour-route-recommendation model. The researchers in [20] created a tour-route-recommendation model that was based on Smart Agent technology. In [21], an individualized tour-route-recommendation model that was based on POI functionality and accessibility was proposed, and it determined tourist physiological and physical conditions as the important reference criteria. The researchers in [22] suggested an individualized tour-route-recommendation model that was based on social networks' geographical context cognition, and it used social relationships and trust networks among tourists as the important indices. The researchers in [23] developed a tour-route-recommendation model that was based on improved collaborative filtering technology. The researchers in [24] designed a tour-route-recommendation algorithm that was based on dynamic clustering to counter the challenge of data scarcity. In [25], a tour-route-recommendation algorithm was designed that was based on deep-interest label mining and association rule clustering. The researchers in [26] also designed a tour-route-recommendation model that was based on a collaborative filtering algorithm. The researchers in [27] suggested a tour-route-recommendation model that was based on geotagging and temporal divisions where the core principle that included user and group ratings as well as time and distance. The researchers in [28] proposed a tour-route-recommendation method that was based on tourist time-space behavioral constraints, and it used temporal and spatial constraints as the important factors. The researchers in [29] proposed a tour-route-recommendation method that was based on a combined recommendation algorithm including hybrid-interest modelling and a heuristic tour-route-planning algorithm. In [30], an energy-aware clustering method was used for mobile application, which provided a

method that efficient routing, resource allocation, and energy management can be achieved through clustering of mobile into local groups. Ref. [31] collected tourists' traveling data on the website and analyzed the tourists' behaviour and, based on the website tourists' mobile data as well as the mined POIs, it set up the tour route recommendation algorithm. Ref. [32] studied the importance of the mobile devices and location-based services. Based on the big data, such as tourism data, location predicting could be realized, which could be used in studying tourists' mobility and the tendency on the traveling behaviour.

According to the literature review, tourism clustering research has predominantly focused on tourist attractions and tourist clustering. As seen in [1–7], clustering algorithms have been used in tourism research for POI extraction, data mining, algorithm modeling, transportation behavior, etc. The other clustering methods in [8–15] indicated that spatial and attribute data of tourist attractions were the main targets that were used to generate proper tourism categories, extract tourist preferences, and recommend appropriate tourist destinations. The studies concerning tourist-attraction data extraction and tour-route algorithms that were used in [16–29] focused on three specific aspects. Refs. [30–32] tended to study the big data that were obtained from social networks on mobile devices and website. The big data could be used as the basic data to do clustering on tourist attractions and tourists or could be used to study the tourists' mobility and traveling behaviour. First, they examined the recommendation algorithm itself, including data scarcity and "cold start" issues. The data scarcity means that, in a database, the most valuable and useful data are missing, or the majority of the data are zero. The "cold start" means that, in a recommendation system, the newly registered users and new added products lack historical data, and they could be hardly recommended to the new registered users. Second, they developed an improved algorithm that was based on traditional recommendation methods such as the collaborative filtering algorithm, where historical data that were extracted from users and groups with similar interests to the current user are identified to customize the recommendations for the current user. Third, they mined historical tourist-interest data to recommend tour routes for the current tourists. Common methods that are used for this process include tourist label, photo, and evaluation data mining. Overall, the existing methods focused on improvements in algorithm performance, historical data, and the improvement on solving the problems such as "cold start" and data scarcity but overlooked tourist needs, attraction attributes, real-world geospatial environment, and tour-route searching, so they have typically yielded fuzzy results that lacked sufficient accuracy.

As indicated above, the challenges in tour-route planning remain. First, the research on tourist needs and tourist-attraction attributes is insufficient, especially in terms of real-world concerns, such as time and cost. Second, since tourist-attraction clustering provides preconditions for matching tourists' interests, there is no effective and reasonable mechanism for urban tourist attraction clustering, and the clustering criterion is merely the spatial distance, neglecting the inner attributes such as tourist attraction classification, popularity, optimal traveling time, and traveling fee. For the traditional AGNES clustering algorithm, a specific tourist's individualized needs and interests were not fully considered. Third, the research on the space-time deduction on the traveling process is insufficient, in which the space-time deduction means a tourist's traveling activity in a whole tour route will be constrained by time, space, and cost, and it is a dynamic deduction process on the traveling cost. The more tourist attractions to be visited, the more time, traveling distance, traveling fee, etc., will be produced. The time and cost play key roles on recommending tour routes. Fourth, under the conditions of fixed time and cost budget, the transportation mode determines the selected tourist attraction quantity and the planned tour route. The existing methods seldom study the mixing transportation modes with tour route planning.

Therefore, this study designed and tested a tour-route-recommendation algorithm that was based on an improved AGNES spatial clustering and space-time deduction model, focusing on precise interest-matching, urban tourist-attraction spatial clustering, space-time deduction of the traveling process, and precise tour route searching based on the transportation mode. Compared with the previous studies, the proposed algorithm has

differences and novelties. First, the AGNES method is not merely and directly used as a clustering tool. In this study, the AGNES was set as the research target and content, whereby the improved AGNES algorithm was developed. It is the precondition of modeling the tour route algorithm. Second, in the process of developing the improved AGNES clustering algorithm, the tourist attractions' feature attributes were set as the critical parameters in forming the clustering criterion function, conforming to the tourist activity in matching tourist interests, while previous studies had only considered the spatial attributes. Third, different from the research line in which the location-based social network was exploited to understand human mobility and people behavior by mining check-in patterns, this research was based on the city tourist attractions' attributes and one tourist's specific interests. The former studies were performed on tourism big data, and they tended to mine the tourists' moving behaviors and find out the potential interested tour routes. The proposed method is a one-one mode in which tourist interests were studied and set as the specific preconditions to extract certain tourist attractions, and then the path-searching algorithm was used to find out the optimal tour route. Thus, they are different in algorithm mechanisms. Fourth, the studies on the tourist attraction and tour route recommendation are based on the fuzzy recommendation, while the proposed algorithm is under the consideration and constraint of the real-world city tourism environment, road conditions, and transportation modes, thus it could find out the global optimal routes that match the tourists' interests within the limited time and space complexity.

Figure 1 shows the research work and the structure of the paper.

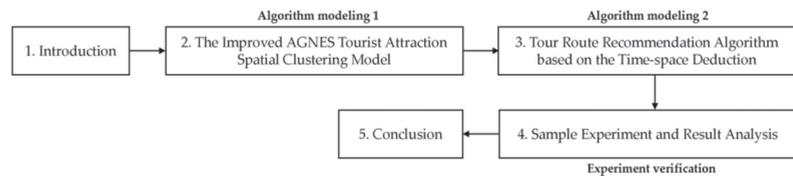


Figure 1. The research work and the structure of the paper.

2. The Improved AGNES Tourist Attraction Spatial Clustering Model

The features and spatial attributes of urban tourist attractions can vary widely. The feature attributes are the characteristics of one tourist attraction that differ from another one, such as tourist attraction classification, popularity, optimal traveling time, traveling fee, etc. The classification labels represent the characteristics or features of one tourist attraction, they determine the tourist attraction's category, and they are typically mined from feature mapping data. The popularity is the average attraction capacity of one tourist attraction, which is determined from the online "big data" sources; for example, "Ctrip", "Fliggy", and "Qunar", among others, provide popularity data for tourist attractions in China. The optimal traveling time and cost stand for the basic time and cost that are needed by the tourists to visit one tourist attraction. Each tourist attraction has various feature attributes that are associated to quantified values. The spatial attributes consider the geospatial location and positioning of a tourist attraction, including the discrete features and the indirectly correlated features. The discrete features should be considered independently for all tourist attractions [33]. Indirectly, the correlated features represent that each tourist attraction is connected with another one by urban roads and tourists can move between two tourist attractions freely. Tourist attraction attributes determine that tourist attractions have a close or distant relationship with each other, bringing different capacities for satisfying tourist interests. The precondition of selecting the tourist attractions to be visited is to confirm the classification that meet the tourist needs and interests. Therefore, the urban tourist attractions should be clustered primarily.

2.1. The Foundation of Tourist Attraction Attribute Label Matrix Model

The preconditions for the clustering algorithm confirmed the tourist attraction attributes and developed the association model that would measure the degree of correlation among the attractions. The degree of correlation among their attributes would be determined by their features as well as by their spatial factors. Thus, the clustering model should combine with the feature attribute factors and the spatial attribute factors [34].

The arbitrary typical tourist attraction in a tourism city is the tourist attraction element $s_{(i)}$, and it belongs to one certain tourist attraction classification. All the elements of $s_{(i)}$ form an entire research range, and it is the tourist attraction research domain S . The domain S contains different types of tourist attractions, and it can be divided into several classifications.

The inner characteristics of one tourist attraction are the feature attributes, and they are noted as $t_{1(i1)}$. The feature attributes influence tourist choices on the interest tendency and intelligent system’s search results of tourist attraction clusters and specific tourist attractions. The factor i_1 is the footnote of the feature attribute. Meanwhile, the tourist-attraction geolocation is the spatial attribute factor $t_{2(i2)}$, and i_2 is the factor’s footnote. The tourist attraction attributes include m number of feature attributes $t_{1(i1)}$ and n number of spatial attributes $t_{2(i2)}$, $i_1 \in (0, m] \subset Z^+$, $i_2 \in (0, n] \subset Z^+$. Each factor $t_{1(i1)}$ or $t_{2(i2)}$ is one feature attribute label and spatial attribute label of $s_{(i)}$, and collectively, the tourist attraction attribute label.

The feature attribute $t_{1(i1)}$ includes $u_{(i1)}$ items of classifying indices $t_{1(i1,j1)}$, $j_1 \in (0, \alpha] \subset Z^+$, j_1 is the footnote of $t_{1(i1,j1)}$, and α is the maximum number of $t_{1(i1)}$. The $1 \times u_{(i1)}$ dimension matrix $\mathbf{t}_{1(i1)}$ is formed by $u_{(i1)}$ items of $t_{1(i1,j1)}$ in the factor $t_{1(i1)}$ and determines the No. $t_{1(i1)}$ feature attribute and tourists’ interest tendency. The tourist attraction feature attribute label vector is $\mathbf{t}_{1(i1)}$. The spatial attribute $t_{2(i2)}$ includes $u_{(i2)}$ items of classifying indices $t_{2(i2,j2)}$, $j_2 \in (0, \alpha] \subset Z^+$, j_2 is the footnote of the indices $t_{2(i2,j2)}$ of $t_{2(i2)}$, and α is the maximum number of $t_{2(i2)}$. The $1 \times u_{(i2)}$ dimension matrix $\mathbf{t}_{2(i2)}$ is formed by $u_{(i2)}$ items of $t_{2(i2,j2)}$ in the factor $t_{2(i2)}$ and determines the No. $t_{2(i2)}$ spatial attribute, and it is the spatial attribute label vector $\mathbf{t}_{2(i2)}$. The classifying index of the spatial attribute is formed to match the attribute label vector and create the attribute matrix. The vector’s rank meets $rank(\mathbf{t}_{1(i1)}) = u_{(i1)}$, and $rank(\mathbf{t}_{2(i2)}) = u_{(i2)}$.

For the tourist attraction attribute label matrix T , it is formed by m number of $\mathbf{t}_{1(i1)}$ and n number of $\mathbf{t}_{2(i2)}$ and determines the tourist attraction’s features and spatial attributes as well as influences the tourists’ interest tendency. The matrix T meets the following conditions: The matrix row is the vector $\mathbf{t}_{1(i1)}$ or $\mathbf{t}_{2(i2)}$. The matrix column is the element of the vector $\mathbf{t}_{1(i1)}$ or $\mathbf{t}_{2(i2)}$. The matrix contains $m + n$ number of rows and α number of columns. The rows from 1 to m relate to the vector $\mathbf{t}_{1(i1)} \sim i_1 \in (0, m] \subset Z^+$, the $m + 1$ to $m + n$ rows relate to the vector $\mathbf{t}_{2(i2)} \sim i_2 \in (0, n] \subset Z^+$. One tourist attraction relates to one matrix element distribution. Equation (1) is the general formula of the matrix T and its element distribution.

$$T = [\mathbf{t}_{1(1)}, \dots, \mathbf{t}_{1(m)}, \mathbf{t}_{2(1)}, \dots, \mathbf{t}_{2(n)}]^T = \begin{bmatrix} t_{1(1,1)} & t_{1(1,2)} & t_{1(1,3)} & \dots & 0 & 0 \\ & & \dots & & \dots & \\ t_{1(m,1)} & t_{1(m,2)} & \dots & & \dots & t_{1(m,\alpha)} \\ t_{2(1,1)} & t_{2(1,2)} & \dots & 0 & \dots & t_{2(1,\alpha)} \\ & & \dots & & & \\ t_{2(m+n,1)} & t_{2(m+n,2)} & \dots & t_{2(m+n,\alpha)} & \dots & 0 \end{bmatrix} \quad (1)$$

The feature attributes and spatial attributes are quantified. The feature attributes include tourist attraction classification $t_{1(1)}$, popularity $t_{1(2)}$, optimal travel time $t_{1(3)}$, and traveling fee $t_{1(4)}$. The spatial attribute mainly relates to the longitude and latitude coordinates $(t_{2(1)}, t_{2(2)}) \sim (l, B)$ of the tourist attraction [35]. The feature attributes $t_{1(i1)}$ and the spatial attributes $t_{2(i2)}$ are quantified, where $t_{1(1)}$ is tourist attraction classification; $t_{1(2)}$ is popularity degree, noted as h_o , $h_o \in (0, 1) \subset R^+$, representing the users’ average

evaluation scores on the website; $t_{1(3)}$ is the optimal travel time, noted as t_b , unit: hour; and $t_{1(4)}$ is the traveling fee (cost), noted as c_o , unit: CNY, ¥ yuan. The spatial attributes include longitude $t_{2(1)} \sim l$ and latitude $t_{2(2)} \sim B$. Each attribute factor includes a specific data value range which forms the tourist attraction feature attribute label vector $\mathbf{t}_{1(i1)}$ and spatial attribute label vector $\mathbf{t}_{2(i2)}$. The classification factor is determined by the tourist attraction's inner attributes and it is the critical index to distinguish different tourist attractions and an important reference for a smart system to select a tourist attraction cluster and specific tourist attractions. The popularity degree represents the average preference of tourists on a tourist attraction $s_{(i)}$. The optimal travel time represents the most suitable time for tourists to visit a tourist attraction $s_{(i)}$. The traveling fee represents the minimum cost for tourists to visit a tourist attraction $s_{(i)}$ such as the fee for the entrance ticket. The formed tourist attraction attribute label matrix \mathbf{T} is $[\mathbf{t}_{1(1)}, \dots, \mathbf{t}_{1(m)}, \mathbf{t}_{2(1)}, \dots, \mathbf{t}_{2(n)}]^T$, each label vector includes the specific index $t_{1(i1,j1)}$ or $t_{2(i2,j2)}$. Quantify the index $t_{1(i1,j1)}$ or $t_{2(i2,j2)}$ as follows, in which the classification factor is also quantified into a specific value.

- \mathbf{T} : $\{t_{1(1)}$: Tourist attraction classification; $t_{1(2)}$: popularity degree; $t_{1(3)}$: the optimal travel time; $t_{1(4)}$: traveling fee; $t_{2(1)}$: longitude; $t_{2(2)}$: latitude};
- $\mathbf{t}_{1(1)}$: $\{t_{1(1,1)}$: nature park (1.00); $t_{1(1,2)}$: humanistic history (2.00); $t_{1(1,3)}$: amusement park (3.00); $t_{1(1,4)}$: leisure shopping (4.00); $t_{1(1,5)}$: modern science and technology (5.00); $t_{1(1,6)}$: artistic aesthetics (6.00)};
- $\mathbf{t}_{1(2)}$: $\{t_{1(2,1)}$: $h_o \in (0, 0.25]$; $t_{1(2,2)}$: $h_o \in (0.25, 0.50]$; $t_{1(2,3)}$: $h_o \in (0.50, 0.75]$; $t_{1(2,4)}$: $h_o \in (0.75, 1.00)$, $t_{1(2,j1)} \in \mathbb{R}^+$ };
- $\mathbf{t}_{1(3)}$: $\{t_{1(3,1)}$: $t_b \in (0, 2.0]$; $t_{1(3,2)}$: $t_b \in (2.0, 4.0]$; $t_{1(3,3)}$: $t_b \in (4.0, 6.0]$; $t_{1(3,4)}$: $t_b \in (6.0, +\infty)$, $t_{1(3,j1)} \in \mathbb{R}^+$ };
- $\mathbf{t}_{1(4)}$: $\{t_{1(4,1)}$: $c_o \in (0, 100]$; $t_{1(4,2)}$: $c_o \in (100, 200]$; $t_{1(4,3)}$: $c_o \in (200, 300]$; $t_{1(4,4)}$: $c_o \in (300, +\infty)$, $t_{1(3,j1)} \in \mathbb{R}^+$ };

When all the feature attribute label vectors $\mathbf{t}_{1(i1)}$ for all the elements $s_{(i)}$ in domain \mathbf{S} are confirmed, the correction parameter for each vector $\mathbf{t}_{1(i1)}$ is then defined to normalize all the values.

The impact of each feature attribute label vector impact on calculating the degree of correlation between the tourist attractions should be in the same order of magnitude, and thus the feature attribute label vector normalized parameter $\delta_{1(i1)}$ is generated, and all the labels are normalized according to a range of $(0, 1]$. According to the range of the vector $\mathbf{t}_{1(i1)}$, each normalized parameter $\delta_{1(i1)}$ is confirmed as follows:

$$\delta_{1(1)} = 0.100, \delta_{1(2)} = 1.000, \delta_{1(3)} = 0.100, \delta_{1(4)} = 0.001.$$

The parameter $\delta_{1(i1)}$ is used to normalize each vector $\mathbf{t}_{1(i1)}$ in the matrix \mathbf{T} to obtain a new normalized matrix \mathbf{T}_δ . As compared to the matrix \mathbf{T} , the elements in the matrix \mathbf{T}_δ are all normalized except for the vector $\mathbf{t}_{2(i2)}$. Equation (2) is the general formula for the matrix \mathbf{T}_δ .

$$\mathbf{T}_\delta = \begin{bmatrix} \delta_{1(1)} \cdot \mathbf{t}_{1(1)}, & \dots, & \delta_{1(m)} \cdot \mathbf{t}_{1(m)}, & \mathbf{t}_{2(1)}, & \dots, & \mathbf{t}_{2(n)} \\ \delta_{1(1)} \cdot t_{1(1,1)} & \delta_{1(2)} \cdot t_{1(1,2)} & \delta_{1(3)} \cdot t_{1(1,3)} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \delta_{1(m)} \cdot t_{1(m,1)} & \delta_{1(m)} \cdot t_{1(m,2)} & \dots & \dots & \dots & \delta_{1(m)} \cdot t_{1(m,\alpha)} \\ t_{2(1,1)} & t_{2(1,2)} & \dots & 0 & \dots & t_{2(1,\alpha)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ t_{2(m+n,1)} & t_{2(m+n,2)} & \dots & t_{2(m+n,4)} & \dots & 0 \end{bmatrix} \quad (2)$$

Based on the tourist attraction attribute label matrix \mathbf{T} and the normalized matrix \mathbf{T}_δ , the tourist attraction research domain \mathbf{S} clustering algorithm is created.

2.2. The Tourist Attraction Domain Clustering Algorithm Based on the Improved AGNES Algorithm

The aim of the tourist attraction domain clustering was to obtain a cluster with a high degree of correlation among the attributes, realizing that the tourist attractions in the same clusters have a high degree of correlation among the attributes while those in different clusters have a low degree of correlation among the attributes, and finally to guide the smart system into precisely matching tourist interests. The clustering process was the automatic process driven by data, and the clustering criteria could differentiate according to the different clustering targets. When a spatial dot is the only a location point in a coordinate system, a traditional clustering algorithm will assume the spatial distance as a singular criterion. Tourist attractions have spatial attributes and feature attributes, and thus the criteria for tourist-attraction clustering should combine both factors.

The k number of elements $s_{(i)}$ in the domain \mathbf{S} are clustered by the clustering algorithm, and the tourist attractions $s_{(i)}$, which have a high degree of correlation among the attributes and are in the same cluster $S_{(i)}$, while the tourist attractions $s_{(i)}$ and $\neg s_{(i)}$, which have a low degree of correlation among the attributes, are in the different clusters $S_{(i)}$ and $\neg S_{(i)}$, $k \in \mathbf{N}$. The cluster's element is noted as $s_{(i,j)}$, i is the footnote of the cluster $S_{(i)}$, j is the footnote of the element in the cluster $S_{(i)}$. In all, it is supposed that the clustering algorithm forms p number of clusters, $p \in \mathbf{N}$ and $p \ll k$. Assume that the cluster $S_{(i)}$ contains $k_{(i)}$ number of elements $s_{(i,j)}$, and $j \in (0, k_{(i)}] \subset \mathbf{Z}^+$, so thus $\sum_{j=1}^p k_{(i)} = k$. The elements $s_{(i,j)}$ in the same cluster $S_{(i)}$ have a high degree of correlation among the attributes, and elements $s_{(i,j)}$ in different clusters $S_{(i)}$ and $\neg S_{(i)}$ have a low degree of correlation among the attributes. An arbitrary cluster $\forall S_{(i)}$ contains at least one element $s_{(i)}$. Arbitrary one element $\forall s_{(i)}$ in the domain \mathbf{S} only belongs to one certain cluster $S_{(i)}$. Clusters $S_{(i)}$ and other cluster $\neg S_{(i)}$ have no intersection, but in the aspect of spatial analysis, the clusters may have a buffer overlap in the city space. The union of all the clusters $S_{(i)}$ is the domain \mathbf{S} , and $i \in (0, p] \subset \mathbf{N}$. In the domain \mathbf{S} , there are at least two clusters, that is $p \geq 2$.

Whether the tourist attraction element $s_{(i)}$ should be absorbed into the cluster $S_{(i)}$ is determined by the objective function $\tilde{\zeta}_{(s_{(i1)}, s_{(i2)})}$ among $s_{(i)}$ and other tourist attractions. The function is determined by several clustering factors, including the feature attribute factors $t_{1(i1)}$ and the spatial attributes factors $t_{2(i2)}$. As to the two independent tourist attractions $s_{(i1)}$ and $s_{(i2)}$, their degree of correlation includes their geospatial relationship and the spatial attributes correlation, and thus their neighborhood relationship is determined by consensus of the two factors. Therefore, the matrix \mathbf{T} and matrix \mathbf{T}_δ both contain the factors classification $t_{1(1)}$, popularity degree $t_{1(2)}$, the optimal travel time $t_{1(3)}$, and the traveling cost $t_{1(4)}$, as well as longitude and latitude $(t_{2(1)}, t_{2(2)}) \sim (l, B)$. The improved Minkowski distance is applied to for the objective function, and the clustering criteria should consider features and spatial attributes simultaneously. The pseudo-code of the process to create the function $\tilde{\zeta}_{(s_{(i1)}, s_{(i2)})}$. (Algorithm 1) is shown as follows.

Algorithm 1: The process to create the function $\tilde{\zeta}_{(s_{(i1)}, s_{(i2)})}$

- 1: **Step 1:** Confirm $\mathbf{T}_{\delta(i)}$ for $s_{(i)}$ in \mathbf{S} .
 - 2: **Step 2:** For the $s_{(i1)}$, extract the non-zero elements $\delta_{1,1(i1)} \cdot t_{1,1(i1,j1)}$ and $t_{2,1(i2,j2)}$ in $\delta_{1,1(i1)} \cdot \mathbf{t}_{1,1(i1)}$ and $\mathbf{t}_{2,1(i2)}$ of $\mathbf{T}_{\delta(i1)}$. Transpose $\mathbf{T}_{\delta(i1)}$ to $\mathbf{T}_{\delta(i1)}^T$.
 - 3: **Step 3:** For the $s_{(i2)}$, extract the non-zero elements $\delta_{1,2(i1)} \cdot t_{1,2(i1,j1)}$ and $t_{2,2(i2,j2)}$ in $\delta_{1,1(i1)} \cdot \mathbf{t}_{1,1(i1)}$ and $\mathbf{t}_{2,1(i2)}$ of $\mathbf{T}_{\delta(i2)}$. Transpose $\mathbf{T}_{\delta(i2)}$ to $\mathbf{T}_{\delta(i2)}^T$.
 - 4: **Step 4:** Confirm the Minkowski distance $d(\mathbf{x}, \mathbf{y})$ as the objective function $\tilde{\zeta}_{(s_{(i1)}, s_{(i2)})}$
-

The Minkowski distance between the two samples \mathbf{x} and \mathbf{y} is shown in Equation (3). The Minkowski distance is used to define the objective function $\tilde{\zeta}_{(s_{(i1)}, s_{(i2)})}$, shown as Equations (4) and (5). According to the function $\tilde{\zeta}_{(s_{(i1)}, s_{(i2)})}$, the norm value of the function is used to judge whether the tourist attractions $s_{(i1)}$ and $s_{(i2)}$ belong to the same cluster. Therefore, the function $\tilde{\zeta}_{(s_{(i1)}, s_{(i2)})}$ value is set as the clustering criterion.

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_p = \left[\sum_{i=1}^n |x_{(i)} - y_{(i)}|^r \right]^{1/r} \quad (3)$$

$$\tilde{\zeta}_{(s_{(i1)}, s_{(i2)})} = \left\| \mathbf{T}_{\delta(i1)}^T - \mathbf{T}_{\delta(i2)}^T \right\|_2 \quad (4)$$

$$\tilde{\zeta}_{(s_{(i1)}, s_{(i2)})} = \left[\sum_{i1=1}^m \left| \delta_{1,1(i1)} \cdot t_{1,1(i1,j1)} - \delta_{1,2(i1)} \cdot t_{1,2(i1,j1)} \right|^2 + \sum_{i2=1}^n \left| t_{2,1(i2,j2)} - t_{2,2(i2,j2)} \right|^2 \right]^{1/2} \quad (5)$$

In the process of generating clusters, the k number of elements $s_{(i)}$ are dynamically stored into one matrix $\mathbf{K}^\wedge(p \times \max k_{(i)})$ in the cluster code sequence by the clustering algorithm. Each row in the matrix dynamically stores the related cluster's elements. When the clustering algorithm ends, all the tourist attraction elements are consistently stored in the matrix $\mathbf{K}(p \times \max k_{(i)})$ according to the cluster code i and cluster's element code j . The matrix row number is p , the column number is $\max k_{(i)}$, in which k number of elements are used to store tourist attractions, while the other $p \times \max k_{(i)} - k$ number of elements are stored as 0. The row rank meets at $\text{rank}(\mathbf{K}^\wedge(p \bullet)) \leq p$ and $\text{rank}(\mathbf{K}(p \bullet)) \leq p$. The column rank meets at $\text{rank}(\mathbf{K}^\wedge(\bullet \max k_{(i)})) \leq \max k_{(i)}$ and $\text{rank}(\mathbf{K}(\bullet \max k_{(i)})) \leq \max k_{(i)}$. The matrix $\mathbf{K}(p \times \max k_{(i)})$ has at least two non-empty rows. Equations (6) and (7) relate to the matrix $\mathbf{K}^\wedge(p \times \max k_{(i)})$ and $\mathbf{K}(p \times \max k_{(i)})$, in the formula, $s_{(i)}^\wedge$ represents the element with random storage value.

$$\mathbf{K}^\wedge(p \times \max k_{(i)}) = \begin{bmatrix} s_{(1)}^\wedge & s_{(2)}^\wedge & \dots & 0 & s_{(i1)}^\wedge \\ s_{(i1+1)}^\wedge & \dots & 0 & \dots & s_{(i2)}^\wedge \\ 0 & \dots & s_{(i)}^\wedge & \dots & 0 \\ s_{(i3)}^\wedge & \dots & 0 & \dots & s_{(k)}^\wedge \end{bmatrix} \quad (6)$$

$$\mathbf{K}(p \times \max k_{(i)}) = \begin{bmatrix} s_{(1,1)} & s_{(1,2)} & \dots & \dots & s_{(1, \max k(1))} \\ s_{(2,1)} & \dots & s_{(2, k(2))} & \dots & s_{(2, \max k(2))} \\ \dots & \dots & s_{(i, j)} & \dots & 0 \\ s_{(p, 1)} & \dots & 0 & s_{(p, \max k(p))} & 0 \end{bmatrix} \quad (7)$$

The tourist attraction clustering objective function $\tilde{\zeta}_{(s_{(i1)}, s_{(i2)})}$ is set as the improved AGNES clustering algorithm criterion. The k number of elements $s_{(i)}$ in the domain \mathbf{S} are clustered into p number of clusters and stored into the matrix $\mathbf{K}(p \times \max k_{(i)})$.

In the improved AGNES clustering algorithm, in a single instance of dot gathering from the bottom to top, a seed point element $s_{(i)}^*$ is chosen as the tourist attraction representing a certain cluster $S_{(i)}$. Take the element $s_{(i)}^*$ as a criterion to calculate and judge the objective function $\tilde{\zeta}_{(s_{(i1)}, s_{(i2)})}$ to confirm another element to be gathered and form the cluster. The tourist attractions that are not the seed point are noted as $\neg s_{(i)}^*$.

In one instance of gathering from bottom to top, if one point $\neg s_{(i)}^*$ belongs to the cluster $S_{(i)}$, the point $\neg s_{(i)}^*$ is absorbed into the cluster $S_{(i)}$, the edge $l(s_{(i)}, \neg s_{(i)})$ connecting $s_{(i)}^*$ and $\neg s_{(i)}^*$ is generated. When the clustering algorithm ends, the $k_{(i)}$ number of tourist attractions as well as the gathered $k_{(i)} - 1$ number of topological edges $l(s_{(i)}, \neg s_{(i)})$ in the cluster $S_{(i)}$ form a cluster structure tree $T_r(S_{(i)})$. The spatial range that is expanded by the tree $T_r(S_{(i)})$ forms the cluster spatial buffer $r_a(S_{(i)})$. The edge $l(s_{(i)}, \neg s_{(i)})$ and the tree $T_r(S_{(i)})$ show the visualized process of the improved AGNES algorithm. The buffer $r_a(S_{(i)})$ is the visualized range for each cluster. Since the objective function $\tilde{\zeta}_{(s_{(i1)}, s_{(i2)})}$ contains both feature attributes and spatial attributes, different buffers $r_a(S_{(i)})$ may intersect. Figure 2 shows the spatial relationship among the cluster $S_{(i)}$ topological edge $l(s_{(i)}, \neg s_{(i)})$, cluster structure tree $T_r(S_{(i)})$, and the cluster spatial buffer $r_a(S_{(i)})$. Figure 2a is an edge $l(s_{(i)}, \neg s_{(i)})$, Figure 2b is the tree $T_r(S_{(i)})$ which is formed by several edges $l(s_{(i)}, \neg s_{(i)})$, and Figure 2c is the buffer $r_a(S_{(i)})$ which is formed by the cluster structure tree $T_r(S_{(i)})$.

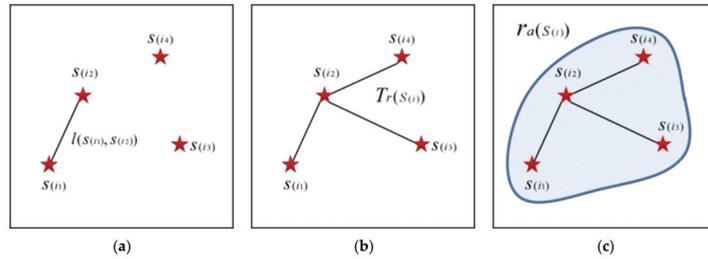


Figure 2. The spatial relationship among the cluster $S_{(i)}$ topological edge $l(s_{(i)}, \neg s_{(i)})$, cluster structure tree $T_{r(S_{(i)})}$, and the cluster spatial buffer $r_{a(S_{(i)})}$. (a) is an edge $l(s_{(i)}, \neg s_{(i)})$, (b) is the tree $T_{r(S_{(i)})}$ that is formed by several edges $l(s_{(i)}, \neg s_{(i)})$, and (c) is the buffer $r_{a(S_{(i)})}$ that is formed by the cluster structure tree $T_{r(S_{(i)})}$.

According to the modeling principle, the improved AGNES clustering algorithm has been created. The smart system will search the optimal tourist attractions and tour routes by the p number of clusters and tourists' interests, time budget, and cost budget, etc. The pseudo-code of the process to create the improved AGNES clustering algorithm (Algorithm 2) is shown as follows:

Algorithm 2: The process to create the improved AGNES clustering algorithm

- Step 1:** Create $\zeta(s_{(i1)}, s_{(i2)})$ to store $S_{to}(\bullet)$ and its descending $S_{to}(\bullet)^d$ with element $S_{to(i,j)}$. The value $\zeta(s_{(i1)}, s_{(i2)})$ is stored in the descending order from the first row first column to the last one.
- Step 2:** Confirm seed point $s_{(i)*}$.
 - Sub-step 1: Take the former e number of elements of the matrix $S_{to}(\bullet)^d$.
 - Sub-step 2: Note $S_{to(maxi,maxj)}^d$ relates to the tourist attractions $a_{(2e-1)}$ and $a_{(2e)}$.
 - Sub-step 3: $a_{(o)}$ relates to the noted $s_{(i)}$ in S .
 - Sub-step 4: Search $S_{to(i,j)}^d$ on $\forall s_{(i1)}$ and $\forall s_{(i2)}$ of $\forall a_{(o1)}$ and $\forall a_{(o2)}$ in $S_{to}(\bullet)^d$.
 - Sub-step 5: The p number of $a_{(i)}$ with the maximum value $S_{to(i,j)}^d$ as $s_{(i)*}$.
- Step 3:** Form the cluster $S_{(i)}$ by $s_{(i)*}$.
 - Sub-step 1: Store the No.1 $s_{(i)*}$ into $K^{\wedge}_{(i,1)}$ in K^{\wedge} .
 - Sub-step 2: Search $\zeta(s_{(i)*}, \neg s_{(i)*})$ and find $\min \zeta(s_{(i)*}, \neg s_{(i)*})$.
- 3: Sub-step 3: Judge $\neg s_{(i)*}$ whether belongs to $s_{(i)}$ cluster $S_{(i)}$.
 - If $\neg s_{(i)*}$ belongs to $S_{(i)}$, store it into the $S_{(i)}$ row; If it does not belong, store into another row.
 - Sub-step 4: Form $l(\neg s_{(i)*}, s_{(i)*})$ with the $s_{(i)*}$ and its cluster $S_{(i)}$ tourist attractions.
- 4: **Step 4:** For the structure tree $T_{r(S_{(i)})}$.
- 5: **Step 5:** Expand the tree $T_{r(S_{(i)})}$ from $l(\neg s_{(i)*}, s_{(i)*})$ with a radius range r and form the cluster spatial buffer $r_{a(S_{(i)})}$.

The proposed AGNES clustering algorithm is significantly different from those that have been used in previous research (see the Introduction section). First, the aim is totally different; the proposed method is to find out the classifications of city tourist attractions, and it tends to extract the correlation among different tourist attractions and calculate the degree of correlation between two tourist attractions, and finally output the tourist attraction clusters. This clustering process is the critical step for tourists' interests matching the tourist attractions' attributes. The previous methods did not concern tourist attractions clustering, and they mainly tended to find out the tourists' clusters, tourists' behaviour, and the relationship between the collaborative economy and tourism, etc. Second, the parameters that were used in developing the AGNES model are different. Besides the spatial attributes, the proposed AGNES algorithm makes improvements on the clustering criterion function by adding tourist attraction's feature attributes, which makes the tourist attraction clustering more logical, since the clusters and tourist attractions are grouped to

match the tourists' interests. The previous methods directly used the AGNES algorithm itself on the basis of spatial attributes such as longitude and latitude. Third, since the proposed AGNES algorithm is an improved method, the detailed steps on modeling the algorithm are provided in the paper. It is an important research content and precondition of the whole research work. In previous studies, the AGNES algorithm is merely a tool that is used by the authors without detailed algorithm modeling process.

3. Tour-Route-Recommendation Algorithm Based on the Space-Time Deduction

The selection of tourist attractions and tour-route design are the two critical factors for any tour itinerary. Tourists must choose the tourist attractions that best match their interests and then plan the most reasonable route based on their selections. Time and cost are always limitations, to which travel and participation significantly contribute. Therefore, with available time and financial expectations as fixed conditions, a "smart" system should be able to recommend attractions that best-match an individual's preferences as well as optimize the transportation route. Since the mode of transportation would be largely influenced by the tourist themselves, it was a crucial factor for consideration when developing our model [36,37].

3.1. Tourist Attraction Reachability Space Model Based on Interest Matrix and Geographical Position

The precondition for the smart system to recommend tourists with tour routes was obtaining the tourist interest data. The interest data were set as the input labels and then matched with the tourist attraction attributes. The capacity of each tourist attraction to satisfy a tourist's interests would be different, and this capacity was defined as the reachable capacity, the value of which would dictate the likelihood of its recommendation by the system. Therefore, creating a reachability space model between the tourist interest data and the tourist attractions was the precondition when searching for tourist attractions that would best satisfy an individual's interests [38,39].

The tourist interest label vector $\mathbf{n}_{1(i_1)}$ and spatial positioning vector $\mathbf{n}_{2(i_2)}$ have the same dimension as the vectors of $\mathbf{t}_{1(i_1)}$ and $\mathbf{t}_{2(i_2)}$, and they represent the tourist-interest data. The variable $\mathbf{n}_{1(i_1)}$ is a $1 \times u_{(i_1)}$ dimension vector. The interest label factor $n_{1(i_1)}$ contains $u_{(i_1)}$ items of different classifying indices $n_{1(i_1, j_1)}$ and $j_1 \in (0, \alpha] \subset \mathbb{Z}^+$. The variable j_1 is the footnote for the index $n_{1(i_1, j_1)}$ of the factor $n_{1(i_1)}$, and α is the maximum number of the index. The variable $\mathbf{n}_{2(i_2)}$ is a $1 \times u_{(i_2)}$ dimension vector. The spatial positioning factor $n_{2(i_2)}$ contains $u_{(i_2)}$ items of different classifying indices $n_{2(i_2, j_2)}$ and $j_2 \in (0, \alpha] \subset \mathbb{Z}^+$, where j_2 is the footnote for the index $n_{2(i_2, j_2)}$ of the factor $n_{2(i_2)}$, and α is the maximum number of the index. The number of vectors $\mathbf{n}_{1(i_1)}$ and $\mathbf{n}_{2(i_2)}$ are m and n .

The starting point of one tour route for the tourist is S_t . The point S_t determines the dimension and specific values of the spatial location vector $\mathbf{n}_{2(i_2)}$. The matrix \mathbf{N} is formed by m number of feature attribute label vectors $\mathbf{n}_{1(i_1)}$ and n number of spatial attribute label vector $\mathbf{n}_{2(i_2)}$ and represents the tourists' interest tendency. The matrix row is the vector $\mathbf{n}_{1(i_1)}$ or $\mathbf{n}_{2(i_2)}$ and the column is the specific element of the vector $\mathbf{n}_{1(i_1)}$ or $\mathbf{n}_{2(i_2)}$. It contains $m + n$ number of rows and α number of columns. The No.1 to No. m rows relate to the vector $\mathbf{n}_{1(i_1)} \sim i_1 \in (0, m] \subset \mathbb{Z}^+$, the No. $m + 1$ to No. $m + n$ rows relate to the vector $\mathbf{n}_{2(i_2)} \sim i_2 \in (0, n] \subset \mathbb{Z}^+$. When the tourist interest data are confirmed, the arbitrary row $\mathbf{n}_{1(i_1)}$ will form one item of an attribute element value $n_{1(i_1, j_1)}$, $j_1 \in (0, \alpha] \subset \mathbb{Z}^+$, and the other elements are 0. The Equation (8) is the general formula \mathbf{N} and its specific elements.

$$\mathbf{N} = \left[\mathbf{n}_{1(1)}, \dots, \mathbf{n}_{1(m)}, \mathbf{n}_{2(1)}, \dots, \mathbf{n}_{2(n)} \right]^T = \begin{bmatrix} n_{1(1,1)} & n_{1(1,2)} & n_{1(1,3)} & \dots & 0 & 0 \\ & & \dots & & \dots & \\ n_{1(m,1)} & n_{1(m,2)} & \dots & & \dots & n_{1(m,\alpha)} \\ n_{2(1,1)} & n_{2(1,2)} & \dots & 0 & \dots & n_{2(1,\alpha)} \\ & & \dots & & & \\ n_{2(m+n,1)} & n_{2(m+n,2)} & \dots & n_{2(m+n,4)} & \dots & 0 \end{bmatrix} \quad (8)$$

The matrix \mathbf{N} elements are related to the matrix \mathbf{T} elements, including tourist classification $\mathbf{n}_{1(1)}$, popularity degree $\mathbf{n}_{1(2)}$, travel time $\mathbf{n}_{1(3)}$, traveling fee $\mathbf{n}_{1(4)}$, longitude $\mathbf{n}_{2(1)}$, and latitude $\mathbf{n}_{2(2)}$.

$\mathbf{n}_{1(1)}$: { $n_{1(1,1)}$: nature park (1.00); $n_{1(1,2)}$: humanistic history (2.00); $n_{1(1,3)}$: amusement park (3.00); $n_{1(1,4)}$: leisure shopping (4.00); $n_{1(1,5)}$: modern science and technology (5.00); $n_{1(1,6)}$: artistic aesthetics (6.00).};

$\mathbf{n}_{1(2)}$: { $n_{1(2,1)}$: $h_o \in (0, 0.25]$; $n_{1(2,2)}$: $h_o \in (0.25, 0.50]$; $n_{1(2,3)}$: $h_o \in (0.50, 0.75]$; $n_{1(2,4)}$: $h_o \in (0.75, 1.00)$ }, $n_{1(2,j_1)} \in \mathbb{R}^+$;

$\mathbf{n}_{1(3)}$: { $n_{1(3,1)}$: $t_b \in (0, 2.0]$; $n_{1(3,2)}$: $t_b \in (2.0, 4.0]$; $n_{1(3,3)}$: $t_b \in (4.0, 6.0]$; $n_{1(3,4)}$: $t_b \in (6.0, +\infty)$ }, $n_{1(3,j_1)} \in \mathbb{R}^+$;

$\mathbf{n}_{1(4)}$: { $n_{1(4,1)}$: $c_o \in (0, 100]$; $n_{1(4,2)}$: $c_o \in (100, 200]$; $n_{1(4,3)}$: $c_o \in (200, 300]$; $n_{1(4,4)}$: $c_o \in (300, +\infty)$ }, $n_{1(4,j_1)} \in \mathbb{R}^+$.

The spatial location vector $\mathbf{n}_{2(i2)}$ of the matrix \mathbf{N} is determined by the longitude and latitude of the point S_i .

The correlation between the tourists' interest and the tourist attraction attributes is determined by the interest quantitative matching objective function $\zeta_{(N,T)}$. Transfer the feature attribute label vector normalized parameter $\delta_{1(i_1)}$ and take it as the parameter to create the function $\zeta_{(N,T)}$, then confirm the tourist-interest data. Traverse $j_1, j_2 \sim (0, \alpha]$, search and extract the non-zero elements $\delta_{1(i_1)} \cdot n_{1(i_1,j_1)}$ and $n_{2(i_2,j_2)}$ in the matrix \mathbf{N} label vector $\delta_{1(i_1)} \cdot \mathbf{n}_{1(i_1)}$ and $\mathbf{n}_{2(i_2)}$. Transpose the matrix \mathbf{N} and generate the matrix \mathbf{N}^T . Create the norm relationship of the Minkowski distance between the tourists' interests and the tourist attractions, shown in the Equations (9) and (10). Calculate the function $\zeta_{(N,T)}$ between the matrix \mathbf{N} and the matrix \mathbf{T} . Use the interest quantitative matching objective function matrix $\mathbf{P}_{\zeta(N,T)}$ to store the function value.

$$\zeta_{(N,T)} = \left\| \mathbf{N}^T - \mathbf{T}^T \right\|_2 \quad (9)$$

$$\zeta_{(N,T)} = \left[\sum_{i1=1}^m \left| \delta_{1(i1)} \cdot n_{1(i1,j1)} - \delta_{1(i1)} \cdot t_{1(i1,j1)} \right|^2 + \sum_{i2=1}^n \left| n_{2(i2,j2)} - t_{2(i2,j)} \right|^2 \right]^{1/2} \quad (10)$$

When the interest data remains the same, the tourist attractions $s_{(i)}$ in different clusters will generate different function values $\zeta_{(N,T)}$. The values $\zeta_{(N,T)}$ are stored in the sequence of the cluster $S_{(i)}$ footnote of the matrix $\mathbf{P}_{\zeta(N,T)}$. The value $\zeta_{(N,T)}$ is stored in the $\mathbf{P}_{\zeta(N,T)}$ in ascending order from the first row and column to the last one. When tourists confirm the interest data, they contain the longitude and latitude of the starting point S_i . Taking the point S_i as the center core, each row of the matrix $\mathbf{P}_{\zeta(N,T)}$ represents the correlation between the tourist attractions and the interest data, and also represents the reachability extent of the tourist attractions.

3.2. The Dynamic Space-Time Deduction Algorithm Based on the Travel Time and Cost

During a city tour, tourists expect to visit several tourist attractions in one day; tourists have different interests and levels of desire to visit various kinds of tourist attractions that will each involve different time investments and associated costs. Therefore, when the time and cost are fixed, the number of tourist attractions to be visited must be finite. According to the tourist attraction reachability model, the smart system would formulate a tour route that would best match tourist interests while meeting the time and cost conditions.

Furthermore, depending on the mode of transportation that was chosen, the goal of saving time and cost could result in better attraction recommendations and optimized route-planning. Since travel time would be directly affected by the mode of transportation and route between attractions, the precondition for the dynamic space–time deduction of the tour-route-recommendation algorithm had the lowest path-searching cost [37,38].

3.2.1. The Shortest-Path-Searching Algorithm Based on the Space-Vector Lattice

After visiting a tourist attraction, tourists will move to the next one. This activity is based on specific activities. First, tourists will use a transportation mode such as walking, cycling, taxi service, etc. Second, they will travel city roads to the destination. Third, the moving process will consume time and cost.

The traffic space between two tourist attractions is the tourist attraction traffic sub-space Φ . The space Φ is the interval from point A to point B , and it is a vector space with coordinates, shown in Figure 3. The left bottom dot of the square Φ is the origin of coordinate. Each line represents an abstract city road. The line intersection $a_{(i)}$ represents the road intersection. In the Figure 3, the space Φ contains all city roads between the two points A and B . The road distance $d_{is}(a_{(i)}, a_{(j)})$ of the edges CD , DE , EF , and CF in the small square $CDEF$ may be different.

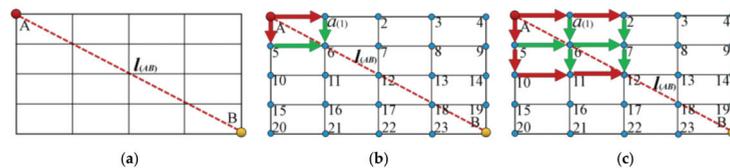


Figure 3. The space-vector lattice between the point A and B to search the shortest path. (a) is the spatial connecting line of the space Φ . (b) is the spatial road and lattice relationship as well as the searching process for the series contained in the square $S_{qu}(A, a_{(1)}, a_{(5)}, a_{(6)})$. (c) is the spatial road-lattice relationship as well as the searching process for the series that is contained in the square $S_{qu}(A, a_{(2)}, a_{(10)}, a_{(12)})$.

Starting from the point A , search the path along the road until the point B is reached; in the whole process, all the searched points are listed in the spatial searching series S_{eq} . The searching series S_{eq} represents a reachable path that is related to a searched distance $d_{is}(S_{eq})$. Figure 3 shows the shortest-path algorithm searching process.

The pseudo-code of the process to create the shortest-path algorithm (Algorithm 3) is shown as follows. This searching mode considers all the city roads and intersections between two points and finds the global minimum value, which may be more precise than the other shortest-path algorithms. The shortest path may reduce time and costs and thus increase the number of tourist attractions to be visited.

3.2.2. The Dynamic Space-Time Deduction Tour-Route-Searching Algorithm

Once the tourist attractions are been identified based on tourist interests, the travel time and cost would determine the number of tourist attractions to be visited and the optimal travel route, all of which would influence a tourist's overall experience. Based on the matrix $\mathbf{P}_{\xi(N,T)}$ and fixed time and cost conditions, the dynamic space-time deduction tour-route-searching algorithm was created. The basic process of the algorithm was as follows: using a daytrip as an example, a tourist confirms the travel time budget t (unit: hour), traveling fee c (unit: CNY ¥ yuan), and then chooses one transportation mode. Starting from the point S_t , search the shortest path between the point S_t and tourist attractions and confirm the travel time using the chosen transportation mode. Iterate the visiting time in the tourist attractions and calculate the travel costs and any entrance or activity fees. Search the minimum time and cost between the point S_t as well as all the tourist attractions and set the related point as the first tourist attraction $K_{(1)}$ to be visited. Starting from the $K_{(1)}$ point, search the

next minimum travel time and cost tourist attraction $K_{(2)}$ until the total travel time or cost reaches the preset time budget t or cost budget c , and then the optimal tour route is defined.

Algorithm 3: The process to create the shortest-path algorithm

- 1: **Step 1:** Make line $l_{(AB)}$ between A and B in Figure 3a.
- 2: **Step 2:** Take the points $a_{(6)}, a_{(12)}, a_{(18)}$
Step 3: Search and compare S_{eq} in $S_{qu}(A, a_{(1)}, a_{(5)}, a_{(6)})$ in Figure 3b.
 Sub-step 1: Search $S_{eq}(A, a_{(1)}, a_{(6)})$, noted as $S_{eq(1)}$, road distance is $d_{is}(S_{eq}(A, a_{(1)}, a_{(6)}))$.
- 3: Sub-step 2: Search $S_{eq}(A, a_{(5)}, a_{(6)})$, noted as $S_{eq(2)}$, road distance is $d_{is}(S_{eq}(A, a_{(5)}, a_{(6)}))$.
 Sub-step 3: Find the minimum one $mind_{is}(S_{eq})_{(1)}$ between A and $a_{(6)}$.
Step 4: Search and compare S_{eq} in $S_{qu}(A, a_{(2)}, a_{(10)}, a_{(12)})$ in Figure 3c.
 Sub-step 1: Search $S_{eq}(S_{eq(1)}, a_{(7)}, a_{(12)}), S_{eq}(S_{eq(1)}, a_{(11)}, a_{(12)}), S_{eq}(A, a_{(1)}, a_{(2)}, a_{(7)}, a_{(12)})$,
- 4: $S_{eq}(S_{eq(2)}, a_{(11)}, a_{(12)}), S_{eq}(S_{eq(2)}, a_{(7)}, a_{(12)}), S_{eq}(A, a_{(5)}, a_{(10)}, a_{(11)}, a_{(12)})$.
 Sub-step 2: Find the minimum one $mind_{is}(S_{eq})_{(2)}$ between A and $a_{(12)}$.
- 5: **Step 5:** Continue searching until the square $S_{qu}(A, a_{(4)}, a_{(20)}, B)$ is finished. Find the minimum one $mind_{is}(S_{eq})_{(4)}$ between A and B .

In a tour, the road interval from the point A to B is a cost iteration sub-unit $Q_{(i)}$. This sub-unit is the basic unit for the dynamic deduction process. The sum of the travel time contains the visiting time of a tourist attraction and the travel time from the point A to B ; it is the time consumption $t_{(i)}$ of the sub-unit. The time $t_{(i)}$ is determined by the tourist attraction B visiting time and the travel time to the B . The sum of traveling costs contains the visiting fees of the tourist attractions and the travel costs from the point A to B ; it is the cost consumption $c_{(i)}$ of the sub-unit. Starting from the point S_t , the tourist passes through n number of sub-units $Q_{(i)}$ and finally deduces to the terminal tourist attraction P ; in this process, the total time and costs are noted as the dynamic deduction time Δt dynamic deduction cost Δc , as shown in the Equation (11):

$$\Delta t = \sum_{i=1}^n t(i) , \quad \Delta c = \sum_{i=1}^n c(i) , \quad i, n \in N \tag{11}$$

A $1 \times k$ dimension vector T_s is used to consistently store tourist attractions that represent the tour route after the searching process. The sequence of the matrix T_s in storing the tourist attractions obeys the algorithm rule, and the empty elements are 0. A $1 \times k$ dimension vector ΔT_s is used to dynamically store the tourist attractions in the searching process, and the empty elements are 0. The pseudo-code of the process to create the tour-route-searching algorithm (Algorithm 4) is as follows:

Algorithm 4: The process to create the tour-route-searching algorithm

- Step 1:** Confirm transportation mode m_o , starting point S_t , time budget t , and cost budget c . Variable $m_o = 1$ represents taking the bicycle, $m_o = 2$ represents taking the taxi, and $m_o = 3$ represents taking the public bus.
- 2: **Step 2:** Determine whether clusters $S_{(i)}$ element $P_{\xi(N,T)}(i, j)$ in matrix $P_{\xi(N,T)}$ meet tourist interests. If meet, keep the best ones.
- 3: **Step 3:** Store the kept Step 2 element $P_{\xi(N,T)}(i, j)$ into ΔT_s .
- 4: **Step 4:** Find the maximum $P_{\xi(N,T)}(i, 1)*$ and confirm $\Phi(S_t, \Delta T_{s(1)})$ and $Q_{(1)}$.
- 5: **Step 5:** Search the shortest path between the points S_t and $\Delta T_{s(1)}$. Calculate $t_{(1)}$ and $c_{(1)}$ of $Q_{(1)}$ according to m_o .
 $\Delta t = t_{(1)}, \Delta c = c_{(1)}$, judge whether $\Delta t \leq t \wedge \Delta c \leq c$ Continue or stop.
- 6: **Step 6:** Confirm $Q_{(i)}$, search the shortest path in $Q_{(i)}$. Calculate $t_{(i)}$ and $c_{(i)}$ of $Q_{(i)}$ according to m_o .
 $\Delta t = \sum_{i=1}^v t_{(i)}, \Delta c = \sum_{i=1}^v c_{(i)}$, judge whether $\Delta t \leq t \wedge \Delta c \leq c$. Continue or stop.
- 7: **Step 7:** Continue searching until $\Delta t > t$ or $\Delta c > c$. Output the tour route.

4. Sample Experiment and Result Analysis

To verify the advantages of the proposed algorithm, the tourism city of Chengdu was selected as the subject of the experiment. The basic thought of the experiment is as follows. First, 15 popular tourist attractions in the Chengdu City were selected. All the tourist attraction feature attributes and spatial attributes were confirmed and quantified. According to the tourist attraction attributes, we used the proposed clustering algorithm to obtain tourist attraction labels and clusters, cluster structure trees, and cluster spatial buffers. Based on these clusters, the tourist-interest data were obtained and the quantified interest-matching objective function matrix was created. According to the tourist time and cost allowances as well as the preferred mode of travel, the tourist attractions and tour routes were analyzed for optimal matches. For the tour-route optimization, the experiment chose two frequently used shortest path searching algorithms as a control group to verify the advantages of our proposed algorithm.

4.1. The Collection Result of the Tourist Attraction Attributes

4.1.1. The Results of the Research Range

The tourist attraction research range of the Chengdu City was as follows:

$S = \{s_{(1)}$: Chunxi Road and Zhongshan Square; $s_{(2)}$: Jinsha Site; $s_{(3)}$: Temple of Marquis Wu; $s_{(4)}$: The People's park; $s_{(5)}$: Wide and Narrow Alley; $s_{(6)}$: East Lake Park; $s_{(7)}$: Wenshu Temple; $s_{(8)}$: Qingyang Taoist Temple; $s_{(9)}$: Wangjiang Park; $s_{(10)}$: Jinniu Wanda; $s_{(11)}$: Tazishan Park; $s_{(12)}$: Eastern Suburb Memory; $s_{(13)}$: SM Square; $s_{(14)}$: Chengdu Zoo; $s_{(15)}$: Raffles Square}.

4.1.2. Analysis and Results of the Feature Attribute and Spatial Attribute

Table 1 shows the quantified feature attributes and spatial attributes of each tourist attraction. The symbol $t_{1(1)}$ represents the classification, $t_{1(2)}$ represents the popularity, $t_{1(3)}$ represents the best travel time, $t_{1(4)}$ represents the traveling fee, $t_{2(1)}$ represents the longitude, and $t_{2(2)}$ represents the latitude.

Table 1. The collected quantified feature attributes and spatial attributes of each tourist attraction.

	$t_{1(1)}$	$t_{1(2)}$	$t_{1(3)}$	$t_{1(4)}$	$t_{2(1)}$	$t_{2(2)}$	$t_{1(1)}$	$t_{1(2)}$	$t_{1(3)}$	$t_{1(4)}$	$t_{2(1)}$	$t_{2(2)}$	$t_{1(1)}$
$s_{(1)}$	0.40	0.92	0.30	0	104.077	30.655	$s_{(9)}$	0.10	0.62	0.20	0	104.092	30.629
$s_{(2)}$	0.20	0.74	0.20	0.07	104.012	30.681	$s_{(10)}$	0.40	0.77	0.30	0	104.074	30.686
$s_{(3)}$	0.20	0.88	0.20	0.05	104.047	30.646	$s_{(11)}$	0.10	0.62	0.20	0	104.122	30.634
$s_{(4)}$	0.10	0.68	0.20	0	104.057	30.657	$s_{(12)}$	0.20	0.72	0.20	0	104.123	30.668
$s_{(5)}$	0.20	0.94	0.20	0	104.053	30.663	$s_{(13)}$	0.40	0.68	0.30	0	104.112	30.668
$s_{(6)}$	0.10	0.45	0.10	0	104.087	30.616	$s_{(14)}$	0.10	0.75	0.30	0.02	104.106	30.710
$s_{(7)}$	0.20	0.65	0.20	0	104.072	30.674	$s_{(15)}$	0.40	0.52	0.20	0	104.067	30.631
$s_{(8)}$	0.20	0.66	0.20	0.01	104.041	30.661							

4.2. The Result of the Clustering and Cluster Visualization

4.2.1. The Results of the Function $\xi_{(s_{(i1)}, s_{(i2)})}$ Values

Based on the Table 1 data, the proposed improved AGNES spatial clustering algorithm was performed to generate tourist attraction clusters. Table 2 shows the analyzed results of the clustering objective function $\xi_{(s_{(i1)}, s_{(i2)})}$ values among the tourist attractions.

Table 2. The analyzed results of the clustering objective function $\xi(s_{(i1)}, s_{(i2)})$ values among tourist attractions.

	$s_{(1)}$	$s_{(2)}$	$s_{(3)}$	$s_{(4)}$	$s_{(5)}$	$s_{(6)}$	$s_{(7)}$	$s_{(8)}$	$s_{(9)}$	$s_{(10)}$	$s_{(11)}$	$s_{(12)}$	$s_{(13)}$	$s_{(14)}$	$s_{(15)}$
$s_{(1)}$	0	0.304	0.235	0.397	0.226	0.594	0.351	0.345	0.437	0.153	0.439	0.304	0.243	0.351	0.413
$s_{(2)}$	0.304	0	0.150	0.145	0.217	0.345	0.129	0.106	0.196	0.244	0.209	0.133	0.262	0.180	0.314
$s_{(3)}$	0.235	0.150	0	0.230	0.080	0.458	0.238	0.224	0.287	0.259	0.293	0.185	0.312	0.213	0.416
$s_{(4)}$	0.397	0.145	0.230	0	0.279	0.256	0.107	0.104	0.075	0.330	0.091	0.127	0.321	0.143	0.341
$s_{(5)}$	0.226	0.217	0.080	0.279	0	0.513	0.291	0.280	0.339	0.283	0.344	0.231	0.348	0.248	0.466
$s_{(6)}$	0.594	0.345	0.458	0.256	0.513	0	0.252	0.261	0.198	0.487	0.201	0.311	0.432	0.374	0.325
$s_{(7)}$	0.351	0.129	0.238	0.107	0.291	0.252	0	0.036	0.115	0.254	0.122	0.087	0.229	0.181	0.242
$s_{(8)}$	0.345	0.106	0.224	0.104	0.280	0.261	0.036	0	0.124	0.253	0.138	0.102	0.236	0.187	0.248
$s_{(9)}$	0.437	0.196	0.287	0.075	0.339	0.198	0.115	0.124	0	0.355	0.030	0.150	0.325	0.185	0.317
$s_{(10)}$	0.153	0.244	0.259	0.330	0.283	0.487	0.254	0.253	0.355	0	0.357	0.235	0.099	0.304	0.275
$s_{(11)}$	0.439	0.209	0.293	0.091	0.344	0.201	0.122	0.138	0.030	0.357	0	0.145	0.324	0.183	0.321
$s_{(12)}$	0.304	0.133	0.185	0.127	0.231	0.311	0.087	0.102	0.150	0.235	0.145	0	0.227	0.153	0.291
$s_{(13)}$	0.243	0.262	0.312	0.321	0.348	0.432	0.229	0.236	0.325	0.099	0.324	0.227	0	0.312	0.197
$s_{(14)}$	0.351	0.180	0.213	0.143	0.248	0.374	0.181	0.187	0.185	0.304	0.183	0.153	0.312	0	0.401
$s_{(15)}$	0.413	0.314	0.416	0.341	0.466	0.325	0.242	0.248	0.317	0.275	0.321	0.291	0.197	0.401	0

4.2.1.1. The Output Result of the Clusters

Based on the objective function values and clustering algorithm, the analysis resulted in three tourist attraction clusters $S_{(1)}$, $S_{(2)}$, and $S_{(3)}$ as follows:

- (1) $S_{(1)}$: $s_{(1)}$ -Chunxi Road and Zhongshan Square, $s_{(10)}$ -Jinniu Wanda, $s_{(13)}$ -SM Square, $s_{(15)}$ -Raffles Square.
- (2) $S_{(2)}$: $s_{(2)}$ -Jinsha Site, $s_{(3)}$ -Temple of Marquis Wu, $s_{(5)}$ -Wide and Narrow Alley, $s_{(7)}$ -Wenshu Temple, $s_{(8)}$ -Qingyang Taoist Temple, $s_{(12)}$ -Eastern Suburb Memory;
- (3) $S_{(3)}$: $s_{(4)}$ -The People’s park, $s_{(6)}$ -East Lake Park, $s_{(9)}$ -Wangjiang Park, $s_{(11)}$ -Tazishan Park, $s_{(14)}$ -Chengdu Zoo.

In the clustering process, the cluster structure trees and cluster spatial buffers were generated, as shown in Figure 4. Figure 4a is the tourist attraction distribution, and Figure 4b–d are the visualization results of the structure trees and spatial buffers for the clusters $S_{(1)}$ – $S_{(3)}$.

4.3. The Output Result of the Tourist Attractions and Tour Route

Considering the daytrip example, we chose two tourists as the research objects. Table 3 shows the attribute label values based on the tourist interests. The last two indices were the longitude and latitude of the starting point for each tourist. The first tourist sample $T_{(1)}$ chose to use a bicycle for transportation, while the second tourist sample $T_{(2)}$ chose to use a taxi service.

4.3.1. The Analyzed Results of the Interest-Matching Objective Function Values

Based on the output cluster results and the tourist interest data, the interest-matching objective function $\xi_{(N,T)}$ values between the tourist interests and each tourist attraction were calculated, as shown in Table 4.

4.3.2. The Sequencing Results of Interest-Matching Objective Function Values

Based on the data shown in Table 4, the results were provided in ascending order values of function $\xi_{(N,T)}$ in the sequence of the clusters, as shown in Table 5.

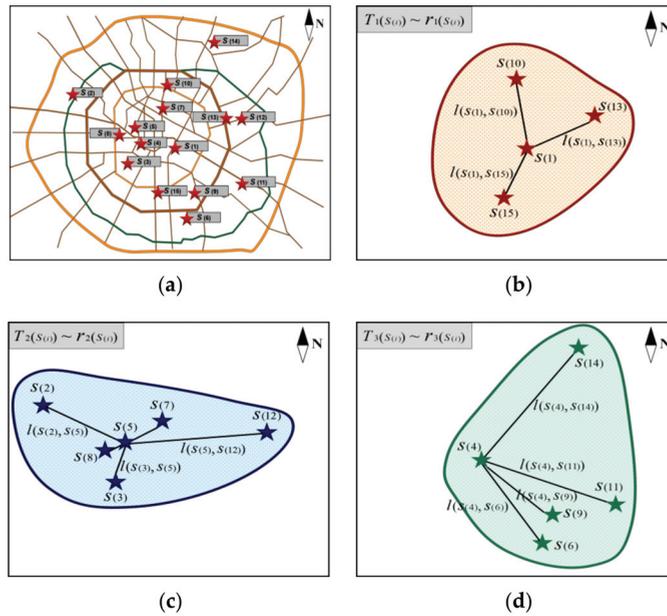


Figure 4. The tourist attraction distribution and clusters, structure trees, and spatial buffers of the clusters. (a) is the tourist attraction distribution. (b–d) are the visualization results of the structure trees and spatial buffers for the clusters $S_{(1)}-S_{(3)}$.

Table 3. The normalization values of tourist samples’ interest labels.

	$n_1(1)$	$n_1(2)$	$n_1(3)$	$n_1(4)$	$n_2(1)$	$n_2(2)$
$T_{(1)}$	0.10	0.90	0.90	0.30	104.052	30.634
$T_{(2)}$	0.20	0.90	1.00	0.50	104.073	30.697

Table 4. The interest-matching objective function $\zeta_{(N,T)}$ values between the tourist samples and each tourist attraction.

$T_{(1)}$					$T_{(2)}$				
$s(1)$	$s(2)$	$s(3)$	$s(4)$	$s(5)$	$s(1)$	$s(2)$	$s(3)$	$s(4)$	$s(5)$
0.736	0.763	0.750	0.793	0.769	0.884	0.924	0.920	0.975	0.945
$s(6)$	$s(7)$	$s(8)$	$s(9)$	$s(10)$	$s(6)$	$s(7)$	$s(8)$	$s(9)$	$s(10)$
0.966	0.809	0.802	0.812	0.748	1.131	0.976	0.969	0.991	0.893
$s(11)$	$s(12)$	$s(13)$	$s(14)$	$s(15)$	$s(11)$	$s(12)$	$s(13)$	$s(14)$	$s(15)$
0.814	0.793	0.770	0.685	0.903	0.992	0.962	0.911	0.868	1.039

Table 5. The interest-matching objective function $\zeta_{(N,T)}$ ascending values between the tourist samples and the cluster tourist attractions.

		$T_{(1)}$					$T_{(2)}$							
$S_{(1)}$		$s_{(1)}$	$s_{(10)}$	$s_{(13)}$	$s_{(15)}$		$S_{(1)}$	$s_{(1)}$	$s_{(10)}$	$s_{(13)}$	$s_{(15)}$			
		0.736	0.748	0.770	0.903			0.884	0.893	0.911	1.039			
$S_{(2)}$		$s_{(3)}$	$s_{(2)}$	$s_{(5)}$	$s_{(12)}$	$s_{(8)}$	$s_{(7)}$	$S_{(2)}$	$s_{(3)}$	$s_{(2)}$	$s_{(5)}$	$s_{(12)}$	$s_{(8)}$	$s_{(7)}$
		0.750	0.763	0.769	0.793	0.802	0.809		0.920	0.924	0.945	0.962	0.969	0.976
$S_{(3)}$		$s_{(14)}$	$s_{(4)}$	$s_{(9)}$	$s_{(11)}$	$s_{(6)}$		$S_{(3)}$	$s_{(14)}$	$s_{(4)}$	$s_{(9)}$	$s_{(11)}$	$s_{(6)}$	
		0.685	0.793	0.812	0.814	0.966			0.868	0.975	0.991	0.992	1.131	

(1) Figure 5a shows the function $\zeta_{(N,T)}$ value distribution of the first tourist in the sequence of the tourist attraction footnotes in the research domain S.

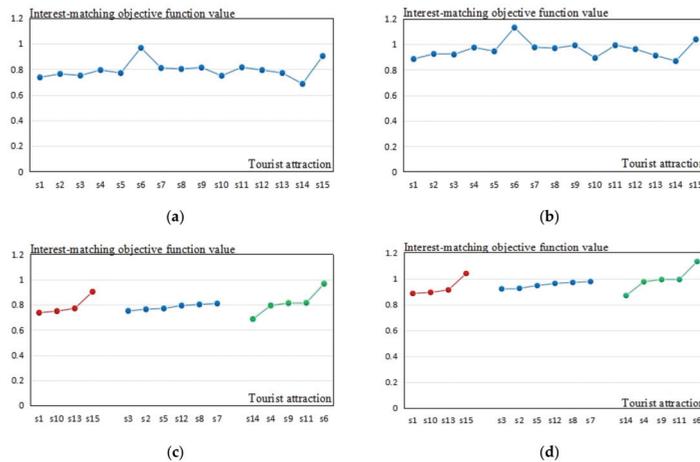


Figure 5. The interest-matching objective function $\zeta_{(N,T)}$ between the tourist samples and the tourist attractions. (a) shows the interest-matching objective function $\zeta_{(N,T)}$ of the first tourist. (b) shows the interest-matching objective function $\zeta_{(N,T)}$ of the second tourist. (c) shows the interest-matching objective function $\zeta_{(N,T)}$ of the first tourist in the cluster sequence. (d) shows the interest-matching objective function $\zeta_{(N,T)}$ of the second tourist in the cluster sequence.

- (2) Figure 5b shows the function $\zeta_{(N,T)}$ value distribution of the second tourist in the sequence of the tourist attraction footnotes in the research domain S.
- (3) Figure 5c shows the function $\zeta_{(N,T)}$ value of the first tourist according to the tourist attraction storage sequence of the interest-matching objective function matrix $P\zeta_{(N,T)}$.
- (4) Figure 5d shows the function $\zeta_{(N,T)}$ value of the second tourist according to the tourist attraction storage sequence of the interest-matching objective function matrix $P\zeta_{(N,T)}$.

In Figure 5c,d, in the cluster sequences, the tourist attraction objective function values in each cluster are listed in the ascending order in which the red curve represents the cluster $S_{(1)}$, the blue curve represents the cluster $S_{(2)}$, and the green curve represents the cluster $S_{(3)}$.

4.3.3. The Results of the Tourist Attractions and Tour-Route Planning

According to the data in Table 3 for a one-day tour, the travel-time allowance for the first tourist sample was 9 h and the cost budget was CNY 300 yuan. The travel-time allowance for the second tourist sample was 11 h and the cost budget was CNY 500 yuan. The first tourist chose to take the bicycle while the second tourist chose to take the taxi.

Based on the proposed algorithm, a potential tourist attraction itinerary and tour route that was based on each tourist sample's interests and their chosen modes of transportation (i.e., bicycle and taxi service, respectively) were identified, as shown in Table 6.

Table 6. The tourist attractions and tour route that best match the tourists' interests.

	S_i	$Q_{(0)}$	$s_{(14)}$	$Q_{(1)}$	$s_{(1)}$	$Q_{(2)}$	$s_{(6)}$	Total		
$T_{(1)}$	0/0	0.80/6.0	3.0/20.0	0.62/4.0	3.0/0	0.35/3.0	1.0/0	8.77/33.0		
	S_i	$Q_{(0)}$	$s_{(14)}$	$Q_{(1)}$	$s_{(1)}$	$Q_{(2)}$	$s_{(6)}$	$Q_{(3)}$	$s_{(3)}$	Total
$T_{(2)}$	0/0	0.27/14.0	3.0/20.0	0.38/22.0	3.0/0	0.32/14.0	1.0/0	0.28/16.0	2.0/50.0	10.25/136

Table 6 shows the results of the tourist attraction element $T_{s(i)}$ of the tour-route searching steady vector T_s and the cost iteration sub-unit $Q_{(i)}$. The values were the required time (unit: hour) and the minimum cost (unit: CNY yuan) to visit the tourist attractions. The values between the two tourist attractions represented the travel time (unit: hour) and minimum travel cost (unit: CNY yuan) in the cost iteration sub-unit $Q_{(i)}$ under the condition of the chosen transportation mode. It also shows the optimal tourist attractions and tour route based on the tourists' requirements.

4.4. The Comparison Results of the Algorithms

To verify the results of the proposed algorithm, a control set of algorithms were conducted under the same experimental conditions and their results compared with those of the proposed algorithm.

4.4.1. Selecting and Confirming of the Control Algorithms

In tourism research, shortest-path algorithms such as the Dijkstra and A* algorithms have typically been used to plan tour routes with the shortest traveling distances. They have the benefits of being easily accessed and applied [40–42]. In addition, the shortest-path algorithms were also constrained by tourism factors such as features and spatial attributes. Once the traveling distances between the tourist attractions have been defined by the city roads and road nodes, the shortest-path algorithms can operate. Since the proposed algorithm's experimental environment conforms to these conditions, the Dijkstra algorithm and the A* algorithm were chosen as controls to plan the travel routes for the sub-unit Φ , and the control group algorithms were defined as Algorithm 1 (A1) and Algorithm 2 (A2). Under the same conditions of the algorithm operating time and the interest data of the two tourist samples, the control group algorithms were used to dynamically search the same tourist attractions, cost iteration sub-units, and tour routes. Their results of were then compared with those of the proposed algorithm (PA), as shown in Table 7, in which the first tourist chose cycling, the second tourist chose a taxi service.

4.4.2. The Comparison Results of the Proposed Algorithm with the Control Algorithms

Table 7 shows the element tourist attractions $T_{s(i)}$ of the steady matrix T_s and the cost iteration sub-units $Q_{(i)}$ under the condition of each algorithm. The values between the two tourist attractions represent the travel time (unit: hour) and minimum moving cost (unit: CNY yuan) in the cost iteration sub-unit $Q_{(i)}$ with the chosen transportation modes. According to Table 7, the Figure 6 curve results were as follows:

Table 7. The tourist attractions and the tour routes that best match tourist interests under the condition of the three algorithms.

	S_t	$Q_{(0)}$	$s_{(14)}$	$Q_{(1)}$	$s_{(1)}$	$Q_{(2)}$	$s_{(6)}$	Total			
PA	$T_{(1)}$	0/0	0.80/6.0	3.0/20.0	0.62/4.0	3.0/0	0.35/3.0	1.0/0	8.77/33.0		
	$T_{(2)}$	0/0	0.27/14.0	3.0/20.0	0.38/22.0	3.0/0	0.32/14.0	1.0/0	0.28/16.0	2.0/50.0	10.25/136
A1	$T_{(1)}$	0/0	0.95/7.5	3.0/20.0	0.73/5.0	3.0/0	0.46/4.0	1.0/0	9.14/36.5		
	$T_{(2)}$	0/0	0.32/15.0	3.0/20.0	0.47/25.0	3.0/0	0.38/17.0	1.0/0	0.38/20.0	2.0/50.0	10.55/147
A2	$T_{(1)}$	0/0	1.00/7.5	3.0/20.0	0.75/5.5	3.0/0	0.45/4.0	1.0/0	9.2/37.0		
	$T_{(2)}$	0/0	0.36/18.0	3.0/20.0	0.43/23.0	3.0/0	0.33/15.0	1.0/0	0.32/19.0	2.0/50.0	10.44/145

- (1) Figure 6a–c shows the deduction and fluctuating tendency of the visiting tourist attraction time, travel time between two tourist attractions, and the total time of the PA, A1, and A2 for the first tourist sample.
- (2) Figure 6d–f shows the deduction and fluctuating tendency of the visiting tourist attraction fee, travel costs between two tourist attractions, and the total costs of the PA, A1, and A2 for the first tourist sample.
- (3) Figure 6g–i shows the deduction and fluctuating tendency of the visiting tourist attraction time, travel time between two tourist attractions, and the total time of the PA, A1, and A2 for the second tourist sample.
- (4) Figure 6j–l shows the deduction and fluctuating tendency of the visiting tourist attraction fee, travel costs between two tourist attractions, and the total costs of the PA, A1, and A2 for the second tourist sample. In each figure, the red and yellow dots represent the total time and total costs, respectively.
- (5) Figure 7a,b shows the comparison of each algorithm on the total time and the total costs for the first tourist under the condition of the first tourist's interest data and the same tourist attractions and cost sub-units.
- (6) Figure 7c,d shows the comparison of each algorithm on the total time and the total costs for the second tourist under the condition of the second tourist's interest data and the same tourist attractions and cost sub-units.

With regard to the computer algorithm optimization, when searching for the shortest route, the Dijkstra algorithm has low efficiency. Compared to the Dijkstra algorithm, the heuristic function is introduced to the A* algorithm, to some extent, the algorithm efficiency was improved. In comparison, the proposed algorithm is based on multiple dot parallel searching, it has higher operating efficiency, and consumes smaller operating space than the Dijkstra algorithm and A* algorithm. Table 8 shows the comparison of the Dijkstra algorithm, A* algorithm, and the proposed algorithm with regard to the time complexity (TC) and space complexity (SC). The data in the table shows the TC and SC examples when the tourist attraction numbers are $n = 4$, $n = 5$, and $n = 6$. The symbol $\rho_{1,1}$ represents the TC ratio between the Dijkstra algorithm and the proposed algorithm, the symbol $\rho_{1,2}$ represents the SC ratio between the Dijkstra algorithm and the proposed algorithm. The symbol $\rho_{2,1}$ represents the TC ratio between the A* algorithm and the proposed algorithm, the symbol $\rho_{2,2}$ represents the SC ratio between the A* algorithm and the proposed algorithm.

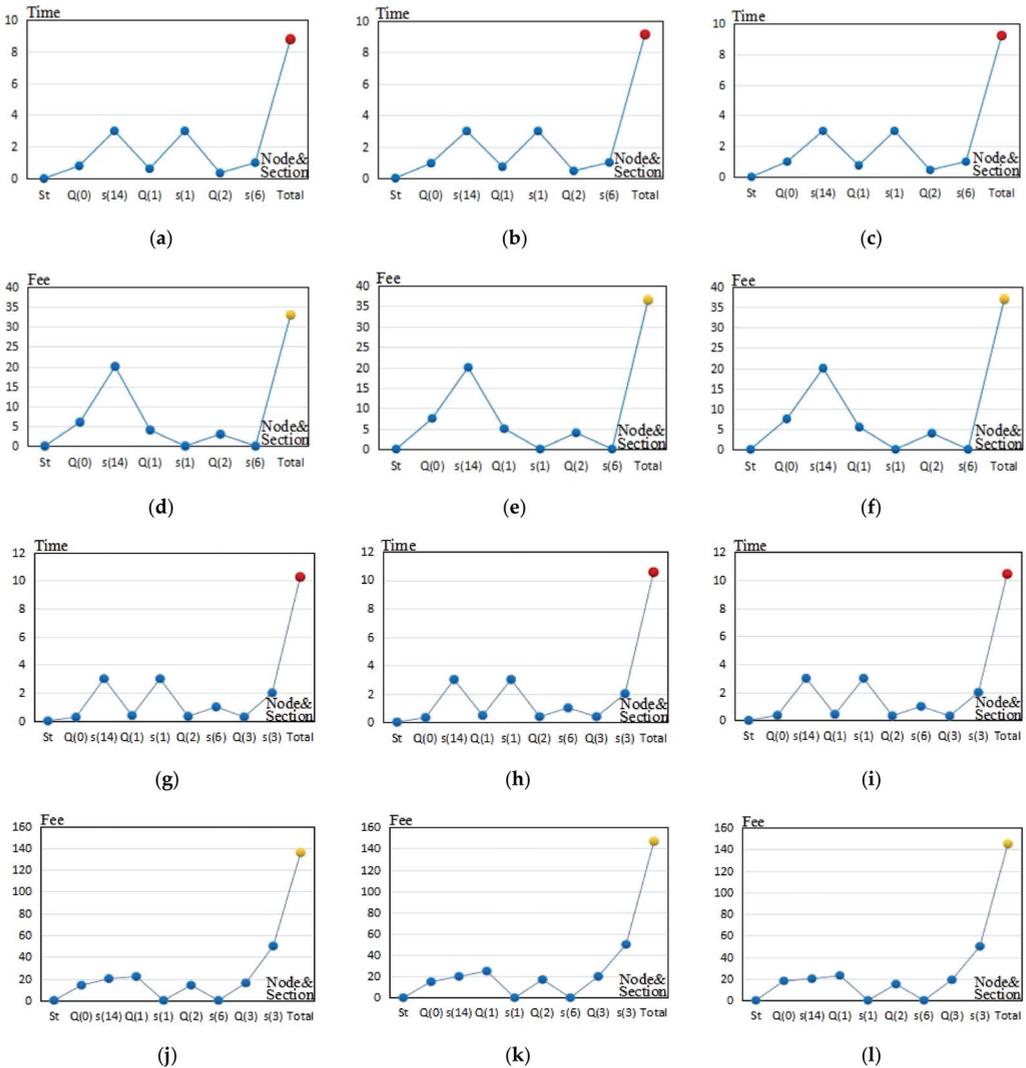


Figure 6. The time- and fee-cost deduction and fluctuating tendency of each algorithm for the two tourist samples. (a–c) are the deduction and fluctuating tendency of the visiting tourist attraction time, travel time between two tourist attractions, and the total time of the proposed algorithm, Algorithm 1, and Algorithm 2, respectively, for the first tourist sample. (d–f) are the deduction and fluctuating tendency of the visiting tourist attraction fee, travel fee between two tourist attractions, and the total costs of the proposed algorithm, Algorithm 1, and Algorithm 2, respectively, for the first tourist sample. (g–i) are the deduction and fluctuating tendency of the visiting tourist attraction time, travel time between two tourist attractions, and the total time of the proposed algorithm, Algorithm 1, and Algorithm 2, respectively, for the second tourist sample. (j–l) are the deduction and fluctuating tendency of the visiting tourist attraction fee, travel fee between two tourist attractions, and the total costs consuming of the proposed algorithm, Algorithm 1, and Algorithm 2, respectively, for the second tourist sample.

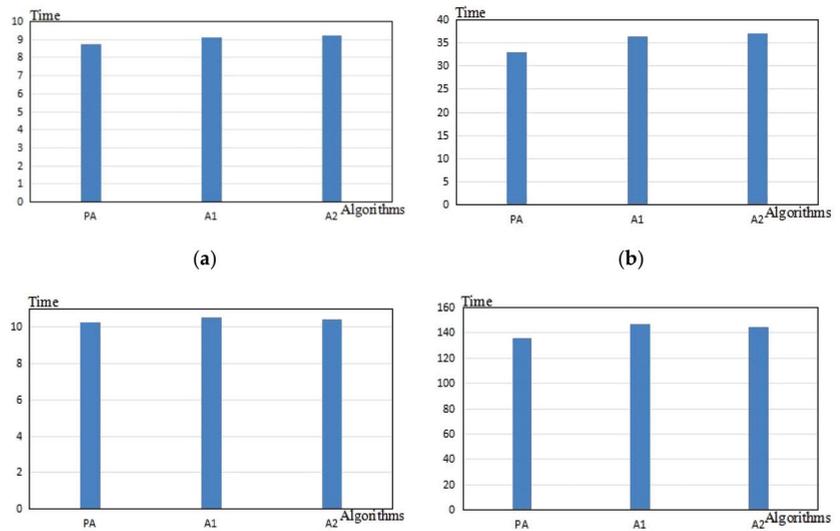


Figure 7. The comparison of the total time and costs of the tour routes for the two tourist samples. (a,b) shows the comparison of each algorithm on the total time and the total costs for the first tourist under the condition of the first tourist’s interest data and the same tourist attractions and cost sub-units. (c,d) shows the comparison of each algorithm on the total time and the total fee cost for the second tourist under the condition of the second tourist’s interest data and the same tourist attractions and cost sub-units.

Table 8. The comparison of the Dijkstra algorithm, A* algorithm, and the proposed algorithm on the aspect of time complexity (TC) and space complexity (SC).

		TC				SC					
	n	$n = 4$	$n = 5$	$n = 6$	ρ	n	$n = 4$	$n = 5$	$n = 6$	ρ	
A1	$O(n \log_2 n + n^2)$	$O(24)$	$O(36.61)$	$O(51.51)$	$\rho_{1,1} > 1$	A1	$O(n)$	$O(4)$	$O(5)$	$O(6)$	$\rho_{1,2} > 1$
A2	$O(n \log_2 n + n)$	$O(12)$	$O(16.61)$	$O(21.51)$	$\rho_{2,1} > 1$	A2	$O(1)$	$O(1)$	$O(1)$	$O(1)$	$\rho_{2,2} > 1$
PA	$O(n \log_2 n)$	$O(8)$	$O(11.61)$	$O(15.51)$		PA	$O(1)$	$O(1)$	$O(1)$	$O(1)$	

4.5. The Analysis and Conclusions of the Experiment Results

4.5.1. The Analysis and Conclusion on the Collection Results of the Tourist Attractions and Tourist Attraction Attributes

After analyzing Section 4.1 and Table 1 data, the following conclusions were reached.

- (1) The tourist attractions that were chosen in the experiment conformed to the preset conditions of the proposed algorithm.
- (2) The tourist attractions were popular tourist attractions in Chengdu City with typical features and spatial attributes. Each tourist attraction had different attributes that affected their capacity to satisfy varying tourist interests, which formed the clustering condition. In addition, they were located across the city, which formed the spatial modeling condition.
- (3) Table 1 data were used to generate clusters and interest-matching objective functions. The data were normalized values that were processed by the feature attribute label vector normalization parameter $\delta_{1(i)}$, and they conformed to the proposed algorithm conditions.

4.5.2. The Analysis and Conclusion of the Results of Clustering and Cluster Visualization

After analyzing Section 4.2, Table 2 data, and Figure 4, the following conclusions were reached.

- (1) In Table 2, the clustering objective function values between the two different tourist attractions were all different. They represented the degree of correlation among the two different tourist attractions.
- (2) Via the proposed algorithm, three tourist attraction clusters were identified.
- (3) The tourist attractions in the same cluster had a high degree of correlation among their attributes and the objective function values were relatively small. The tourist attractions in different clusters had a low degree of correlation among their attributes and the objective function values were relatively large.
- (4) The Figure 4 shows the formed three clusters, cluster structure trees, and cluster spatial buffers that were constrained by the proposed algorithm.
 - ① Figure 4a shows the distribution of all the tourist attraction samples with note labels. They were spatially discrete.
 - ② As to the inner tree structure of the clusters: in Figure 4b, the topological connecting lines among tourist attractions formed the first structure tree and it indicated the searching process of the first cluster. In Figure 4c, the topological connecting lines among the tourist attractions formed the second structure tree, and it indicated the searching process of the second cluster. In Figure 4d, the topological connecting lines among the tourist attractions formed the third structure tree, and it indicated the searching process of the third cluster.
 - ③ As to the structure of the cluster buffer: in Figure 4b, the closed brown space was the first cluster spatial buffer and indicated the spatial range of the first cluster. In Figure 4c, the closed blue space was the second cluster spatial buffer and indicated the spatial range of the second cluster. In Figure 4d, the closed green space was the third cluster spatial buffer and indicated the spatial range of the third cluster.
- (5) Since the proposed algorithm combined spatial attributes, the three structure trees and buffers each had different shapes and topology tendencies, which visually indicated that different clusters not only had discrepancies in feature attributes but had larger discrepancies in their spatial attributes. In addition, the three clusters had spatial overlap in the city range. It indicated that the correlation relationship among the tourist attractions in the clusters as relative but not isolated. The formed tour routes could pass through different clusters.

4.5.3. The Analysis and Conclusion on the Results of the Tourist Attractions and Tour Route

After analyzing Section 4.3, Tables 3–6 data, and Figure 5, the following conclusions were reached.

- (1) Table 3 shows the normalized data of the two tourist interest labels. It indicated that the two tourists had completely different interests. In addition, the starting points for the two tourists were different. According to the proposed algorithm, the precondition in Table 3 differentiated the results of the two tourists and followed the experimental logic.
- (2) Table 4 shows the interest-matching objective function values between the two tourist samples with each tourist attraction.
 - ① The values were different due to Table 3 preconditions and the operation of the proposed algorithm. It indicated that each tourist attraction's capacity on satisfying tourist's interests would be different. The tourist attraction that had the stronger capacity would be preferentially selected as the tour-route tourist attraction.

- ② Upon further interpretation, the smaller the value was, the closer the tourist attraction attributes were to the tourist's interests and the tourist attraction would be more likely to satisfy the tourist. On the contrary, the bigger the value was, the more remote the tourist attraction attributes were to the tourist's interests and the tourist attraction would be less likely to satisfy the tourist.
- (3) Table 5 was deduced from Table 4. It shows that each tourist attraction was stored in the matrix $\mathbf{P}_{\xi(N,T)}$ in the ascending order of the interest-matching objective function value. It indicated the capacity sequence of each tourist attraction on satisfying tourist interests.
- (4) Figure 5 was deduced from Tables 3–5, and it indicated that the interest-matching objective function values had the fluctuate tendency and the tourist attraction capacities varied with the footnotes in each cluster.

As to the first tourist, the interest-matching objective function values are shown in the Figure 5a,c:

- (1) Tourist attraction $s_{(6)}$: East Lake Park had the highest matching function value and we interpreted that it had the lowest capacity for satisfying the tourist's interests.
- (2) Tourist attraction $s_{(14)}$: Chengdu Zoo had the lowest matching function value, and we interpreted that it had the highest capacity for satisfying the tourist's interests.
- (3) In the cluster $S_{(1)}$, tourist attraction $s_{(1)}$: Chunxi Road and Zhongshan Square had the highest capacity on satisfying the tourist's interests; The tourist attraction $s_{(15)}$: Raffles Square had the lowest capacity for satisfying the tourist's interests.
- (4) In the cluster $S_{(2)}$, tourist attraction $s_{(3)}$: Temple of Marquis Wu had the highest capacity on satisfying the tourist's interests; The tourist attraction $s_{(7)}$: Wenshu Temple had the lowest capacity for satisfying the tourist's interests.
- (5) In the cluster $S_{(3)}$, tourist attraction $s_{(14)}$: Chengdu Zoo had the highest capacity for satisfying the tourist's interests; The tourist attraction $s_{(6)}$: East Lake Park had the lowest capacity for satisfying the tourist's interests.

As to the second tourist, the interest-matching objective function values are shown in the Figure 5b,d.

- (1) Tourist attraction $s_{(6)}$: East Lake Park had the highest matching function value, and we interpreted that it had the lowest capacity for satisfying the tourist's interests.
- (2) Tourist attraction $s_{(14)}$: Chengdu Zoo had the lowest matching function value, and we interpreted that it had the highest capacity for satisfying the tourist's interests.
- (3) In the cluster $S_{(1)}$, tourist attraction $s_{(1)}$: Chunxi Road and Zhongshan Square had the highest capacity for satisfying the tourist's interests; The tourist attraction $s_{(15)}$: Raffles Square had the lowest capacity for satisfying the tourist's interests.
- (4) In cluster $S_{(2)}$, tourist attraction $s_{(3)}$: Temple of Marquis Wu had the highest capacity for satisfying the tourist's interests; The tourist attraction $s_{(7)}$: Wenshu Temple had the lowest capacity for satisfying the tourist's interests.
- (5) In cluster $S_{(3)}$, tourist attraction $s_{(14)}$: Chengdu Zoo had the highest capacity for satisfying the tourist's interests; The tourist attraction $s_{(6)}$: East Lake Park had the lowest capacity for satisfying the tourist's interests.
- (6) Table 6 indicated the tour route output results that were based on Tables 3–5 preconditions. The following conclusions were reached. (result interpretation 5)
 - ① The tourist attractions of the two tour routes all matched the tourist interests.
 - ② The recommended tour route for the first tourist was 8.77 h long and cost CNY 33 yuan. We interpreted that the proposed algorithm's tour route results conformed to the tourist's requirements.
 - ③ The recommended tour route for the second tourist was 10.25 h long and cost CNY 136 yuan. We interpreted that the proposed algorithm's tour route conformed to the tourist's requirements.
 - ④ The total time and costs were within the ranges of the tourists' allowances and met their needs. We interpreted that the algorithm was feasible and accurate.

4.5.4. The Analysis and Conclusion on the Comparison Result of the Algorithms

After analyzing Section 4.4, Table 7, Table 8, and Figures 6 and 7, the following conclusions were reached.

- (1) The controls that were used were the Dijkstra and A* algorithms, as they have both been used extensively for shortest-path calculations and for planning optimal tour routes. Therefore, the control group algorithms were feasible, accessible, and comparable.
- (2) Due to the different preconditions on tourist interests, starting points, time, costs, and their chosen modes of transportation, the first tourist had three recommended tourist attractions while the second tourist had four recommendations. We interpreted that when the preconditions changed, the results of the proposed algorithm and the control algorithms changed as well.
- (3) All three algorithms produced fluctuating time durations and costs for visiting various tourist attractions and traveling between two tourist attractions. Each algorithm resulted in different values for these variables. The differences are caused by the preconditions of the tourists' needs, tourist attraction attributes, and city geospatial environment, and were also caused by the three algorithms' different performances.
 - ① The tour routes by the Dijkstra and A* algorithms were less efficient and more expensive than those by the proposed algorithm. We interpreted that the proposed algorithm had an advantage on saving time and costs when planning tour routes, as compared to the controls.
 - ② From the Table 8, it can be concluded that the three algorithms had different performances. On the aspect of computer algorithm performance, when searching the shortest tour route, the proposed algorithm had much lower time complexity and space complexity than the Dijkstra algorithm, while it had much lower time complexity than the A* algorithm and had the same dimension of space complexity with the A* algorithm. Through the mathematical calculating, the ratio ρ was obtained. When the tourist attraction number n was larger than 2, the ratios $\rho_{1,1}$, $\rho_{1,2}$, $\rho_{2,1}$, and $\rho_{2,2}$ were all larger than 1. It can be concluded that when tourist attractions are confirmed in the searching process on the shortest tour route, the Dijkstra and A* algorithm always consumed higher time complexity and space complexity than the proposed algorithm, and the Dijkstra algorithm always consumed higher time complexity than the proposed algorithm while the A* algorithm consumed the same dimension of space complexity with the proposed algorithm.
 - ③ Under the condition of the small tourist attraction data set, the proposed algorithm relied on an exhaustive method, and thus it found global optimal solutions. The Dijkstra and A* algorithms rely on local "greedy" search methods, they might easily converge on a local optimal solution and consume more time complexity and space complexity. In one complete tour route, the larger number of tourist attraction is, the more computer operating time and computer space will be required. That is, the weaker the algorithm performance is, the more time complexity and space complexity will be needed to search the optimal solution. In the experiment, when the three algorithms were carried out under the same computer operating times, the proposed algorithm would find out the optimal tour route more quickly, while the control group algorithms might not find out the optimal one since the Dijkstra algorithm and the A* algorithm's performances were not better than the proposed algorithm with regard to time complexity and space complexity, especially when the tourist attraction number is sufficiently large, the time and space consuming gap would be rapidly widened. Thus, under the conditions of the identical limited operating time and space consumption, the Dijkstra and A* algorithm are inferior to finding out the optimal solution, or even could not find it out and converge on a local optimal solution. In other words, if the Dijkstra algorithm or the A* algorithm are set as the embedded algorithm of the smart tourism

system, they can also find out the optimal tour route, but they will consume more computer operating time and space. In all, the proposed algorithm had a better performance than the Dijkstra and A* algorithms in searching optimal tour routes.

(4) In Figure 7, the following conclusions were reached.

- ① With regard to the first tourist, the proposed algorithm route was 8.77 h long and cost CNY 33 yuan. The Dijkstra algorithm route was 9.14 h long and cost CNY 36.5 yuan. The A* algorithm route was 9.2 h long and cost CNY 37 yuan. We interpreted that the proposed algorithm was superior to the control algorithms.
- ② With regard to the second tourist, the proposed algorithm route was 10.25 h long and cost CYN 136 yuan. The Dijkstra algorithm route was 10.55 h long and cost CYN 147 yuan. The A* algorithm route was 10.44 h long and cost CYN 145 yuan. We interpreted that the proposed algorithm was superior to the control algorithms.
- ③ For the first tourist, the time duration of the tour routes that were recommended by the Dijkstra and A* algorithms both exceeded the nine hours, and thus the results did not conform to the tourist's allowance. In this aspect, we interpreted that the control algorithms were inferior to the proposed algorithm.

5. Conclusions

5.1. Contribution

Based on the current challenges in tour-route planning and attraction recommendations, this study designed a tour-route planning and recommendation algorithm that was based on an improved AGNES spatial clustering and space-time deduction model. This model improved interest-matching, urban-tourist-attraction clustering, space-time deduction, and tour-route planning based on various modes of transportation. By combining the tourist attraction features and spatial attributes, the improved AGNES tourist attraction clustering algorithm was created, and the cluster structure trees, cluster spatial buffers, and clusters were generated. All the tourist attractions with a high degree of correlation among the attributes were clustered together. Based on the tourist-interest data, the interest-matching objective function was created. This function reflected each tourist attraction's capacity for satisfying the tourist's interests, which formed the precondition when planning the tour route. Under the constraint conditions of time and cost allowance, the proposed algorithm searched for the optimal tourist attractions to match the tourist interests as well as considered the optimal tour route. The resultant tour routes met the tourists' needs and interests. Based on the comparison results, the proposed algorithm had advantages when compared to the controls. The proposed algorithm reduced the costs and time investment for tour-route planning. The improved AGNES clustering algorithm considered spatial distance and various tourist attraction attributes. The proposed algorithm integrated mixed (i.e., preferred) transportation modes for different optimized results. Tour-route planning that was based on space-time deduction was an innovative method that not only considered the time and cost constraints, but also considered the shortest traveling distance between two tourist attractions. Therefore, the resultant tour routes satisfied the tourist's interests and reduced the time and costs that were invested by tourists.

5.2. Addressing Challenges for Research

Smart mobile devices have become part of daily life, and, for many applications, activities and events are planned using smart mobile devices. Mobile planning is the key to ensuring efficient routing, resource allocation, and energy management. For example, the researchers in [30] considered that efficient routing, resource allocation, and energy management could be achieved through clustering of mobile nodes into local groups. In the study, a clustering scheme was developed to prolong the network lifetime by distributing energy consumption among clusters. In [31], a novel travel route recommendation system

was proposed that collected tourist on-site travel behavior data automatically regarding a specific POI that was based on smart phone and Internet of Things technologies. The tour-route-recommendation algorithm was then created to search and rank the tangible travel routes. The researchers in [32] considered that the prevalence of smart mobile devices and location-based services would lead to an increasing volume of mobility data. Based on big mobile data, it proposed a method for accurately predicting the next location of a traveling object.

In tourism activities, tourists' traveling behaviors also generate massive amounts of data on mobile devices. How to appropriately and accurately use these data is a future challenge for tourism research. Mobile data could be used in tourism data mining, tourist attraction location, tourist interest tendency research, tourism facility evaluations, tour-route planning, and recommendations, etc. It has been deemed the most important, challenging, and valuable research field for the future. How to precisely optimize mobile data acquisition, mine interest data, match tourists' needs, search optimal solutions, etc., are challenges that should be addressed.

5.3. Limitation and Future Work

When searching the tour routes, the proposed algorithm sets the transportation mode, time allowance, and costs as the constraint conditions. However, the proposed algorithm still has some drawbacks and limitations. First, the AGNES clustering algorithm itself has its limitation in efficiency, accuracy, and space complexity. Second, in the tour-route algorithm, the transportation modes were relatively fixed, whereas tourists might choose different transportation modes in the tour process. Third, the proposed method did not involve mobile data; we provided a method under the condition of city tourist attractions' attributes, tourists' specific interests, and an urban tourism environment. Therefore, additional research could expand and validate our proposed algorithm further. First, more precise tourist attraction clustering methods could be studied, which could refine and better target the clustering results based on tourist interests. The clustering objective function criteria and model procedure could be refined further as well. The criteria to select the parameters could add more factors to satisfy more individualized interests. Second, the transportation mode selection for the whole tour should be more flexible and random, which could then consider tourist selection tendency on different cost deduction sub-units between two tourist attractions. In further research, we will study random transportation mode selection in different sub-units, and a more individualized tour-route-searching algorithm will be designed and proposed. Third, mobile data should be used to mine tourist interests and to integrate specialized interests. To some extent, a smart tourism recommendation system could be set up by mining historical tourists' data and find related knowledge.

Author Contributions: Conceptualization, Xiao Zhou, Jiangpeng Tian and Mingzhan Su; methodology, Xiao Zhou and Jiangpeng Tian; validation and formal analysis, Mingzhan Su and Jiangpeng Tian; investigation, data resources and data processing, Mingzhan Su; writing—original draft preparation, Xiao Zhou; writing—review and editing, Xiao Zhou, Jiangpeng Tian, and Mingzhan Su; visualization, Xiao Zhou and Mingzhan Su; supervision, Xiao Zhou and Jiangpeng Tian; project administration and funding acquisition, Jiangpeng Tian; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China(Grant No.2017YFB0503503), the National Natural Science Foundation of China(Grant No.41701457), the Military "Double Key" construction project(Grant No.2021KY05), and the Leshan Science and Technology Project (Grant No.20RKX0007 and No.20ZRKX006).

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 thank the postdoctoral innovation practice base of Sichuan province of Leshan vocational and Technical College and Computer Science postdoctoral mobile station of Sichuan University. Meanwhile, we thank the editors and reviewers for their valuable comments.

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

References

- Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2009; Volume 344.
- Sano, A.V.D.; Imanuel, T.D.; Calista, M.I.; Nindito, H.; Condrobimo, A.R. The Application of AGNES Algorithm to Optimize Knowledge Base for Tourism Chatbot. In Proceedings of the 2018 International Conference on Information Management and Technology, Bina Nusantara University, Jakarta, Indonesia, 3–5 September 2018; pp. 65–68.
- Rocha, J.M.P.; Alcaraz, J.A.S.; Rodriguez, R.G.; Sifuentes, O.J.P.; Espinal, A.; Figueroa, M.A.S. Application of Agglomerative and Partitioning Algorithms for the Study of the Phenomenon of the Collaborative Economy within the Tourism Industry. *J. Autom. Mob. Robot. Intell. Syst.* **2020**, *14*, 81–86.
- Valls, A.; Gibert, K.; Orellana, A.; Antón-Clavé, S. Using ontology-based clustering to understand the push and pull factors for British tourists visiting a Mediterranean coastal destination. *Inf. Manag.* **2018**, *55*, 145–159. [[CrossRef](#)]
- Loiseau, T.J.; Djebali, S.; Raimbault, T.; Branchet, B.; Chareyron, G. Characterization of daily tourism behaviors based on place sequence analysis from photo sharing websites. In Proceedings of the 2017 IEEE International Conference on Big Data, Boston, MA, USA, 11–14 December 2017; pp. 2760–2765.
- Yi, Y.T. Design of Intelligent Recommendation APP for Ecotourism Routes Based on Popular Data Clustering of Points of Interest. In Proceedings of the Second International Conference on Smart Electronics and Communication, Chennai, Tamil Nadu, India, 7–9 October 2021; pp. 1270–1273.
- Hu, S.; Chau, P.M.; Weng, J.; Lin, P.; Zhou, W. Extending the Theory of Planned Behavior to Explore the Influence of Residents' Dependence on Public Transport. *IEEE Access* **2021**, *9*, 137224–137240. [[CrossRef](#)]
- Zahra, A.M.; Hamed, V.N.; Javad, S. Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Syst. Appl.* **2021**, *167*, 114324.
- Retnosari, M.; Wasino; Arisandi, D. Tourism Destination Grouping Systems With Partitioning Clustering Method from Geotagged Photo using Android. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *852*, 012160. [[CrossRef](#)]
- Majewska, J.; Truskolaski, S. Cluster-mapping procedure for tourism regions based on geostatistics and fuzzy clustering: Example of Polish districts. *Curr. Issues Tour.* **2019**, *22*, 2365–2385. [[CrossRef](#)]
- Darmawiguna, I.G.M.; Pradnyana, G.A.; Santyadiputra, G.S. The Development of Integrated Bali Tourism Information Portal using Web Scrapping and Clustering Methods. *J. Phys. Conf. Ser.* **2019**, *1165*, 012010. [[CrossRef](#)]
- Deng, N.; Liu, J. Where did you take those photos? Tourists' preference clustering based on facial and background recognition. *J. Destin. Mark. Manag.* **2021**, *21*, 100632. [[CrossRef](#)]
- Lee, H.; Kang, Y. Mining tourists' destinations and preferences through LSTM-based text classification and spatial clustering using Flickr data. *Spat. Inf. Res.* **2021**, *29*, 825–839. [[CrossRef](#)]
- Rodriguez-Echeverria, J.; Semajski, I.; Van Gheluwe, C.; Ochoa, D.; Ijben, H.; Gautama, S. Density-Based Spatial Clustering and Ordering Points Approach for Characterizations of Tourist Behaviour. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 686. [[CrossRef](#)]
- Kirilenko, A.P.; Stepchenkova, S.; Hernandez, J.M. Comparative clustering of destination attractions for different origin markets with network and spatial analyses of online reviews. *Tour. Manag.* **2018**, *72*, 400–410. [[CrossRef](#)]
- Hong, M.; Jung, J.J. Multi-criteria tensor model consolidating spatial and temporal information for tourism recommendation. *J. Amb. Intel. Smart. Environ.* **2020**, *13*, 5–19. [[CrossRef](#)]
- Han, S.; Liu, C.; Chen, K.; Gui, D.; Du, Q. A Tourist Attraction Recommendation Model Fusing Spatial, Temporal, and Visual Embeddings for Flickr-Geotagged Photos. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 20. [[CrossRef](#)]
- Zheng, W.; Liao, Z.; Lin, Z. Navigating through the complex transport system: A heuristic approach for city tourism recommendation. *Tour. Manag.* **2020**, *81*, 104162. [[CrossRef](#)]
- Seo, J.C.; Hwan, R.K.; Yong, L.J.; Dong, J.K. Deep Learning-based Tourism Recommendation System using Social Network Analysis. *Int. J. Internet Broadcasting Commun.* **2020**, *12*, 113–119.
- Borade, J.; Makwana, S.; Gupta, P.; Lanjewar, S. Intelligent Agent for Tourism Recommendation System. *Int. J. Res. Eng. Technol.* **2018**, *7*. [[CrossRef](#)]
- Santos, F.; Almeida, A.; Martins, C.; Gonçalves, R.; Martins, J. Using POI functionality and accessibility levels for delivering personalized tourism recommendations. *Comput. Environ. Urban Syst.* **2017**, *77*, 101173. [[CrossRef](#)]
- Zhang, Z.; Pan, H.; Xu, G.; Wang, Y.; Zhang, P. A Context-Awareness Personalized Tourist Attraction Recommendation Algorithm. *Cybern. Inf. Technol.* **2016**, *16*, 146–159. [[CrossRef](#)]
- Ai, J. Research on individualized travel route recommendation based on improved collaborative filtering technology. *Mod. Electron. Technol.* **2019**, *42*, 182–186.

24. Xiao, C.; Xia, K.; Qiao, Y.; Zhang, Y. Tourism route recommendation based on dynamic clustering. *J. Comput. Appl.* **2017**, *37*, 2395–2400.
25. Song, W. Research on Mobile Tourism Route Recommendation Model Based on Optimized Social Tags and Association Rules Algorithm. Master's Thesis, Guilin University of Technology, Guilin, China, 2019.
26. Wu, J. Research and Implementation of Personalized Travel Recommendation System Based on Collaborative Filtering Recommendation. Master's Thesis, Beijing Jiaotong University, Beijing, China, 2017.
27. Chen, G. Research and Implementation of Group Travel Recommendation Algorithm Based on Geo-Tag and Time Factor. Master's Thesis, Beijing University of Posts and Telecommunications, Beijing, China, 2019.
28. Chen, Y.; Wu, S.; Dong, C. Tourist route recommendation under the constraints of personal time and space behavior. *J. Green Sci. Technol.* **2018**, *11*, 100.
29. Li, D. Design and Implementation of Travel Recommendation System Based on Hybrid Recommendation Algorithm. Master's Thesis, Zhejiang University, Hangzhou, Zhejiang Province, China, 2018.
30. Comito, C.; Talia, D.; Trunfio, P. An Energy-Aware Clustering Scheme for Mobile Applications. In Proceedings of the 2011 IEEE 11th International Conference on Computer and Information Technology, Paphos, Cyprus, 31 August–2 September 2011; pp. 15–22.
31. Bin, C.; Gu, T.; Sun, Y.; Chang, L.; Sun, L. A Travel Route Recommendation System Based on Smart Phones and IoT Environment. *Wirel. Commun. Mob. Comput.* **2019**, *2019*, 7038259. [[CrossRef](#)]
32. Monreale, A.; Pinelli, F.; Trasarti, R.; Giannotti, F. WhereNext: A location predictor on trajectory pattern mining. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July 2009.
33. Zhou, X.; Xu, C.; Kimmons, B. Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Comput. Environ. Urban Syst.* **2015**, *54*, 144–153. [[CrossRef](#)]
34. Zhao, S.; Zhao, T.; King, I.; Lyu, M.R. Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation. In Proceedings of the 26th International Conference on World Wide Web Companion, Geneva, Switzerland, 3–7 April 2017; pp. 153–162.
35. Young, C.; Young, U.R.; Kyeong, K. A Recommender System based on Personal Constraints for Smart Tourism City. *Asia Pac. J. Tour. Res.* **2021**, *26*, 440–453.
36. Jabbarpour, M.R.; Noor, R.M.; Khokhar, R.H. Green vehicle traffic routing system using ant-based algorithm. *J. Netw. Comput. Appl.* **2015**, *58*, 294–308. [[CrossRef](#)]
37. Li, D.; Yu, Z.; Luo, W.; Hu, Y.; Che, X.; Yuan, L. Optimal route searching with multiple dynamical constraints—A geometric algebra approach. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 172. [[CrossRef](#)]
38. Lin, Q.; Chao, Y.; Yang, N. A travel routes recommendation algorithm based on user's time constraints. *Geogr. Geo-Inf. Sci.* **2017**, *25*–29.
39. Damos, M.A.; Zhu, J.; Li, W.; Hassan, A.; Khalifa, E. A novel urban tourism path planning approach based on a multiobjective genetic algorithm. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 530. [[CrossRef](#)]
40. Cergibozan, Ç.; Tasan, A.S. Tourist Route Planning with a Metaheuristic Approach. In *Closing the Gap between Practice and Research in Industrial Engineering*; Springer: Cham, Switzerland, 2017; pp. 193–199.
41. Fitriansyah, A.; Parwati, N.W.; Wardhani, D.R.; Kustian, N. Dijkstra's Algorithm to Find Shortest Path of Tourist Destination in Bali. *J. Phys. Conf. Ser.* **2019**, *1338*, 012044. [[CrossRef](#)]
42. Liang, Y.; Wang, L. Applying genetic algorithm and ant colony optimization algorithm into marine investigation path planning model. *Soft Comput.* **2019**, *24*, 8199–8210. [[CrossRef](#)]

Article

Spatiotemporal Dynamic Analysis of A-Level Scenic Spots in Guizhou Province, China

Yuanhong Qiu ^{1,2}, Jian Yin ^{1,2,3,*}, Ting Zhang ¹, Yiming Du ¹ and Bin Zhang ^{1,2}

- ¹ College of Big Data Application and Economic, Guizhou University of Finance and Economics, Huaxi District, Guiyang 550025, China; yuanhongq@mail.gufe.edu.cn (Y.Q.); zhangti@mail.gufe.edu.cn (T.Z.); DuYiming@mail.gufe.edu.cn (Y.D.); gzcdbin@mail.gufe.edu.cn (B.Z.)
- ² Center for China Western Modernization, Guizhou University of Finance and Economics, Huaxi District, Guiyang 550025, China
- ³ School of Water Conservancy and Civil Engineering, Northeast Agricultural University, Xiangfang District, Harbin 150050, China
- * Correspondence: jiany@mail.gufe.edu.cn

Abstract: A-level scenic spots are a unique evaluation form of tourist attractions in China, which have an important impact on regional tourism development. Guizhou is a key tourist province in China. In recent years, the number of A-level scenic spots in Guizhou Province has been increasing, and the regional tourist economy has improved rapidly. The spatial distribution evolution characteristics and influencing factors of A-level scenic spots in Guizhou Province from 2005 to 2019 were measured using spatial data analysis methods, trend analysis methods, and geographical detector methods. The results elaborated that the number of A-level scenic spots in all counties of Guizhou Province increased, while in the south it developed slowly. From 2005 to 2019, the spatial distribution in A-level scenic spots were characterized by spatial agglomeration. The spatial distribution equilibrium degree of scenic spots in nine cities in Guizhou Province was gradually developed to reach the “relatively average” level. By 2019, the kernel density distribution of A-level scenic spots had formed the “two-axis, multi-core” layout. One axis was located in the north central part of Guizhou Province, and the other axis ran across the central part. The multi-core areas were mainly located in Nanming District, Yunyan District, Honghuagang District, and Xixiu District. From 2005 to 2007, the standard deviation ellipses of the scenic spots distribution changed greatly in direction and size. After 2007, the long-axis direction of the ellipses gradually formed a southwest to northeast direction. We chose elevation, population density, river density, road network density, tourism income, and GDP as factors, to discuss the spatiotemporal evolution of the scenic spots’ distribution with coupling and attribution analysis. It was found that the river, population distribution, road network density, and the A-level scenic spots’ distribution had a relatively high coupling phenomenon. Highway network density and tourist income have a higher influence on A-level tourist resorts distribution. Finally, on account of the spatiotemporal pattern characteristics of A-level scenic spots in Guizhou Province and the detection results of influencing factors, we put forward suggestions to strengthen the development of scenic spots in southern Guizhou Province and upgrade the development model of “point-axis network surface” to the current “two-axis multi-core” pattern of tourism development. This study can explain the current situation of the spatial development of tourist attractions in Guizhou Province, formulate a regulation mechanism of tourism development, and provide a reference for decision-making to boost the high-quality development of the tourist industry.

Citation: Qiu, Y.; Yin, J.; Zhang, T.; Du, Y.; Zhang, B. Spatiotemporal Dynamic Analysis of A-Level Scenic Spots in Guizhou Province, China. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 568. <https://doi.org/10.3390/ijgi10080568>

Academic Editors: Andrea Marchetti, Angelica Lo Duca and Wolfgang Kainz

Received: 23 June 2021

Accepted: 20 August 2021

Published: 23 August 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: A-level scenic spots; spatiotemporal evolution; trend analysis; Geodetector

1. Introduction

The planning and development of tourist attractions have become a key link to promote the growth of the local tourist economy. As the important material carrier of tourist supply, tourist scenic spots provide a material basis for the development of the regional

tourism industry. Their projection in geographical space shows the spatial attributes and mutual relations of tourist activities, which influences and promotes the development of regional tourism resources and the tourist economy [1]. In order to strengthen the quality assessment and management of tourist scenic spots, the Chinese government has formulated the A-level scenic spot planning (http://zwgk.mct.gov.cn/zfxgkml/zcfg/gfxwj/202012/t20201204_906214.html, accessed on 16 June 2021). The grade is divided into A, AA, AAA, AAAA and AAAAA. A-level scenic spots refer to scenic spots that can receive tourists, have the functions of sightseeing and entertainment, and have a relatively complete management system. A-level scenic spots must have a visitor center, basic visitor services, tourist consultations, tourist complaints, and management of all kinds of tourist affairs within the service radius of the visitor center and the visitor center itself. The rating of the scenic spot includes eight aspects: tourist transportation (14%), sightseeing (21%), tourist safety (8%), health (14%), post and telecommunications services (3%), tourist shopping (5%), comprehensive management (19.5%), and protection of resources and environment (15.5%); the rating agency will give a score on each aspect [2]. According to the standard “Classification and Evaluation of Tourist Areas (Spots) Quality Grade” (GB/T 17775–1999), the score of “service quality and environment evaluation system”, “landscape quality evaluation system”, and “tourist opinion evaluation system”, the level of participating scenic spots is divided [2].

Guizhou Province is a distinctive mountain tourist area in Southwest China. Its unique karst landform and climate characteristics make tourist scenic spots diverse in Guizhou. In the Fourteenth Five-Year Plan for National Economic and Social Development of Guizhou Province and the Outline of the Vision of the Year 2035, it is proposed to actively promote the construction and upgrading of scenic spots [3]. Recently, owing to the booming of tourist industry in Guizhou, it is urgent to explore the development process of scenic spots, especially A-level scenic spots. The study of the scenic spots’ spatial distribution and its influencing factors plays a positive role in formulating tourism planning, promoting traffic development, and alleviating the environmental pressure caused by tourism [4,5].

At present, scholars’ studies on the evolution pattern of scenic spots mainly focus on the network of scenic spots [6], the spatio-temporal influence of scenic spots on tourists’ behavior and emotion [7,8], demand prediction [9,10], optimization of tourists’ tourism experience [11–13], landscape changes of scenic spots [14], environmental impact [15–20] and intelligent tourism [21]. Since the China Tourism Administration issued the “A-level scenic spot assessment standard” in 2002, how to rationally plan and develop the A-level scenic spots has become a hot spot for domestic scholars to study scenic spots. From the perspective of research content, it mainly focuses on the spatial planning of scenic spots [22], spatial structure and optimization [23–27], influence factors of scenic spot distribution [28–35], and so on. In terms of research methods, exploratory spatial data analysis and geographic detector were mainly applied. For example, Liu and Hao [5] researched the influencing factors of the spatial distribution evolution of scenic spots in Shanxi Province with the help of a geographic detector model. Peng and Huang [21] analyzed the popularity distribution of Beijing’s scenic spots under different temporal and weather contexts. Li and Zhang [28] systematically sorted out the spatial distribution characteristics and influencing factors of 1010 scenic spots in the Yellow River Basin, China. Tang and Sun [29] explored the spatial layout of scenic spots in Beijing–Tianjin–Hebei urban agglomeration and its influencing factors by using spatial data analysis, the Gini coefficient, and geographical detector methods. Jia and Hu [34] used the average nearest neighbor index, kernel density analysis, and geographic detector to analyze the spatial distribution evolution and the influence mechanism of A-level scenic spots in the middle reaches of the Yangtze River through exploratory spatial data analysis and geographic detector. Lu and Zhang [35] explored the spatial distribution characteristics, differentiation trend, and driving mechanisms of A-level scenic spots. From the perspective of time scale, some studies analyzed the spatial evolution characteristics of scenic spot distribution from a single time node to a continuous-time point [23–36]. From the perspective of the study region, the studies

covered the evolution of the spatial pattern of national A-level scenic spots at the national, urban agglomeration, provincial, and urban levels [23–36]. Liu and Wang [37] analyzed the spatial distribution characteristics of scenic spots in Guizhou Province, China, while they lacked research on the evolution of the spatial pattern of tourism scenic spots.

Guizhou is a province with unique tourism characteristics in China, its tourism income accounts for a very large proportion of GDP. In recent years, the number of scenic spots in Guizhou Province has developed rapidly, but the research on scenic spots in Guizhou is very lacking. The current analysis methods for scenic spots distribution contain one or more of exploratory spatial data analysis, density analysis, direction analysis, spatial coupling analysis, and influencing factor analysis, but lack comprehensive analysis. Tourism is a complex spatial process. Multi angle analysis is more conducive to the exploration of the spatiotemporal dynamic processes.

Against this backdrop, we selected the A-level scenic spots in Guizhou, established a database of A-level scenic spots from 2005 to 2019 and explored the spatiotemporal distribution characteristics and influencing factors of the A-level scenic spots in Guizhou Province, by using direction analysis, the Gini coefficient method, trend analysis, and geographical detector. The aim is to reveal the evolution of the scenic spots' distribution law, clear the driving mechanism of its spatial dynamic characteristics, and put forward suggestions for optimizing the layout of scenic spots, in order to provide decision support for the upgrading of A-level scenic spots and promoting the development of regional tourism quality of Guizhou Province.

2. Materials and Methods

2.1. Data and Area

Guizhou Province is located in the hinterland of the southwest of China, which owns 88 counties and districts (Figure 1). There have been many ethnic minorities living in Guizhou for generations, and the ethnic culture is profound. Guizhou is an important transportation hub in Southwest China, and it is a world-famous mountain tourist destination with a livable climate, good ecological environment, and tourism conditions. Affected by the South Asian monsoon, Guizhou Province has distinct dry and wet seasons. It is a typical low-latitude plateau climate, warm and humid. Due to many clouds throughout the year, it has less sunshine and more cloudy days, an obvious rainy season, abundant precipitation, and the rainy and hot periods are mostly concentrated in summer. The average annual precipitation is 682–1134 mm, and the average annual temperature is 14–16 °C [38]. By 2020, it had seven national 5A-level scenic spots, including famous scenic spots such as Huangguoshu Waterfall, Loong Palace, Zhenyuan Ancient Town, Qingyan Ancient Town, National Forest Park of Azalea, and Mount Fanjing.

The data of scenic spots mainly contained the distribution, rating, and geographical coordinates of A-level scenic spots in Guizhou Province from 2005 to 2019, and the influencing factors included tourist income, GDP, the river system, altitude, population distribution, and vegetation coverage rate. With the help of the Baidu API coordinate pickup system, particle coordinates of each scenic spot were calibrated as the spatial position of the scenic spot. Among them, A-level scenic spots in the Guizhou directory data mainly came from the culture and tourism section of the Guizhou hall official website (<http://whhly.guizhou.gov.cn/>, accessed on 16 June 2021); part of the scenic spot data was from the municipal state tourism administration network. The administrative boundaries, digital elevation, normalized difference vegetation index (NDVI), and population distribution came from the resources and environmental science and data center of the Chinese Academy of Sciences (<http://www.resdc.cn>, accessed on 16 June 2021). The river data and road data were from an open-source map website (<http://www.openstreetmap.org>, accessed on 16 June 2021); the tourism income data came from the macroeconomic database (<http://hgk.guizhou.gov.cn/index.vhtml#>, accessed on 16 June 2021).

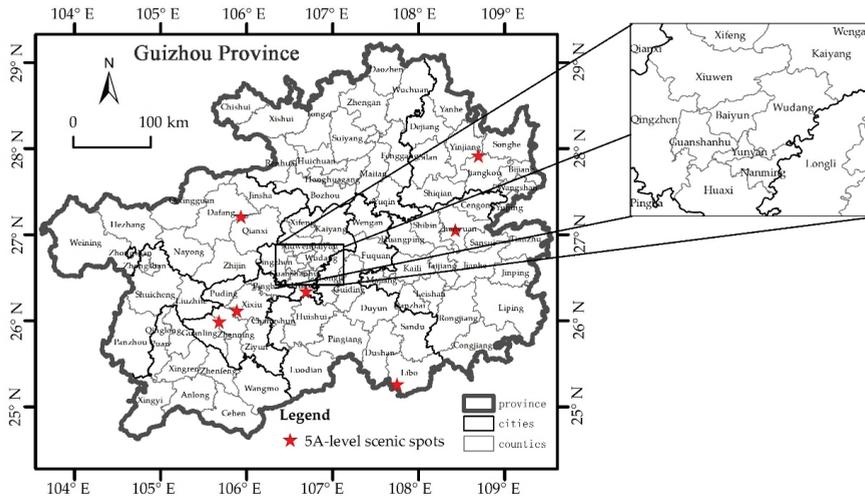


Figure 1. Map of the counties located in Guizhou Province.

2.2. Methodology

In the study, the average nearest neighbor index and Gini coefficient were used to calculate the equilibrium degree of the spatial distribution of scenic spots, then the standard deviation ellipse and trend analysis were used to calculate the spatial distribution trend of scenic spots. Kernel density analysis was used to calculate the difference of the spatial distribution of scenic spots, the factors affecting the spatial distribution of scenic spots were analyzed in the geographical detector according to the kernel density analysis results. The following is the workflow of the study (Figure 2).

2.2.1. Average Nearest Neighbor Index

In this study, A-level scenic spots in Guizhou Province were taken as point-like targets. The nearest neighbor index is a measurement method to measure the actual point-like distribution based on the condition of random distribution. The nearest neighbor analysis can determine the attributes of point pattern more accurately and objectively [5]. The clustering degree of A-level scenic spots in Guizhou each year was obtained by analyzing the data of A-level scenic spots over the years with the average nearest neighbor index.

$$H_1 = \frac{\sum_{i=1}^n x_i}{n}, H_2 = \frac{1}{2\sqrt{n/S}}, H' = \frac{H_1}{H_2} \quad (1)$$

where, N represents the number of scenic spots, S represents the area of Guizhou Province, and H_1 denotes the average nearest distance of each scenic spot, H_2 represents the theoretical nearest proximity distance. H' is the ratio of H_1 to H_2 , that is, the average nearest neighbor index. When $H' > 1$, the locations of scenic spots are evenly distributed. When $H' < 1$, scenic spots are clustered and distributed.

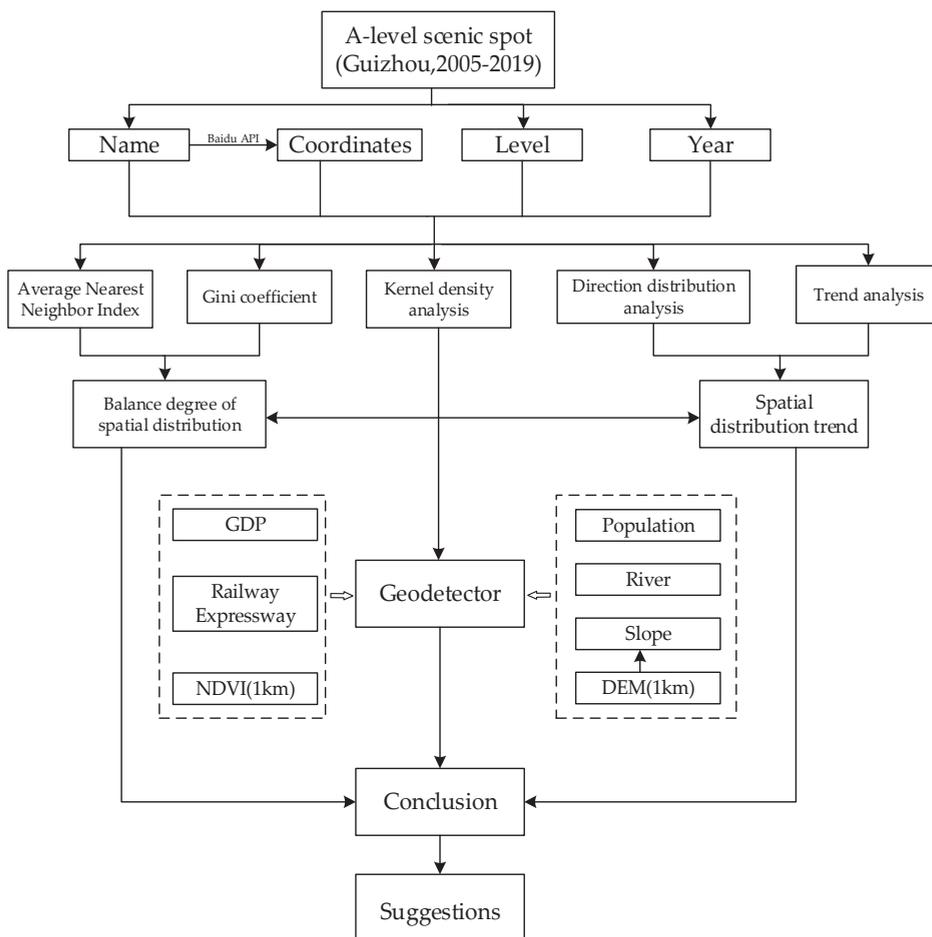


Figure 2. Workflow of the study.

2.2.2. Gini Coefficient

The Gini coefficient was originally used as a common indicator to measure regional economic income differences, and was later improved by relevant scholars and applied to the measurement of geographic spatial distribution. In this study, the Gini coefficient algorithm proposed by Zhang [39] was used to ensure the accuracy of the measurement of spatial distribution balance degree of A-level scenic spots in nine cities and states of Guizhou Province. According to Liu’s paper, the Gini coefficient values of scenic spots can be divided into different equilibrium types of the spatial distribution of scenic spots [40].

$$G = 1 - \frac{2 \sum_{i=1}^{d=1} W_i + 1}{d} \tag{2}$$

where, d is the number of cities and state scenic spots. G is the Gini coefficient from 0 to 1, and G was also close to 1, indicating that the balance of distribution of A-level scenic spots in Guizhou Province was smaller.

2.2.3. Kernel Density Analysis

The spatial distribution density of regional elements can clearly reflect their spatial dispersion or agglomeration characteristics and the change of this form. The spatial distribution density of regional elements is usually expressed by kernel density estimation method [24–27]. Kernel density clearly reflected the spatial dispersion and agglomeration characteristics of A-level scenic spots in Guizhou Province. Then, the evolution law of its characteristics was obtained by analyzing the annual kernel density of scenic spots. The analytical formula for kernel density is

$$f_z(k) = \frac{1}{nx} \sum_{i=1}^n h\left(\frac{k - K_i}{x}\right) \quad (3)$$

where, n represents the number of sample points, $h(\)$ represents the kernel function, $x > 0$ and represents the bandwidth, and $(k - K_i)$ represents the distance from k to the event K_i . This formula was tested many times, and the data selection bandwidth was 3 km to more intuitively reflect the spatial distribution of tourism resources.

2.2.4. Direction Distribution Analysis

Direction distribution can reflect the degree of dispersion and evolution of the spatial distribution of scenic spots from time and space dimensions. The standard deviation ellipse which is a common direction distribution analysis method, was employed to reflect the spatial distribution characteristics and the spatial distribution variation of research elements [25–30]. This method can reflect spatial characteristics such as the centrality, distribution, and directionality of the spatial distribution of scenic spots in each year by using index parameters such as the center, long axis, short axis, and azimuth of the standard deviation ellipse.

2.2.5. Geodetector

The geographical detector was originally based on the geographical perspective proposed by Wang [41]. This study used the Wang's Geodetector model for calculation [42,43]. An algorithm is about detecting the spatial difference of the influence factors on the dependent variables [42–46].

$$q = 1 - \frac{\sum_{x=1}^L N_x \sigma_x^2}{N \sigma_x^2} \quad (4)$$

where, the q value represents the influence degree of each detection factor on the distribution of A-level scenic spots in Guizhou. L represents the variable stratification, that is, classification or partition; N_x and N represent the number of units in layer h and the entire area, respectively. σ^2 refers variance.

2.3. Data Preprocessing

2.3.1. Data of Scenic Spots

Based on the collected data, we built a space database containing the name, grade, counties and cities, geographic coordinates, and evaluation time of scenic spot grade in Guizhou from 2005 to 2019. Figure 3 shows the spatial distribution of A-level scenic spots in 2005, 2009, 2013, 2016, and 2019 as examples.

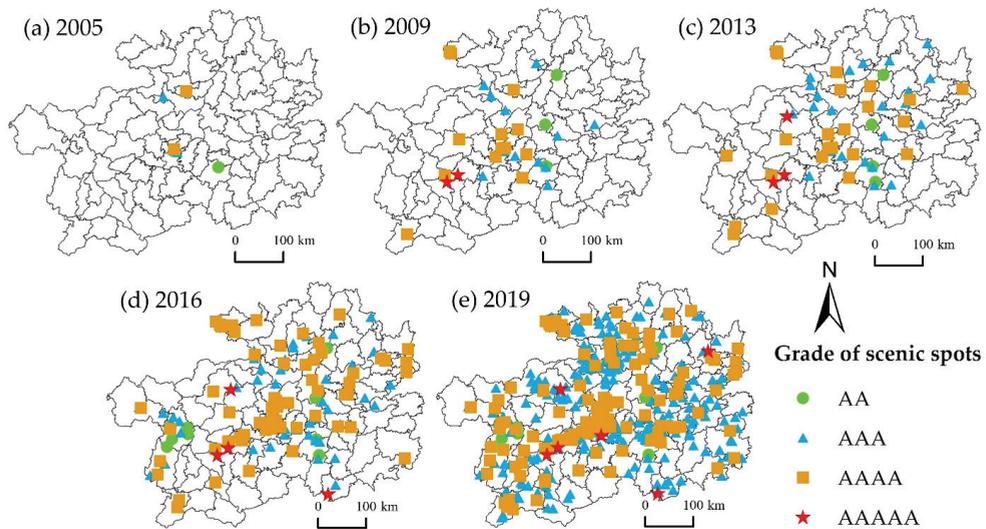


Figure 3. Spatial distribution of A-level scenic spots in Guizhou Province, (a) Spatial distribution of A-level scenic spots in 2005, (b) Spatial distribution of A-level scenic spots in 2009, (c) Spatial distribution of A-level scenic spots in 2013, (d) Spatial distribution of A-level scenic spots in 2016, (e) Spatial distribution of A-level scenic spots in 2019.

2.3.2. Geodetector Data

In the explanatory variables of the Geodetector model, tourism income and Gross Domestic Product (GDP) were the statistical data of 88 counties (districts) in Guizhou, while altitude, population density, river density, and highway network density were vector data. The kernel density of scenic spots in Guizhou Province over the years was the explained variable of the model. All data in the model were reclassified using the natural breakpoint method. Considering the long construction cycle of scenic spots, the dependent variables in the model were all data with one lag period.

3. Results

3.1. Spatial Distribution Characteristics of A-Level Scenic Spots

3.1.1. A-Level Scenic Spots Development

According to the line chart of the number of A-level scenic spots in Guizhou Province in Figure 4, the number of A-level scenic spots increased from 6 to 406 during 2005–2019. As can be seen from the line chart, the number of A-level scenic spots increased little from 2005 to 2011, and the number of scenic spots only increased by 35 in six years. From 2011 to 2015, the growth rate of the number of A-level scenic spots increased slowly. With the improvement of Guizhou's tourism policy, market, and system, the number of A-level scenic spots increased steadily. From 2016 to 2019, Guizhou Province clearly proposed to improve in the three "long boards" of big data, big ecology and big tourism. Scenic spots had "blowout" growth, the scenic spots increased, respectively, by 73, 122, and 80 in 2017, 2018, and 2019.

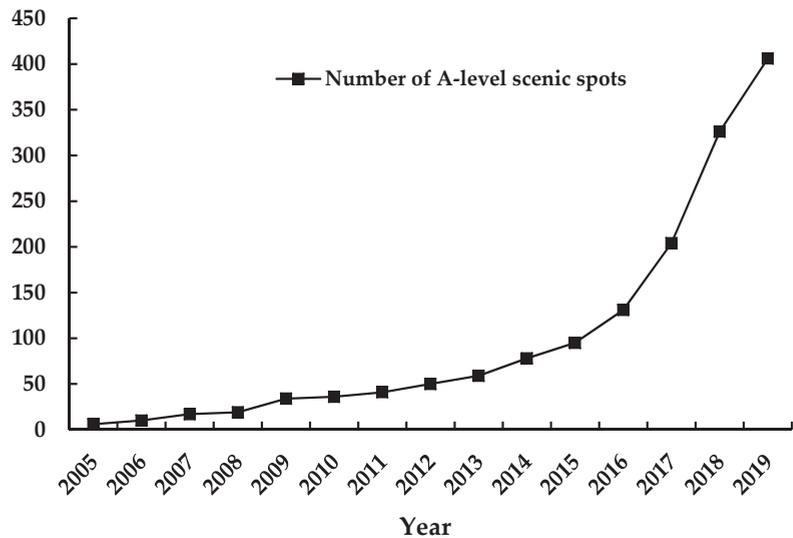


Figure 4. Changes in the number of A-level scenic spots in Guizhou Province (2005–2019).

3.1.2. Evolution of Spatial Distribution Types

Table 1 shows the analysis results of the average nearest neighbor index of A-level scenic spots in Guizhou Province over the years. p values were less than 0.05, that is, they passed the significance test in 95% of cases. From 2005 to 2019, the nearest neighbor index was less than 1, and the spatial distribution types of scenic spots were all concentrated. From 2005 to 2009, the value of the nearest neighbor index increased gradually, from 0.356 to 0.717. From 2009 to 2016, the value of the nearest neighbor index increased in an “M” shape, showing an overall upward trend. In 2012, the value of the nearest neighbor index reached the highest value of 0.854, and in 2017–2019, the value changed slightly, all around 0.81, which was relatively stable.

Table 1. The nearest neighbor index of the average spatial distribution of A-level scenic spots in Guizhou Province.

Year	Average Observation Distance (m)	Expected Mean Distance (m)	Nearest Neighbor Index	p Values	Type of Spatial Distribution
2005	30,488.682	85,675.502	0.356	0.003	agglomeration
2006	33,731.608	66,363.959	0.508	0.003	agglomeration
2007	24,046.061	50,898.833	0.472	0.000	agglomeration
2008	21,825.061	48,145.476	0.453	0.000	agglomeration
2009	25,812.507	35,990.910	0.717	0.002	agglomeration
2010	24,552.270	34,976.877	0.702	0.001	agglomeration
2011	24,341.123	32,774.823	0.743	0.002	agglomeration
2012	25,332.103	29,678.865	0.854	0.048	agglomeration
2013	22,627.417	27,321.609	0.828	0.012	agglomeration
2014	20,253.575	23,762.109	0.852	0.013	agglomeration
2015	17,576.858	21,531.311	0.816	0.001	agglomeration
2016	14,396.817	18,335.664	0.785	0.000	agglomeration
2017	11,953.753	14,693.227	0.814	0.000	agglomeration
2018	9444.529	11,623.140	0.813	0.000	agglomeration
2019	8467.087	10,415.240	0.813	0.000	agglomeration

3.1.3. Equilibrium of Spatial Distribution

Table 2 shows the calculation results of the Gini coefficient. The spatial distribution equilibrium degree of A-level scenic spots in Guizhou has changed greatly, developing towards the “relatively equality” type [40]. From 2005 to 2019, the Gini coefficient values showed a downward trend on the whole, but the coefficient values were all greater than 0.2 (below 0.2 is the “absolute equality”), with the maximum coefficient value of 0.741 in 2005 and the minimum coefficient value of 0.2 in 2017. For the A-level scenic spots, the space distribution equilibrium degree was mainly “inequality” from 2005 to 2006, the degree was “relative inequality” from 2007 to 2011, the degree was “reasonable” from 2012 to 2016, and the degree was “relative equality” after 2017. The development process of spatial distribution equilibrium degree could be divided into three periods: “great disparity—relatively reasonable—relative equality”.

Table 2. Gini coefficient calculation results of A-level scenic spots in Guizhou.

Year	The Gini Coefficient	Type of Equilibrium Degree	Year	The Gini Coefficient	Type of Equilibrium Degree
2005	0.741	Inequality	2013	0.399	Relatively reasonable
2006	0.667	Inequality	2014	0.379	Relatively reasonable
2007	0.549	Relative inequality	2015	0.386	Relatively reasonable
2008	0.573	Relative inequality	2016	0.327	Relatively reasonable
2009	0.542	Relative inequality	2017	0.215	Relative equality
2010	0.556	Relative inequality	2018	0.249	Relative equality
2011	0.493	Relative inequality	2019	0.285	Relative equality
2012	0.359	Relatively reasonable			

According to the line chart of the Gini coefficient change of grade A-level scenic spots in Guizhou Province from 2005 to 2019 (Figure 5), the Gini coefficient dropped sharply in the three years from 2005 to 2007, which indicates that the regional gap of the scenic spots’ spatial distribution balance was narrowing. From 2007 to 2011, the Gini coefficient was stable at around 0.55, and the balance of spatial distribution of scenic spots was not significantly improved. In 2012, the Gini coefficient dropped below 0.4, reaching 0.359. From 2013 to 2016, the Gini coefficient was in the range of 0.3–0.4. In 2017, the Gini coefficient dropped to the lowest level over the years, reaching 0.215. From 2018 to 2019, the Gini coefficient rose and finally stabilized at 0.285. In general, the Gini coefficient of the spatial distribution of A-level scenic spots in the nine cities and states of Guizhou Province showed a decreasing trend from 2005 to 2017 and an increasing trend from 2018 to 2019. Moreover, the Gini coefficient showed a significant decrease in 2012 and 2017. With the steady growth of the number of A-level scenic spots in Guizhou Province, the spatial distribution of the scenic spots gradually shifted from highly concentrated in 2005 to balanced development. However, due to the rapid growth of A-level scenic spots after 2012 and 2017, the spatial distribution of the scenic spots showed the characteristics of small concentration.

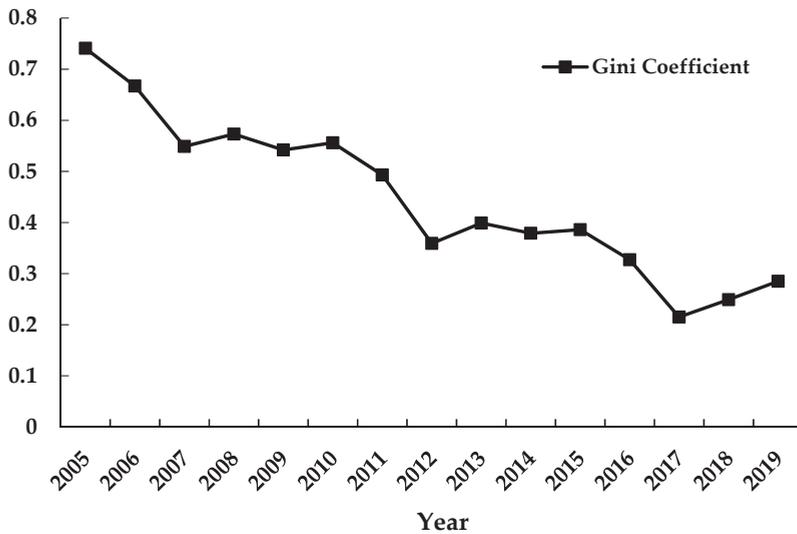


Figure 5. Change process of the Gini coefficient (2005–2019).

3.2. Spatial Distribution Evolution Processes of A-Level Scenic Spots

3.2.1. Density Change Process

The data of A-level scenic spots in Guizhou Province in 2005, 2009, 2013, 2016, and 2019 were selected for the density analysis. The density evolution of scenic spots each year was investigated to reveal the evolution law of spatial nuclear density of A-level scenic spots in Guizhou Province. The result is shown in Figure 6.

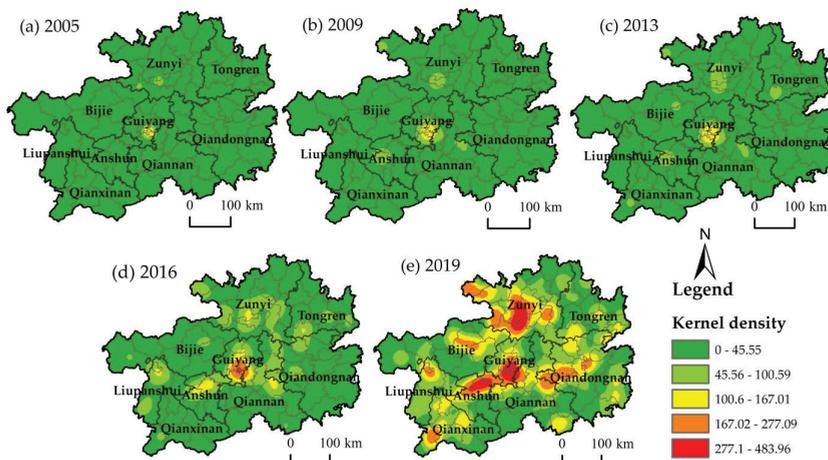


Figure 6. Kernel density map of spatial distribution of A-level scenic spots in Guizhou Province, (a) Kernel density of spatial distribution of A-level scenic spots in 2005, (b) Kernel density of spatial distribution of A-level scenic spots in 2009, (c) Kernel density of spatial distribution of A-level scenic spots in 2013, (d) Kernel density of spatial distribution of A-level scenic spots in 2016, (e) Kernel density of spatial distribution of A-level scenic spots in 2019.

According to Figure 6, the spatial distribution of A-level scenic spots in Guizhou had obvious changes in the kernel density map. In 2005, there were three high-density areas in Guizhou Province, which were located in Guiyang City, Zunyi City, and the junction of Zunyi–Bijie City, and the spatial core density of the scenic spots was relatively small on the whole. In 2009, the high-density area of A-level scenic spots was expanded. Compared with 2005, two high-density areas were added at the junction of Anshun–Liupanshui City and the junction of Qiannan Zhou–Qiandongnan Prefecture, while the high-density area at the junction of Zunyi–Bijie disappeared. In 2013, the main core density area in Guiyang City and Zunyi City was still expanding. Compared with 2009, three high-density areas were added in Bijie, Tongren, and Qianxinan Prefecture. At this time, high-density areas of scenic spots appeared in all nine cities and prefectures in Guizhou Province. In 2016, one main core area and three secondary core areas appeared, and the Zunyi, Qiannan and Anshun core areas also gradually formed. Other high-density areas expanded significantly, and the overall spatial pattern of “one axis and multiple cores” was formed. The “one axis” was located in the central part of Guizhou, spanning Anshun, Guiyang, southern Guizhou, and southeast Guizhou. By 2019, the spatial distribution of A-level scenic spots in Guizhou Province had formed the feature of “two-axis, multi-core”. One axis is located in the north of Guizhou Province, ranging from Bijie to Zunyi. The other axis crosses the middle of Guizhou, along with the distribution of “Anshun–Guiyang–Qiannan–Qiandongnan”. The core areas were mainly distributed in Liupanshui, Xingyi, Qiandongnan, and Zunyi. The southern and eastern parts of Guizhou Province are mostly low-density areas.

3.2.2. Directional Distribution

Figure 7 and Table 3, respectively, represent the standard deviation ellipse plot and its attribute table after directional distribution analysis. According to Figure 7, the overall spatial distribution of A-level scenic spots in Guizhou Province showed the obvious southwest to northeast trend. On the whole, the coverage of the ellipse tended to expand. This trend was more obvious after 2007. The standard deviation ellipses in 2005 and 2006 showed significant morphological differences compared with other years. The ellipse range of standard deviation in 2005 includes parts of Guiyang, Zunyi, Bijie, and Qiannan Prefecture. In 2019, it overlapped with some areas of all nine cities in Guizhou Province. In 2005, the standard deviation ellipse was located in the central part of Guizhou Province, its main axis was in the north–south direction, which indicated that the scenic spot expanded and developed more greatly in the north–south direction than in the southeast and northwest directions. In 2006, the shape of the standard deviation ellipse was close to the circle, which indicated that the expansion and development direction of the scenic spot was relatively uniform. From 2017 to 2019, the ellipse centered on Guiyang, the capital of Guizhou Province, and expanded mainly along the east–west direction. According to the migration map of the ellipse center (Figure 7), the migration scope of the center was small from 2005 to 2019, and the centers were almost located in Guiyang.

According to the structure calculated in Table 3, the standard deviation ellipse area increased significantly in 2006, 2007, and 2009. The standard deviation ellipse area in 2006 increased by 132.27% compared with that in 2005, 21.16% compared with that in 2006, and 34.49% compared with that in 2008. From 2005 to 2019, the standard deviation ellipse area increased year by year, except that there was no change in 2011. The minimum standard deviation ellipse area in 2005 was 9121.892 km², the standard deviation ellipse area reached the maximum of 70589.659 km² in 2019. The area of the standard deviation ellipse increased year by year, with an increase of 61467.767 km² from 2005 to 2019. From the point of view of the scenic spot distribution center, the central location was moving in Xiuwen County, Kaiyang County, Wudang District, and Baiyun District of Guiyang City. The standard deviation ellipse center of most years was located in Kaiyang County from 2005 to 2019.

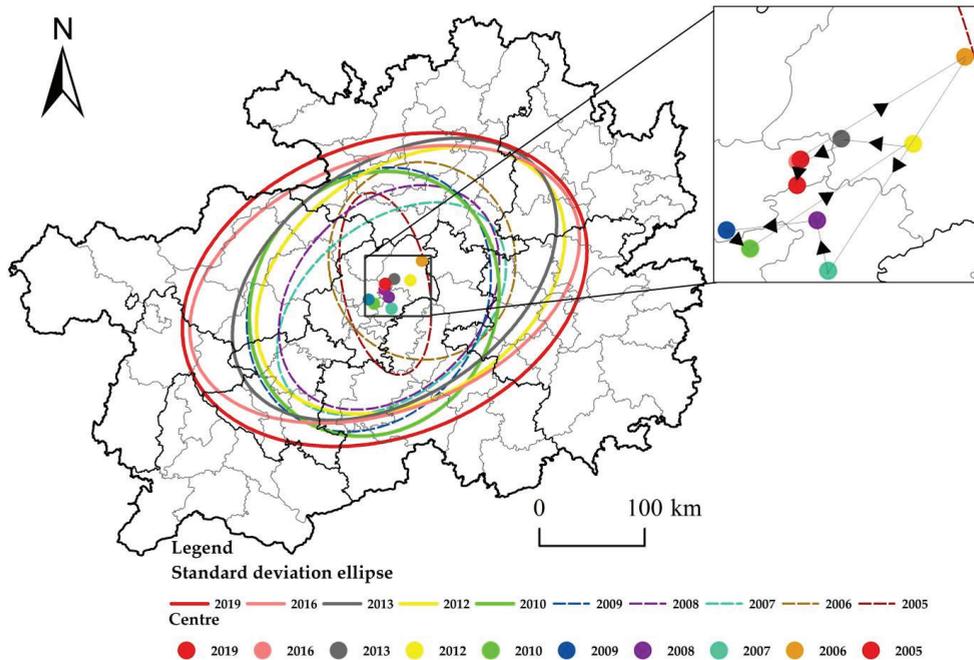


Figure 7. Standard deviation ellipse of the spatial distribution of A-level scenic spots in Guizhou Province.

Table 3. Calculation results of the standard deviation ellipse.

Year	The Area of the Ellipse Per Square Kilometers	X-Axis Standard Deviation	Y-Axis Standard Deviation	Center Position	Rotation Angle
2005	9121.892	36.767	78.980	Xiuwen County	168.590
2006	21,187.468	79.757	84.564	Kaiyang County	171.932
2007	25,671.735	111.409	73.352	Kaiyang County	50.423
2008	27,092.553	79.496	108.487	Wudang District	43.288
2009	36,438.113	97.806	118.595	Xiuwen County	32.793
2010	37,758.833	103.422	116.219	Baiyun District	29.979
2011	37,758.833	103.422	116.219	Wudang District	29.979
2012	44,658.750	143.451	99.102	Kaiyang County	56.846
2013	48,231.594	154.432	99.421	Kaiyang County	54.761
2014	52,959.536	161.470	104.407	Kaiyang County	64.671
2015	54,510.220	162.984	106.467	Kaiyang County	64.768
2016	58,713.324	173.958	107.442	Xiuwen County	68.437
2017	66,098.563	184.158	114.257	Wudang District	66.537
2018	69,953.237	183.143	121.589	Wudang District	69.048
2019	70,589.659	179.377	125.272	Xiuwen County	68.204

According to Figure 7 and Table 3, the standard deviation ellipse area was minimum and the scenic spots were along the north and south direction in 2005. This showed that the development scope of scenic spots in Guizhou Province was increasing. Since 2007, the development direction of scenic spots gradually formed a trend of extending along the southwest and northeast, and tended to disperse along the northwest and southeast. Both the short axis and the long axis of the standard deviation ellipse showed a “growing trend in fluctuations” during the process of change from 2005 to 2019.

3.2.3. Spatial Differentiation Characteristics

Trend surface analysis can intuitively show the general trend of the number of A-level scenic spots in the spatial layout of each county (district) in Guizhou Province [47]. The results are shown in Figure 8, where x and y axes point to the east and north directions respectively, and the z-axis represents the number of A-level scenic spots in each county of Guizhou Province.

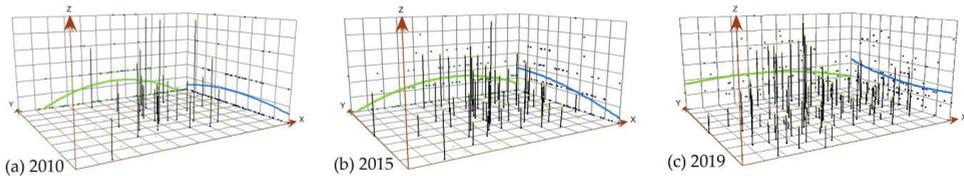


Figure 8. The spatial distribution trend surface analysis plots of A-level scenic spots in Guizhou, (a) 2010, (b) 2015, (c) 2019.

In the east–west direction, the A-level scenic spots in 2010 and 2015 presented an inverted U-shaped distribution, and the growth rate of the number of scenic spots was greater in the central counties. In 2019, the distribution of A-level scenic spots tended to be uniform, and the number of A-level scenic spots in the western counties such as Xingyi, Panzhou, and Shuicheng increased significantly. The trend curve in the east–west direction was more flat than in the north–south direction. In 2010, the trend curve was relatively flat, and its central part was slightly higher. In 2015, the distribution of scenic spots showed a “parabola” in the form of higher in the north and lower in the south. The growth rates of scenic spots in northern counties such as Xishui, Renhuai, and Chishui were greater. In 2019, the distribution of scenic spots showed a “concave” pattern with a high level in the north and a low level in the south, but the overall number of scenic spots was significantly higher than that of 2015, and the growth rate of the number of scenic spots in the south was higher than that in the north. On the whole, A-level county scenic spots in Guizhou Province were distributed in the east–west direction, and the curve gradually changed from a relatively steep inverted U-shaped curve to a gentle curve with high height in the west and low height in the east. In the north–south direction, the steepness of the trend curve gradually changed from a gentle curve to a steep curve, and finally formed a concave curve with high height in the north and low value in the south.

3.3. Factors Influencing the Spatial Distribution of A-Level Scenic Spots

The spatial distribution of A-level scenic spots in Guizhou Province changed significantly from 2005 to 2019. By the end of 2019, there had been 18 scenic spots selected into the fifth batch of the Chinese national representative catalog of intangible cultural heritage.

The distribution of scenic spots is mainly affected by natural and cultural factors. Both natural resources and cultural resources are important driving forces for the development and construction of scenic spots, and the density of scenic spots in resource rich areas will increase accordingly. The study was based on the perspective of time and space, and summarized the evolution law of the quantitative and spatial differentiation characteristics of A-level scenic spots in Guizhou Province. In addition, we discussed the coupling relationship between natural and human factors and the distribution of scenic spots and put forward some suggestions for the optimization of decision-making of the development planning and layout of scenic spots in Guizhou on the basis of detecting the impact of various factors on the development and construction of scenic spots in Guizhou Province.

3.3.1. Coupling Analysis of Natural/Human Elements and Scenic Spots Distribution

The distribution of the A-level scenic spots is greatly affected by the factors of topography and altitude, and the scattered topography can create a stronger visual impact and appreciation. Water and vegetation are also important elements of scenic spots, and

different water and vegetation landscapes create natural scenic spots with different characteristics. Guizhou Province is located in the Yunnan–Guizhou Plateau in Southwest China with an average altitude of 1100 m. More than 50% of the area is karst landform. The unique climate of “one mountain has four seasons and ten miles with different days” has become one of the natural advantages for the development of regional scenic spots. Based on the terrain slope of Guizhou (Figure 9a), the central and northern parts of the terrain are relatively gentle and the surrounding terrain is relatively steep. There was no obvious coupling phenomenon between the newly added scenic spots and the slope area in 2019. Guizhou is the main birthplace of the Yangtze River Basin and the Pearl River Basin, with a dense river network and broad watershed, and the distribution of A-level scenic spots is highly coincident with the 3 km buffer zone of the rivers (Figure 9b). The NDVI was high in the southeast and northern regions and low in the central and western regions in 2018 (Figure 9c). The A-level scenic spots in Guizhou in 2018 and the new A-level scenic spots in 2019 were mainly distributed in areas with high NDVI value.

The development of the regional economy is the basis to promote the development of tourism and is also a powerful guarantee to strengthen the construction of tourism infrastructure. From 2005 to 2019, the GDP of Guizhou Province increased from 2005 billion yuan to 16,769 billion yuan. Rapid economic development promotes the rapid development of tourism, and the proportion of tourism income makes the tourist industry become one of the indispensable key industries to promote the economic development of Guizhou Province. By 2019, the total tourism income of Guizhou Province had jumped to third place in China, and the added value of tourism income had increased to 11.6% of the province’s GDP. The spatial distribution of scenic spots was not only closely related to topography, rivers, and economic development but also the population distribution and road network were important factors affecting the distribution of scenic spots. In this study, GDP, population, and traffic were selected to conduct coupling analysis with the distribution of A-level scenic spots. Based on the spatial distribution chart of GDP in 2018 (Figure 9d), it can be seen that the areas with high GDP value in Guizhou Province mainly appeared in Guiyang and Zunyi. However, there was no significant coupling between the A-level scenic spots in 2018 and the new A-level scenic spots in 2019. The areas with high population in 2018 were mostly concentrated in central and northern Guizhou, and the A-level scenic spots in 2018 and the new A-level scenic spots in 2019 were also mostly concentrated in the same areas (Figure 9e). It can be seen in Figure 9e that the railway network mainly runs through Guizhou along the east–west direction and extends northward and westward in the centre of Guizhou. The railway density in southern Guizhou is relatively low. There was a high spatial distribution correlation between A-level scenic spots and the railway network. The expressway distribution in Guizhou Province is more uniform, more intensive in the middle and north. The distribution of the areas with concentrated population density, dense railway, and expressway network in Guizhou Province is consistent with that of the areas with dense A-level scenic spots. The A-level scenic spots are built in the areas with high road network accessibility, which will also increase the accessibility of scenic spots, thus improving the travel time efficiency of tourists, and become one of the advantages of attracting tourists.

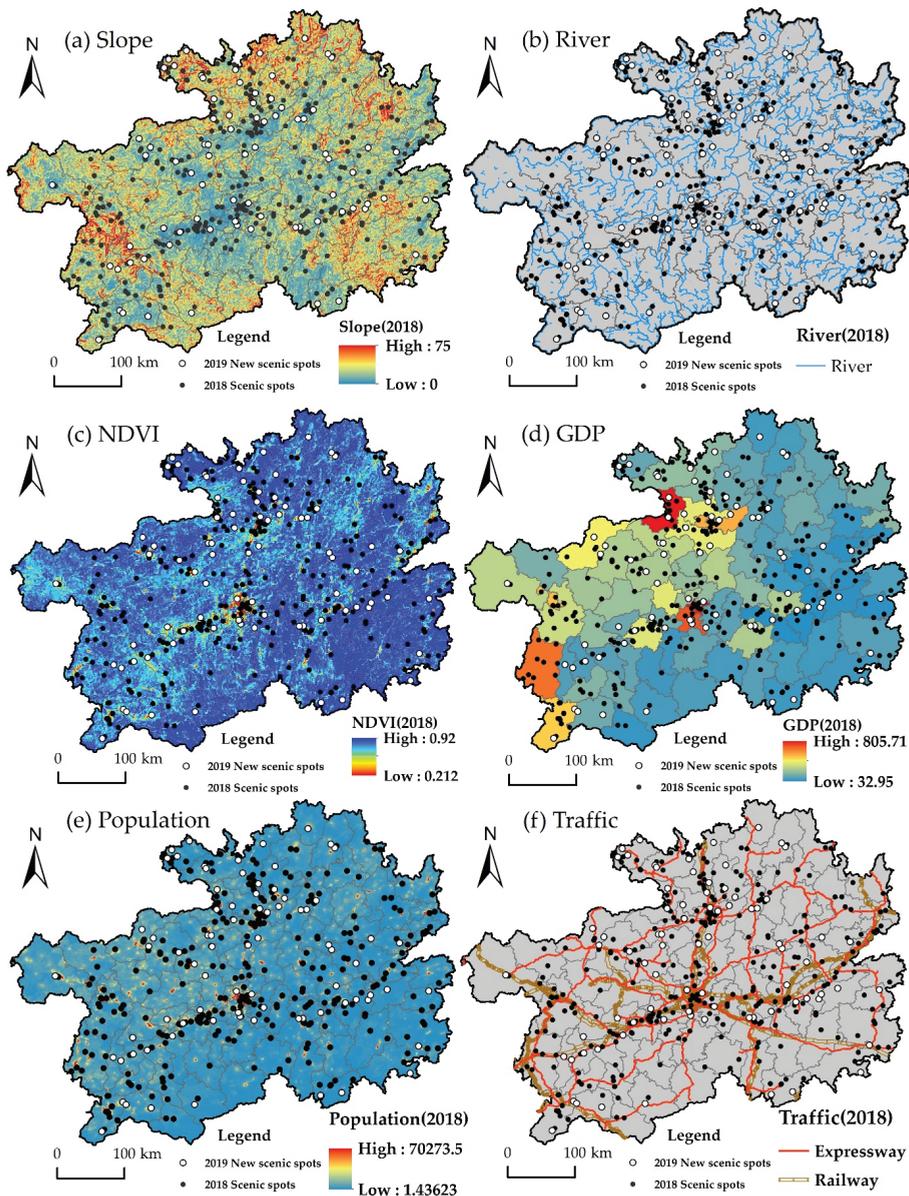


Figure 9. The typical factors influencing the spatial distribution of A-level scenic spots in Guizhou, (a) Slope, (b) Rivers, (c) NDVI, (d) GDP, (e) Population, (f) Traffic.

According to the topography of Guizhou Province, the elevation is divided into four grades: below 791 m, 792–1169 m, 1170–1682 m and 1683–2885 m (Table 4). From 2015 to 2019, the number of A-level scenic spots at the four different elevations increased. In the areas with an elevation below 791 m, the proportion of scenic spots decreased from 2015 to 2016, then, the number of scenic spots increased gradually from 2016 to 2019. In the areas with an elevation of 792–1169 m, the proportion of A-level scenic spots decreased slowly

year by year. In the areas with an elevation of 1170–1682 m, the proportion of A-level scenic spots increased gradually. In the areas with an elevation of 1683–2885 m, the number of A-level scenic spots increased firstly and then decreased, showing an inverted U-shaped downward trend. Generally, the areas with an altitude of less than 791 m and 1170–1682 m gradually became the preferred areas for the construction of A-level scenic spots. The development degree of A-level scenic spots in areas with an altitude of 792–1169 m and 1683–2885 m decreased.

Table 4. Distribution of A-level scenic spots at different elevations.

Year	<791 m		792–1169 m		1170–1682 m		1683–2885 m	
	The Number of Scenic Spots	Proportion						
2015	28	29.47%	43	45.26%	20	21.05%	4	4.21%
2016	33	25.19%	55	41.98%	29	22.14%	14	10.69%
2017	57	27.94%	76	37.25%	51	25.00%	20	9.80%
2018	92	28.22%	124	38.04%	82	25.15%	28	8.59%
2019	116	28.57%	151	37.19%	110	27.09%	29	7.14%

3.3.2. Analysis of Detection Factor Interaction Results

Coupling analysis failed to quantify the influence of various factors on the distribution of scenic spots. A geographical detector model [41] was used to explore the influence mechanism of the spatial distribution of A-level scenic spots from 2013 to 2019. The p values of all influencing factors in the measurement results were less than 0.01, which means that all factors passed the significance test. Table 5 showed that six detection factors influenced the A-level scenic spots development and construction, the road network density and tourism income factor explanatory power (q) averaged over 20%, GDP and the altitude factor explanatory power averaged around 10%, population density and river density factor explanatory power averaged small, under 5%.

Table 5. Detection results of the spatial evolution influencing factors of A-level scenic spots in Guizhou Province.

Year	The Altitude	The Population Density	River Density	Road Network Density	Tourism Income	GDP
2013	0.103	0.028	0.010	0.130	0.253	0.085
2014	0.093	0.027	0.012	0.169	0.238	0.054
2015	0.064	0.033	0.013	0.201	0.240	0.100
2016	0.066	0.028	0.021	0.249	0.241	0.095
2017	0.068	0.037	0.013	0.287	0.220	0.174
2018	0.078	0.037	0.013	0.325	0.217	0.156
2019	0.080	0.038	0.018	0.277	0.199	0.135
Mean	0.079	0.032	0.014	0.234	0.230	0.114

Both the density of the road network and tourism income have a great influence on the spatial layout of scenic spots. The factor explanatory power of road network density increased year by year from 2013 to 2018, reaching more than 20% after 2015, and reached its highest value in 2018, which was 32.5%. The explanatory power of the tourism income factor from 2013 to 2019 was above 20%, and its influence degree fluctuated slightly with the year and finally decreased. Tourism income feeds into local economic growth.

The influence of GDP on the distribution of A-level scenic spots showed an inverted U-shaped growth, increasing from 2013 to 2017 and decreasing from 2017 to 2019. The influence of altitude on the distribution of A-level scenic spots showed a V-shaped growth, and the explanatory power of the factor reached its highest in 2013, 10.3%, and showed a significant decline from 2014 to 2015, and began to rise from 2016.

The influence of population density on the spatial distribution of A-level scenic spots showed a fluctuating upward trend. The explanatory power of population density showed a W-shaped fluctuation from 2013 to 2017 and changed to stabilize after 2017. The influence

of river density on the spatial distribution of A-level scenic spots in Guizhou showed a fluctuating upward trend. From 2013 to 2015, the impact was relatively stable. The impact was significantly higher in 2016 than in 2015. The impact began to decline in 2017.

During the study period, the influences of altitude, river density, and population density on the spatial distribution of A-level scenic spots in Guizhou Province were relatively stable. The q value of road network density and GDP increased significantly, indicating that these two influencing factors had a significant increase in the distribution of A-level scenic spots. The influences of tourism income on the distribution of A-level scenic spots decreased slightly. Compared with other detection factors, road network density and tourism income had a higher influence on the spatial distribution of A-level scenic spots.

4. Discussion and Suggestions

4.1. Discussion

From 2005 to 2019, the number growth of A-level scenic spots in Guizhou Province can be divided into three stages: the early development period from 2005 to 2010, the moderate development period during 2011–2015, and the rapid development period after 2016. According to the calculation through the average nearest neighbor index, the spatial distribution of A-level scenic spots in Guizhou Province has been concentrated for many years. During China's 13th Five Year Plan period, Guizhou Province launched the integration policy of 'big tourism', 'big data' and 'big ecology'. The number of scenic spots has ushered in a blowout development, and tourism income has also achieved considerable growth. Tourism development in Southern Guizhou still lags behind relatively.

The Gini coefficient of the spatial distribution of A-level scenic spots in Guizhou Province showed a significant downward trend from 2005 to 2017. The Gini coefficient increased slightly from 2017 to 2019. It showed that the equilibrium degree of the scenic spot distribution was developing towards equality from 2005 to 2017, and the equilibrium degree was closest to absolute equality in 2017. While the spatial distribution equilibrium of scenic spots tended to be inequality from 2017 to 2019, with the characteristics of small-scale concentration.

Based on the spatial distribution kernel density of A-level scenic spots in Guizhou Province from 2005 to 2019, the number and scope of high-density areas in the scenic spots increased year by year, forming the spatial distribution characteristics of "one axis and two cores" in 2017 and "two axes and multiple cores" in 2019. According to the standard deviation ellipse of the distribution of A-level scenic spots in Guizhou Province, the size of the ellipse increased year by year, and Guiyang was always in the center of the ellipse. The long axis direction of the standard deviation ellipse changed significantly from 2005 to 2007, and the long axis direction from 2007 to 2019 was mainly southwest to northeast.

In the east–west direction, the number of county-level scenic spots in Guizhou Province was a gentle curve with high in the West and low in the East. In the north–south direction, the steepness of the trend curve was a concave curve with high in the north and low in the south. The distributions of rivers and scenic spots showed coupling phenomenon. The distributions of slope, NDVI, and scenic spots showed a significant coupling phenomenon. The population distribution, the road network, and the scenic spot distribution were highly correlated. In recent years, with the continuous improvement of China's GDP, the government's investment in transportation and tourism has promoted the development of regional tourism and the construction of scenic spots [48–50]. At the same time, the development of tourism promotes the development of transportation and regional economy [48–50]. It has also driven GDP growth and local investment.

With the help of Geodetector, it was found that the road network density and tourism income had a strong impact on the distribution of A-level scenic spots. The density of road network will directly affect the accessibility of scenic spots, thus affecting the tourism planning of tourists in selecting scenic spots, and then affect the maintenance income and brand effect of scenic spots. Tourism income will stimulate local attention and investment

in the tourist industry, and further affect the construction and development of local scenic spots. The influence of altitude on the distribution density of scenic spots in the early years was stronger than that in the later years. The possible reason is that in the early years, the construction of scenic spots with high altitude was difficult and relatively inexperienced, while in the later years, with the progress of technology, the construction difficulty was no longer a large problem in determining the construction of scenic spots. With the growth and change of the regional economy, the mode of economic growth will gradually promote the tourism industry. Therefore, the impact of GDP on the distribution of scenic spots is increasing year by year. The population density is mainly affected by the city, and scenic spots are mainly used as a tourist destination for non-local visitors. The rivers are distributed widely and evenly in Guizhou Province. So, the impact of population density and river distribution on the spatial distribution of scenic spots is relatively weak.

4.2. Suggestions

Based on the analysis of the spatiotemporal evolution characteristics and influencing factors of A-level scenic spots in Guizhou Province, combined with the regional resource endowment, we put forward suggestions for the development, construction, and layout optimization of scenic spots in Guizhou Province.

According to the evolution characteristics of the spatial distribution of A-level scenic spots, the lagging development of scenic spots in the south of Guizhou is the important problem existing in the development of scenic spots in Guizhou Province. Traffic conditions are one of the main factors affecting scenic spot planning. However, the road network density is small in the south of Guizhou. Therefore, strengthening the traffic construction of Qianxinan, Qiannan, and Qianzhongnan can effectively improve the regional accessibility of the three prefectures, strengthen the convenience for tourists in the southern scenic spots, promote the growth of local tourism, and promote the development and construction of scenic spots in the region. The tourism resource endowment in the southern region has not been effectively developed. There are abundant river valleys with excellent water resources. Using river resources to build hydrological scenic spots can be one of the effective ways to develop southern scenic spots.

The construction of A-level scenic spots is an important starting point for tourism development, regional coordination, and urban-rural integration. Therefore, Guizhou should make full use of the regional resources advantages to optimize the layout of the scenic spots. At present, A-level scenic spots in Guizhou present a spatial layout of "two axes and multiple cores". Guiyang has always been the core of scenic spot planning and development. Guizhou can take the advantage of the "two axis and multi-core" pattern, combined with the development mode of "point axis and network", explore relying on the big data technology advantage platform to expand the radiation scope of each A-level scenic spot, so as to strengthen inter regional tourism industry cooperation and eliminate a regional tourism market fortress, promote the development of global tourism in Guizhou Province, and finally realize the effective allocation and rational utilization of tourism resources.

5. Conclusions

This paper studied the temporal and spatial evolution characteristics of A-level scenic spots in Guizhou Province from 2005 to 2019, including spatial distribution, density, balance degree, temporal change trend, and direction characteristics, and analyzed the natural and human factors influencing the scenic spots' distribution qualitatively and quantitatively.

Overall, the A-level scenic spots in Guizhou Province have shown a good development trend in recent years. However, the development in southern Guizhou province is less optimistic. The rapid growth in the number of A-level scenic spots led to small-scale agglomeration in spatial distribution from 2017 to 2019. Guiyang has always been the center of A-level scenic spots planning in Guizhou Province. The kernel density distribution of A-level scenic spots in Guizhou Province forms the "two-axis, multi-core" layout.

The road network density, tourism income, and GDP had a higher influence on the A-level scenic spots distribution. As time goes by, the influence of terrain height on scenic spot construction was gradually reduced. The area with an altitude of 1170 m to 1682 m has gradually become the first choice for the construction of scenic spots in Guizhou Province. Because of the unique terrain and water system in Guizhou Province, population distribution and rivers have little impact on the planning and construction of A-level scenic spots. Finally, we have provided some suggestions for scenic spot layout optimization in Guizhou Province on the basis of the perspective of regional resource endowment and scenic spot spatial layout.

Author Contributions: Conceptualization, Jian Yin; methodology, Jian Yin and Yuanhong Qiu; validation, Jian Yin, Yuanhong Qiu and Ting Zhang; formal analysis, Yuanhong Qiu; investigation, Jian Yin; resources, Ting Zhang and Bin Zhang; data curation, Yuanhong Qiu; writing—original draft preparation, Yuanhong Qiu and Yiming Du; writing—review and editing, Jian Yin; visualization, Jian Yin and Yuanhong Qiu. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the MOE (Ministry of Education in China) Liberal Arts and Social Sciences Foundation (Grant No. 19YJCZH228), and the Scientific Research Project of Guizhou University of Finance and Economics (2020ZXS08). The authors are grateful to the reviewers for their help and thought-provoking comments.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We thank the editors and the anonymous reviewers for their valuable comments and suggestions.

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

References

1. Wang, H.Q.; Yuan, J.D.; Meng, X.J. Spatial Distribution and Its Influencing Factors of Level-A Scenic Spots in Northeast China. *Sci. Geogr. Sin.* **2017**, *37*, 895–903. [CrossRef]
2. Ma, X.L.; Yang, X.J. A Correlating Analysis on High Grades Tourism Area: Take 4A Grade Tourism Area (Spots) for Example. *J. Northwest Univ. Natl. Sci. Ed.* **2014**, *32*, 233–237. [CrossRef]
3. The Fourteenth Five-Year Plan for the National Economic and Social Development of Guizhou Province and the Outline of the Long-range Goals in 2035. *Guizhou Daily*. 27 February 2021. Available online: http://www.szb.gzrbs.com.cn/pc/layout/202102/27/node_01.html (accessed on 16 June 2021).
4. Weng, G.M.; Li, J.P.; Yang, X.P.; Li, C.H. Research Trends of Tourism Environment Carrying Capacity at Home and Abroad in Recent 20 Year. *Geogr. Geo-Inf. Sci.* **2021**, *37*, 106–111. [CrossRef]
5. Liu, M.; Hao, W. Spatial Distribution and Its Influencing Factors of National A-level Tourist Attractions in Shanxi Province. *Acta Geogr. Sin.* **2020**, *75*, 878–888. [CrossRef]
6. Kang, S.; Lee, G.; Kim, J.; Park, D. Identifying the Spatial Structure of the Tourist Attraction System in South Korea Using GIS and Network Analysis: An Application of Anchor-point Theory. *J. Destin. Mark. Manag.* **2018**, *9*, 358–370. [CrossRef]
7. Padilla, J.J.; Kavak, H.; Lynch, C.J.; Gore, R.J.; Diallo, S.Y. Temporal and Spatiotemporal Investigation of Tourist Attraction Visit Sentiment on Twitter. *PLoS ONE* **2018**, *13*, e0198857. [CrossRef]
8. Wang, C.; Zhang, J.H.; Yu, P.; Hu, H. The Theory of Planned Behavior as a Model for Understanding Tourists' Responsible Environmental Behaviors: The Moderating Role of Environmental Interpretations. *J. Clean. Prod.* **2018**, *194*, 425–434. [CrossRef]
9. Heng, C.; Gang, L.; Haiyan, S. Modelling the Interdependence of Tourism Demand: The Global Vector Autoregressive Approach. *Chin. J. Tour. Res.* **2017**, *31*, 1–13. [CrossRef]
10. Samitas, A.; Asteriou, D.; Polyzos, S.; Kenourgios, D. Terrorist incidents and tourism demand: Evidence from Greece. *Tour. Manag. Perspect.* **2018**, *25*, 23–28. [CrossRef]
11. Zhang, L.; Wang, Y.P.; Sun, J.; Yu, B. The Sightseeing Bus Schedule Optimization under Park and Ride System in Tourist Attractions. *Ann. Oper. Res.* **2019**, *273*, 587–605. [CrossRef]
12. Fyall, A.; Leask, A.; Barron, P.; Ladkin, A. Managing Asian attractions, Generation Y and face. *J. Hosp. Tour. Manag.* **2017**, *32*, 35–44. [CrossRef]
13. Jensen, O.; Li, Y.; Uysal, M. Visitors' Satisfaction at Managed Tourist Attractions in Northern Norway: Do On-site Factors Matter? *Tour. Manag.* **2017**, *63*, 277–286. [CrossRef]

14. Zuo, L.; Zhang, J.; Zhang, R.D.J.; Zhang, Y.Y.; Hu, M.; Zhuang, M.; Liu, W. The Transition of Soundscapes in Tourist Destinations from the Perspective of Residents' Perceptions: A Case Study of the Lugu Lake Scenic Spot, Southwestern China. *Sustainability* **2020**, *12*, 1073. [[CrossRef](#)]
15. Day, J.; Chin, N.; Sydnor, S.; Cherkauer, K. Weather, Climate, and Tourism Performance: A Quantitative Analysis. *Tour. Manag. Perspect.* **2013**, *5*, 51–56. [[CrossRef](#)]
16. Smith, T.; Fitchett, J.M. Drought Challenges for Nature Tourism in the Sabi Sands Game Reserve in the Eastern Region of South Africa. *Afr. J. Range Forage Sci.* **2020**, *37*, 107–117. [[CrossRef](#)]
17. Atzori, R.; Fyall, A.; Miller, G. Tourist Responses to Climate Change: Potential Impacts and Adaptation in Florida's Coastal Destinations. *Tour. Manag.* **2018**, *69*, 12–22. [[CrossRef](#)]
18. Susanto, J.; Zheng, X.; Liu, Y.; Wang, C. The Impacts of Climate Variables and Climate-related Extreme Events on Island Country's Tourism: Evidence from Indonesia. *J. Clean. Prod.* **2020**, *276*, 9. [[CrossRef](#)]
19. Kanwal, S.; Rasheed, M.I.; Pitafi, A.H.; Pitafi, A.; Ren, M.L. Road and Transport Infrastructure Development and Community Support for Tourism: The Role of Perceived Benefits, and Community Satisfaction. *Tour. Manag.* **2020**, *77*, 10. [[CrossRef](#)]
20. Li, L.S.; Yang, F.X.; Cui, C. High-speed Rail and Tourism in China: An Urban Agglomeration Perspective. *Int. J. Tour. Res.* **2019**, *21*, 45–60. [[CrossRef](#)]
21. Peng, X.; Huang, Z. A Novel Popular Tourist Attraction Discovering Approach Based on Geo-Tagged Social Media Big Data. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 216. [[CrossRef](#)]
22. Zhan, M.S.; Zhu, J.H. Spatial Planning of Qingyun Mountain Scenic Spot in Benxi, Northeast China based on Ecological Sensitivity Assessment. *Chin. J. Appl. Ecol.* **2019**, *30*, 2352–2360. [[CrossRef](#)]
23. Xie, Z.H.; Wu, B.H. Tourism Spatial Structure of Resources-based Attractions in China. *Sci. Geogr. Sin.* **2008**, *28*, 748–753. [[CrossRef](#)]
24. Li, J.X.; Xie, D.T.; Wang, S. Characteristics of the Spatial Pattern of Tourist Attractions of Grade A in Chongqing and Their Evolution. *J. Southwest Univ. Nat. Sci.* **2021**, *3*, 153–163. [[CrossRef](#)]
25. Zhu, H.; Chen, X.L. Space Distribution Structure of A-grade Scenic Spot in China. *Sci. Geogr. Sin.* **2008**, *28*, 607–615. [[CrossRef](#)]
26. Zhang, H.T.; Shi, T.T.; Bao, H. The Spatial Structure Characteristics of China's 5A-Level Tourist Attractions. *J. Huaqiao Univ. Philos. Soc. Sci.* **2019**, *4*, 80–90. [[CrossRef](#)]
27. Wang, W.X.; Xie, S.Y. The Spatial Pattern of the A-grade Tourist Sites in Hubei Province and Its Optimization. *Areal Res. Dev.* **2012**, *31*, 124–128. [[CrossRef](#)]
28. Li, D.H.; Zhang, X.Y.; Lu, L. Spatial Distribution Characteristics and Influencing Factors of High-Level Tourist Attractions in the Yellow River Basin. *Econ. Geogr.* **2020**, *40*, 70–80. [[CrossRef](#)]
29. Tang, C.C.; Sun, M.Y.; Wan, Z.W. Spatial Distribution Characteristics of High-Level Scenic Spots and Its Influencing Factors in Beijing-Tianjin-Hebei Urban Agglomeration. *Econ. Geogr.* **2019**, *39*, 204–213. [[CrossRef](#)]
30. Liu, M.; Hao, W.; Zhang, F.R. The Spatial Distribution and Influence Factors of A-Level Scenic Spots in Shanxi Province. *Econ. Geogr.* **2020**, *40*, 231–240. [[CrossRef](#)]
31. Wu, Q.; Ma, H.L.; Pan, W.K. Study on the Spatio-Temporal Evolution and Its Influencing Factors of A-level Tourist Attractions in Guangdong Province. *J. Nat. Sci. Hunan Norm. Univ.* **2021**, *44*, 26–33, 40. [[CrossRef](#)]
32. Wu, Q.; Li, X.; Wu, L. Distribution Pattern and Spatial Correlation of A-Grade Tourist Attractions in Hunan Province. *Econ. Geogr.* **2017**, *37*, 193–200. [[CrossRef](#)]
33. Hu, W.X.; Liang, X.T.; Sang, Z.Y. Analysis on the Characteristics and Causes of the Spatial-temporal Evolution of 3 A and above Tourist Attractions in Shanxi Province. *J. Arid Land Resour. Environ.* **2020**, *34*, 187–194. [[CrossRef](#)]
34. Yao, Y.J.; Jing, H.; Liu, D.J. Spatial Evolution and Influence Mechanism of A-Level Scenic Spots in Urban Agglomeration in the Middle Reaches of the Yangtze River. *Econ. Geogr.* **2019**, *39*, 198–206. [[CrossRef](#)]
35. Lu, B.Y.; Zhang, E.W.; Ming, Q.Z. Spatial Evolution Characteristics and Driving Mechanism of A-class Tourist Attractions in Yunnan Province, China. *Mt. Res.* **2019**, *37*, 879–890. [[CrossRef](#)]
36. Jing, Z.Z.; Han, J.; Liu, W.B. Spatial Distribution Characteristics and Accessibility of A-Level Scenic Spots in Shanxi Province. *Resour. Dev. Mark.* **2021**, *37*, 200–207. [[CrossRef](#)]
37. Liu, K.; Wang, K.; Fan, W.Q.; Cui, R.Y.; Li, C. Study on Spatial Distribution Characteristics and Influencing Mechanism of Tourism Scenic Spots in Ethnic Mountain Areas—Taking Guizhou Province as an Example. *J. Nat. Sci. Hunan Norm. Univ.* **2019**, *42*, 17–25. [[CrossRef](#)]
38. Wang, M.; Jiang, C.; Sun, O.J. Spatially Differentiated Changes in Regional Climate and Underlying Drivers in Southwestern China. *J. For. Res.* **2021**. [[CrossRef](#)]
39. Zhang, J.H. A Convenient Method to Calculate Gini Coefficient. *J. Shanxi Agric. Univ. Soc. Sci. Ed.* **2007**, *6*, 275–278, 283. [[CrossRef](#)]
40. Liu, L.M.; Lv, J. Research on Spatial Form Evolution of Inbound Tourism Destination in Inner Mongolia. *J. Arid Land Resour. Environ.* **2016**, *30*, 203–208. [[CrossRef](#)]
41. Wang, J.F.; Xu, C.D. Geodetector: Principle and Prospect. *Acta Geogr. Sin.* **2017**, *72*, 116–134. [[CrossRef](#)]
42. Wang, J.F.; Gao, B.B.; Stein, A. The spatial statistic trinity: A generic Framework for Spatial Sampling and Inference. *Environ. Model. Softw.* **2020**, *134*, 104835. [[CrossRef](#)]

43. Wang, J.F.; Xu, C.D.; Hu, M.G.; Li, Q.X.; Yan, Z.W.; Jones, P. Global Land Surface Air Temperature Dynamics since 1880. *Int. J. Climatol.* **2018**, *38*, e466–e474. [[CrossRef](#)]
44. Yang, J.T.; Song, C.; Yang, Y.; Xu, C.D.; Guo, F.; Xie, L. New Method for Landslide Susceptibility Mapping Supported by Spatial Logistic Regression and GeoDetector: A Case Study of Duwen Highway Basin, Sichuan Province, China. *Geomorphology* **2019**, *324*, 62–71. [[CrossRef](#)]
45. Hu, M.G.; Xu, C.D.; Wang, J.F. Spatiotemporal Analysis of Men Who Have Sex with Men in Mainland China. *JMIR mHealth uHealth* **2020**, *8*, e14800. [[CrossRef](#)]
46. Zhu, L.J.; Meang, J.J.; Zhu, L.K. Applying Geodetector to Disentangle the Contributions of Natural and Anthropogenic Factors to NDVI Variations in the Middle Reaches of the Heihe River Basin. *Ecol. Indic.* **2020**, *117*, 12. [[CrossRef](#)]
47. Xu, D.; Huang, R.; Huang, Z.F.; Tang, F.J. Spatial-Temporal Pattern and Spillover Effects of Tourism Flows in Chinese Cities from 2001 to 2015. *Geogr. Geo-Inf. Sci.* **2020**, *36*, 135–142. [[CrossRef](#)]
48. Zhou, M.; Liu, X.Q.; Pan, B.; Yang, X.; Wen, F.H.; Xia, X.H. Effect of Tourism Building Investments on Tourist Revenues in China: A Spatial Panel Econometric Analysis. *Emerg. Mark. Financ. Trade* **2017**, *53*, 1973–1987. [[CrossRef](#)]
49. Zuo, B.; Huang, S.S. Revisiting the Tourism-Led Economic Growth Hypothesis: The Case of China. *J. Travel Res.* **2018**, *57*, 151–163. [[CrossRef](#)]
50. Zhang, X.; Song, C.; Wang, C.W.; Yang, Y.L.; Ren, Z.P.; Xie, M.Y.; Tang, H.H. Socioeconomic and Environmental Impacts on Regional Tourism across Chinese Cities: A Spatiotemporal Heterogeneous Perspective. *ISPRS Int. J. Geo-Inf.* **2020**, *10*, 410. [[CrossRef](#)]

Article

Spatiotemporal Evolution and Trend Prediction of Tourism Economic Vulnerability in China's Major Tourist Cities

Chengkun Huang^{1,†}, Feiyang Lin^{2,†}, Deping Chu^{1,3,*}, Lanlan Wang¹, Jiawei Liao¹ and Junqian Wu⁴

¹ College of Tourism, Fujian Normal University, Fuzhou 350108, China; qsx20191055@student.fjnu.edu.cn (C.H.); qsx20180979@student.fjnu.edu.cn (L.W.); 132032018094@student.fjnu.edu.cn (J.L.)

² School of International Education, Tianjin University, Tianjin 300350, China; linfeiyang0123@tju.edu.cn

³ The Higher Educational Key Laboratory for Smart Tourism of Fujian Province, Fuzhou 350108, China

⁴ West Center for Economic Research, Southwestern University of Finance and Economics, Chengdu 610074, China; wujunqian@swufe.edu.cn

* Correspondence: chudeping@fjnu.edu.cn

† These authors contributed equally to this work.

Abstract: The evaluation and trend prediction of tourism economic vulnerability (TEV) in major tourist cities are necessary for formulating tourism economic strategies scientifically and promoting the sustainable development of regional tourism. In this study, 58 major tourist cities in China were taken as the research object, and an evaluation index system of TEV was constructed from two aspects of sensitivity and adaptive capacity. On the basis of the entropy weight method, TOPSIS model, obstacle diagnosis model, and BP neural network model, this study analyzed the spatiotemporal patterns, obstacle factors, and future trends of TEV in major tourist cities in China from 2004 to 2019. The results show three key findings: (1) In terms of spatiotemporal patterns, the TEV index of most of China's tourist cities has been on the rise from 2004 to 2019. Cities throughout the coast of China's Yangtze River Delta and the Pearl River Delta urban agglomeration show high vulnerability, whereas low vulnerability has a scattered distribution in China's northeast, central, and western regions. (2) The proportion of international tourists out of total tourists, tourism output density, urban industrial sulfur dioxide emissions per unit area, urban industrial smoke and dust emission per unit area, and discharge of urban industrial wastewater per unit area are the five major obstacles affecting the vulnerability degree of the tourism economy. (3) According to the prediction results of TEV from 2021 to 2030, although the TEV of many tourist cities in China is increasing year by year, cities with low TEV levels occupy the dominant position. Research results can provide reference for tourist cities to prevent tourism crises from occurring and to reasonably improve the resilience of the tourism economic system.

Keywords: tourism economic vulnerability; spatiotemporal evolution; obstacle factors; trend prediction; major tourist cities

Citation: Huang, C.; Lin, F.; Chu, D.; Wang, L.; Liao, J.; Wu, J. Spatiotemporal Evolution and Trend Prediction of Tourism Economic Vulnerability in China's Major Tourist Cities. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 644. <https://doi.org/10.3390/ijgi10100644>

Academic Editors: Andrea Marchetti, Angelica Lo Duca and Wolfgang Kainz

Received: 22 July 2021

Accepted: 21 September 2021

Published: 25 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The concept of "vulnerability" originated from natural science research; it is used to characterize the ability of a system or system combination to withstand and recover from risk events [1]. In the early stages, the vulnerability concept was mainly applied to the assessment of natural disasters such as floods and droughts or ecosystems such as forests and coasts [2–5]. With the gradual integration and penetration of the natural and social systems, the interaction between the natural environment and human social activities has become increasingly obvious [6], and the relevant research on vulnerability has gradually extended to the social and economic fields [7,8]. As one of the important components of the concept and connotation of vulnerability, economic vulnerability refers to the bearing capacity of the regional economy due to the impact of unexpected events in the process of

development [9,10]. Economic vulnerability was first proposed by Briguglio in the 1990s and has been gradually deepened in subsequent studies [11]. At present, it has become an important indicator to measure whether the development of a regional or urban economic system is healthy and stable [12].

With the rapid development of China's social economy and the improvement of people's living standards, tourism has been gradually positioned as a "strategic pillar industry and modern service industry," playing an increasingly important role in regional economic development. However, as a typical sensitive industry, tourism will be greatly impacted by financial crises [13], political conflicts [14], social disturbances [15], public health events [16], and natural disasters [17] in the context of the integration of global trade of services. This is especially true in areas where economic development is highly dependent on tourism; although these areas have gained huge profits through the vigorous development of tourism, the instability of tourism will inevitably bring about regional economic shocks, and regional economic development is generally vulnerable to hidden worries [18]. Nowadays, with the increasing role of tourism in national political communication, economic development, and residents' well-being, as well as the pursuit of regional sustainable development goals, the research on TEV is receiving growing attention [19,20].

TEV refers to the inherent property wherein the structure and function of the tourism economy system are easily damaged due to the restriction of its own property and the inability to adapt to various disturbances inside and outside the system [21]. TEV is usually divided into two types: "endogenous" and "exogenous". Endogenous vulnerability is formed under the constraints of certain economic systems and tourism resources and cannot be eliminated by conscious actions, such as policy combining [22,23]. Exogenous vulnerability is a result of "non-systemic causes" from the external environment, such as earthquakes, public health events, financial crises, and social disturbances, which have contingent and sudden characteristics [22,23]. In general, the literature on TEV mainly focused on the following two aspects: (1) The analysis of TEV under the impact of crisis events; such studies focus on the impact of some emergencies on the tourism economy from the perspective of crisis management and take the impact degree of the crisis as the basis for assessing vulnerability. For example, Huang et al. analyzed the long-term impact of the Wenchuan Earthquake on inbound tourists in Sichuan and found a significant increase in inbound tourists after the earthquake, with a "blessing in disguise" effect [24]. Pham et al. used the tourism satellite account approach and tourism CGE model to effectively measure the changes and impacts of COVID-19 on the core and related industries of Australia's inbound tourism [25]. In addition, the recovery and development of the tourism economy in the context of crisis events is also an important research topic [26]. Gurtner used the case of Bali to illustrate that after a tourism crisis, the government, industry, community, and other tourism stakeholders need to strengthen cooperation and adopt a wide range of new strategies to deal with the changing destination environment and potential challenges in the future [27]. Raki et al. discussed the role of active and proactive tourism recovery strategies in improving the well-being of tourists, improving the profitability of companies, and reducing employee turnover under the impact of COVID-19 [28]. (2) Assessment of the TEV of typical tourist destinations; this kind of research focuses on the evaluation of tourism economic system shock resistance of various types of tourism destinations. Research on islands, countries, typical tourism cities, national regions, and other traditional tourist destinations is prioritized using the entropy weight method, TOPSIS model, obstacle degree model, and geographical detectors and comprehensive quantitative analysis of vulnerability degree; research contents include TEV measurement, spatiotemporal pattern evolution, and influence factors [29–33].

It can be seen from the above analysis that the existing literature is still mostly limited to discussing the TEV of individual typical tourist destinations. However, with the rapid development of China's tourism industry, a global analysis of TEV in major tourist cities on a national scale is urgently required to optimize the regional pattern of tourism development. In addition, the existing literature usually measures regional TEV in previous

years but lacks predictive research on the regional TEV in the future. Such discussion is more conducive to grasping the evolutionary trend of TEVs in order to rationally plan relevant strategies to reduce the regional TEV. In view of this, 58 major tourist cities in China were selected as study areas for this paper. Our objectives were as follows: (1) Clarify the spatiotemporal evolution of TEV in major tourist cities in China. (2) Explore the main obstacles affecting TEV in major tourist cities in China. (3) Forecast the evolution trend of TEV of major tourist cities in China in the next 10 years.

In this study, we first assessed the level of TEV of each city from 2004 to 2019 based on the case studies of 58 major tourist cities in China, using the entropy weight method and TOPSIS model. Then, the obstacle diagnosis model was used to analyze the obstacle factors affecting TEV in major tourist cities in China. Finally, the BP neural network model was used to predict the evolutionary trend of TEV in major tourist cities in China in the future. The research conclusions are of great significance for the detailed understanding of TEV and the future evolutionary trend of major tourist cities in China in the context of high-quality development. These results can provide a reference for regional tourism crisis prevention and effectively enhance the resilience of the urban tourism economy.

The structure of this study can be divided into five parts. The first part is the Introduction, which introduces the research background, research objectives, existing research results, and the value of this research. The second part is the Materials and Methods, which establishes the evaluation index system of TEV, explains the data sources, and introduces the application logic of research methods. The third part is the Results, which expounds on the spatiotemporal evolution of the TEV of major tourist cities in China, the obstacle factors affecting the TEV, and the future evolutionary trend of the TEV. The fourth part is the Discussion, which summarizes the spatiotemporal characteristics and future evolutionary trends of TEV in major tourist cities in China, and puts forward countermeasures to improve the resilience of the urban tourism economy. The fifth part is the Conclusion, which shows the highlights of the results and limitations of the study.

2. Materials and Methods

2.1. Study Area

A tourist city considers tourism development to be an important goal that has a prominent function after a certain period of accumulation [34]. The Yearbook of China Tourism Statistics has recorded long-term tracking statistics on the tourism development of 60 major tourist cities in China. However, due to the lack of statistical data of Yanbian and Lhasa, among the 60 major tourist cities, this study selected only 58 cities as research objects in this study, as shown in Figure 1. These major tourist cities not only have prominent tourism functions, evident progress in the city's tourism construction, and enjoy high popularity at home and abroad; they also have large differences in their urban population on an economic scale, wide regional coverage, and diverse urban types. These characteristics make them suitable for exploring the urban TEV.

2.2. Research Framework

Figure 2 shows the implementation framework of this study, which mainly includes three steps. First, on the basis of relevant research, the evaluation index system of TEV was constructed from the two dimensions of sensitivity and adaptive capacity. Second, the data needed for this study were collected from various statistical yearbooks of China. Finally, the econometric correlation model and spatial visualization methods were used to present the research results.

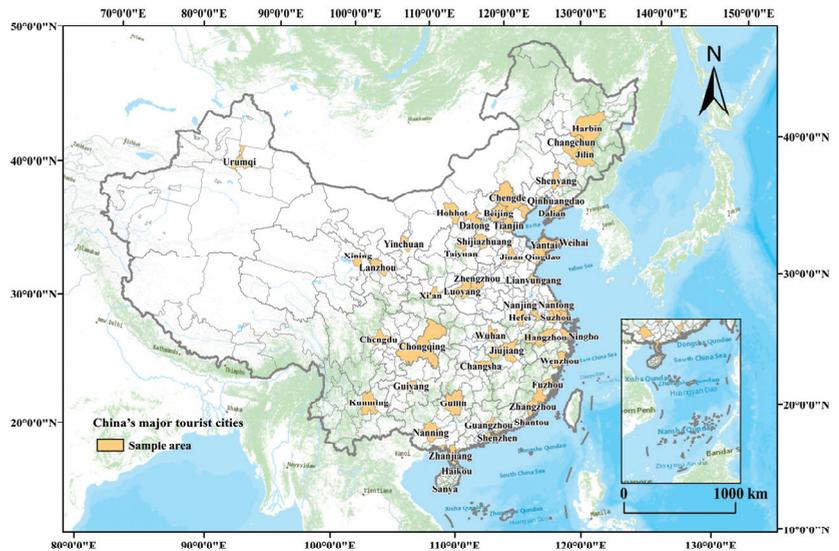


Figure 1. Spatial distribution of 58 major tourist cities in China.

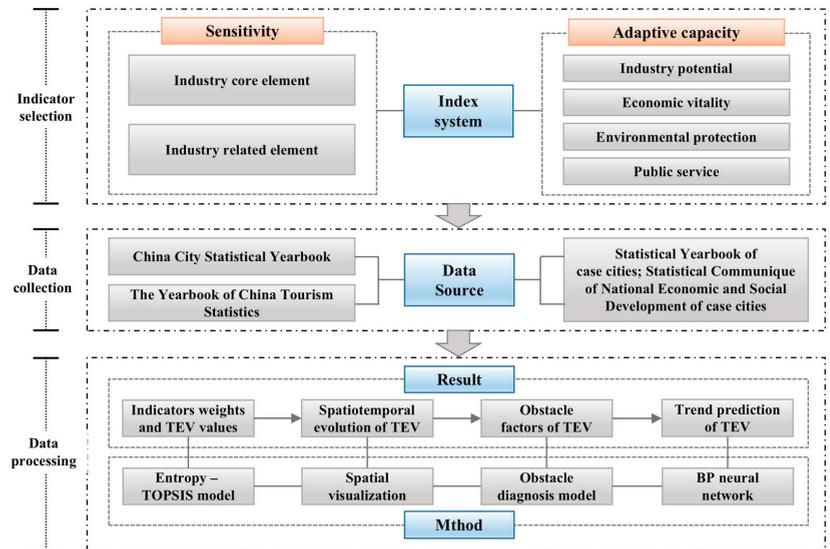


Figure 2. Research framework.

2.3. Index System Construction

Polsky et al. constructed the vulnerability assessment system of “exposure, sensitivity, and adaptive capacity” in 2007, which provided a solid theoretical basis for vulnerability research [35]. After that, scholars in different fields continued to apply and expand the vulnerability theory model based on it, among which the two dimensions of “sensitivity and adaptive capacity” have been gradually taken as the core dimension of the vulnerability assessment of tourism [23,36]. Sensitivity refers to the ability of a system to withstand damage in the case of internal disorder and external impact [35]. The weaker the sensitivity is, the less vulnerable a system is to damage. Adaptive capacity refers to the ability of

a system to quickly adjust from a crisis situation to a safe and stable situation [35]. The stronger the adaptive capacity is, the stronger the self-maintenance ability of a system and its ability to quickly recover from adverse effects. Sensitivity and adaptive capacity determine the vulnerability of a system in the interaction. The interaction between sensitivity and adaptive capacity determines the vulnerability of the tourism economic system. When the tourism economy has a high vulnerability, it indicates that the tourism economy has a poor anti-crisis ability, which reduces the speed at which the tourism economy can recover to a stable state; otherwise, the economic system is more secure.

As the sensitivity and adaptive capacity of the tourism economy are multiple structural variables, they involve complex economic environmental factors. To reflect the degree of TEV of major tourist cities in China in a comprehensive way, the evaluation index system proposed in this study was constructed as follows. First, the construction methods and contents involved in the existing research on the index system of TEV were fully utilized for reference [21,23,32,36]. Second, the accessibility of the data of each indicator was ensured. Finally, the index system can be applied to different types of tourist cities in China. On the basis of the above considerations, this study combined the basic elements of the tourism industry, social economy, finance, infrastructure construction, and ecological environment, and a total of 27 indicators from the two aspects of sensitivity and adaptive capacity were selected to construct an evaluation index system for the TEV of major tourist cities in China. Table 1 shows the specific indicators.

Table 1. Index system of TEV.

System	Primary Indicator	Secondary Indicator	Unit	Attribute	Serial Number
Sensitivity	Industry core element	Proportion of total earnings from tourism in GDP	—	Positive	S1
		Elasticity coefficient of tourism to GDP growth	—	Positive	S2
		Proportion of international tourists out of total tourists	%	Positive	S3
		Proportion of tourism foreign exchange income in total tourism revenue	%	Positive	S4
		Proportion of total tourism revenue in tertiary industry output value	%	Positive	S5
		Tourism output density	10,000 CNY/km ²	Positive	S6
	Industry-related element	Discharge of urban industrial wastewater per unit area	10,000 tons/km ²	Positive	S7
		Urban industrial sulfur dioxide emissions per unit area	ton/km ²	Positive	S8
		Urban industrial smoke and dust emission per unit area	ton/km ²	Positive	S9
		Urban registered unemployment rate	%	Positive	S10
Industry potential	Annual growth rate of total tourism income	%	Negative	A1	
	Annual growth rate of total tourist arrivals	%	Negative	A2	
Economic vitality	GDP per capita	CNY	Negative	A3	
	GDP growth rate	%	Negative	A4	
	Fixed asset investment per capita	10,000 CNY	Negative	A5	
	Per capita year-end deposit balance of financial institutions	CNY	Negative	A6	
Adaptive capacity	Environmental protection	Green coverage rate in built-up areas	%	Negative	A7
		Urban green space per capita	m ² /person	Negative	A8
		Household harmless garbage disposal rate	%	Negative	A9
		Centralized sewage treatment rate	%	Negative	A10
Public service	Per capita postal revenue	CNY	Negative	A11	
	The number of doctors per 10,000 people	Person	Negative	A12	
	Per capita expenditure in local general public budgets	Yuan	Negative	A13	
	Per capita paved road area at the end of the year	km ²	Negative	A14	
	Number of buses per 10,000 people	Bus	Negative	A15	
	Number of subscribers with broadband internet access	10,000 households	Negative	A16	
	Per capita revenue from telecommunications services	CNY	Negative	A17	

In terms of sensitivity, TEV is not only affected by the core elements within the tourism industry; it is also closely related to the external elements of the tourism industry. According to the viewpoints of scholars, the sensitivity is positively correlated with TEV; that is, the higher the sensitivity, the higher the TEV, and vice versa [23,37]. Therefore, all indicators attribute of sensitivity should be positive. In this study, the sensitivity index of TEV was mainly constructed from two levels of industry core elements (S1–S6) and industry-related elements (S7–S10), including 10 specific indicators. Among them, S1, S2,

S5, and S6 were mainly used to reflect the dependence of urban economic development on the tourism industry; due to the instability of the tourism industry, the higher the dependence proportion, the higher the vulnerability of the urban tourism economy. S3 and S4 mainly reflect the dependence of the urban tourism industry on inbound tourism development. Inbound tourism has many potential uncertainties and is more susceptible to various unexpected factors than domestic tourism. Therefore, the higher the dependency ratio, the higher the vulnerability of the urban tourism economy. S7, S8, and S9 mainly reflect the level of environmental quality of the tourist destination; The higher the pollution level, the higher the vulnerability of the urban tourism economy. S10 mainly reflects the employment situation of tourist cities; if the unemployment rate is higher, it indicates that urban economic development is at a low stage, and the vulnerability of the urban tourism economy is higher.

In terms of adaptive capacity, when the urban tourism economic system is impacted, the development potential of the urban tourism industry and the construction level of the city in terms of economy, ecology, and public services are particularly important for coping with the crisis. According to the viewpoints of scholars, the adaptive capacity is negatively correlated with TEV; that is, the higher the adaptive capacity, the lower the TEV, and vice versa [23,37]. Therefore, the indicators attribute of adaptive capacity should all be negative. In this study, the indicators of the adaptive capacity of the tourism economic system were mainly constructed from four aspects of the industrial potential, economic vitality, environmental protection, and public service of urban tourism, including 17 specific indicators. A1 and A2 reflect the growth capacity of the regional tourism industry and the attraction of urban tourism, respectively; the higher the growth rate of the total tourism income and the total number of tourists received, the stronger the adaptive capacity of the tourism economy and the lower the TEV. A3, A4, A5, and A6 objectively reflect the city's overall economic strength and economic development potential. The higher the GDP per capita, GDP growth rate, fixed asset investment per capita, and per capita year-end deposit balance of financial institutions, the higher the level of urban economic development, and the lower the TEV [36]. A7, A8, A9, and A10 reflect the environmental protection level of the city; the more green space and the higher the garbage and sewage treatment rate, the higher the anti-risk response-ability of the city's tourism economy and the lower the TEV. A11, A12, A13, A14, A15, A16, and A17 reflect the city's public service levels in terms of postal services, medical services, transportation, and communication; the better the public service level, the stronger the city's ability to deal with tourism emergencies and the lower the corresponding TEV.

2.4. Data Sources

The data sources of this study mainly include the following two aspects. First, data on the economy, environment, and public services of 58 major tourist cities in China, from 2004 to 2019, mainly came from the China City Statistical Yearbook (CCSY). CCSY is an annual publication reflecting the social and economic development of Chinese cities. Each issue contains major statistics on the social and economic development of Chinese cities at all levels in the previous year. Detailed statistics of the development data of 58 major tourist cities in China can be found in CCSY. If some of the data could not be found in the CCSY, Statistical Yearbooks (SY) of each tourist city were searched to supplement the data in this study. Second, data on the tourism industry and other aspects of 58 major tourism cities in China, from 2004 to 2019, mainly came from the SY of each city, Statistical Communiqué of National Economic and Social Development (SCNESD), and the Yearbook of China Tourism Statistics (YCTS). In addition, in order to enhance comparability, some of the data were processed by secondary calculations. The data sources for each case city were detailed in Table A1 (Appendix A).

2.5. Research Methods

2.5.1. The Weights of Indicators Were Calculated by the Entropy Weight Method

As an objective weight assignment method, the entropy weight method determines the weight based on the variation degree of the data, which can effectively eliminate the interference of human factors and has strong objectivity and reliability. In view of this, the method in this study was used to calculate the weight of each of the 27 indicators in the TEV index system. The formula for each step of the model is as follows [38,39]:

- (1) Set the original evaluation matrix as:

$$X = (x_{it})_{m \times n} \quad (1)$$

In the formula, x_{it} represents the original value of the t -th index in the i -th sample; $i = 1, 2, \dots, m$, where m is the sample number; $t = 1, 2, \dots, n$; n is the number of indicators. It should be noted that the sample number m in this study is 928, which is composed of 16 years of data (2004–2019) for 58 major tourist cities in China. In addition, the number of indicators n in this study is 27, and they are the indicators in Table 1.

- (2) Standardize the above original evaluation matrix to form a standardized matrix:

$$Y = (y_{it})_{m \times n} \quad (2)$$

where y_{it} represents the standardized value of the t -th indicator in the i -th sample. Among them, the positive indicators are $y_{it} = (x_{it} - x_{min}) / (x_{max} - x_{min})$, and the negative indicators are $y_{it} = (x_{max} - x_{it}) / (x_{max} - x_{min})$.

- (3) Use the entropy weight method to obtain the weights of indicators. The specific calculation formula is as follows:

$$w_t = (1 - E_t) / \left(m - \sum_{t=1}^m E_t \right) \quad (3)$$

$$p_{it} = y_{it} / \sum_{i=1}^n y_{it} \quad (4)$$

$$E_t = -\frac{1}{\ln n} \sum_{i=1}^n p_{it} \ln p_{it} \quad (5)$$

In the formula, i is the sample reference, and t is the indicator reference. p_{it} represents the feature proportion, E_t represents the information entropy, and w_t represents the weight of the t -th indicator.

2.5.2. TOPSIS Model Was Used to Calculate the Values of Sensitivity, Adaptive Capacity, and TEV

On the basis of the indicator weight, the TOPSIS model can be used to calculate the value of each evaluation object. The calculation principle of the TOPSIS model is to calculate the distance between each evaluation object and the optimal (inferior) solution, and then determine the relative approximation degree between the evaluation object and the ideal solution, so as to calculate the evaluation value. In this study, it was introduced to calculate the annual value of the TEV of 58 major tourist cities in China from 2004 to 2009, as well as the values of sensitivity and adaptive capacity of the components of TEV. In this study, the formula of the TOPSIS model used in the calculation of sensitivity, adaptive capacity, and TEV is the same, with the only differences being the indicator types. The calculation formula for each step of the TOPSIS model is as follows [38,39]:

- (1) Construct the weighting matrix:

$$S = Y \times W_t \quad (6)$$

where Y is the matrix obtained after standardized processing in the entropy weight method mentioned above, and W_t is the weight of indicator t .

- (2) Determine the optimal solution S_i^+ and the worst solution S_i^- for the t -th indicator:

$$\begin{aligned} S_i^+ &= \max\{S_{1t}, S_{2t}, \dots, S_{mt}\} \\ S_i^- &= \min\{S_{1t}, S_{2t}, \dots, S_{mt}\} \end{aligned} \quad (7)$$

- (3) Calculate the Euclidean distance between the optimal (inferior) solution and the positive (negative) ideal solution. i is the sample reference, and t is the indicator reference:

$$R_i^+ = \sqrt{\sum_{t=1}^n (S_i^+ - S_{it})^2}; R_i^- = \sqrt{\sum_{t=1}^n (S_i^- - S_{it})^2} \quad (8)$$

- (4) Calculate the proximity C_i :

$$C_i = \frac{R_i^-}{R_i^+ + R_i^-} \quad (9)$$

In the formula, i is the sample reference. The C_i value is within (0, 1). The higher the value of C_i is, the better the evaluation object, and vice versa.

It should be noted that when calculating sensitivity, the indicator t in the formula contains S1–S10, a total of 10 indicators. When calculating adaptive ability, the indicator t in the formula includes A1–A17, a total of 17 indicators. When calculating TEV, the indicator t in the formula includes S1–S10 and A1–A17, a total of 27 indicators.

2.5.3. The Main Factors Affecting TEV Were Detected by the Obstacle Diagnosis Model

The obstacle diagnosis model can effectively analyze and identify the obstacles that affect the development level of the regional system elements and has been widely used in many fields. In this study, two problems can be clarified by introducing the obstacle degree model. First, it is clear which of the 27 indicators of TEV have a major impact on TEV. Second, the obstacle factors affecting TEV in different cities are clearly different. The formula is as follows [40]:

$$M_{it} = \frac{R_t \times P_{it}}{\sum_{t=1}^n (R_t \times P_{it})} \times 100\% \quad (10)$$

In the formula, i is the sample reference, and t is the indicator reference. M_{it} is the obstacle degree of the t -th indicator to the ecological tourism security in i samples; R_t is the weight of each indicator, representing the contribution degree of the obstacle factors. $P_{it} = 1 - y_{it}$ represents the deviation between indicators and development goals, and y_{it} is the standardized value of each indicator. In addition, it should be noted that there may be deviations in evaluation results caused by accidental factors in a single year. Therefore, for the diagnosis results of the obstacle factors in 58 major tourist cities in China, the 16-year average, from 2004 to 2019, was used to obtain the diagnosis results.

2.5.4. The Evolution Trends of TEV Were Predicted by the BP Neural Network Model

- (1) Model setting

The BP neural network, also known as the error-back propagation neural network, has been developed into the most important and widely used artificial neural network algorithm owing to its advantages such as flexible structure design, multiple training algorithms, and good operability [41]. The structure of the BP neural network is a multilayer forward neural network, with an input layer, several hidden layers, and an output layer (Figure 3). Neural networks are connected by links, each of which has a weight. Weight is the basic form of the neural network, and artificial neurons learn by constantly adjusting these weights. The process of a neural network involves the following steps [42]. The first is the selection framework; the second is deciding what kind of learning algorithm to use. Finally, the neural network is trained, which involves initializing the weight of the network and changing the weight value through a series of training steps.

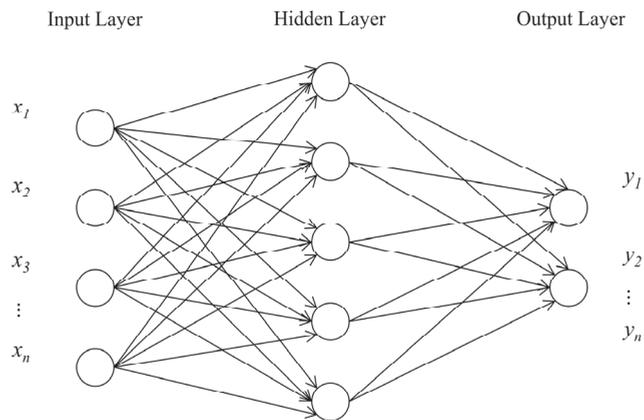


Figure 3. Architecture of the BP neural network.

(2) Model building

The BP neural network with a three-layer structure was adopted. The input variable is the year corresponding to TEV index, the middle is the hidden layer, and the output variable is TEV index. The number of neurons in the hidden layer was determined by experiments. According to the number of neurons in the input layer and the output layer, the number of neurons in the hidden layer was tentatively determined as 8–12. By comparing the prediction errors of different hidden layer networks, the number of hidden layer neurons was finally set as 10.

(3) Initial data processing and parameter setting

To prevent neurons from reaching the saturation state, the sample data were first normalized. MATLAB programming was used to normalize the sample data to the interval of 0–1, according to the positive and negative properties of the indicators. These data were taken as the input, and the standardized TEV was taken as the output data to form a training sample for the BP neural network. When the transfer function of the intermediate layer is an S-shaped tangent function, and the transfer function of the output layer is a linear function, the prediction result is optimized. Considering that the function trainlm converges quickly, and the training error of the network is relatively small, the LM algorithm was selected for training, the maximum training times were set as 1000, the target accuracy was set as 0.0001, and the learning rate was set as 0.01.

(4) Model training and testing

The data of 58 major tourist cities in China, from 2004 to 2019, were trained separately. During the training, the sample data were randomly divided into two groups according to the proportions of 80% and 20% and used as training and test data, respectively. Figure 4 shows the regression accuracy of the neural network model. The correlation coefficient R^2 of the test samples was higher than 0.95, and the average error rate was 1.49%, showing a good fitting effect. Therefore, the neural network model can be used to better predict the TEV of China's major tourist cities from 2021 to 2030.

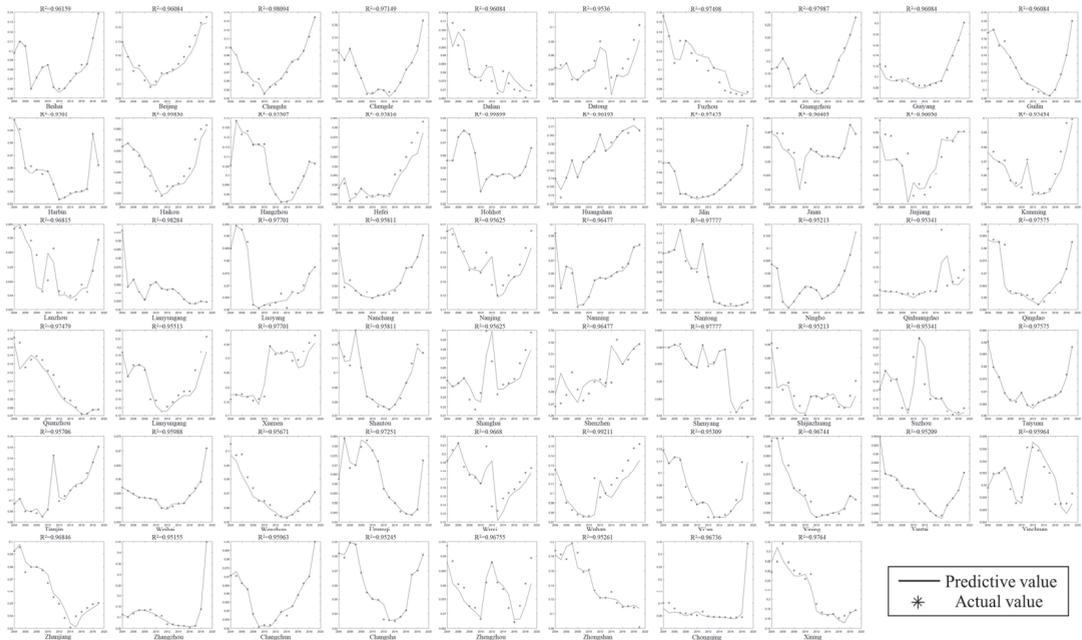


Figure 4. Accuracy tests of the BP neural network model.

3. Results

3.1. Spatiotemporal Evolution of Urban TEV

3.1.1. The Evaluation of Each Indicator Weight in Urban TEV

In this study, the weights of 27 indicators were calculated by using the entropy weight method, and the calculation results are shown in Table 2. Among them, S1–S10 are the indicators of the “sensitivity” part of the TEV, A1–A17 are indicators of “adaptive capacity” part of the TEV.

Table 2. The Weight of each indicator.

Serial Number	Weight	Serial Number	Weight	Serial Number	Weight
S1	0.0603	S10	0.0580	A9	0.0006
S2	0.0006	A1	0.0003	A10	0.0411
S3	0.1984	A2	0.0002	A11	0.0007
S4	0.0868	A3	0.0010	A12	0.0036
S5	0.0514	A4	0.0045	A13	0.0011
S6	0.1340	A5	0.0004	A14	0.0013
S7	0.1290	A6	0.0010	A15	0.0005
S8	0.1043	A7	0.0016	A16	0.0003
S9	0.1168	A8	0.0014	A17	0.0008

3.1.2. Spatiotemporal Changes of Urban TEV

Figure 5 shows the sensitivity dimension of the urban TEV. During the period from 2004 to 2007, cities with high sensitivity values were mainly distributed in economically developed regions such as Shanghai, Suzhou, Shenzhen, Zhuhai, and Guangzhou in China’s Yangtze River Delta and Pearl River Delta. In addition, Tianjin, located in Northern China, has a high sensitivity value. During the period from 2008 to 2011, the sensitivity values of all cities in this stage were basically similar to those in the previous stage, and only a few cities’ sensitivity values changed. For example, Shanghai was added as one of

the cities with the highest sensitivity values, whereas the sensitivity of Tianjin declined in this stage. During the period from 2012 to 2015, among the cities with high sensitivity values, the value of Zhuhai declined, whereas the value of Taiyuan rose sharply and became one of the cities with the highest sensitivity values. In addition, the sensitivity values of Huangshan, Qinhuangdao, and other cities rose to a higher level. During the period from 2016 to 2019, the cities with the highest sensitivity were Shanghai, Xiamen, Zhangzhou, and Shenzhen, among which the sensitivity value of Zhangzhou increased the most. In addition, compared with the previous stage, the sensitivity of Chongqing has also been greatly improved.

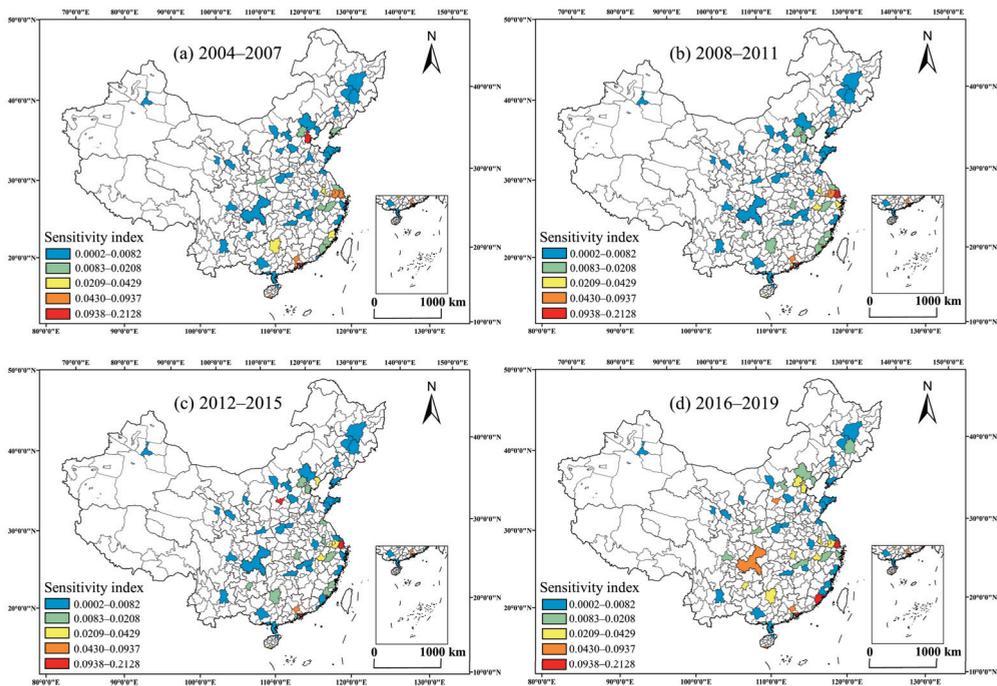


Figure 5. Spatiotemporal evolution of the sensitivity index ((a) Average sensitivity index for 2004–2007 were shown; (b) Average sensitivity index for 2008–2011 were shown; (c) Average sensitivity index for 2012–2015 were shown; (d) Average sensitivity index for 2016–2019 were shown).

Figure 6 shows the dimension of adaptive capacity of the urban TEV. During the period from 2004 to 2007, Harbin, Jilin, Chengde, Xining, Luoyang, Chongqing, Changsha, Huangshan, Guiyang, Quanzhou, and Nanning had the highest adaptive capacities. During the period from 2008 to 2011, the adaptive capacity of all cities as a whole declined significantly. Meanwhile, at this stage, Urumqi, Xining, Wenzhou, Guiyang, Beihai, Shantou, and other cities had the highest adaptive capacity. During the period from 2012 to 2015, the adaptive capacity of all cities decreased further on the whole. At this stage, the adaptive capacity of Lianyungang, Nanjing, and Xining were at their highest level. Compared with the previous stage, the adaptive capacity of many cities decreased significantly. During the period from 2016 to 2019, the adaptive capacity of all cities decreased further compared with the previous period. Only Harbin and Nanjing had a high level of adaptive capacity. The adaptive capacity of Lianyungang, Nanning, Xining, and other cities decreased significantly in this stage.

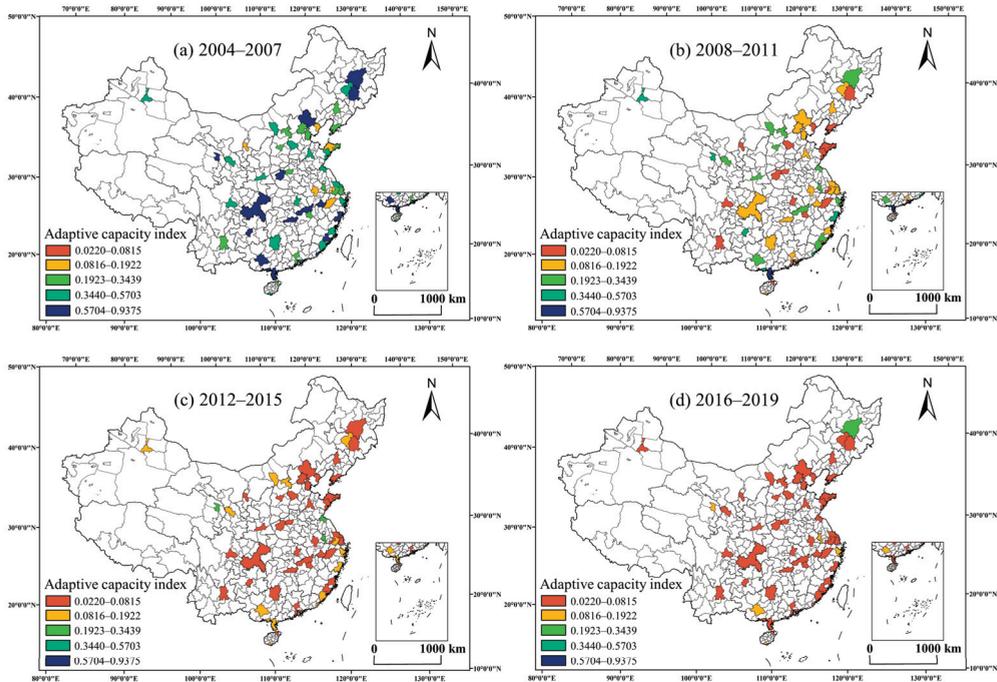


Figure 6. Spatiotemporal evolution of the adaptive capacity index ((a) Average adaptive capacity index for 2004–2007 were shown; (b) Average adaptive capacity index for 2008–2011 were shown; (c) Average adaptive capacity index for 2012–2015 were shown; (d) Average adaptive capacity index for 2016–2019 were shown).

As shown in Figure 7, During the period from 2004 to 2007, cities with a high TEV were mainly distributed in the economically developed areas along the eastern coast of China, including Shanghai, Shenzhen, Zhuhai, Guangzhou, Zhongshan, and Tianjin. During the period from 2008 to 2011, Shanghai, Shenzhen, and Zhuhai still had the highest TEV, while the TEVs of the tourist cities in central and western China generally decreased. During the period from 2012 to 2015, the coastal cities of Shanghai, Xiamen, and Shenzhen had the highest TEV. The TEV of Ningbo, Fuzhou, Quanzhou, Zhangzhou, Shantou, and other coastal cities had decreased. The TEVs in the Middle and western regions of China generally had little change, and only the TEV in Taiyuan increased. During the period from 2016 to 2019, the TEV in Jilin, Chengde, Datong, Chengdu, Chongqing, Guilin, Beihai, and other cities increased significantly compared with the previous stage.

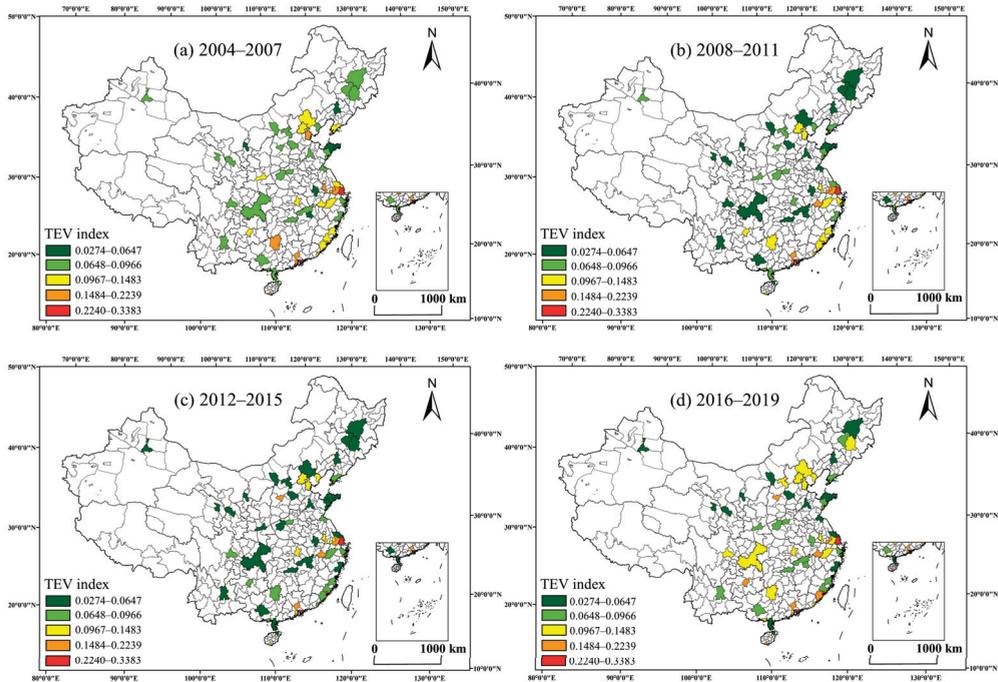


Figure 7. Spatiotemporal evolution of the TEV index ((a) Average TEV index for 2004–2007 were shown; (b) Average TEV index for 2008–2011 were shown; (c) Average TEV index for 2012–2015 were shown; (d) Average TEV index for 2016–2019 were shown).

3.2. Obstacle Factor Diagnosis of TEV

Considering the space limitation, we only screened out the top five main obstacle factors of each city for display and explanation (Table 3).

Table 3 shows that in the obstacle factors of the TEV of China's major tourist cities, S3 (proportion of international tourists out of total tourists), S6 (tourism output density), S8 (urban-industrial sulfur dioxide emissions per unit area), S9 (urban-industrial smoke and dust emission per unit area), and S7 (discharge of urban industrial wastewater per unit area) are the five indicators with the highest occurrence frequency. Thus, the five factors were the top five most critical factors affecting TEV values. In all 58 major tourist cities in China, S3 was the greatest obstacle factor affecting TEV, and the obstacle degree of all the other cities was above 0.2 except for Shenzhen, Taiyuan, and Zhuhai. The second obstacle factor of most cities was S6, and the obstacle degree was between 0.1234 and 0.1650. However, the second obstacle factor of a few cities, such as Guangzhou, Xiamen, Shanghai, and Shenzhen, was S9, and the obstacle degree was between 0.1276 and 0.1362. The third obstacle factors affecting TEV in 58 major tourist cities were S8 and S9. Among these 58 cities, S8 was the obstacle factor in 27 cities, and the obstacle degree was between 0.0989 and 0.1308. By comparison, S9 was the obstacle factor in 31 cities, and the obstacle degree was between 0.1109 and 0.1276. The fourth obstacle factor affecting TEV was similar to the third obstacle factor, and the fourth obstacle factors in most cities were mainly S8 and S9. In 27 cities, the fourth obstacle factor was S8, and the obstacle degree ranged from 0.1033 to 0.1252. In 22 cities, the fourth obstacle factor was S9, and the obstacle degree ranged from 0.1124 to 0.1259. In addition, heterogeneity was observed in some cities. For example, the fourth obstacle factor in Guangzhou, Xiamen, and Shanghai was S6, with the obstacle degree being between 0.0886 and 0.1228; whereas the fourth obstacle factor in Guiyang, Luoyang, Ningbo, Qinhuangdao, Shenzhen, and Taiyuan was S7, with an

obstacle degree between 0.0987 and 0.1116. The fifth obstacle factor of 49 cities was S7, and the obstacle degree ranged from 0.0838 to 0.1196. In addition, the distribution of the fifth obstacle factor in other cities was scattered, among which the fifth obstacle factor in Guiyang, Luoyang, Ningbo, and Taiyuan was S8, and the obstacle degree ranged from 0.1034 to 0.1097. The fifth obstacle factor in Suzhou and Wuxi was S8, and the obstacle degrees were 0.0892 and 0.0957, respectively. The fifth obstacle factor in Xiamen was S10 (urban registered unemployment rate), and the obstacle degree was 0.0720. The fifth obstacle factor in Shenzhen was S6, and the obstacle degree was 0.0952. The fifth obstacle factor in Qinhuangdao was S9, and the obstacle degree was 0.1077.

3.3. Prediction of the Evolution Trend of Urban TEV in the Next 10 Years

In this study, the vulnerability indices of the tourism economy from 2004 to 2019 were taken as sample data and imported into the trained network model to obtain the vulnerability values of the tourism economy from 2021 to 2030; Figure 8 shows the results. During the period of 2021–2030, although the TEV of many major tourist cities in China increases year by year, the cities with low TEV levels still occupy the dominant position. In this period, the cities with high TEV levels will be Shenzhen, Xiamen, Shanghai, and Zhuhai. These cities are all located in the eastern coastal zone of China, and the average values of their TEV will be 0.2911, 0.2621, 0.2510, and 0.2092, respectively. Low-level TEV cities are mostly concentrated in the northeast and western regions of China, such as Yinchuan, Lanzhou, Harbin, and Hohhot, and the average TEV are 0.0310, 0.0483, 0.0513, and 0.0531, respectively. In general, the TEV of high-level and low-level regions differ greatly, indicating that the TEV of major tourist cities in China have strong spatial heterogeneity during this period. The cities with high TEV are mostly distributed in the Yangtze River Delta and Pearl River Delta urban agglomerations along the eastern coast of China, whereas the cities with low TEV are scattered in the northeast, central, and western regions of China. This spatial feature is similar to the existing situation explored above.

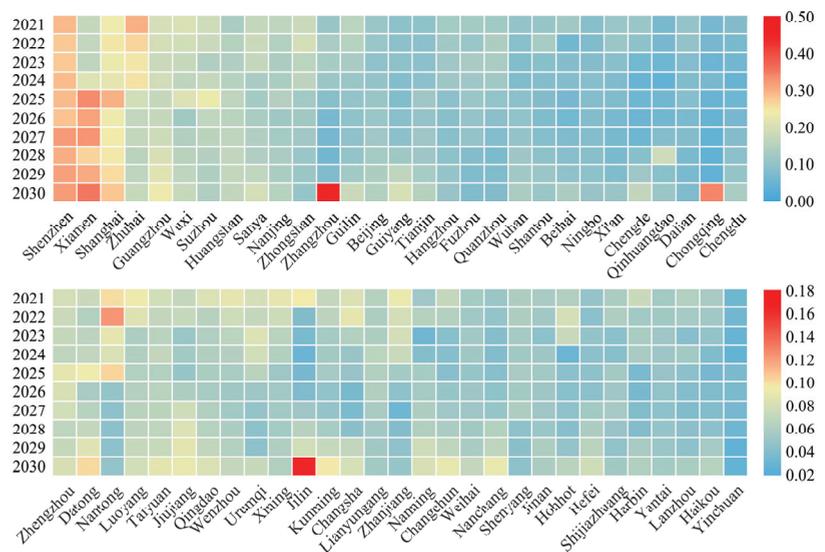


Figure 8. Predicted index of TEV in 2021–2030.

Table 3. Obstacle factors and obstacle degree of TEV of major tourist cities in China.

City	Level 1	Level 2	Level 3	Level 4	Level 5	City	Level 1	Level 2	Level 3	Level 4	Level 5
Beihai	S3(0.2238)	S6(0.1445)	S9(0.1146)	S8(0.1137)	S7(0.1076)	Qingdao	S3(0.2215)	S6(0.1408)	S9(0.1159)	S8(0.1136)	S7(0.1065)
Beijing	S3(0.2295)	S6(0.1234)	S8(0.1216)	S9(0.1203)	S7(0.1117)	Quanzhou	S3(0.2255)	S6(0.1517)	S9(0.1191)	S8(0.1189)	S7(0.1046)
Chengdu	S3(0.2257)	S6(0.1369)	S8(0.1164)	S9(0.1161)	S7(0.1047)	Sanya	S3(0.2328)	S6(0.1455)	S9(0.1308)	S9(0.1259)	S7(0.1119)
Chengde	S3(0.2238)	S6(0.1528)	S8(0.1201)	S9(0.1171)	S7(0.1112)	Xiamen	S3(0.2370)	S9(0.1276)	S8(0.1107)	S6(0.0989)	S10(0.0720)
Dalian	S3(0.2268)	S6(0.1474)	S9(0.1185)	S8(0.1170)	S7(0.1006)	Shantou	S3(0.2208)	S6(0.1382)	S9(0.1155)	S8(0.1083)	S7(0.0965)
Datong	S3(0.1508)	S6(0.1508)	S9(0.1149)	S8(0.1132)	S7(0.1098)	Shanghai	S3(0.2528)	S9(0.1301)	S8(0.0989)	S6(0.0886)	S7(0.0850)
Fuzhou	S3(0.2229)	S6(0.1497)	S9(0.1171)	S8(0.1162)	S7(0.1106)	Shenzhen	S3(0.2215)	S6(0.1445)	S9(0.1135)	S8(0.1134)	S7(0.1068)
Guangzhou	S3(0.2110)	S9(0.1288)	S8(0.1229)	S6(0.1228)	S7(0.1063)	Shijiazhuang	S3(0.1948)	S9(0.1362)	S8(0.1285)	S7(0.0987)	S6(0.0952)
Guiyang	S3(0.2304)	S6(0.1384)	S9(0.1177)	S7(0.1107)	S8(0.1097)	Suzhou	S3(0.2250)	S6(0.1481)	S9(0.1143)	S8(0.1093)	S7(0.1034)
Guilin	S3(0.2220)	S6(0.1565)	S8(0.1245)	S9(0.1212)	S7(0.1148)	Taiyuan	S3(0.2421)	S6(0.1422)	S9(0.1244)	S8(0.1053)	S4(0.0892)
Harbin	S3(0.2209)	S6(0.1484)	S8(0.1180)	S9(0.1141)	S7(0.1086)	Tianjin	S3(0.1902)	S6(0.1477)	S9(0.1175)	S7(0.1116)	S8(0.1095)
Haikou	S3(0.2157)	S6(0.1400)	S8(0.1173)	S9(0.1133)	S7(0.1063)	Weihai	S3(0.2183)	S6(0.1360)	S9(0.1217)	S8(0.1111)	S7(0.1082)
Hangzhou	S3(0.2344)	S6(0.1480)	S8(0.1243)	S9(0.1243)	S7(0.1002)	Wenzhou	S3(0.2182)	S6(0.1413)	S9(0.1135)	S8(0.1123)	S7(0.1063)
Hefei	S3(0.2204)	S6(0.1407)	S8(0.1154)	S9(0.1132)	S7(0.1060)	Urumqi	S3(0.2223)	S6(0.1446)	S8(0.1162)	S9(0.1157)	S7(0.1065)
Hohhot	S3(0.2196)	S6(0.1468)	S8(0.1132)	S9(0.1130)	S7(0.1079)	Wuxi	S3(0.2163)	S6(0.1485)	S9(0.1139)	S8(0.1133)	S7(0.1079)
Huangshan	S3(0.2314)	S6(0.1583)	S8(0.1290)	S9(0.1245)	S7(0.1178)	Wuhan	S3(0.2338)	S6(0.1252)	S9(0.1161)	S8(0.1033)	S4(0.0957)
Jilin	S3(0.2242)	S6(0.1498)	S8(0.1183)	S9(0.1150)	S7(0.1078)	Xi'an	S3(0.2302)	S6(0.1303)	S9(0.1175)	S8(0.1122)	S7(0.1008)
Jinan	S3(0.2201)	S6(0.1405)	S9(0.1119)	S8(0.1103)	S7(0.1046)	Xining	S3(0.2258)	S6(0.1418)	S9(0.1174)	S8(0.1163)	S7(0.1068)
Jiujiang	S3(0.2248)	S6(0.1502)	S8(0.1174)	S9(0.1161)	S7(0.1087)	Yantai	S3(0.2209)	S6(0.1475)	S9(0.1119)	S8(0.1101)	S7(0.1066)
Lanzhou	S3(0.2205)	S6(0.1468)	S9(0.1156)	S8(0.1153)	S7(0.1093)	Yancheng	S3(0.2208)	S6(0.1449)	S9(0.1145)	S8(0.1127)	S7(0.1063)
Lianyungang	S3(0.2185)	S6(0.1460)	S8(0.1127)	S9(0.1124)	S7(0.1062)	Zhanjiang	S3(0.2158)	S6(0.1452)	S9(0.1109)	S8(0.1104)	S7(0.1037)
Luoyang	S3(0.2228)	S6(0.1463)	S8(0.1154)	S9(0.1144)	S7(0.1063)	Zhangzhou	S3(0.2186)	S6(0.1479)	S8(0.1160)	S9(0.1141)	S7(0.1066)
Nanchang	S3(0.2256)	S6(0.1476)	S9(0.1155)	S7(0.1091)	S8(0.1090)	Changchun	S3(0.2170)	S6(0.1607)	S8(0.1276)	S9(0.1243)	S7(0.0928)
Nanjing	S3(0.2197)	S6(0.1414)	S8(0.1144)	S9(0.1127)	S7(0.1021)	Changsha	S3(0.2212)	S6(0.1451)	S8(0.1169)	S9(0.1138)	S7(0.1083)
Nanning	S3(0.2353)	S6(0.1309)	S9(0.1195)	S8(0.1105)	S7(0.0910)	Zhengzhou	S3(0.2199)	S6(0.1411)	S8(0.1161)	S9(0.1139)	S7(0.1072)
Nanning	S3(0.2234)	S6(0.1480)	S8(0.1182)	S9(0.1152)	S7(0.1075)	Zhongshan	S3(0.2245)	S6(0.1372)	S9(0.1125)	S8(0.1052)	S7(0.1018)
Nantong	S3(0.2227)	S6(0.1473)	S9(0.1146)	S8(0.1121)	S7(0.1010)	Chongqing	S3(0.2103)	S6(0.1413)	S9(0.1162)	S8(0.1085)	S7(0.0838)
Ningbo	S3(0.2278)	S6(0.1412)	S9(0.1183)	S7(0.1041)	S8(0.1034)	Zhuhai	S3(0.2342)	S6(0.1650)	S9(0.1272)	S8(0.1252)	S7(0.1181)
Qinhuangdao	S3(0.2261)	S6(0.1479)	S8(0.1115)	S7(0.1084)	S9(0.1077)		S3(0.1987)	S6(0.1454)	S9(0.1276)	S8(0.1150)	S7(0.1060)

4. Discussion

4.1. Internal Logic of Spatiotemporal Evolution of TEV

The tourism industry is highly sensitive due to location variability, complexity, and comprehensiveness. Under the influence of various factors, such as economy, society, and nature, TEV has formed significant regional differences [43]. During the study period, the cities with high TEV values are mainly distributed in the eastern region of China, with Shanghai, Shenzhen, Zhuhai, and other economically developed cities as typical representatives. These cities are located in the center of China's economy, with convenient transportation and frequent business and trade exchanges at home and abroad. Owing to the high proportion of the regional tourism industries scale and a large number of inbound tourists, the regional tourism economy faces a higher risk of external impact, to some extent; thus, presenting strong vulnerabilities [44].

The cities with low TEV in China are widely distributed in the northeast, central, and western regions, and they are characterized by a contiguous distribution. First, Harbin, Jilin, and Changchun in northeast China are important old industrial bases. In recent years, the development speed of tourism has been slow compared with that of other regional central cities with developed tertiary industries. In addition, due to the remote geographical location, fewer long-distance tourists, weak ability to earn foreign exchange in tourism, and low dependence on the tourism industry, the cities show a low TEV [36]. Second, cities such as Urumqi, Yinchuan, and Lanzhou are located in the underdeveloped areas in the west of China; thus, the level of social and economic development is relatively weak. In addition, the status of the local tourism industry is not outstanding, and tourism visibility and attraction are not high. As a result, the development level of the tourism industry is low, the industrial correlation is not strong, and the tourism economy is weak [23]. Finally, due to the geographical location, ecological environment, and socioeconomic characteristics of the central Chinese cities, the multiplier effect and ripple effect of the tourism industry are relatively weak, and they do not occupy a dominant position in the economic structure so they exhibit low TEV levels. The above analysis shows that the vulnerability of China's tourism economy generally still follows the distribution characteristics dominated by the economy, which echoes the previous research conclusions to a certain extent [23].

Overall, During the period from 2004 to 2011, the TEV of most tourism cities showed a decreasing trend year by year. At this stage, the tourism industry has not yet formed a perfect system, the growth of the tourism market is flat, and the tourism economy has not formed enough scale to cause strong economic sensitivity. Moreover, the tourism incentive policy accelerates the influx of tourism enterprises and the construction of tourism facilities, which makes the growth rate of regional tourism economic strain capacity higher than the sensitivity of the tourism economy [45]. During the period from 2012 to 2019, the TEV of most tourism cities showed a slight upward trend, which is closely related to the imbalance of industrial structure caused by the rapid growth of the tourism industry and the external dependence caused by international tourism income.

4.2. Obstacle Factors Affecting TEV

On the basis of the obstacle diagnosis model, this study measures the obstacle factors that affect TEV. The results show that the proportion of international tourists out of total tourists is the most influential factor, which is mainly due to the many unstable factors in the international environment, such as natural disasters, economic crisis, and social unrest. A series of factors may have a strong impact on inbound tourism and impact the whole tourism economic system [46,47]. Tourism output value density is the second major factor affecting TEV. According to Sun, due to the high sensitivity and low resistance of the tourism economic system to internal and external environmental disturbances, compared with other industries, it is very easy to lose the original structure, state, and functional attributes of the tourism economic system in internal and external environmental disturbances, thus leading to a fluctuating and unstable state [48]. The high density of the

tourism industry in the local area will magnify this inherent defect to a certain extent [49]. In addition, urban-industrial sulfur dioxide emissions per unit area, urban industrial smoke and dust emission per unit area, and discharge of urban industrial wastewater per unit area of the three environmental factors have a great impact on TEV. This reflects that the development of the tourism industry has higher requirements for the local ecological environment, which is consistent with the views of Fei et al. [50]. At the same time, it also highlights the characteristics of the tourism economy as an “eco-socioeconomic” composite system [51]. The three are interdependent, adapt to each other, and penetrate, blend, and interact in their development [52]. How to realize the coordinated development of ecology, culture, and economy is a subject to be discussed in the future.

4.3. Trend Prediction of TEV

The prediction accuracy of the BP neural network model established in this study is more than 95%. Therefore, the prediction method proposed in this study is applicable to the development analysis of China’s urban TEV and can provide an important theoretical basis for the development and decision-making of the tourism industry. According to the evolution trend, TEV values in China’s major tourist cities will continue to show an increasing trend in the next 10 years. However, the rise of the TEV will obviously bring many adverse effects, so how to reasonably regulate TEV to achieve sustainable development of the tourism economy is an urgent issue to be discussed at present. For some scholars, TEV is accumulated by two forms of environmental stress: endogenous and exogenous [22]. The endogenous vulnerability factors are formed by the activities of the tourism economic system, including the irrationality of the internal structure of the tourism market structure, tourism income structure, tourism organization structure, tourism investment structure, and tourism product structure. Exogenous environmental stress is the abrupt change and gradual change of external environmental factors of the tourism economic system, such as the political environment, economic environment, natural environment, and tourism industry policy. This understanding means that to reduce TEV, we need to start from two aspects of internal structure optimization and external policy regulation.

5. Conclusions

At present, the rapid development of China’s tourism industry plays an evident role in promoting economic and social development. However, due to the inherent nature of the tourism industry, it is vulnerable to the impact of the internal and external environment. Therefore, promoting the sustainable development of the regional tourism economy is an objective requirement to evaluate and forecast the TEV in major tourist cities. Using panel data from 2004 to 2019, a comprehensive evaluation index system for TEV was constructed in this study, which used 58 major tourist cities in China as the research objects. The TEV was reasonably measured by using entropy weight method, TOPSIS model, obstacle diagnosis model, and BP neural network model. Finally, the spatiotemporal pattern, obstacle factors, and future trend of TEV were discussed.

The contributions of this study to the literature are as follows. Limited by the difficulty of obtaining statistical data, existing studies mostly compare the state of TEV in different cities from a horizontal perspective, and there is a lack of studies on the evolution process and mechanism of TEV in different cities from a vertical perspective [36]. In this study, panel data of a longer time scale were used to predict the evolutionary trend of TEV in the future, which can provide scientific reference for different tourism cities to formulate targeted tourism economic development policies. In addition, prediction is the basis of decision-making, but the traditional statistical methods have a strong assumption of the data distribution law, so the problem of random interference in the economic system has not been addressed [53]. In the prediction of the TEV, the existing time series analysis method can only reflect the linear law with a strong tendency and cannot describe the nonlinear characteristics. In this study, BP neural network was used to build a prediction model, which can excavate and predict the regularity of time series indicators. The application of

this method not only enriches the research system of vulnerability methods but also has reference significance for other fields.

The findings of this study have several practical implications for the development of the urban tourism economy. First, in terms of the spatiotemporal pattern of evolution, cities with high TEV are mostly distributed in the eastern coastal urban agglomerations of China, while cities with low TEV are scattered in the northeastern, central, and western regions of China. This is the result of tourism industry dependence and is closely related to location, economy, nature, and other factors. Therefore, promoting the coordination of urban infrastructure, industrial structure, and the ecological environment should become an important measure of urban construction. Second, this study found that the five obstacle factors that have the greatest impact on the vulnerability of the urban tourism economy are the proportion of international tourists out of total tourists, tourism output density, urban industrial sulfur dioxide emissions per unit area, urban industrial smoke and dust emission per unit area, and discharge of urban industrial wastewater per unit area. This shows that accelerating the adjustment of economic structure and the transformation of economic mode [54], as well as the purification and discharge of waste gas, centralized treatment of hazardous waste and wastewater, and the improvement of tourists' awareness of environmental protection, should become key issues to reduce TEV. Third, in the next 10 years, TEV of major tourist cities in China will increase, which is the result of the accumulation of endogenous structure and the stress of the exogenous environment. In order to effectively reduce the vulnerability of tourism economic development, we can adjust the orientation of the tourism industry development, highlight the driving effect of tourism association, and build a multi-pillar industry system.

Although this study measured and analyzed the spatiotemporal evolution, obstacle factors, and future trends of the TEV of China's major tourist cities, it has limitations. The index system of this research is constructed on the basis of the general characteristics of all typical tourist cities in the dataset. However, due to the large area of China, cities in different regions in the natural environment, and social-cultural differences, the index system will ignore the heterogeneity between different cities, leading to uncertain factors. Future research may construct an inter-city differentiated index evaluation system according to the unique properties of each city. Such an evaluation system may make the measurement results more accurate. In addition, on the basis of the BP neural network, this study prefigured the time series evolution of the future TEV of China's major tourist cities well. The overall model shows high precision, but some cases show a poor-fitting effect. Scholars point out that combination-prediction is better than single prediction [55]. Therefore, building a variety of prediction models for comparison, such as a GM (1,1) prediction model, linear regression prediction model, and time series prediction model supplemented by the BP neural network model to make decisions, presents a promising direction for future research.

Author Contributions: Conceptualization, Deping Chu; methodology, Chengkun Huang and Feiyang Lin; software, Chengkun Huang and Feiyang Lin; validation, Chengkun Huang, Feiyang Lin, and Deping Chu; formal analysis, Chengkun Huang; investigation, Chengkun Huang, Feiyang Lin, Lanlan Wang, and Jiawei Liao; data curation, Chengkun Huang, Feiyang Lin, Lanlan Wang, and Jiawei Liao; writing—original draft preparation, Chengkun Huang and Feiyang Lin; writing—review and editing, Chengkun Huang, Feiyang Lin, Deping Chu, and Junqian Wu; visualization, Chengkun Huang and Feiyang Lin; supervision, Deping Chu and Junqian Wu; project administration, Deping Chu and Junqian Wu; funding acquisition, Deping Chu. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Fujian Province, China (grant number 2018J01743).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

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

Appendix A

Table A1. Data sources for each case city.

City	Data Source	City	Data Source
Beihai	CCSY 2005–2020; Beihai SY 2005–2020; Beihai SCNESD 2004–2019; YCTS 2005–2018	Qingdao	CCSY 2005–2020; Qingdao SY 2005–2020; Qingdao SCNESD 2004–2019; YCTS 2005–2018
Beijing	CCSY 2005–2020; Beijing SY 2005–2020; Beijing SCNESD 2004–2019; YCTS 2005–2018	Quanzhou	CCSY 2005–2020; Quanzhou SY 2005–2020; Quanzhou SCNESD 2004–2019; YCTS 2005–2018
Chengdu	CCSY 2005–2020; Chengdu SY 2005–2020; Chengdu SCNESD 2004–2019; YCTS 2005–2018	Sanya	CCSY 2005–2020; Sanya SY 2005–2020; Sanya SCNESD 2004–2019; YCTS 2005–2018
Chengde	CCSY 2005–2020; Chengde SY 2005–2020; Chengde SCNESD 2004–2019; YCTS 2005–2018	Xiamen	CCSY 2005–2020; Xiamen SY 2005–2020; Xiamen SCNESD 2004–2019; YCTS 2005–2018
Dalian	CCSY 2005–2020; Dalian SY 2005–2020; Dalian SCNESD 2004–2019; YCTS 2005–2018	Shantou	CCSY 2005–2020; Shantou SY 2005–2020; Shantou SCNESD 2004–2019; YCTS 2005–2018
Datong	CCSY 2005–2020; Datong SY 2005–2020; Datong SCNESD 2004–2019; YCTS 2005–2018	Shanghai	CCSY 2005–2020; Shanghai SY 2005–2020; Shanghai SCNESD 2004–2019; YCTS 2005–2018
Fuzhou	CCSY 2005–2020; Fuzhou SY 2005–2020; Fuzhou SCNESD 2004–2019; YCTS 2005–2018	Shenyang	CCSY 2005–2020; Shenyang SY 2005–2020; Shenyang SCNESD 2004–2019; YCTS 2005–2018
Guangzhou	CCSY 2005–2020; Guangzhou SY 2005–2020; Guangzhou SCNESD 2004–2019; YCTS 2005–2018	Shenzhen	CCSY 2005–2020; Shenzhen SY 2005–2020; Shenzhen SCNESD 2004–2019; YCTS 2005–2018
Guiyang	CCSY 2005–2020; Guiyang SY 2005–2020; Guiyang SCNESD 2004–2019; YCTS 2005–2018	Shijiazhuang	CCSY 2005–2020; Shijiazhuang SY 2005–2020; Shijiazhuang SCNESD 2004–2019; YCTS 2005–2018
Guilin	CCSY 2005–2020; Guilin SY 2005–2020; Guilin SCNESD 2004–2019; YCTS 2005–2018	Suzhou	CCSY 2005–2020; Suzhou SY 2005–2020; Suzhou SCNESD 2004–2019; YCTS 2005–2018
Harbin	CCSY 2005–2020; Harbin SY 2005–2020; Harbin SCNESD 2004–2019; YCTS 2005–2018	Taiyuan	CCSY 2005–2020; Taiyuan SY 2005–2020; Taiyuan SCNESD 2004–2019; YCTS 2005–2018
Haikou	CCSY 2005–2020; Haikou SY 2005–2020; Haikou SCNESD 2004–2019; YCTS 2005–2018	Tianjin	CCSY 2005–2020; Tianjin SY 2005–2020; Tianjin SCNESD 2004–2019; YCTS 2005–2018
Hangzhou	CCSY 2005–2020; Hangzhou SY 2005–2020; Hangzhou SCNESD 2004–2019; YCTS 2005–2018	Weihai	CCSY 2005–2020; Weihai SY 2005–2020; Weihai SCNESD 2004–2019; YCTS 2005–2018
Hefei	CCSY 2005–2020; Hefei SY 2005–2020; Hefei SCNESD 2004–2019; YCTS 2005–2018	Wenzhou	CCSY 2005–2020; Wenzhou SY 2005–2020; Wenzhou SCNESD 2004–2019; YCTS 2005–2018
Hohhot	CCSY 2005–2020; Hohhot SY 2005–2020; Hohhot SCNESD 2004–2019; YCTS 2005–2018	Urumqi	CCSY 2005–2020; Urumqi SY 2005–2020; Urumqi SCNESD 2004–2019; YCTS 2005–2018
Huangshan	CCSY 2005–2020; Huangshan SY 2005–2020; Huangshan SCNESD 2004–2019; YCTS 2005–2018	Wuxi	CCSY 2005–2020; Wuxi SY 2005–2020; Wuxi SCNESD 2004–2019; YCTS 2005–2018
Jilin	CCSY 2005–2020; Jilin SY 2005–2020; Jilin SCNESD 2004–2019; YCTS 2005–2018	Wuhan	CCSY 2005–2020; Wuhan SY 2005–2020; Wuhan SCNESD 2004–2019; YCTS 2005–2018
Jinan	CCSY 2005–2020; Jinan SY 2005–2020; Jinan SCNESD 2004–2019; YCTS 2005–2018	Xi'an	CCSY 2005–2020; Xi'an SY 2005–2020; Xi'an SCNESD 2004–2019; YCTS 2005–2018
Jiujiang	CCSY 2005–2020; Jiujiang SY 2005–2020; Jiujiang SCNESD 2004–2019; YCTS 2005–2018	Xining	CCSY 2005–2020; Xining SY 2005–2020; Xining SCNESD 2004–2019; YCTS 2005–2018
Kunming	CCSY 2005–2020; Kunming SY 2005–2020; Kunming SCNESD 2004–2019; YCTS 2005–2018	Yantai	CCSY 2005–2020; Yantai SY 2005–2020; Yantai SCNESD 2004–2019; YCTS 2005–2018
Lanzhou	CCSY 2005–2020; Lanzhou SY 2005–2020; Lanzhou SCNESD 2004–2019; YCTS 2005–2018	Yinchuan	CCSY 2005–2020; Yinchuan SY 2005–2020; Yinchuan SCNESD 2004–2019; YCTS 2005–2018
Lianyungang	CCSY 2005–2020; Lianyungang SY 2005–2020; Lianyungang SCNESD 2004–2019; YCTS 2005–2018	Zhanjiang	CCSY 2005–2020; Zhanjiang SY 2005–2020; Zhanjiang SCNESD 2004–2019; YCTS 2005–2018
Luoyang	CCSY 2005–2020; Luoyang SY 2005–2020; Luoyang SCNESD 2004–2019; YCTS 2005–2018	Zhangzhou	CCSY 2005–2020; Zhangzhou SY 2005–2020; Zhangzhou SCNESD 2004–2019; YCTS 2005–2018
Nanchang	CCSY 2005–2020; Nanchang SY 2005–2020; Nanchang SCNESD 2004–2019; YCTS 2005–2018	Changchun	CCSY 2005–2020; Changchun SY 2005–2020; Changchun SCNESD 2004–2019; YCTS 2005–2018
Nanjing	CCSY 2005–2020; Nanjing SY 2005–2020; Nanjing SCNESD 2004–2019; YCTS 2005–2018	Changsha	CCSY 2005–2020; Changsha SY 2005–2020; Changsha SCNESD 2004–2019; YCTS 2005–2018
Nanning	CCSY 2005–2020; Nanning SY 2005–2020; Nanning SCNESD 2004–2019; YCTS 2005–2018	Zhengzhou	CCSY 2005–2020; Zhengzhou SY 2005–2020; Zhengzhou SCNESD 2004–2019; YCTS 2005–2018
Nantong	CCSY 2005–2020; Nantong SY 2005–2020; Nantong SCNESD 2004–2019; YCTS 2005–2018	zhongshan	CCSY 2005–2020; zhongshan SY 2005–2020; zhongshan SCNESD 2004–2019; YCTS 2005–2018
Ningbo	CCSY 2005–2020; Ningbo SY 2005–2020; Ningbo SCNESD 2004–2019; YCTS 2005–2018	Chongqing	CCSY 2005–2020; Chongqing SY 2005–2020; Chongqing SCNESD 2004–2019; YCTS 2005–2018
Qinhuangdao	CCSY 2005–2020; Qinhuangdao SY 2005–2020; Qinhuangdao SCNESD 2004–2019; YCTS 2005–2018	Zhuhai	CCSY 2005–2020; Zhuhai SY 2005–2020; Zhuhai SCNESD 2004–2019; YCTS 2005–2018

References

1. Timmerman, P. *Vulnerability, Resilience and the Collapse of Societies: A Review of Models and Possible Climatic Applications*; Institute for Environmental Studies Toronto, University of Toronto: Toronto, ON, Canada, 1981.
2. Yamano, H.; Kayanne, H.; Yamaguchi, T.; Kuwahara, Y.; Yokoki, H.; Shimazaki, H.; Chikamori, M. Atoll island vulnerability to flooding and inundation revealed by historical reconstruction: Fongafale Islet, Funafuti Atoll, Tuvalu. *Glob. Planet. Chang.* **2007**, *57*, 407–416. [[CrossRef](#)]
3. Toni, F.; Holanda, E. The effects of land tenure on vulnerability to droughts in Northeastern Brazil. *Glob. Environ. Chang.* **2008**, *18*, 575–582. [[CrossRef](#)]

4. Wilson, K.; Newton, A.; Echeverría, C.; Weston, C.; Burgman, M. A vulnerability analysis of the temperate forests of south central Chile. *Biol. Conserv.* **2005**, *122*, 9–21. [[CrossRef](#)]
5. Ho, H.C.; Knudby, A.; Chi, G. Spatiotemporal analysis of regional socio-economic vulnerability change associated with heat risks in Canada. *Appl. Geogr.* **2018**, *95*, 61–70. [[CrossRef](#)] [[PubMed](#)]
6. Liu, W.; Li, L.; Chen, L.; Wen, M.; Wang, J.; Yuan, L.; Liu, Y.; Li, H. Testing a comprehensive volcanic risk assessment of tenerife by volcanic hazard simulations and social vulnerability analysis. *ISPRS Int. Geo-Inf.* **2020**, *9*, 273. [[CrossRef](#)]
7. Susan, L.; Cutter, B.J.; Boruff, W.; Lynn, S. Social vulnerability to environmental hazards. *Soc. Sci. Q.* **2003**, *84*, 242–261. [[CrossRef](#)]
8. Adrianto, L.; Matsuda, Y. Developing economic vulnerability indices of environmental disasters in small island regions. *Environ. Impact Assess. Rev.* **2002**, *22*, 393–414. [[CrossRef](#)]
9. Wim, N.; Amelia, S.-P.; Mark, M. Measuring vulnerability: An overview and introduction. *Oxf. Dev. Stud.* **2009**, *37*, 183–191. [[CrossRef](#)]
10. Altimari, A.; Balzano, S.; Zezza, G. Measuring economic vulnerability: A structural equation modeling approach. In *Statistical Learning of Complex Data*; Greselin, F., Deldossi, L., Bagnato, L., Vichi, M., Eds.; Springer: Cham, Switzerland, 2019; pp. 95–102.
11. Briguglio, L. Small island developing states and their economic vulnerabilities. *World Dev.* **1995**, *23*, 1615–1632. [[CrossRef](#)]
12. Briguglio, L.; Cordina, G.; Farrugia, N. Economic vulnerability and resilience: Concepts and measurements. *Oxf. Dev. Stud.* **2009**, *37*, 229–247. [[CrossRef](#)]
13. Chan, V.K.Y. The Impact of the Global Financial Crisis on the Entertainment Tourism Industry: A Financial Engineering Case Study of Macao from 2007 to 2010. *Syst. Eng. Procedia* **2011**, *1*, 323–329. [[CrossRef](#)]
14. Zhou, B.; Zhang, Y.; Zhou, P. Multilateral political effects on outbound tourism. *Ann. Touris. Res.* **2021**, *88*, 103184. [[CrossRef](#)]
15. Kılıçlar, A.; Uşaklı, A.; Tayfun, A. Terrorism prevention in tourism destinations: Security forces vs. civil authority perspectives. *J. Destin. Mark. Manag.* **2018**, *8*, 232–246. [[CrossRef](#)]
16. Altuntas, F.; Gok, M.S. The effect of COVID-19 pandemic on domestic tourism: A DEMATEL method analysis on quarantine decisions. *Int. J. Hosp. Manag.* **2021**, *92*, 102719. [[CrossRef](#)]
17. Rosselló, J.; Becken, S.; Santana-Gallego, M. The effects of natural disasters on international tourism: A global analysis. *Tour. Manag.* **2020**, *79*, 104080. [[CrossRef](#)]
18. Croes, R.; Ridderstaat, J.; Bak, M.; Zientara, P. Tourism specialization, economic growth, human development and transition economies: The case of Poland. *Tour. Manag.* **2021**, *82*, 104181. [[CrossRef](#)]
19. Wang, Z.D. Dynamic relationship between tourism, economic growth, and environmental quality. *J. Sustain. Tour.* **2018**, *26*, 1928–1943. [[CrossRef](#)]
20. Canh, N.P.; Thanh, S.D. Domestic tourism spending and economic vulnerability. *Ann. Tour. Res.* **2020**, *85*, 103063. [[CrossRef](#)]
21. Su, F.; Chen, Y.; Zhang, P. Vulnerability assessment of tourism city's economic system based on the set pair analysis: A case study of Zhoushan city. *Sci. Geogr. Sin.* **2013**, *33*, 538–544. (In Chinese)
22. Li, F. The fragibility of tourism economy: Definition, formation mechanism and frame analysis. *East China Econ. Manag.* **2013**, *27*, 76–81. (In Chinese)
23. Ma, H.Q.; Lian, Q.W.; Lun, Y.C. Spatial differentiation of tourism economic system vulnerability based on BP neural network in different provinces of China. *Res. Sci.* **2019**, *41*, 2248–2261. (In Chinese) [[CrossRef](#)]
24. Huang, L.; Yin, X.; Yang, Y.; Luo, M.; Huang, S. “Blessing in disguise”: The impact of the Wenchuan earthquake on inbound tourist arrivals in Sichuan, China. *J. Hosp. Tour. Manag.* **2020**, *42*, 58–66. [[CrossRef](#)]
25. Pham, T.D.; Dwyer, L.; Su, J.-J.; Ngo, T. COVID-19 impacts of inbound tourism on Australian economy. *Ann. Tour. Res.* **2021**, *88*, 103179. [[CrossRef](#)]
26. Backer, E.; Ritchie, B. VFR Travel: A viable market for tourism crisis and disaster recovery? *Int. J. Tour. Res.* **2017**, *19*, 400–411. [[CrossRef](#)]
27. Gurtner, Y. Returning to paradise: Investigating issues of tourism crisis and disaster recovery on the island of Bali. *J. Hosp. Tour. Manag.* **2016**, *28*, 11–19. [[CrossRef](#)]
28. Raki, A.; Nayer, D.; Nazifi, A.; Alexander, M.; Seyfi, S. Tourism recovery strategies during major crises: The role of proactivity. *Ann. Tour. Res.* **2021**, 103144. [[CrossRef](#)]
29. Moghal, Z.; O'Connell, E. Multiple stressors impacting a small island tourism destination-community: A nested vulnerability assessment of Oistins, Barbados. *Tour. Manag. Perspect.* **2018**, *26*, 78–88. [[CrossRef](#)]
30. Yin, P.; Liu, S.; Duan, P. Analysis on the vulnerability and obstacle indicators in island-type tourism destination—Take Zhoushan city for example. *Ecol. Geogr.* **2017**, *37*, 234–240. (In Chinese)
31. Jia, Y.; Hu, J.; Xie, S.; Qiao, H.; Liu, D. Vulnerability and influence mechanisms of social-ecological system in poor mountainous tourism destinations. *Hum. Geogr.* **2021**, *36*, 155–164. (In Chinese)
32. Liang, Z.; Xie, L. On the vulnerability of economic system of traditional tourism cities—A case from Guilin. *Tour. Trib.* **2011**, *26*, 40–46. (In Chinese)
33. Yang, Y.; Wang, R. Research on vulnerability assessment of tourism economic system and the path optimization in border ethnic Region: A case study of Korean autonomous prefecture of Yanbian. *Areal Res. Dev.* **2015**, *34*, 104–109. (In Chinese)
34. Huang, C.; Lin, F.; Chu, D.; Wang, L.; Liao, J.; Wu, J. Coupling relationship and interactive response between intensive land use and tourism industry development in China's major tourist cities. *Land* **2021**, *10*, 697. [[CrossRef](#)]

35. Polsky, C.; Neff, R.; Yarnal, B. Building comparable global change vulnerability assessments: The vulnerability scoping diagram. *Glob. Environ. Chang.* **2007**, *17*, 472–485. [[CrossRef](#)]
36. Su, F.; Chu, Y.; Zhang, P. Analysis of economic vulnerability and obstacle factors of typical tourism cities in China. *Ecol. Geogr.* **2013**, *33*, 189–194. (In Chinese)
37. Tian, L.; Tian, Y.; Yang, Y. Study on tourism economic system's vulnerability assessment of Dali Prefecture based on TOPSIS. *Res. Dev. Mark.* **2017**, *33*, 1529–1534. (In Chinese)
38. Chen, Y.; Zhu, M.; Lu, J.; Zhou, Q.; Ma, W. Evaluation of ecological city and analysis of obstacle factors under the background of high-quality development: Taking cities in the Yellow River Basin as examples. *Ecol. Indic.* **2020**, *118*, 106771. [[CrossRef](#)]
39. Li, J.; Chen, Y.; Yao, X.; Chen, A. Risk management priority assessment of heritage sites in China based on entropy weight and TOPSIS. *J. Cult. Herit.* **2021**, *49*, 10–18. [[CrossRef](#)]
40. Wang, D.; Shen, Y.; Zhao, Y.; He, W.; Liu, X.; Qian, X.; Lv, T. Integrated assessment and obstacle factor diagnosis of China's scientific coal production capacity based on the PSR sustainability framework. *Resour. Policy* **2020**, *68*, 101794. [[CrossRef](#)]
41. Li, S.; Wang, Y.; Hao, M.; Jiang, D.; Xun, Z.; Zhou, R. Site selection of digital signage in Beijing: A combination of machine learning and an empirical approach. *ISPRS Int. Geo-Inf.* **2020**, *9*, 217. [[CrossRef](#)]
42. Wang, W.; Tang, R.; Li, C.; Liu, P.; Luo, L. A BP neural network model optimized by Mind Evolutionary Algorithm for predicting the ocean wave heights. *Ocean Eng.* **2018**, *162*, 98–107. [[CrossRef](#)]
43. Broshi-Chen, O.; Mansfeld, Y. A wasted invitation to innovate? Creativity and innovation in tourism crisis management: A QC&IM approach. *J. Hosp. Tour. Manag.* **2021**, *46*, 272–283. [[CrossRef](#)]
44. Gnanngnon, S.K. Trade openness and structural vulnerability in developing countries. *J. Ecol. Stud.* **2016**, *43*, 70–89. [[CrossRef](#)]
45. Deng, T.; Hu, Y.; Ma, M. Regional policy and tourism: A quasi-natural experiment. *Ann. Tour. Res.* **2019**, *74*, 1–16. [[CrossRef](#)]
46. Wang, Y.-S. The impact of crisis events and macroeconomic activity on Taiwan's international inbound tourism demand. *Tour. Manag.* **2009**, *30*, 75–82. [[CrossRef](#)]
47. Hanon, W.; Wang, E. Comparing the impact of political instability and terrorism on inbound tourism demand in Syria before and after the political crisis in 2011. *Asia Pac. J. Tour. Res.* **2020**, *25*, 651–661. [[CrossRef](#)]
48. Sun, G. Some responses to the doubt of whether tourism develops with certain rules. *Tour. Sci.* **2008**, *22*, 1–4. (In Chinese)
49. Espiner, S.; Becken, S. Tourist towns on the edge: Conceptualising vulnerability and resilience in a protected area tourism system. *J. Sustain. Tour.* **2014**, *22*, 646–665. [[CrossRef](#)]
50. Fei, J.; Lin, Y.; Jiang, Q.; Jiang, K.; Li, P.; Ye, G. Spatiotemporal coupling coordination measurement on islands' economy-environment-tourism system. *Ocean Coast. Manag.* **2021**, *212*, 105793. [[CrossRef](#)]
51. Tang, F.; Tao, W. Establish a composite system for sustainable tourism development. *J. Dongbei Univ. Financ. Econ.* **2001**, 28–30. (In Chinese)
52. Jia, Y.; Hu, J.; Liu, D. Spatial-temporal pattern of the coordinated development of ecology-culture-tourism of China's provinces. *World Reg. Stud.* **2021**, *30*, 620–631. (In Chinese)
53. Pan, D.; Shi, C. The forecasting method and application of the total retail sales of consumer goods in China. *Stat. Decis.* **2015**, 96–98. (In Chinese)
54. Demir, E.; Gozgor, G.; Paramati, S.R. To what extent economic uncertainty effects tourism investments? Evidence from OECD and non-OECD economies. *Tour. Manag. Perspect.* **2020**, *36*, 100758. [[CrossRef](#)]
55. Zou, F.; Zou, D. A study on forecast of ecotourism demand in Hunan province based on the comparison of five models. *Res. Environ. Yangtze Basin* **2020**, *29*, 2406–2417. (In Chinese)

Article

Why Is Green Hotel Certification Unpopular in Taiwan? An Analytic Hierarchy Process (AHP) Approach

Yen-Cheng Chen ¹, Ching-Sung Lee ^{2,*}, Ya-Chuan Hsu ² and Yin-Jui Chen ²

¹ Department of Applied Science of Living, Chinese Culture University, Taipei 11114, Taiwan; cyc4@g.pccu.edu.tw

² Department of Restaurant, Hotel and Institutional Management, Fu Jen Catholic University, New Taipei City 242062, Taiwan; gueylin.hotel@msa.hinet.net (Y.-C.H.); stay.hotel88@msa.hinet.net (Y.-J.C.)

* Correspondence: 038614@mail.fju.edu.tw

Abstract: The main purpose of this study was to investigate the factors that discouraged Taiwan hoteliers from applying for green hotel certification. The analytic hierarchy process (AHP) method was used to perform a weighted analysis that comprehensively identified important hindering factors based on information from hotel industry, government, academic, and consumer representatives. Overall, in order of importance, the five dimensions of hindering factors identified by these experts and scholars were hotel internal environment, consumers' environmental protection awareness, environmental protection incentive policy, hotel laws and regulations policy, and hotel external environment. Among the 26 examined hindering factor indices, the three highest-weighted indices overall for hoteliers applying for green hotel certification were as follows: environmental protection is not the main consideration of consumers seeking accommodations, lack of support by investment owners (shareholders), and lack of relevant subsidy incentives. The major contribution of this study is that hoteliers can understand important hindering factors associated with applying for green hotel certification; therefore, strategies that can encourage or enhance the green certification of hotels can be proposed to improve corporate image in the hotel industry, implement social responsibility in this industry, and obtain consumers' approval of and accommodation-willingness for green hotels.

Keywords: green hotel; corporate social responsibility; green hotel certification

Citation: Chen, Y.-C.; Lee, C.-S.; Hsu, Y.-C.; Chen, Y.-J. Why Is Green Hotel Certification Unpopular in Taiwan? An Analytic Hierarchy Process (AHP) Approach. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 255. <https://doi.org/10.3390/ijgi10040255>

Academic Editors: Andrea Marchetti and Wolfgang Kainz

Received: 30 January 2021

Accepted: 5 April 2021

Published: 10 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Green business practices have become very popular with the wave of green and sustainable issues in recent years. Throughout the world, enterprises are adopting a variety of environmentally sustainable activities while managing their business operations [1,2]. One motivation for these changes is that many individuals and corporate customers consider the company's sustainable environmental performance when making purchasing decisions [3,4]. Of course, other reasons also exist, such as government supervision requirements, social responsibility requirements, and mandatory implementation of green practices in enterprises [5–8].

Taiwan is an island area composed of Taiwan Island and 121 small islands. Development on islands is affected by their remoteness, limited natural resources, small markets, marginal decision-making centers, unique internal structure, and vulnerability to natural disasters. The islands of various countries in the world, especially small islands, are geographically isolated from the mainland, resulting in differences in climate, topography, and physical environment, and each has its own natural and cultural characteristics. Taiwan is surrounded by the sea; its fishery resources are rich, the ecological environment is well preserved, and the natural landscape is dominated by ocean features, which constitute the greatest attractions of island tourism. Taiwan has unique natural resources and is an island with the potential to develop international ecological tourism. To help Taiwan move

towards green and sustainable development, this research analyzed expert data to identify the obstacles to applications for green hotel certification.

The hotel industry used to be regarded as a chimney-free industry [9–11]. However, with the increased range and level of services, energy consumption, amounts of waste and wastewater, and chemical emissions now have a considerable impact on the environment [12,13]. Most studies on green hotels in Taiwan areas have examined green hotel-related issues from the perspectives of consumers, whereas the difficulties in obtaining green hotel certification from the perspective of the hotelier have rarely been discussed [14,15]. Therefore, how hotels in Taiwan area make green changes and how they reduce energy consumption and damage to the environment, as well as the difficulties encountered in applying for green hotel certification, remain important and unanswered issues in the sustainable operation of green hotels in Taiwan.

Sustainability is currently one of the major priorities of tourism all over the world. One aspect of sustainable tourism is green management. According to the trends in the hotel industry, accommodation facilities have recognized that adopting green practices is beneficial [16]. The Environmental Protection Administration of Taiwan launched the “Green Mark” for the hotel industry in 2008 in response to the global demand for environmental protection and sustainable development; however, there were very few applicants. Therefore, the government re-launched the “Green Hotels” project with lower thresholds for environmental protection conditions in 2011 to shorten and reduce the review time and investment and increase the application willingness of hoteliers. However, by the end of February 2019, among 13,268 hotels in the tourism industry, only 1450 hotels, less than 10%, became green hotels [17]. There are only 64 hotels with a green mark certification [18]. This result indicates that hoteliers in Taiwan generally hold a reserved attitude towards green hotels, worry about the investment and service quality of green facilities, do not know how to proceed, or even do not understand the market benefit. What are the important factors that hinder the application for green hotel mark certification, and why is the number of hotels that obtain green mark certification not increasing as expected?

To address these questions, this study used the analytic hierarchy process (AHP) method to analyze the level of importance of each factor hindering the application for green hotel certification based on questionnaire surveys of industrial experts, government unit experts, scholars, and consumers. Finally, quantitative ranking of the level of importance of each hindering factor of application for green hotel mark certification, identified through the expert questionnaire surveys, was performed to provide a reference to aid in increasing the number of applications for green hotel certification in the future.

2. Background and Related Works

2.1. Impact of Data Science and Geographical Location on Taiwan’s Green Certified Hotels

In recent years, Taiwanese governmental units have integrated tourism and information technology, promoted sustainable green tourism information services, integrated and established a “Taiwan Green Certificate Hotel Database” [18], and promoted tourism business models. At the same time, mastering the development trend of cloud technology combined with social media and mobile technology and gradually integrating various artificial intelligence tourism services is an important policy of Taiwan’s tourism official unit. The O2O (online to offline) model has been widely used in Taiwan’s tourism industry. Green hotels are trying to use a variety of online channels, including online travel agency (OTA) and hotel websites, to show their green certification to attract customers to their hotels. The OTA website provides certified green hotel geographical distribution information and marketing activities, helps introduce new customers to the green hotel, and provides information to enrich the OTA website (Figure 1). Therefore, applying for green hotel certification and attracting guests through OTA have become top priorities for hotel operators in Taiwan.



Figure 1. Green Hotel OTA website in Taiwan.

Factors affecting the location of green hotels include traffic conditions, geographical conditions, natural landscapes, and geographical location, and they have a certain impact on green hotels [19]. Popovic et al. [20] pointed out that the hotel's location should be considered as follows: geographical environment: beautiful and comfortable climate, cultural attraction, recreational opportunities, and surrounding environmental characteristics; accessibility: convenient transportation; natural limitations: topography and slope, hydrology, geology, plants and wildlife; environmental management: urban area division, building regulations, comfort and convenience, current land use, restrictions on future land and building changes. Fang et al. [21] found that hotel location factors include following: transaction advantages: the hotel's location close to the tourist destination; landscape factors: the landscape and public facilities near the area; convenience: including time and distance, highways, and railway connectivity; hotel environment: appropriateness of the surrounding environment.

Therefore, most hotels with green certification in Taiwan are concentrated in metropolitan areas (Table 1). Additionally, a city's geographical location will become an important key factor for green hotel certification because of the relevant subsidy incentives of cities in Taiwan, strict environmental laws and regulations for geographical locations, regional restrictions in cities, and land cost considerations.

Table 1. Distribution of Green Hotel Certificates in Taiwan.

Geographical Location	Certificates Level	Total
Urban	Gold	10
	Silver	15
	Copper	17
Country	Gold	4
	Silver	8
	Copper	10

Source: Environmental Protection Administration Executive Yuan, R.O.C. (Taiwan) [18].

2.2. Development of Green Hotel Certification

In the hotel and tourism field, consumers' understanding of the sustainable development of hotels is also increasing [6,16,22,23]. Therefore, hotel operators and managers

also recognize that hotels should actively participate in sustainable operations and environmental protection to attract customers with increasing green consciousness [24,25]. According to the global sustainable travel report released by [Bookings.com](#), 65% of global travelers expressed their intention to stay in green hotels [26]. In response to the higher expectations of these environmentally conscious consumers regarding “green accommodations”, many hotels have actively adopted green and sustainable environmental protection practices [16,23,27–29]. Most hotel operators have a positive awareness of green hotels and environmentally friendly labels and have a willingness to implement them. They agree that green hotels can help improve the hotel’s image and energy-saving effects and are willing to implement environmental protection measures to help reduce hotel costs [30–32].

Green hotel certification is intended to provide a series of environmentally friendly standards and encourage the hotel industry to increase its environmental performance. Of approximately 140 green certification institutions for hotels, 50 eco-labels focus on green hotel certification [33]. The green hotel certification schemes vary; however, the majority of certifications include the following components: water, energy, waste, sustainable procurement, biodiversity conservation, community engagement, and architecture and design [34]. Reviewing the green hotel certification program is very important to understanding the key structural components of hotel environmental management.

The well-known Hilton hotels launched the corporate responsibility plan, “Travel with Purpose”, in 2011. This plan includes not only social impacts but also environmental impacts to focus on effectively reducing energy, water, and waste output in environmental management [35]. TripAdvisor provides information about hotel environmental practices. The company negotiated with international expert organizations for sustainable development, the United Nations Environment Programme, and the International Tourism Organization to develop the TripAdvisor GreenLeaders program. The GreenLeaders program includes 6 components: energy, water, waste, purchasing, site, and innovation and education. The evaluation items include towel and linen reuse, electric car recharge stations, and solar panels. Hotels can apply for the TripAdvisor GreenLeaders program through completion of a self-evaluation survey. Hotels can achieve 1 of the 4 badge levels based on the environmental protection practice level: bronze, silver, gold, or platinum. More than 1000 hotels have obtained the TripAdvisor GreenLeaders award, including all major brands and many independent hotels. In addition, hotel customers can comment on the green practices of hotels to ensure the integrity of the program [36,37].

Green Seal is a nonprofit organization that provides environmental certification standards. These standards represent responsible choices for company purchasers and consumers to promote more effective sustainable development. Green Seal released 33 standards covering 400 product and service categories [38]. Green Seal regulates requirements for hotels and lodging properties. The Green Seal standard GS-33 for lodging properties has three levels: bronze, silver, and gold. Hotels can apply for green certification. Since its first launch in 1999, this green seal standard has become a method to help hotels to improve their environmental practices and develop into environmental leaders. This standard focuses on energy conservation, pollution prevention, waste minimization, water conservation, and freshwater resource management [39].

3. Materials and Methods

This study used the AHP method to analyze important factors that hinder hoteliers from applying for green hotel certification. In the study design, green hotel scholars and experts and hoteliers that already obtained green hotel certifications were first interviewed through meetings. In addition, previous related literature was collected using the literature analysis method to prepare the AHP questionnaire. Next, AHP was performed to categorize and rank important factors that hindered hoteliers from applying for green hotel certification.

3.1. Sampling Design

The sample selection of this study is based on the proportion of geographical distribution, and the invitation process is to invite participants through Email or telephone. As was explained to experts, the survey items were filled in from a professional perspective. Moreover, because most selected experts support the topic of green hotels, they were very enthusiastic to respond to the invitation and make suggestions. Because this research topic was professional, actual surveys were performed based on the research content to target hotel and environmental protection professionals with more than 15 years of experience in four fields—hotel industry professionals (five people), government unit professionals (4 people), academic professionals (four people), and consumer professionals (four people)—to determine the weight of each item. All of the above experts are professionals who are engaged in or have contact with fields closely related to this research topic. The survey was completed in 2020. Expert background information is provided in Table 2 below.

Table 2. Description of Participating Experts by the AHP method.

Items	Category	Professional Background
1	Consumer Representative	Consumer who has stayed in eco-hotels more than 10 times
2	Consumer Representative	General Manager, Housekeeping Supplier
3	Consumer Representative	General Manager, Enterprise Decorators
4	Consumer Representative	Cabin Chief, National
5	Government representative	Executive, Tourism Bureau, M.O.T.C. Republic of China (Taiwan)
6	Government representative	Executive, Tourism Bureau, M.O.T.C. Republic of China (Taiwan)
7	Government representative	Executive, Environmental Protection Administration, Executive Yuan, Republic of China (Taiwan)
8	Government representative	Executive, Department of Tourism
9	Academic Representative	Professor; Fields of research: Leisure and Tourism, Brand Image
10	Academic Representative	Professor; Fields of research: Green Hotel, Human Resources Management in the Hospitality Industry
11	Academic Representative	Professor; Fields of research: Corporate Sustainability, Corporate Social Responsibility
12	Academic Representative	Professor; Fields of research: Strategic Management of Ecotourism, Public Relations
13	Industrial Representative	22 years of experience working in hospitality industry; Director, International Hotel Chain
14	Industrial Representative	20 years of experience working in hospitality industry; the hotel has eco-hotel certification in Taiwan
15	Industrial Representative	16 years of experience working in hospitality industry; the hotel has eco-hotel certification in Taiwan
16	Industrial Representative	15 years of experience working in hospitality industry; the hotel has eco-hotel certification in Taiwan
17	Industrial Representative	15 years of experience working in hospitality industry; the hotel has eco-hotel certification in Taiwan

Before the study was officially conducted, all indices that hindered green hotel certification were explained to the respondents in detail to avoid confusion and to effectively establish the understanding of respondents on each index and their relationship. After invalid questionnaires such as those with missing answers were excluded, consistency statistical verification was performed. The results showed that recovered questionnaires all conformed to the standard of the consistency ratio (CR) value lower than 0.10. Therefore, there were, in total, 17 copies of valid recovered questionnaires. This study used the post-event method to calculate weighted values of each index in valid recovered questionnaires

based on the AHP guidelines. In addition, individual and overall weights were analyzed based on their professional attribute categories to evaluate each index.

3.2. Application of the AHP Method for Indicators That Hindered Green Hotel Certification

The AHP method is mainly applied in uncertain conditions and decision-making issues with many evaluation criteria. The application scope of AHP is very diverse, especially in planning, prediction, judgement, resource allocation, and portfolio trials [40]. AHP analyses and divides complicated questions into several hierarchies to establish a hierarchical structure with mutual influence. It decomposes step-by-step from high hierarchies to low hierarchies. Through quantitative judgement, AHP simplifies and improves the previous decision-making procedures of decision makers who relied on instinct to obtain priority-weighted values of all schemes. The hierarchical relationship can provide a logical approach to evaluation for decision makers to select appropriate schemes. Schemes with higher priority-weighted values have higher priority orders of acceptance. Therefore, the risk of mistakes in decision making is reduced. The procedure of the AHP method is divided into eight steps [41].

1. Decision-making issues are identified, and evaluation indicators are listed.

The definition of the research topic is determined, the scope of the problems is analyzed and defined, and the purpose of decision making is confirmed. Next, opinions of experts and decision makers are integrated. The relevant evaluation criteria of the decision-making problems are listed, the criteria are defined, and the criteria are categorized into different hierarchies.

2. The hierarchical structure is constructed

All viewpoints of decision makers are repeatedly amended, using the group discussion method or referencing relevant literature and expert opinions, and summarized to establish the target-scheme hierarchical structure.

3. Pairwise comparisons are performed for evaluation

After the hierarchical structure is established, the next step is the evaluation task. Indices in the upper hierarchy are mainly used as the baseline. Pairwise comparisons between the relative importance of the index to the upper hierarchy of any two indices at the same hierarchy are performed. If there are n indices, $n(n - 1)/2$ pairwise comparisons should be performed to determine the relative importance among all indices at the same level of the hierarchy.

4. The matrix at each level of the hierarchy is developed according to step 3 to construct all evaluation matrices.

This study targeted sub-hierarchies of all hierarchies to perform pairwise comparisons to obtain all evaluation matrices. Evaluation matrices of all hierarchies are constructed according to the following Formula (1):

$$[A_k] = [a_{ij}] = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{bmatrix}, k = 1, 2, \dots, n, \quad (1)$$

where A_k is the evaluation matrix at each hierarchy;

i is the hierarchical code;

j is the index code;

k is the expert code; and

a is the matrix of each hierarchy.

5. Pairwise comparison matrices are constructed, priority vectors are calculated, and consistency is examined.

According to evaluation data obtained in step 3, a pairwise comparison matrix is constructed and is called the positive reciprocal matrix. Based on the constructed pairwise comparison matrix, the eigenvector and the maximized eigenvalue (λ_{max}) are calculated. Next, the values of differences between the maximized eigenvalue and n indices are converted into the consistency index (CI). The ratio between indices in the evaluation matrix is measured and used as the reference for whether the pairwise comparison matrix is acceptable. The consistency is examined using the following Formula (2):

$$CI = (\lambda_{max} - n) / (n - 1). \quad (2)$$

The consistency index of the randomly produced positive reciprocal matrix is the random index (RI). Using the above CI and RI , the consistency ratio of the pairwise comparison matrix is obtained, $CR = CI/RI$.

6. All hierarchies are subject to steps 3–5 and connected according to each hierarchy.
7. The total priority weight of the overall hierarchy is calculated.

The relative weights of indices in all hierarchies are integrated to calculate the total priority weight of the overall hierarchy. The calculated weight represents the relative priority order of all decision-making schemes corresponding to the decision-making target. This study targeted experts in four individual fields (industry experts, government unit representatives, scholars, and consumers) to separately calculate the index weights of all hierarchies to evaluate the priority order of all indices and determine the quantitative ranking of the levels of importance of hindering factors of application for green hotel mark certification.

8. The consistency of the overall hierarchy is evaluated.

The CR of the overall hierarchy is mainly the consistency index of the hierarchy (CIH) divided by the random index of the hierarchy (RIH). Therefore, the consistency ratio of the overall hierarchy (CRH) should be smaller than 0.10. If this standard is not met, evaluation should be amended again to improve the CR . In summary, the CRH was smaller than 0.10. The AHP questionnaire design of this research is based on the environmental hotel-related literature and a survey of experts (list of experts is shown in Table 1). A total of 26 indicators that affect the application for environmental hotel certification are classified into five major categories: hotel internal environment, hotel external environment, hotel laws and regulations policy, environmental protection incentive policy, and consumers' environmental awareness, as shown in Figure 2.

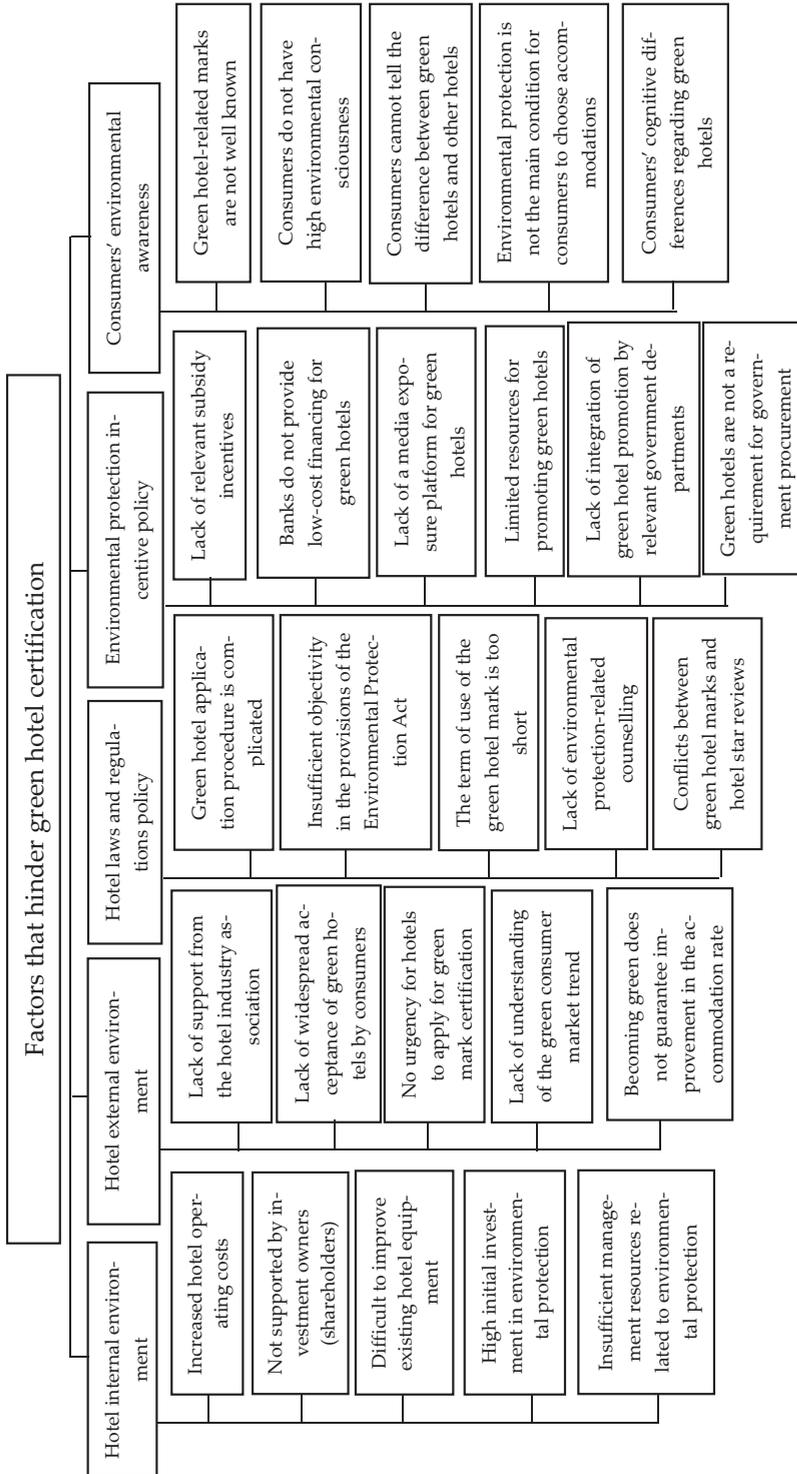


Figure 2. The hierarchical structure of hindering factors of green hotel certification.

4. Results

4.1. AHP Analysis of Hindering Indices of Green Hotel Certification by Scholars and Experts

Table 3 shows that “hotel internal environment” (0.262) was the most important index, followed by “consumers’ environmental awareness” (0.258) and “environmental protection incentive policy” (0.204). These results indicated that scholars and experts overall considered that the hotel internal environment was the index with the highest level of hindrance in green hotel certification.

Table 3. Analysis of hindering indices of application for green hotel certification considered by scholars and experts.

Dimension	Index	Weight	Relative Weight	Ranking
Hotel internal environment 0.262	Increased hotel operating costs	0.181	0.037	11
	Not supported by investment owners (shareholders)	0.345	0.093	2
	Difficult to improve existing hotel equipment	0.189	0.049	6
	High initial investment in environmental protection	0.223	0.060	4
	Hotel has insufficient management resources related to environmental protection	0.063	0.014	24
Hotel external environment 0.106	Lack of support from the hotel industry association	0.097	0.008	26
	Lack of widespread acceptance of green hotels by consumers	0.305	0.037	12
	No urgency for hotels to apply for green mark certification	0.131	0.011	25
	Lack of understanding of the green (environmental protection) consumer market trend	0.119	0.015	23
	Becoming green does not guarantee improvement in the accommodation rate	0.348	0.033	14
Hotel laws and regulations policy 0.170	Green hotel application procedure is complicated	0.311	0.049	7
	Insufficient objectivity in the provisions of the Environmental Protection Act	0.128	0.021	22
	The term of use of the green hotel mark is too short	0.116	0.023	20
	Lack of environmental protection-related counselling mechanisms	0.251	0.042	10
	Conflicts between green hotel marks and hotel star reviews	0.193	0.031	15
Environmental protection incentive policy 0.204	Lack of relevant subsidy incentives	0.322	0.075	3
	Banks do not provide low-cost financing for green hotels	0.112	0.021	21
	Lack of media exposure platform for green hotels	0.121	0.024	19
	Limited resources for promoting green hotels	0.176	0.028	18
	Lack of integration of green hotel promotion by relevant government departments	0.162	0.028	17
	Green hotels are not a requirement for government procurement	0.107	0.031	16
Consumers’ environmental awareness 0.258	Green hotel-related marks are not well known	0.171	0.044	8
	Consumers do not have high environmental consciousness	0.147	0.050	5
	Consumers cannot tell the difference between green hotels and other hotels	0.141	0.035	13
	Environmental protection is not the main condition for consumers to choose accommodations	0.375	0.099	1
	Consumers’ cognitive differences regarding green hotels	0.166	0.043	9

Regarding the index weights within each individual dimension, in the hotel internal environment dimension, “not supported by investment owners (shareholders)” had the highest weight (0.345), followed by “high initial investment in environmental protection” (0.223) and “difficult to improve existing hotel equipment” (0.189). In the hotel external environment dimension, “becoming green does not guarantee improvement in the accommodation rate” had the highest weight (0.348), followed by “lack of widespread acceptance of green hotels by consumers” (0.305) and “no urgency for hotels to apply for green mark certification” (0.131). In the hotel laws and regulations policy dimension, “green hotel application procedure is complicated” had the highest weight (0.311), followed by “lack of environmental protection-related counselling mechanisms” (0.251) and “conflicts between green hotel marks and hotel star reviews” (0.193). In the environmental protection incentive

policy dimension, “lack of relevant subsidy incentives” had the highest weight (0.322), followed by “limited resources for promoting green hotels” (0.176) and “lack of integration of green hotel promotion by relevant government departments” (0.162). In the consumers’ environmental awareness dimension, “environmental protection is not the main condition for consumers to choose accommodations” had the highest weight (0.375), followed by “green hotel-related marks are not well known” (0.171) and “consumers’ cognitive differences regarding green hotels” (0.166).

Furthermore, the relative weights of all indices were determined. Overall, the scholar and expert representatives expressed that “environmental protection is not the main condition for consumers to choose accommodations” (0.099) had the highest relative weight among all indices. This result indicated that these representatives believed that the most critical factor hindering green hotel certification was this index, followed by “not supported by investment owners (shareholders)” (0.093) and “lack of relevant subsidy incentives” (0.075).

4.2. Analysis of the Relative Weighted Ranking of Indices That Hindered Green Hotel Certification Considered by All Scholars and Experts

In the analysis of the relative weighted ranking of indices that hindered green hotel certification, this study ranked the relative weights considered by scholars and experts in all fields and performed comprehensive analyses targeting consumer representatives, government representatives, academic representatives, and industry representatives to understand the concentration level of ranking of all indices. Table 4 shows that the results were mainly divided into: 1. high-ranking concentration indices (indices concentrated in the top five of all fields), 2. low-ranking concentration indices (indices concentrated in the bottom five of all fields), and 3. differential ranking indices (indices ranked in both the top five and the bottom five).

In the high-ranking indices, “environmental protection is not the main condition for consumers to choose accommodations” was a high-ranking concentration index. The results showed that consumer representatives and industry representatives all considered that the relative weight of this index was the overall number 1 in ranking, and academic representatives considered it number 3. The second was “not supported by investment owners (shareholders)”. The results showed that government representatives and academic representatives all considered that the relative weight of this index is the overall number 1 in ranking, and industry representatives considered it number 4. The next was “lack of relevant subsidy incentives”. The results showed that government representatives and academic representatives all considered that the relative weight of this index is the overall number 2 in ranking, and consumer representatives considered it number 3.

In the low-ranking indices, “lack of support from the hotel industry association” was the low-ranking concentration index. The results showed that consumer representatives and academic representatives all considered that the relative weight of this index was last in the overall ranking (26th place), and industry representatives considered it to be second to last (25th place), followed by “lack of understanding of the green (environmental protection) consumer market trend”. The results showed that academic representatives considered that the relative weight of this index was the second to last in the overall ranking (25th place), consumer representatives considered it to be third to last (23rd place), and government representatives considered it to be fourth to last (22nd place). The next was “no urgency for hotels to apply for green mark certification”. The results showed that government representatives considered that the relative weight of this index was second to last in the overall ranking (25th), and consumer representatives considered it to be third to last (24th place).

Table 4. Analysis of rankings of the relative weights of indices that hindered green hotel certification, as indicated by all scholars and experts.

Dimension	Index	Ranking			
		Consumer Representatives	Government Representatives	Academic Representatives	Industry Representatives
Hotel internal environment	Increased hotel operating costs	12	20	7	10
	Not supported by investment owners (shareholders)	21	1	1	4
	Difficult to improve existing hotel equipment	18	5	18	2
	High initial investment in environmental protection	6	6	13	3
Hotel external environment	Hotel has insufficient management resources related to environmental protection	25	21	24	14
	Lack of support from the hotel industry association	26	17	26	25
Hotel laws and regulations policy	Lack of widespread acceptance of green hotels by consumers	14	26	9	5
	No urgency for hotels to apply for green mark certification	24	25	20	17
	Lack of understanding of the green consumer market trend	23	22	25	15
	Becoming green does not guarantee improvement in the accommodation rate	13	16	14	13
Environmental protection incentive policy	Green hotel application procedure is complicated	8	4	21	7
	Insufficient objectivity in the provisions of the Environmental Protection Act	17	14	19	19
	The term of use of the green hotel mark is too short	5	15	22	23
	Lack of environmental protection-related counselling mechanisms	20	3	15	8
Consumers' environmental awareness	Conflicts between green hotel marks and hotel star reviews	2	24	10	20
	Lack of relevant subsidy incentives	3	2	2	16
	Banks do not provide low-cost financing for green hotels	11	11	23	24
	Lack of media exposure platform for green hotels	19	7	16	22
	Limited resources for promoting green hotels	22	9	8	18
	Lack of integration of green hotel promotion by relevant government departments	10	13	12	21
Consumers' environmental awareness	Green hotels are not a requirement for government procurement	4	8	17	26
	Green hotel-related marks are not well known	16	12	5	11
	Consumers do not have high environmental protection consciousness	7	23	4	12
	Consumers cannot tell the difference between green hotels and other hotels	15	18	11	6
Consumers' environmental awareness	Environmental protection is not the main condition for consumers to choose accommodations	1	10	3	1
	Consumers' cognitive differences regarding green hotels	9	19	6	9

In the differential ranking of indices, “green hotels are not a requirement for government procurement” had the largest difference in the ranking of overall indices. The results indicated that consumer representatives considered that the relative weight of this index to be ranked fourth among the overall indices, and industry representatives considered it to be in last place (26th place). The next was “conflicts between green hotel marks and hotel star reviews”. The results indicated that consumer representatives considered that the relative weight of this index overall ranked second, and government representatives considered it to rank third to last (24th place). The next was “Lack of widespread acceptance of green hotels by consumers”. Industry representatives considered that the relative weight of this index ranked fifth among all indices, and government representatives considered it to be ranked last (26th place).

In summary, this study compared the ranking in the fields of consumer representatives, government representatives, academic representatives, and industry representatives and showed that the ranking of the relative weights of some indices had high degrees of concentration, showing that scholars and experts in all fields agreed on the degree of hindrance of that index, whereas there were opposite opinions on some indices. Therefore, green hotel promotion involves different fields and scholars. Experts in all fields had different subjective experiences on green hotel application hindrance and had different opinions on the levels of hindrance.

5. Discussion

Green hotels are a current trend in tourism accommodation, and the distribution of green hotels is related to the environmental protection of tourism-related geography. The indicators revealed in this research can provide a reference for increasing applications for green hotels and illustrate the importance of regional geography to the environmental ecology. This study analyzed five important index dimensions: hotel internal environment, hotel external environment, hotel laws and regulations policy, environmental protection incentive policy, and consumers’ environmental awareness. The study results are summarized and discussed below.

The “not supported by investment owners (shareholders)” index, “difficult to improve existing hotel equipment” index, and “high initial investment in environmental protection” index ranked second, sixth, and fourth, respectively, in the overall ranking of indices by scholars and experts. These results indicated that the thinking of hoteliers and the recognition of hotels’ senior executives were important factors affecting the willingness of hoteliers to participate in green hotel certification. “Not supported by investment shareholders” and “high initial investment in environmental protection” were important factors hindering hoteliers from applying for green marks. This result is consistent with Chan et al. [42], Iorgulescu [43], and Moon et al. [44]. Although the hotel industry has an incentive to invest in environmentally friendly hotels, it is currently adopting a wait-and-see attitude due to input cost considerations and uncertainty in output performance.

These five indices—“lack of support from the hotel industry association”, “lack of widespread acceptance of green hotels by consumers”, “no urgency for hotels to apply for green mark certification”, “lack of understanding of the green (environmental protection) consumer market trend”, and “becoming green does not guarantee improvement in the accommodation rate”—in the analysis of overall ranking by overall scholars and experts were all not in the top 10, whereas the ranking of “hotel external environment dimension” was ranked last among five dimensions.

The “green hotel application procedure is complicated” index ranked 7th in the analysis of overall rankings of indices by scholars and experts. The results indicated that the certification process of many green hotel marks in Taiwan is too complicated, and the content of green mark requirements is not clear. This result is consistent with Suryawan and Aris [45], Nelson et al. [46], and Sharma et al. [47]. Many hoteliers expressed that the label certification procedures are too complicated, and they are not clear about the content of the green label items.

The “lack of relevant subsidy incentives” index ranked third in the analysis of overall ranking of indices by scholars and experts. This result is consistent with Heras-Saizarbitoria et al. [48] and dos Santos et al. [49]. The increase in incentives is a very important planning consideration for groups targeted by the policy. The government should start with subsidies and environmental protection tax incentives to directly encourage hoteliers to implement green hotels, and relevant incentive measures by the government are an important hindering factor affecting hoteliers from applying for green marks.

The “environmental protection is not the main condition for consumers to choose accommodations” and “consumers do not have high environmental protection consciousness” indices ranked first and fifth in the analysis of the overall rankings of indices by scholars and experts, respectively. The results indicated that when consumers choose hotels for accommodations, green hotels will not be listed as a factor in purchasing decisions. Consumers consider price and only have economic motivation [50–52].

6. Conclusions

The “environmental protection is not the main condition for consumers to choose accommodations” index in the analysis of indices by consumer representatives and hotel industry representatives was ranked 1st in both rankings. Therefore, hoteliers understand very clearly that consumers still use the economic factor as the priority in the accommodation selection environment [53]. Hoteliers should target environmental behaviors such as consumers bringing their own toiletries to give substantial price discounts in the practical operations. The “not supported by investment owners (shareholders)” index ranked first in the analysis of indices by government representatives and academic representatives. Therefore, the major decision makers of the hotels are still owners (shareholders). The attitude of owners (shareholders) is an important factor determining whether the promotion of government policy succeeds or fails. It is recommended that government units seek assistance from hotel associations in all counties and cities to promote green hotels and strengthen education about the environment in the ventures undertaken by owners (shareholders). In addition, counselling and promotion can be immediately performed when hotels apply for certification. In addition, hotels planned and established using the standard of green hotels should be opened. Furthermore, when hotels actually obtain green hotel mark certification after formal operations, they can also apply for subsidies to increase the application willingness of hoteliers. The contributions of this research are as follows: through the analysis of the actual data indicators of the tourism experts in this research, we know that the current official government tourism-related units in the Taiwan region recognize the policy effectiveness of the environmentally friendly hotel label system. In discussing the obstacles to the environmental labelling of hotels, this study collected opinions from tourism experts, and the expert data presented by AHP can be used by official tourism organizations and tour operators in Taiwan as a reference, and the impact of various related factors can be explored and prioritized.

7. Limitations and Future Studies

Experts in the fields in this study were limited to the Taiwanese area. It is recommended that researchers in the future expand the sources and categories of experts to make them more representative; thus, the study results will not be influenced by the regions in which the experts are located. This study only targeted the hindering factors of green mark certification in the Taiwanese area. Because the implementation time of green hotels in various countries around the world is earlier than that in Taiwan, it is suggested that subsequent researchers target hindrance of implementation of green hotels in different countries to perform in-depth studies and compare the results with the factors hindering the implementation of green hotels in Taiwan. In addition, we believe that the comparison from the perspectives of hoteliers and the perspectives of consumers would help us better understand the factors affecting the promotion of green hotel marks.

Author Contributions: Conceptualization, Ching-Sung Lee; Data curation, Yin-Jui Chen; Formal analysis, Ching-Sung Lee; Investigation, Ya-Chuan Hsu and Yin-Jui Chen; Methodology, Yen-Cheng Chen and Ching-Sung Lee; Project administration, Yen-Cheng Chen and Ya-Chuan Hsu; Software, Ya-Chuan Hsu; Validation, Yin-Jui Chen; Writing—review & editing, Yen-Cheng Chen. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Science and Technology, Taiwan: 109-2410-H-030-047-.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

- Gast, J.; Gundolf, K.; Cesinger, B. Doing business in a green way: A systematic review of the ecological sustainability entrepreneurship literature and future research directions. *J. Clean. Prod.* **2017**, *147*, 44–56. [CrossRef]
- Sharma, V.K.; Chandna, P.; Bhardwaj, A. Green supply chain management related performance indicators in Agro industry: A review. *J. Clean. Prod.* **2017**, *141*, 1194–1208. [CrossRef]
- Guyader, H.; Ottosson, M.; Witell, L. You can't buy what you can't see: Retailer practices to increase the green premium. *J. Retail. Consum. Serv.* **2017**, *34*, 319–325. [CrossRef]
- Zhang, H.; Yang, F. On the drivers and performance outcomes of green practices adoption. *Ind. Manag. Data Syst.* **2016**, *116*, 2011–2034. [CrossRef]
- Govindan, K.; Diabat, A.; Madan Shankar, K. Analyzing the drivers of green manufacturing with fuzzy approach. *J. Clean. Prod.* **2015**, *96*, 182–193. [CrossRef]
- Verma, V.K.; Chandra, B.; Kumar, S. Values and ascribed responsibility to predict consumers' attitude and concern towards green hotel visit intention. *J. Bus. Res.* **2019**, *96*, 206–216. [CrossRef]
- Miisso, R.; Andreopoulou, Z.; Cesaretti, G.P.; Hanna, S.S.; Tzoulis, I. Sustainable development and green tourism: New practices for excellence in the digital era. *J. Int. Bus. Entrep. Dev.* **2018**, *11*, 65–74. [CrossRef]
- Crespi, G.; Becchio, C.; Buso, T.; Corgnati, S.P. Environmental performances in green labels for hotels—A critical review. In *International Symposium: New Metropolitan Perspectives*; Crespi, G., Becchio, C., Buso, T., Corgnati, S.P., Eds.; Springer: Cham, Switzerland, 2020; pp. 1176–1186.
- Gupta, A.; Dash, S.; Mishra, A. All that glitters is not green: Creating trustworthy ecofriendly services at green hotels. *Tour. Manag.* **2019**, *70*, 155–169. [CrossRef]
- Jiang, Y.; Gao, Y. Factors that influence potential green hotel customers' decision-making process—Evidence from China. *J. China Tour. Res.* **2019**, *15*, 1–23. [CrossRef]
- Yadav, R.; Balaji, M.S.; Jebarajakirthy, C. How psychological and contextual factors contribute to travelers' propensity to choose green hotels? *Int. J. Hosp. Manag.* **2019**, *77*, 385–395. [CrossRef]
- Chen, H.; Bernard, S.; Rahman, I. Greenwashing in hotels: A structural model of trust and behavioral intentions. *J. Clean. Prod.* **2019**, *206*, 326–335. [CrossRef]
- Kularatne, T.; Wilson, C.; Månsson, J.; Hoang, V.; Lee, B. Do environmentally sustainable practices make hotels more efficient? A study of major hotels in Sri Lanka. *Tour. Manag.* **2019**, *71*, 213–225. [CrossRef]
- Yeh, S.S.; Guan, X.; Chiang, T.Y.; Ho, J.L.; Huan, T.C.T.C. Reinterpreting the theory of planned behavior and its application to green hotel consumption intention. *Int. J. Hosp. Manag.* **2021**, *94*, 102827. [CrossRef]
- Han, H. Theory of green purchase behavior (TGPB): A new theory for sustainable consumption of green hotel and green restaurant products. *Bus. Strategy Environ.* **2020**, *29*, 2815–2828. [CrossRef]
- Scholz, P.; Voracek, J. Organizational culture and green management: Innovative way ahead in hotel industry. *Meas. Bus. Excell.* **2016**, *20*, 41–52. [CrossRef]
- Tourism Bureau; Ministry of Transportation and Communications. Traveling Industry Related Statistics. Available online: <https://admin.taiwan.net.tw/FileUploadCategoryListC003330.aspx?CategoryID=a7cf8979-fa4b-4473-98eb-86467e19dbbb&appName=FileUploadCategoryListC003330> (accessed on 20 March 2021).
- Environmental Protection Administration. Executive Yuan; R. O. C. Green Hotel Annual Statistics. Available online: <https://greenliving.epa.gov.tw/GreenLife/walksing2013/Statistics.aspx> (accessed on 20 March 2021).
- Mandić, A.; Petrić, L. The impacts of location and attributes of protected natural areas on hotel prices: Implications for sustainable tourism development. *Environ. Dev. Sustain.* **2021**, *23*, 833–863. [CrossRef]
- Popovic, G.; Stanujkic, D.; Brzakovic, M.; Karabasevic, D. A multiple-criteria decision-making model for the selection of a hotel location. *Land Use Policy* **2019**, *84*, 49–58. [CrossRef]
- Fang, L.; Li, H.; Li, M. Does hotel location tell a true story? Evidence from geographically weighted regression analysis of hotels in Hong Kong. *Tour. Manag.* **2019**, *72*, 78–91. [CrossRef]

22. Choi, H.M.; Kim, W.G.; Kim, Y.J.; Agmapisarn, C. Hotel environmental management initiative (HEMI) scale development. *Int. J. Hosp. Manag.* **2019**, *77*, 562–572. [CrossRef]
23. Scholz, P. Implementation of the green management on an example of the first class accommodation facilities in Prague. In Proceedings of the 4th Biannual CER Comparative European Research Conference, London, UK, 26–30 October 2015; pp. 9–12.
24. Kim, W.G.; Li, J.; Han, J.S.; Kim, Y. The influence of recent hotel amenities and green practices on guests' price premium and revisit intention. *Tour. Econ.* **2015**, *23*, 577–593. [CrossRef]
25. Ting, C.T.; Hsieh, C.M.; Chang, H.P.; Chen, H.S. Environmental consciousness and green customer behavior: The moderating roles of incentive mechanisms. *Sustainability* **2019**, *11*, 819. [CrossRef]
26. Booking.com. Booking.com Reveals 2017 Sustainable Travel Intentions, Goals and Considerations. Available online: <https://news.booking.com/bookingcom-reveals-2017-sustainable-travel-intentions-goals-and-considerations> (accessed on 18 January 2021).
27. Chou, C.J. Hotels' environmental policies and employee personal environmental beliefs: Interactions and outcomes. *Tour. Manag.* **2014**, *40*, 436–446. [CrossRef]
28. Han, H.; Lee, J.S.; Trang, H.L.T.; Kim, W. Water conservation and waste reduction management for increasing guest loyalty and green hotel practices. *Int. J. Hosp. Manag.* **2018**, *75*, 58–66. [CrossRef]
29. Park, J.; Kim, H.J.; McCleary, K.W. The impact of top management's environmental attitudes on hotel companies' environmental management. *J. Hosp. Tour. Res.* **2014**, *38*, 95–115. [CrossRef]
30. Liu, G.; Tan, Y.; Li, X. China's policies of building green retrofit: A state-of-the-art overview. *Build. Environ.* **2020**, *169*, 106554. [CrossRef]
31. Jaykumar, P. The triumphs & challenges in the path of green hotel housekeeping. *J. Serv. Res.* **2020**, *20*, 21–37.
32. Hou, H.; Wu, H. Tourists' perceptions of green building design and their intention of staying in green hotel. *Tour. Hosp. Res.* **2021**, *21*, 115–128. [CrossRef]
33. Green Hotel World. Global Green Hotel Density. Available online: <https://www.greenhotelworld.com> (accessed on 18 January 2021).
34. Siobhan, O.N. Know How Guide to Sustainable Hotel Certification Schemes. Available online: <http://www.greenhotelier.org/know-how-guides/know-how-guide-to-sustainable-hotel-certification-schemes/> (accessed on 18 January 2021).
35. Hilton. Environmental Impact. Available online: <https://cr.hilton.com/environment/> (accessed on 18 January 2021).
36. Hasek, G. Three Years After Launch, Tripadvisor Greenleaders Tops 11,000 Participants. Available online: <http://www.greenlodgingnews.com/three-years-after-launch-tripadvisor-greenleaders/> (accessed on 18 January 2021).
37. TripAdvisor. Your Eco-Friendly Stay Starts Here. Available online: <https://www.tripadvisor.com/GreenLeaders> (accessed on 18 January 2021).
38. Green Seal. Certification Checklists. Available online: <http://www.greenseal.org/GreenBusiness/Certification/Getcertified/CertificationChecklists.aspx> (accessed on 18 January 2021).
39. Green Seal. Get Certified. Available online: <http://www.greenseal.org/GreenBusiness/Certification/GetCertified.aspx> (accessed on 18 January 2021).
40. Harker, P.T.; Vargas, L.G. The theory of ratio scale estimation: Saaty's analytic hierarchy process. *Manag. Sci.* **1987**, *33*, 1383–1403. [CrossRef]
41. Chen, Y.C.; Yu, T.H.; Tsui, P.L.; Lee, C.S. A fuzzy AHP approach to construct international hotel spa atmosphere evaluation model. *Qual. Quant.* **2014**, *48*, 645–657. [CrossRef]
42. Chan, E.S.W.; Okumus, F.; Chan, W. What hinders hotels' adoption of environmental technologies: A quantitative study. *Int. J. Hosp. Manag.* **2020**, *84*, 102324. [CrossRef]
43. Iorgulescu, M.C. An insight into green practices and eco-labels in the hotel industry. In Proceedings of the 6th BASIQ International Conference on New Trends in Sustainable Business and Consumption, Messina, Italy, 4–6 June 2020; Bucharest: ASE. pp. 1164–1171.
44. Moon, C.; Batle, J.; Morales, M. Eco-innovation in the hotel industry: A review of seven hotels using the UNWTO 2030 sustainability guidelines. In Proceedings of the International Conference on Tourism Research, Valencia, Spain, 27–28 March 2020; pp. 365–366.
45. Suryawan, I.G.B.; Aris, I. Strengthening environmental law policy and its influence on environmental sustainability performance: Empirical studies of green constitution in adopting countries. *Int. J. Energy Econ. Policy* **2020**, *10*, 132–138. [CrossRef]
46. Nelson, K.M.; Partelow, S.; Stäbler, M.; Graci, S.; Fujitani, M. Tourist willingness to pay for local green hotel certification. *PLoS ONE* **2021**, *16*, e0245953. [CrossRef] [PubMed]
47. Sharma, T.; Chen, J.; Liu, W.Y. Eco-innovation in hospitality research (1998–2018): A systematic review. *Int. J. Contemp. Hosp. Manag.* **2020**, *32*, 913–933. [CrossRef]
48. Heras-Saizarbitoria, I.; Boiral, O.; Allur, E.; García, M. Communicating environmental management certification: Signaling without signals? *Bus. Strategy Environ.* **2020**, *29*, 422–431. [CrossRef]
49. Dos Santos, R.A.; Méxas, M.P.; Meiriño, M.J.; Sampaio, M.C.; Costa, H.G. Criteria for assessing a sustainable hotel business. *J. Clean. Prod.* **2020**, *262*, 121347. [CrossRef]
50. Rawashdeh, A.A.; Al-Ababneh, M.M. Hotel guests' perceptions of environmentally friendly practices in Jordan. *J. Environ. Manag. Tour.* **2021**, *12*, 107–120. [CrossRef]
51. Chung, K.C. Green marketing orientation: Achieving sustainable development in green hotel management. *J. Hosp. Mark. Manag.* **2020**, *29*, 722–738. [CrossRef]

52. Dani, R.; Tiwari, K.; Negi, P. Ecological approach towards sustainability in hotel industry. *Mater. Today Proc.* **2021**. [[CrossRef](#)]
53. Yu, C.C.; Liu, C.C. Applying fuzzy AHP to evaluate the key selection criteria for green hotel operation managers. *J. Bus. Manag. Sci.* **2021**, *9*, 12–16. [[CrossRef](#)]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

ISPRS International Journal of Geo-Information Editorial Office
E-mail: ijgi@mdpi.com
www.mdpi.com/journal/ijgi



MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34

www.mdpi.com



ISBN 978-3-0365-5030-5