

Received 20 September 2022, accepted 18 October 2022, date of publication 28 October 2022, date of current version 23 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3217916

RESEARCH ARTICLE

Chinese Tourists in Malaysia: An Analysis of Spatio-Temporal Behavior Based on Tourism Digital Footprints

ZHENG CAO^{1,2}, HENG XU¹, AND BRIAN TEO SHENG XIAN²

¹School of Management, Henan University of Technology, Zhengzhou, Henan 450001, China

²Graduate School of Management, Management and Science University, Shah Alam, Selangor Darul Ehsan 40100, Malaysia

Corresponding authors: Heng Xu (xuheng@haut.edu.cn) and Brian Teo Sheng Xian (brian_teo@msu.edu.my)

ABSTRACT Travel digital footprints can be used to analyze tourists' spatio-temporal behavior to customize travel plans and recommendations. This study focuses on Chinese tourists in Malaysia as the research subject and uses the tourism digital footprint of Chinese tourists' travel texts on Qunar.com as the primary data source. It connects traditional quantitative analysis with complex network analysis to study tourists' behavior, time pattern, and complex network effects. The research results were as follows: (1) there was an uneven distribution of core tourism nodes with structural holes in Malaysia, which form a network pattern of imbalanced power and fierce internal competition; (2) the digital footprints of Chinese tourists were mainly concentrated in traditional tourist hot spots in Malaysia; (3) besides the competition between domestic islands, Thailand and Singapore are Malaysia's main competitors for Chinese tourists. These results provide helpful information for the tourism management departments of Malaysia to improve their marketing and development efforts directed for Chinese tourists.

INDEX TERMS Chinese tourists, spatial-temporal behavior, Malaysia, tourism digital footprints, complex network.

I. INTRODUCTION

Under human-centered development, studying tourists' spatio-temporal behavior has attracted increasing attention from scholars in tourism, urban planning, geography, and other fields [1]. It is crucial to predict and analyze tourism markets to understand how time and space affect the travel patterns of tourists [2]. It can also help planners assess the attractiveness and carrying capacity of tourist destinations. In addition, it provides necessary insight into tourism mobility for local and national governments and improves transport policies according to tourist distribution and travel distance. Thus, reducing environmental pollution to promote sustainable development. The lack of proper tourism activity data has always been a significant constraint in tourism research [3], [4]. However, as the Internet continues to develop, tourists' "digital footprint" has generated a large

amount of data, making it possible to acquire customer-centered personal spatial-temporal behavior and contextual information. This information has long-time sequences, large quantities, high precision, and many other advantages, which can effectively reduce the cost and inconvenience caused by offline questionnaires and interviews [5]. This helps researchers understand tourists' spatio-temporal behavior patterns in destinations and their interactions with the tourism environment. These advantages make the "digital footprint" a vital data source for studying tourists' spatio-temporal behavior [6].

According to a barometer compiled by the United Nations World Tourism Organization, Malaysia has been one of the most popular tourist destinations in the Asia-Pacific region in recent years. In 2018, 25.83 million international tourists visited Malaysia. In 2019, the number rose to 26.1 million, and the tourist revenue reached RM86.14 billion. Currently, the tourism industry in Malaysia has become the second-largest source of foreign exchange income for the Malaysian

The associate editor coordinating the review of this manuscript and approving it for publication was Yichuan Jiang¹.

economy, and it is still maintaining a year-on-year growth trend. Due to the COVID-19 pandemic since 2020, the tourism industry in Malaysia has been facing unprecedented challenges (Ministry of Tourism and Culture, 2019). With the implementation of the Belt and Road Initiative, China and Malaysia have presented close cooperation and considerable development prospects in trade and tourism. China is Malaysia's third-largest tourist source market. The number of Chinese tourists to Malaysia has increased from 0.94 million in 2008 to 2.94 million in 2018, an increase of 200% over the past decade. In 2019, the tourism revenue of Chinese tourists in Malaysia was RM12.3 billion, accounting for 14% of the total tourism revenue of Malaysia [7]. Analyzing Chinese tourists' tourism behavior in Malaysia is crucial for further developing tourism markets, improving tourism services, and strengthening the bilateral cultural exchanges and cooperation. It can also provide a sufficient reference for government departments, destination planners, and tourism service providers to exploit proper facilities in specific destinations.

In the current research on tourism in Malaysia, researchers mainly concentrate on the relationship between tourism and the economy [8]. Secondly, scholars pay more attention to the sustainable development of tourism in Malaysia and the protection of the local natural environment. Besides, they also focus on heritage tourism [9] and Muslim tourism [10]. There is very little research on Chinese tourists, and the current research focuses on medical tourism [7], [11] and economic factors that affect Chinese tourists' travel demand to Malaysia [12]. There is no literature on the temporal and spatial behavior of Chinese tourists in Malaysia.

In this context, this paper discusses the tourism digital footprint, represented by tourists' online Chinese travel texts as the primary data source. It analyzed the spatio-temporal behavior characteristics of Chinese tourists by mining the information of tourists in the two dimensions of time and space. The structure of this paper is as follows: the second section reviews the most relevant literature reference on tourism digital footprints and tourists' spatial-temporal behavior; the third section describes the research design (data and methods) applied in this paper, including the combination of traditional quantitative methods and complex network analysis; the fourth section introduces the results, and the corresponding suggestions are put forward; Finally, the conclusion summarizes the main points of this paper according to the temporal and spatial behavior characteristics of Chinese tourists in Malaysia.

II. LITERATURE REVIEW

As China improves its national economy and relaxes its outbound tourism policies, Chinese tourists have become the major participants in the global tourism market [13]. However, due to the uniqueness of Chinese culture, Chinese tourists' behavior is often different from that of other international tourists [14]. Therefore, Chinese tourists visiting tourist destinations have shown strong behavioral diversity. Therefore, a strong behavioral diversity has been shown

by Chinese tourists in tourist destinations [15]. Currently, Chinese tourists have become typical research subjects in the tourism behavior literature. Researchers conducted an in-depth exploration of the aspects of demand [16], motivation [3], [17], perception of destinations [18] and travel characteristics [19], [20]. In recent years, travel social media platforms have become increasingly popular in China. Therefore, more and more Chinese tourists begin to write and share their travel diaries on the platform. In particular, the travel itinerary shared by tourists makes the analysis of Chinese tourists' spatio-temporal behavior more accurate than traditional data sources, such as statistical yearbooks, which can bring a new perspective to the detailed study of Chinese tourists' behavior [21]. In the following sections, tourists' spatio-temporal behavior and travel digital footprints are reviewed and summarized.

A. TOURISTS' SPATIO-TEMPORAL BEHAVIOR

Tourists' spatio-temporal behavior refers to tourists' spatial movement behavior and time allocation behavior in the travel process from the origin to the destination and back to the origin. The analysis of tourists' spatio-temporal characteristics can be presented from the three dimensions of time, space, and flow (Table 1). A common type of research is classifies tourists according to their temporal characteristics to compare their spatial behavior information, such as analyzing tourism demand in Spain according to seasonal changes [28]. In the example of Nanjing, three tourism flow networks corresponded to different travel durations [24], which proves the importance of time in tourism study. Regarding the selection of indicators, the Gini coefficient, the coefficient of variation, Theil index, and entropy were the most frequently used indicators to study tourists' temporal behavior. The other common study analyzed the spatial-temporal changes in tourists' behavior from a geospatial perspective, such as extracting popular tourist attractions in Beijing through spatial clustering [22]; analyzing the movement of Chinese tourists to ASEAN through the gravity center model, and tourism market model [26]. In the research methods, most researchers chose to adopt the Markov chains [29], cluster analysis [30], logistic-regression and general log-linear models [31] to analyze the tourist flow.

However, most of the above studies are qualitative, only analyzing the types of tourism flow under static conditions, without quantitative verification of the dynamic network characteristics of tourism flow. Therefore, scholars have begun to reveal the spatio-temporal characteristics of tourism through the study of tourism flows, mainly concentrating on the flow of tourists and the network patterns and characteristics generated by tourism flows, and have summarized relevant flow laws and behavior patterns [32], [33]. The researchers analyzed tourism flow density, centrality, Core-Periphery Model, and cohesive subgroups. For example, the network model of Chinese tourists' tourism flow in Japan was proposed using network density, node center degree, and other indicators [23]. However, these research methods

TABLE 1. Summary of methodologies used in the field of tourism digital footprints to study the spatial and temporal behavior characteristics of tourists.

Author (Year)	Data sources	Measure	Conclusions
(Peng and Zhou, 20[18]) [22]	Flickr geo-tagged images within the urban area of Beijing from 2005 to 2016.	Spatial clustering and text mining approaches	This novel approach can distinguish adjacent high-density areas when discovering popular tourist attractions and has better adaptability in the case of an uneven density distribution as compared to the traditional DBSCAN method.
(Zeng, 2021) [23]	430 itineraries from travel services and 458 itineraries from free independent tourists' trip dairies	Social Network Analysis	232 nodes and 981 ties constitute the Chinese tourist flow network, and the structure of this network is complex.
(Jin, Cheng et al., 2018)[24]	1424 valid samples of complete trips involving 45 attractions within Nanjing in 2012	Social network analysis and a revised spatial interaction model	The power law of distance decay in intra-urban tourist mobility.
(Qin, Song, et al., 2019) [25]	Flickr geo-tagged images covering the entire Chinese country from 2011 to 2017	Markov chain and community detection	Explore the spatial characteristics of China's inbound tourist flow, the spatial patterns of tourist movement, and the tourist destination cities group.
(Liu, Yu, et al., 2019) [26]	The number of tourists from 9 ASEAN countries from 2002 to 2016	Gravity model and market form model of the tourism system	The gravity centers of tourist flow from China to ASEAN from 2002 to 2016 amass within Thailand Gulf and Kompong Som Bay. The tourist flow gravity centers generally take a southward-northward-southward pattern.
(Yao, Shi, et al., 2021) [27]	Open global positioning systems (GPS) trajectory data at Yuanmingyuan Park from January 2014 to August 2020.	Python and ArcGIS	Hot spots of tourist passing, hot spots of tourist gathering, high average dwelling time areas, and tourist interest areas

concentrate on the explicit flow and ignore potential and internal network structure. Therefore, it cannot fully present the spatio-temporal behavior characteristics of tourists in tourists' destinations and lack the breadth and depth of pattern monitoring. The community detection indicators in complex networks can make up for this, which are widely used in many fields, such as social networks [34], [35], [36], computer

networks [37], [38], and biology and ecology [39], [40]. Community detection is a typical method for identifying subgroups with strong connections in complex networks. For tourism research, community discovery provides a method to divide the network into multiple communities to maximize the connection within the same community and minimize the connection between different communities [25]. This is the key to understanding the complex network structure in virtual space and, more importantly, in geographical space [41].

B. TRAVEL DIGITAL FOOTPRINT

In 2008, American scholars Girardin and others defined the concept of “digital footprint” [42]. It provides a new method for collecting tourist flow data and a new research perspective for studying tourist mobility [43]. The use of travel digital footprints can accurately and quickly analyze the routes and emotions of tourists, which reveals the behavior law of tourists. Travel digital footprints can be collected from different types of data sources, such as GPS [27], mobile network data [44], geo-labeled photos [45] and User Generated Content data [46], [47].

In recent years, travel notes have become increasingly popular among User Generated Content (UGC) data sources. As a new type of travel data voluntarily shared by tourists, online travel diaries include traditional text and image information and record tourists' location information [43]. These timely and strong coverage of tourism digital footprints provides tourists with obvious space and time labels, which can achieve accurate positioning and completely restore the time traces of tourists [1]. Therefore, tourism digital footprint has become a new measurement method for studying the spatiotemporal characteristics of tourism flow in the era of intelligent tourism, providing a new tool perspective for studying the spatiotemporal characteristics of tourism flow [45]. It can help reveal the spatio-temporal characteristics of tourist flow, and accurately restore the selection process of tourist destination. For example, online travel notes were used to study the spatial behavior characteristics of tourists, tourism flows, and patterns of temporal heterogeneity [24], [46]. However, most scholars obtain the location information of tourists using travel text extraction. This method may lead to the following problems. First, travel notes reflect tourist attractions more than administrative locations. After extracting the specific location, the text needs to rearrange and summarize the specific location of each scenic spot, which requires too much work. Second, tourists generally write travel notes based on their memories after traveling, which are highly random. Many notes are not written according to the chronological order of travel, leading to incomplete and inaccurate descriptions of tourism flows [48]. The above will result in errors in the data and affect analysis results. With the development of location-based services (LBS), an increasing number of social media platforms have integrated location-based service modules into travel note-writing tools so that tourists can directly share their travel itineraries to improve the reading experience of travel blogs [49]. These itineraries were

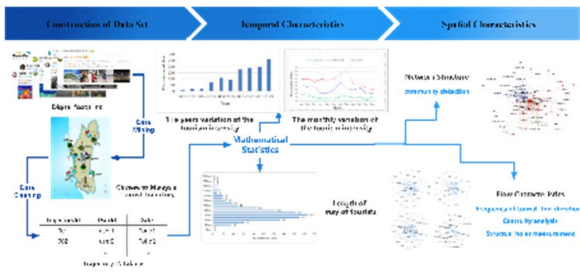


FIGURE 1. The research framework on spatial-temporal behavior.

directly edited by tourists and included rich spatio-temporal label information, which can provide a detailed description of tourists’ itineraries [50]. These data overcome the deficiencies in traditional data (statistical yearbooks, questionnaires, etc.) and avoid the problems of information redundancy and record deviation, which are common to other types of digital footprint data sources.

Therefore, this paper utilized tourists’ itineraries with geographic information as the source of digital footprint data and used the indicators of complex networks such as centrality, structural hole, and community detection to construct a research framework from the three dimensions of time, space, and flow to carry out an in-depth study on tourists’ spatial-temporal behavior.

III. METHODOLOGY

A. RESEARCH FRAMEWORK ON SPATIAL-TEMPORAL BEHAVIOR

This paper proposes a research framework for tourists’ spatio-temporal behavior analysis based on itinerary text data containing geographic location information. This research framework (Figure 1) combines traditional quantitative analysis with complex network analysis. First, intelligent data acquisition and processing rules are designed to construct the data set. Second, to comprehensively analyze the characteristics of spatial-temporal networks, the research contents of several detection methods (Flow direction statistics and complex network analysis) were divided into two parts: (1) community detection (to analyze the clusters formed by tourism destinations and analyze the “implicit” spatial network characteristics); (2) frequent flow direction, centrality and structural holes (to analyze the “explicit” flow characteristics of tourist flow between attractions).

B. STATISTICS OF FLOW DIRECTIONS

Tourist flow direction refers to the movement sequence of tourists between two attractions, such as “From Kuala Lumpur to Malacca”. The frequency formula of flow direction is as follows:

$$P_i = \frac{v_i}{V} \times 100\% \tag{1}$$

where P_i is the frequency of the flow direction i ; v_i is the quantity of the flow direction i , that is, the occurrence number of the flow direction i in the sequence data; V is the total flow

of all flow directions i , that is, the number of tourists’ movements between two attractions in the data. By calculating and sorting the frequency of each flow direction in the data, the frequent directions of tourist flows can be obtained.

C. COMPLEX NETWORK ANALYSIS

Complex network analysis was used to analyze human movement or behavior and reveal spatial structure or interaction. If a user travels from one city to another, the two cities are considered to interact with each other. All cities are connected by tourism flows, and a network structure can be constructed with vertices representing cities and edges representing interactive relationships [46]. On this basis, the complex network analysis method was adopted to study the spatial network structure of tourist destinations, and various internal relationship characteristics of tourism flows can be found [43]. Therefore, this study employed Ucinet and Netdraw software to conduct quantitative processing and visual analysis of complex network analysis. It utilizes community discovery, centrality analysis, and structural hole measurement to quantitatively measure the spatial behavior patterns, network structure, and node characteristics of Chinese tourists to Malaysia. Then, the network structure was visualized using the analysis results.

1) COMMUNITY DETECTION

In complex network theory, community refers to a dense subnetwork within a more extensive network [51], meaning they are more closely related to in-group nodes than to out-group nodes [52]. Community detection provides a typical method for understanding the spatial structure of complex networks in geographic space [53]. We detected tourism communities by establishing connections between different locations according to tourists’ large numbers of digital footprints. Such a community reflects the compactness of tourists’ spatial movements [54]. Generally, nodes that are more closely connected within the same network are more likely to form clusters. The calculation formula of cluster partition density is as follows:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{2}$$

In this formula, A_{ij} represents the connection between the node i and the node j . $k_i = \sum_{ij} A_{ij}$ represents the sum of weights of all edges connected to the node i , C_i and C_j represents the clustered index of the node i , and the node j . And the following function, $\delta(c_i, c_j)$ represents whether the node i and the node j are in the same cluster (If they are in the same cluster, the function is equal to 1; if they are not, the function is equal to 0). Finally, $m = \frac{1}{2} \sum A_{ij}$ represents the sum of connection weights of the whole network.

2) CENTRALITY ANALYSIS

Centrality analysis can quantify the position and center position of each node in the measurement network, and is an important means of quantifying the rights in the social

network. The important nodes in the measured network are represented by degree centrality, betweenness centrality, and closeness centrality. Degree centrality was divided into indegree and outdegree centrality, representing the direct association between the target node and other nodes; the higher the value, the more well-known and widely popular these sites are among tourists, and they are necessary places for tourists to carry out travel activities. Betweenness centrality indicates the degree of the target node’s control over other nodes. The higher the degree of betweenness centrality, the higher the irreplaceability of these sites in the tourist flow network, as an important “bridge” for tourists to transfer to other sites, so other sites should enhance the connection with these sites to increase more tourist flow. Closeness centrality was used to measure the degree of closeness between nodes [32]. Higher values indicate relatively higher overall accessibility and more convenient transportation, which can better attract tourists. The details can be found in the following formula:

a: DEGREE CENTRALITY

In-degree centrality formula:

$$C_{AD,in}(i) = \sum_j r_{ij,in} \tag{3-1}$$

Out-degree centrality formula:

$$C_{AD,out}(i) = \sum_j r_{ij,out} \tag{3-2}$$

where $r_{ij, in}$ and $r_{ij, out}$ represent the directional relationship between the nodes i and j , i.e., the number of tourists from attraction i to attraction j or the opposite; the former indicates that j flows to i , and the latter indicates that i flows to j , and n is the number of nodes in the tourist flow network.

b: BETWEENNESS CENTRALITY FORMULA

$$C_{AB}(i) = \sum_j \sum_k \frac{g_{jk}(i)}{g_{jk}}, \quad j \neq k \neq i, j < k \tag{3-3}$$

where g_{jk} is the number of paths that the traveler reaches the node k from the node j , $g_{jk}(i)$ is the number of paths by node i in the paths from the node j to the node k , and n is the number of nodes in the tourist flow network.

c: CLOSENESS CENTRALITY FORMULA

$$C_{AP}^{-1}(i) = \sum_j g_{ij} \tag{3-4}$$

where g_{ij} is the number of paths from the node i to the node j , and n is the number of nodes in the tourist flow network.

3) STRUCTURAL HOLES MEASUREMENT

In 1992, Burt proposed the theory of “Structural holes” [55]. The calculation of structural holes enables the identification of bottleneck problems in regional tourist flow [56]. The dominant nodes of structural holes generally have strong regional competitive advantages, and are less affected by the tourist flow of the surrounding nodes. They are irreplaceable,

and there are significant differences in accessibility between them and surrounding nodes. For this end, Burt proposed the use of effective size and constraint as metrics of structural holes in social networks. These metrics are widely used and are thus employed here [32].

a: EFFECTIVE SIZE

Effective size refers to the non-redundant factor of the node in the network; that is, the effective size of one node is equal to the individual network scale of the node minus the network redundancy. The individual network scale of a node is the number of nodes contained in its neighborhood. The redundancy degree is equal to the average degree of other nodes in the individual network member of the node; therefore, the effective size is equivalent to the individual network scale minus the average degree of all network members of the node. Effective size can measure the whole influence of this node, which can be used to measure the importance of structural hole nodes: the higher the effective size, the more obvious the competitive advantage of the target node in attracting visitors and the higher the number of visitors. The calculation formula is as follows:

$$ES_i = \sum_j^n \left(1 - \sum_q^n p_{iq} m_{jq}\right), \quad \neq i, j, \text{ where}$$

$$p_{iq} = \frac{(z_{iq} + z_{qi})}{\sum_j^n (z_{ij} + z_{ji})}, \quad i \neq j$$

$$m_{jq} = \frac{(z_{jq} + z_{qj})}{(z_{jk} + z_{kj})}, \quad j \neq k \tag{4-1}$$

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

b: CONSTRAINT

Constraint refers to the ability of a node to use structural holes in its network. The dependence of the node on other nodes is regarded as the evaluation criterion. The larger the value, the stronger the constraint and the more obvious the competitive disadvantage of the target node in attracting visitors, the lower the number of visitors.

$$CT_i = \sum_j^n (p_{ij} + \sum_q^n p_{iq} p_{qi})^2, \quad q \neq i, j \tag{4-2}$$

where p_{ij} is the proportional relationship between node i and node j ; p_{jq} is the proportional relationship between node i and node q ; p_{qj} is the proportional relationship between node q and node j ; The calculation method of the proportional relationship between nodes is the same as equation (4-1). n is the number of nodes in the tourist flow network.

D. STUDY AREA

Malaysia is located in Southeast Asia and consists of Malaya to the south of the Malay Peninsula and Sarawak and Sabah to the north of Kalimantan Island. Being rich in tourism resources is an advantage for Malaysia, making it a tourist spot for tourists worldwide [57]. According to the Malaysian Department of Statistics, tourism accounts for 15.9% of its Gross Domestic Product (GDP) in 2019 (Department of Statistics, Malaysia, 2020). However, the COVID-19 outbreak in 2020 exerted an enormous negative impact on Malaysia’s tourism industry, and the Malaysian government then introduced a wide range of policy measures to attract more Chinese tourists [58]. It is of great significance for the sustainable development of Malaysia’s tourism industry by considering Malaysia as a case study to clarify the spatial characteristics of its tourism and the strengths and weaknesses of its development.

E. DIGITAL FOOTPRINT DATA

This paper focused on the Chinese website Qunar (<http://www.qunar.com>) as its data source. Qunar.com is the most famous travel media in China. It includes tourism strategies, reviews, and other information on more than 60000 tourism destinations worldwide, with more than 130 million registered users. Users can share their travel experiences here on the website, whose relevant information is relatively complete and updated in time [32].

This study used Python’s web crawler technology and collected 1990 Malaysian travel routes voluntarily shared by Chinese tourists as initial data. Considering the amount of data issues, it was decided to expand the time period from 2010 to 2019, so all 1,990 data can be used. Due to the epidemic’s impact in early 2020, strict measures were taken by all countries to restrict inbound and outbound travel to curb the spread of the disease [59]. As a result, no new data on visiting Malaysia has been generated since January 2020.

Considering the possibility of information errors and logic errors in the initial data, we designed the following data cleaning rules: 1)exclude the travel records with incomplete itineraries; 2)if the travel records involve places outside Malaysia, only the point of interesting (POI) records of the latest entry into or departure from Malaysia shall be retained; 3)remove the data of repeated scenic spots (if POI numbers of the same scenic spot appear in the data in sequence, it is considered that the user did not leave the scenic spot, so the redundant records were deleted). Finally, 1,148 valid samples were obtained based on primary data screening and sorted to build the digital footprint database of Chinese tourists in Malaysia, including tourist names, travel dates, travel nodes, length of stay, and other data. Finally, 49 pieces of node information for Malaysia and 3,270 edges between them were obtained.

According to the order of visits, the actual daily trips of tourists were split into directed nodes. For example, Kuala Lumpur - Penang - Langkawi - Kuala Lumpur was

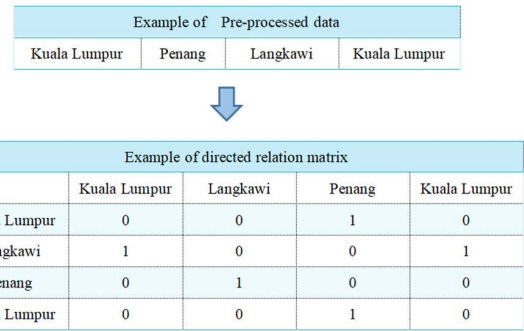


FIGURE 2. Example of the data processing procedure.

divided into Kuala Lumpur - Penang, Penang - Langkawi, and Langkawi - Kuala Lumpur. If there was a direct flow between nodes, it was denoted as 1 and if there was no direct flow, it was denoted as 0 to construct the data matrix of the flow direction of the travel nodes. By cleaning the travel information and extracting the data of node paths, a 49×49 multi-valued directed relation matrix was obtained in this study (figure 2). The characteristics of Malaysia’s tourism flow network structure were analyzed using this matrix.

IV. RESULTS

A. TEMPORAL CHARACTERISTICS

Based on the premise that “the number of travel notes is positively correlated with the flow of tourists”. this paper probed the regularities of the time distribution of Chinese tourists in Malaysia. Time is a type of widespread data that objectively records the states of the observed characteristic values at different times or time points. The time involved is mainly divided into two types: travel period, including the distribution of traveling years and months; the other is the length of stay, presented in the form of the number of days of traveling.

The itinerary text released by tourists reflects the specific travel time. As shown in figure 3, the number of Chinese tourists has been rising steadily, with a relatively small but still growing number from 2010 to 2010 to 2014. This is mainly due to the Malaysian government’s ninth Five-Year Plan for Tourism Development, which for the first time regards tourism as one of the four priorities for economic development [26]. The number of Chinese tourists dropped suddenly in 2015 was a consequence of the global economic slowdown and unfavourable domestic events in Malaysia such as the worst flood in the past 30 years that has affected several states in early 2015; the earthquakes in Ranau, Sabah in middle of 2015 altogether with the lingering effects of the MH370 and M17 incidents. In their efforts to counter the adverse effect of MH370 tragedy as well as the kidnapping incident in Sabah towards the Chinese visitors [12]. During times of political instability, hazards, and economic crises, Malaysia has made many changes to the Chinese market, such as relaxing visa requirements, introducing electronic visa measures, and changing marketing strategies to attract

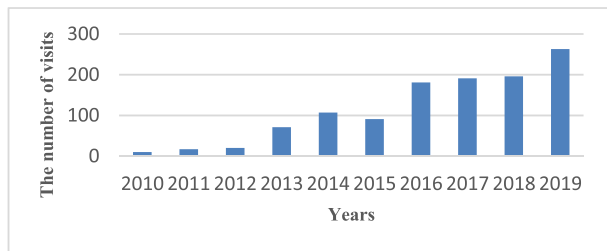


FIGURE 3. The years variation of the tourism intensity of Chinese tourists in Malaysia.

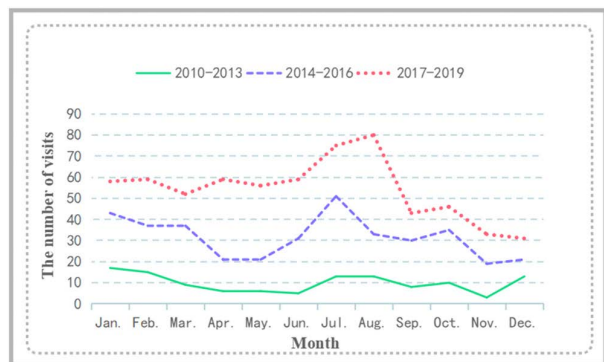


FIGURE 4. The monthly variation of the tourism intensity of Chinese tourists in Malaysia.

Chinese tourists [60]. Therefore, the number of Chinese tourists has increased annually since 2016.

As shown in figure 4, the tourist itinerary from 2010 to 2019 is divided into three periods:2010-2013, 2014-2016 and 2017-2019, and is marked according to the monthly time period, which studies the overall distribution features and characteristics of Chinese tourists in Malaysia. The number of monthly visits in these three time periods is basically the same,with little change. Figure 4 shows that the number of travel notes was the highest in July and August, followed by January and February, which were the peak tourist seasons. This is because July and August are the summer months in China, coinciding with the summer holiday for students. Thus, there are more groups of students and groups of parents with their children deciding to travel during this period. January and February are winter months in China, and the weather during this period is cold. Chinese tourists prefer to spend their holidays in places with warm climates. Meanwhile, this period meets the Chinese New Year, and more people tend to travel with their families.

Figure 5 can be obtained by combing and comparing the travel notes. Temporal behavior characteristics of tourists can be divided into two types: The first was the sightseeing tour, in which the visitors stay 5 to 7 days, and they accounted for 42%. The primary purpose of this kind of outbound tourist is to visit the classic areas of Malaysia, and they will choose one to two famous tourist destinations to visit. They either choose the island tour, as shown in travel notes like “Independent travel to Langkawi”, or just the city tour, as seen from travel notes like “Tour to Kuala Lumpur and

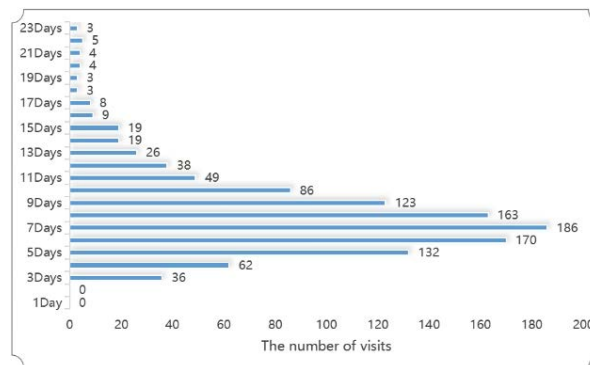


FIGURE 5. Length of stay of Chinese visitors.

TABLE 2. Tourist spot visited.

Popularity Rank	City Name	Percentage (%)
1	Kuala Lumpur	26
2	Tawau	15
3	Kota Kinabalu	12
4	Malacca	12
5	Semporna	10
6	Penang	9
7	Langkawi	7

Malacca”. The second was the in-depth tour, during which the visitors spend 8 to 10 days, accounting for 29%. For example, in travel notes like “eight days’ in-depth delicacy enjoying and touring in Malaysia” and “Independent spring travel in Malaysia”. Most of these tourists travel extensively and deeply in Malaysia. There are usually more than three tourism nodes, including urban and island tourism.

B. SPATIAL CHARACTERISTICS

1) FREQUENT DIRECTIONS OF TOURIST FLOWS

The data shows that in the 3270 edges, 452 of them were produced by returning from and leaving for other overseas countries.After removing the 452 edges generated by Malaysia’s neighboring countries and undergoing a screening process, 2818 pieces of data on tourist flows within Malaysia were obtained. We calculated the number of Chinese tourists visiting each destination to determine the most popular destinations in Malaysia that are frequently visited by Chinese tourists. As shown in table 2, there are seven destinations most visited by Chinese tourists. Kuala Lumpur was the most popular among Chinese tourists because it is the capital of Malaysia and serves as a major transportation hub for tourists. Chinese tourists usually visit Kuala Lumpur and then transfer to other places. Second was Tawau, which is often used as a springboard for tourists to and from Semporna because of its airport. Overall, traditional destination cities remain the most popular for Chinese tourists, which shows that Chinese tourists may not know enough about Malaysia’s tourism resources. At the same time, and there may also be time constraints: visitors stayed for relatively short periods, so they were more likely to visit the best-known places.

The online travel itinerary text data showed a total of 1146 visitors’ tourist activities between the scenic spots.

TABLE 3. Frequencies of tourist flow directions in Malaysia.

source	Target	Percentage (%)
Kuala Lumpur	Malacca	9.41
Malacca	Kuala Lumpur	8.95
Tawau	Semporna	7.48
Semporna	Tawau	7.34
Kota Kinabalu	Tawau	5.17
Kuala Lumpur	Penang	5.09
Tawau	Kota Kinabalu	4.38
Penang	Kuala Lumpur	3.74
Kuala Lumpur	Langkawi	3.35
Langkawi	Kuala Lumpur	3.24
Kuala Lumpur	Kota Kinabalu	3.10

According to formula (1), these movements were sorted according to the frequency of tourist flow of each scenic spot, and the passenger flow ranking of Malaysia was obtained. Table 3 shows the top 10 tourist flow directions. Frequent tourist flow direction reflects the connection between scenic spots. The tourist flow relationship between Kuala Lumpur and Malacca was the most noteworthy (two-way tourist flow accounted for 18.36% of the total tourist flow). The second was the tourist flow relationship between Tawau and Semporna, accounting for 14.82%. The passenger flow between Kota Kinabalu and Tawau accounted for 9.55%, between Kuala Lumpur and Penang accounts for 8.83%, and between Kuala Lumpur and Langkawi accounted for 6.59%. This shows that these routes were the core routes for Chinese tourists in Malaysia.

2) COMMUNITY DETECTION

Community detection method was adopted to divide destination cities into different tourist groups based on the linking strength between cities in the inbound tourist flow network. As shown in Figure 6, Malaysia is divided into three communities. Community A was city tour from Kuala Lumpur to various major cities in Malaysia. The most popular cities are Malacca, Penang, and Langkawi. Community A can be further divided into cultural tours (e.g., Malacca, Kudat, Kuching) and natural tours (e.g., Sarawak, Langkawi, and Pahang). Both Communities B and C were island tours. Community B started in Kuala Lumpur or Kota Kinabalu, went to Tawau, turned to Semporna, and headed for each island. Community C started from Kuala Lumpur, flowed to Terengganu, and transferred to Pulau Redang and the other two islands. Because both B and C are island tours, competition exists. Generally, Chinese tourists would choose one of these two. The data showed that more Chinese tourists chose community B. The reasons can be inferred from the Chinese travel notes: (1) the Malaysian government developed Pulau Redang in 2000, which was long ago. Hence, the facilities are relatively outdated compared with those in Semporna. (2) The scenery of Semporna is much better, and its popularity is higher in China thanks to the Internet and vigorous promotion by travel agencies. Therefore, more Chinese tourists tend to choose Semporna for island tours.

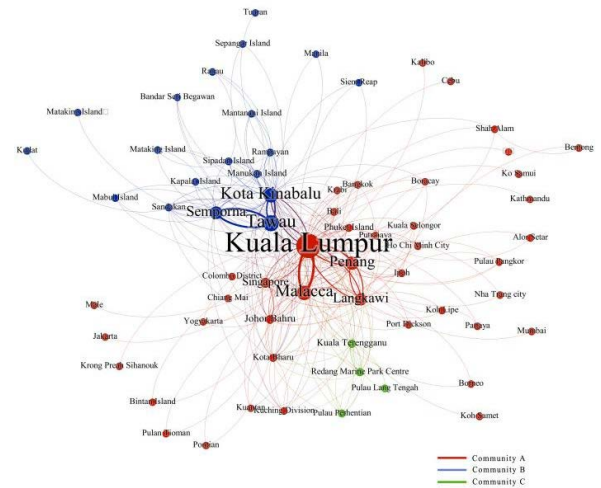


FIGURE 6. Spatial distributions of the communities detected in the network.

Pulau Redang is well known to us because of the film “SUMMER Tea”. You know, the film was released in 2000, so the island was developed very early and the hardware facilities were very old. The rooms are large; therefore, they are suitable for both children and the elderly.

2017.10

I have been to both, and I recommend Semporna. There is only sea around Pulau Redang, and the boat journey takes one hour. The coral and fish are ordinary, and almost no starfish can be found. There are both mountains and sea around Semporna, and it takes one hour to get there by boat.

2018.7

Absolutely Semporna! There is no seafood to eat in Pulau Redang, and there is a Marine park around it. Fishing is prohibited within several nautical miles around the island; therefore, the island’s food is fried chicken, fried chicken, fried chicken, steak, steak, and steak! So, if you want to have a happy dining, of course, you should go to Semporna. When returning from island hopping and snorkeling at 4 or 5 p.m., fishermen sold seafood on the dock. Two lobsters can be bought with only 100 ringgits, and the nearby processing shops are not charging high prices.

2019.09

First, I personally think that snorkeling in Semporna is more beautiful than that in Pulau Redang. Second, if you live in the town of Semporna, you can have great seafood, snork in the most beautiful islands and walk on the beach.

2020.09

TABLE 4. Alternative sites in Malaysia.

Network name	Country	Indegree
Singapore	Singapore	128
Phuket	Thailand	24
Bali	Indonesia	17
Bangkok	Thailand	5
Krabi	Thailand	4
Rebecca Island	Thailand	4
Boracay Island	Philippine Islands	3

If you have already visited many islands and have a reasonably high demand for scenery, I recommend the Semporna.

2018.05

In terms of island tours, besides the competition between communities B and C, there was competition for Chinese tourist resources from neighboring countries in community A. The data show that in the 3270 edges, 452 of them and 25 POIs were produced by returning from and leaving for other overseas countries. Many Chinese tourists traveled to the islands in countries such as Singapore, Thailand, Indonesia, and the Philippines after visiting Community A, rather than choosing the islands within Malaysia. Table 4 lists the top seven locations with the most frequent flows. The results show that Singapore, Thailand, and Indonesia were the uppermost alternative places for Malaysia, consistent with the results presented by F.Habibi et al. [61], and R. Suppiah et al. [62].

3) CENTRALITY ANALYSIS OF TOURIST FLOW NETWORK

According to the Chinese network text data, a flow matrix between attractions was constructed for the social network analysis. Malaysia’s tourist flow network structure is illustrated according to the online travel diary data. Node centrality metrics (in-degree centrality, out-degree centrality, betweenness centrality, and closeness centrality) were calculated according to Equations (3)–(5) using UCINET6. The sample results for the most relevant attractions are shown in Table5. Figure 7 shows the tourism flow network according to the value of the degree centrality of these nodes. The size of the nodes in Fig.7 represents the level of node centrality, while the thickness of the connections between nodes indicates the volume of tourist flows.

By comparing Figure 7(a) and 7(b),the calculated results of the outdegrees and indegrees of the Malaysian tourism flow network nodes were the same. The value of degree centrality of Kuala Lumpur, Kota Kinabalu, Malacca, Semporna, and Penang ranked in the top five. Kuala Lumpur’s value of centrality of outdegree and indegree was the highest,which indicated that Kuala Lumpur’s core position was prominent, and the degree of radiation was more significant than that of aggregation.

As seen in Figure 7(c),from the betweenness centrality, Kuala Lumpur, Kota Kinabalu, Putrajaya, Semporna, and Langkawi ranked in the top five,which shows that these five

TABLE 5. Calculation results of the node centrality.

Id	Indegree	Outdegree	Closeness Centrality	Betweenness Centrality
Kuala Lumpur	21.00	20.00	0.65	291.72
Kota Kinabalu	18.00	16.00	0.62	238.79
Malacca	13.00	12.00	0.56	93.51
Semporna	13.00	12.00	0.55	100.14
Kota Bharu	11.00	9.00	0.51	35.98
Penang	11.00	13.00	0.59	92.44
Tawau	10.00	8.00	0.51	33.94
Johor Bahru	8.00	9.00	0.53	70.46
Langkawi	8.00	12.00	0.57	96.51
Sandakan	8.00	7.00	0.50	41.09
Redang island	7.00	6.00	0.45	8.41
Kuala Terengganu	5.00	7.00	0.46	66.65
Manukan Island	5.00	5.00	0.43	7.32
Putrajaya	5.00	4.00	0.47	105.25
Ipoh	4.00	6.00	0.49	36.86
Pulau Perhentian	4.00	5.00	0.45	2.40
Rampayan	4.00	5.00	0.42	34.82
Sipadan Island	4.00	6.00	0.48	13.35
Shah Alam	3.00	1.00	0.32	36.19
Kapalai Island	2.00	2.00	0.38	0.00
Kuala Selangor	2.00	4.00	0.46	0.00
Kuantan	2.00	2.00	0.41	0.00
Mantanani Island	2.00	2.00	0.39	0.00
Mataking Island	2.00	0.00	0.00	0.00
Port Dickson	2.00	2.00	0.41	0.00
Pulau Pangkor	2.00	2.00	0.43	0.00
Alor Setar	1.00	1.00	0.32	0.33
Bentong	1.00	1.00	0.25	0.33
Kuching	1.00	5.00	0.47	0.50
Kudat	1.00	0.00	0.00	0.00
Pontian	1.00	0.00	0.00	0.00
Pulau Tioman	1.00	0.00	0.00	0.00
Pulau Lang Tengah	1.00	1.00	0.32	0.00
Sepangar Island	1.00	0.00	0.00	0.00
Tuaran	1.00	0.00	0.00	0.00
Ranau	0.00	3.00	0.41	0.00

attractions have the highest “irreplaceability” and dispersion ability of tourist flows in the tourist flow network. However,the betweenness centrality values between every scenic spot varied greatly among attractions and the difference between the maximum and minimum values was 291.72. The influence range of control force between core nodes and edge nodes varies greatly, which leads to unbalanced tourism development in different regions. The attraction with the highest betweenness centrality value was Kuala Lumpur, a “core intermediary” in the Malaysian tourism flow network. In comparison, the betweenness centrality was zero for the following areas, including Kapalai Island, Kuala Selangor,

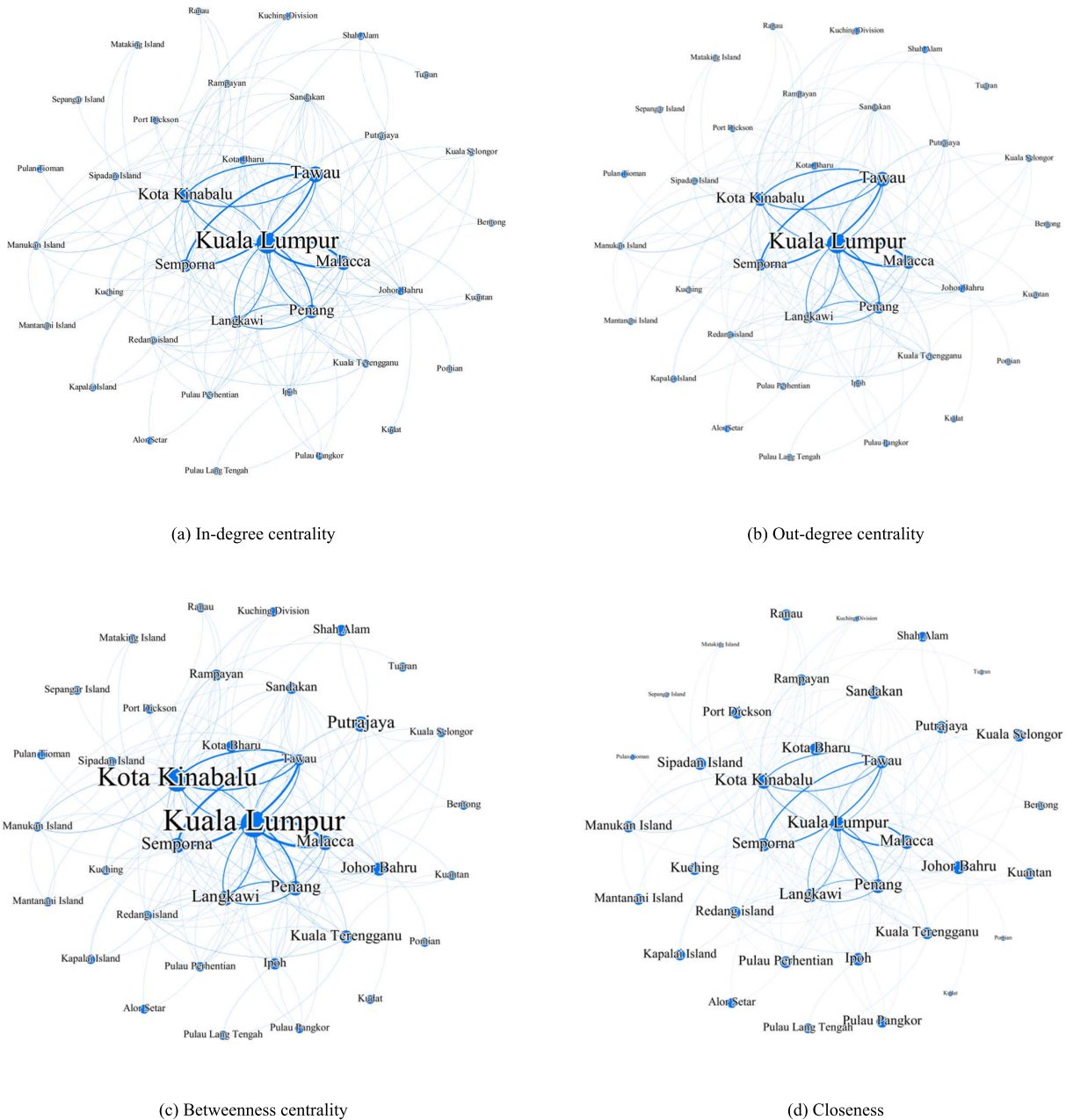


FIGURE 7. Network structure of tourist flows in Malaysia based on the calculated node centrality.

Kuantan, Mantanani Island, Matakang Island, Port Dickson, Pulau Pangkor, Kudat, Pontian, Pulau Tioman, Pulau Lang Tengah, Sepangar Island, Tuaran, and Ranau, which made them marginal in the Malaysian tourism flow network.

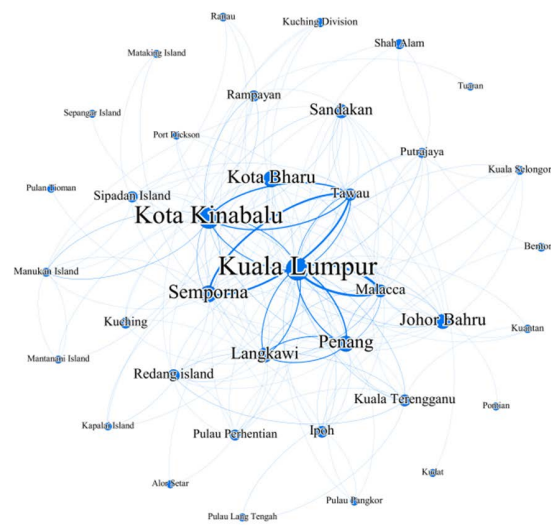
As shown in Figure 7(d), the closeness centrality values of the nodes of the Malaysian tourism flow network were more evenly distributed than that of the degree centrality and betweenness centrality. Kuala Lumpur, Kota Kinabalu, Penang, Langkawi, and Malacca had higher closeness centrality values. The locations with the lowest closeness centrality values were Matakang Island, Kudat, Pontian, Pulau Tioman, Sepangar Island, and Tuaran. This showed that the

overall accessibility of most tourist attractions in Malaysia was relatively high, had good independence, and was close to the network center. However, the construction of tourism transportation needs to be strengthened in marginal areas.

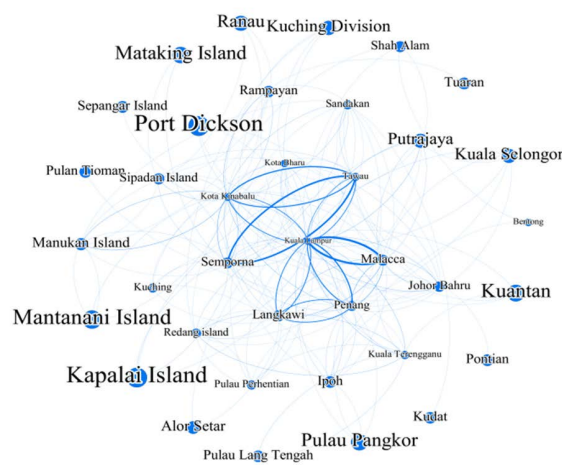
In conclusion, (1) Kuala Lumpur, Kota Kinabalu, Malacca, Semporna, Penang, and other places are the core nodes of the Malaysian tourism flow network. These places are the earliest developed scenic spots in Malaysia, and they are also the most promoted places through cooperation with Chinese travel agencies. With convenient transportation, perfect tourism infrastructure, and large passenger flow, it is a popular destination for Chinese tourists. (2) The distribution of nodes

TABLE 6. Calculation results of the structural hole indicators.

	Attractions	EffSize	Constra
1	Kuala Lumpur	22	0.82
2	Kota Kinabalu	20	0.808
3	Malacca	16	0.401
4	Penang	15	0.674
5	Semporna	15	0.734
6	Kota Bharu	14	0.748
7	Langkawi	14	0.625
8	Johor Bahru	13	0.692
9	Sandakan	10	0.772
10	Tawau	9	0.671
11	Ipoh	8	0.648
12	Kuala Terengganu	8	0.671
13	Rampayan	8	0.563
14	Redang island	8	0.717
15	Sipadan Island	8	0.588
16	Manukan Island	7	0.279
17	Putrajaya	7	0.476
18	Kuching	6	0.749
19	Pulau Perhentian	6	0.693
20	Kuala Selangor	5	1.296
21	Shah Alam	4	0.9
22	Kuantan	3	1.622
23	Pulau Pangkor	3	1.532
24	Ranau	3	1.42
25	Alor Setar	2	1.125
26	Bentong	2	0.5
27	Kapalai Island	2	1.966
28	Mantanani Island	2	1.799
29	Mataking Island	2	1.578
30	Port Dickson	2	1.986
31	Kudat	1	1
32	Pontian	1	1
33	Pulan Tioman	1	1
34	Pulau Lang Tengah	1	1
35	Sepangar Island	1	1
36	Tuaran	1	1



(a) Effective size



(b) Constraint

FIGURE 8. Network structure of tourist flows in Malaysia based on the calculated structural hole indicator.

in the tourism flow network in Malaysia is not balanced; The streamlining of the tourism flow network is mainly “controlled” by these core nodes. Most other nodes are highly “depending” on these core nodes. During traveling, Chinese tourists take these nodes as the core tourist areas, and evacuate and gather from them. (3) The node centrality values of the neighboring core nodes had lower centrality values. Therefore, it can be assumed that the core nodes have certain constraints on the tourism flows of the surrounding attractions (i.e. it may be a structural hole in the tourism flow network).

4) STRUCTURAL HOLES MEASUREMENT OF TOURIST FLOW

Ucinet software was used to analyze the related indicators of the structural holes, and the calculated results are shown

in Table 6. The size of the nodes in Fig.8 represents the level of the structural hole indicators, and the width of the connections between the nodes represents the size of tourist flow.

Comparing the results of the structural hole indicators in Figure 8(a) and (b), the effective size and constraint were higher in Kuala Lumpur and Kota Kinabalu. It showed that they were in the center of the network, playing the role of “bridge”, which can organically connect the tourism of other cities into a network. It also be affected by other popular, surrounding, and nearby attractions. For example, tourists to Kuala Lumpur generally visit Malacca and Penang to

disperse, and Kuala Lumpur has formed a structural hole with Malacca and Penang. Kuala Lumpur strongly constrains tourism in Kota Kinabalu because Kota Kinabalu is situated in East Malaysia, and Kuala Lumpur is located in West Malaysia. Both cities have direct flights from China. Most of the tourists who go to East Malaysia choose to fly directly to Kota Kinabalu, so there was a certain competition between these two places. The surrounding nodes such as Kudat, Pontian, Pulau Tioman, Pulau Lang Tengah, Sepangar Island, and Tuaran have low effective value and high constraint value, which indicates that the development of tourist flow in these surrounding areas was limited and that attractions were at a disadvantage in the competition to attract tourists.

V. CONCLUSION

This paper discussed the temporal pattern and complex network of tourists' behavior as the research object, utilized the tourism itinerary text of Qunar to construct a tourist trajectory database, and applied the mathematical statistics and complex network analysis methods to analyze the spatial-temporal behavior of Chinese tourists in Malaysia in detail. The main conclusions are as follows.

- 1) As Malaysia has become increasingly popular with Chinese tourists as a tourist destination, the number of tourists is increasing annually. However, emergencies and national policies have a great influence on the outbound travel of Chinese tourists, and there is a clear difference in time for Chinese tourists to visit Malaysia. Summer and winter were the peak seasons for Chinese tourists to visit Malaysia. They usually stay for 5 - 7 days and visit one or two famous tourist destinations in Malaysia. Most of them will choose traditional and popular destinations.
- 2) Malaysia's tourism space is divided into three communities. In Malaysia, apart from competition among domestic islands, Thailand and Singapore were the main competitors for Chinese tourists. How to deal with the competition among islands, how to compete with other countries, and better retain Chinese tourists are particularly important for Malaysia's future tourism construction.
- 3) Kuala Lumpur, Kota Kinabalu, Malacca, Semporna, and Penang, etc., are the core destinations of Malaysia's tourism flow network. Chinese tourists usually use these spots as centers for gathering and spreading. Kuala Lumpur plays a central role, connecting other destinations around the network and having a strong agglomeration and diffusion effect on the tourism spatial pattern.
- 4) Kuala Lumpur, Kota Kinabalu, Penang, Langkawi, Malacca have convenient transportation. These nodes are closely linked to other nodes, and have the largest number of tourist destination combinations, which have obvious radiation effects on other tourist destinations. In the planning of tourism routes, the core position of key nodes and the advantages of the

distribution center should continue to be exerted to promote the flow and dissemination of tourists, broaden the range of tourism routes, and drive the scale with the brand.

- 5) The higher level of structural holes in the tourism flow network hindered the tourism development of other attractions in this region to some extent. Malaysia's marketing and tourism investment in China mainly focuses on developed traditional tourism regions that lack motivation for endogenous development. Many regions, including Kapalai Island, Kuala Selangor, Kuantan, Mantanani Island, Matakang Island, Port Dickson, Pulau Pangkor, Kudat, Pontian, Pulau Tioman, Pulau Lang Tengah, Sepangar Island, Tuaran, Ranau, etc. were experiencing relatively slow growth in tourism. It should improve its competitive advantage and promote balanced development of the tourism flow network by increasing transport accessibility and strengthening links to the core area.

A. THEORETICAL IMPLICATIONS

This paper proposes a research framework for analyzing the spatio-temporal behavior characteristics of tourists based on tourists' digital footprint data of online tourism itineraries. This framework combines the traditional spatial quantitative analysis method with complex network analysis, providing a new perspective for the future study of tourists' spatio-temporal behavior. The framework combines community discovery with centrality and structural holes and analyzes the tourism spatial pattern from "implicit" and "explicit" aspects. It provides a comprehensive research framework for tourism spatial patterns and enriches theoretical research on the tourism flow network structure.

B. POLICY IMPLICATIONS

Based on the research results above, combined with the outbreak of the COVID-19 pandemic in 2020, Malaysia's tourism is faced with a sharp decline in tourist sources, homogeneous competition, periodic operating difficulties, etc. This paper proposes following tourism policy suggestions:

- 1) The government should give the highest priority to the role of core cities. On the one hand, the government needs to deeply explore and improve the cultural connotations and quality of urban tourism products. The government must actively guide tourists to transfer and spread to the next destination and develop more suitable tourism routes for short-term travel.
- 2) the government should conduct institutional and technological innovation to further optimize the tourism ecosystem, including policies, environment, infrastructure and cultural resources, etc., which are conducive to tourism development in the Chinese market. During the epidemic, Chinese tourists could not leave the country. Malaysia can create an immersive tourism experience using 5G, virtual reality, and other cross-application technologies. Showing unique and

eye-catching cultural relics, natural scenery, and folk customs of Malaysia to Chinese short-video users via live videos of Malaysian tourism. Based on the existing urban tourism, the country should vigorously promote emerging or potential tourist destinations, take advantage of resources, ensure sustainable development and enhance Malaysia's international tourism competitiveness.

- 3) the Malaysian government should focus on the formulation and implementation of marketing strategies in the Chinese market, integrate regional resources, enrich the perception of tourist destinations, refine the tourism market to meet the needs of different tourist groups, guide the diverse development of featured tourism products, and further promote regional collaborations. The government should fully use a digital marketing strategy and carry out spatial drainage through product mix. Meanwhile, relying on big data on tourists' consumption behavior, digital precision marketing must be carried out to effectively attract potential Chinese tourists.

VI. LIMITATIONS AND FURTHER STUDY

Limitations of the research framework and suggestions for subsequent research are as follows: First, there are some deficiencies in the digital footprint data source we used. Duggan points out that young and educated travelers are more likely to use online travel sites. We must combine online travel data with official survey data in the future. Meanwhile, the existing data can be supplemented by questionnaires and other methods to better conduct corresponding analysis. Second, With regard to the infrastructure changes, which were not specifically considered in the course of the study, due to this study scope limitations, this is a shortcoming in our study and can be further explored in future studies. Third, the research framework lacks information about tourists' non-spatial behavior. This study only uses POI data of tourists' itineraries and does not analyze the contents of tourism notes. This information can help us analyze the deep-seated reasons for forming tourism spatial patterns and the driving factors of tourists' behavior. That will be the next step. Finally, the COVID-19 pandemic has heavily impacted the global tourism industry. How to meet the needs of tourists during the pandemic and provide an innovative way for the full recovery of the tourism industry in the post-pandemic era is also an area we need to study further.

REFERENCES

- [1] D. Xinjun et al., "Exploration and practice of tourist behavior research based on international time geography," *Econ. Geogr.*, vol. 36, no. 8, pp. 183–188, 2016.
- [2] B. A. Peterson, E. E. Perry, M. T. J. Brownlee, and R. Sharp, "The transient nature of concentrated use at a national park: A spatiotemporal investigation into visitor behavior," *J. Outdoor Recreation Tourism*, vol. 31, Sep. 2020, Art. no. 100310.
- [3] D. G. Pearce, "An integrative framework for urban tourism research," *Ann. Tourism Res.*, vol. 28, no. 4, pp. 926–946, Jan. 2001.
- [4] A. A. Lew and B. McKercher, "Trip destinations, gateways and itineraries: The example of Hong Kong," *Tourism Manage.*, vol. 23, no. 6, pp. 609–621, Dec. 2002.
- [5] Y. Liu, J. Shi, and M. Jian, "Understanding visitors' responses to intelligent transportation system in a tourist city with a mixed ranked logit model," *J. Adv. Transp.*, vol. 2017, pp. 1–13, Jan. 2017, doi: 10.1155/2017/8652053.
- [6] W. Chen and E. Al, "Research on temporal and spatial behavior of domestic tourists based on tourism digital footprint—Taking Wuyishan scenic spot as an example," *Tourism Manage.*, vol. 75, no. 5, pp. 52–64, 2020.
- [7] X. Chen, I. Calder, and B. Mak, "China's second-child baby boom and fertility tourism: Strategic considerations for Malaysia," *J. Destination Marketing Manage.*, vol. 15, Mar. 2020, Art. no. 100377, doi: 10.1016/j.jdmm.2019.100377.
- [8] P. Mura and R. Tavakoli, "Tourism and social capital in Malaysia," *Current Issues Tourism*, vol. 17, no. 1, pp. 28–45, Jan. 2014, doi: 10.1080/13683500.2012.718320.
- [9] Z. B. A. Hassan, M. A. K. B. Jailani, and F. A. Rahim, "Assessing the situational analysis of heritage tourism industry in Melaka," *Proc.-Social Behav. Sci.*, vol. 130, pp. 28–36, May 2014.
- [10] G. M. Ghani, "Tourist arrivals to Malaysia from muslim countries," *Tourism Manage. Perspect.*, vol. 20, pp. 1–9, Oct. 2016, doi: 10.1016/j.tmp.2016.06.003.
- [11] T. H. Cham, Y. M. Lim, B. C. Sia, J. H. Cheah, and H. Ting, "Medical tourism destination image and its relationship with the intention to revisit: A study of Chinese medical tourists in Malaysia," *J. China Tourism Res.*, vol. 17, no. 2, pp. 1–29, 2020.
- [12] C.-H. Puah, S.-H. Huan, and F.-T. Thien, "Determinants of Chinese demand for tourism in Malaysia," *Bus. Econ. Horizons*, vol. 14, no. 3, pp. 501–512, 2018, doi: 10.15208/beh.2018.36.
- [13] X. Jin and Y. Wang, "Chinese outbound tourism research: A review," *J. Travel Res.*, vol. 55, no. 4, pp. 440–453, 2016.
- [14] Y. Liu, K. Huang, J. Bao, and K. Chen, "Listen to the voices from home: An analysis of Chinese tourists' sentiments regarding Australian destinations," *Tourism Manage.*, vol. 71, pp. 337–347, Apr. 2019.
- [15] H. Han, S. Kim, and F. E. Otoo, "Spatial movement patterns among intra-destinations using social network analysis," *Asia Pacific J. Tourism Res.*, vol. 23, no. 8, pp. 806–822, Aug. 2018, doi: 10.1080/10941665.2018.1493519.
- [16] Y. Ma, A. Hardy, and C. S. Ooi, "Researching Chinese tourists on the move," *J. China Tourism Res.*, vol. 16, no. 2, pp. 214–229, Apr. 2020, doi: 10.1080/19388160.2019.1607794.
- [17] X. Fu, L. Cai, and X. Lehto, "Framing Chinese tourist motivations through the lenses of confucianism," *J. Travel Tourism Marketing*, vol. 34, no. 2, pp. 149–170, Feb. 2017, doi: 10.1080/10548408.2016.1141156.
- [18] J. Wen, T. Ying, D. Nguyen, and S. Teo, "Will tourists travel to post-disaster destinations? A case of 2019 Australian bushfires from a Chinese tourists' perspective," *Tourism Recreation Res.*, vol. 45, no. 3, pp. 420–424, Jul. 2020.
- [19] L. Wang, D. Fong, and R. Law, "Travel behaviors of mainland Chinese visitors to Macao," *J. Travel Tourism Marketing*, vol. 33, no. 6, pp. 854–866, Jul. 2016, doi: 10.1080/10548408.2015.1068266.
- [20] Y. Cai, J. Ma, and Y.-S. Lee, "How do Chinese travelers experience the Arctic? Insights from a hedonic and eudaimonic perspective," *Scandin. J. Hospitality Tourism*, vol. 20, no. 2, pp. 144–165, Mar. 2020.
- [21] M.-Y. Wu and P. L. Pearce, "Tourism blogging motivations: Why do Chinese tourists create little 'lonely planets'?" *J. Travel Res.*, vol. 55, no. 4, pp. 537–549, Apr. 2016.
- [22] X. Peng and Z. Huang, "A novel popular tourist attraction discovering approach based on geo-tagged social media big data," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 7, p. 216, Jul. 2017.
- [23] B. Zeng, "Pattern of Chinese tourist flows in Japan: A social network analysis perspective," *Tourism Spaces*, vol. 7, pp. 42–64, Jul. 2021.
- [24] C. Jin, J. Cheng, and J. Xu, "Using user-generated content to explore the temporal heterogeneity in tourist mobility," *J. Travel Res.*, vol. 57, no. 6, pp. 779–791, Jul. 2018.
- [25] J. Qin, C. Song, M. Tang, Y. Zhang, and J. Wang, "Exploring the spatial characteristics of inbound tourist flows in China using geotagged photos," *Sustainability*, vol. 11, no. 20, p. 5822, Oct. 2019.
- [26] L. Yaping and E. Al, "Analysis on the moving track of the center of gravity of China's tourism flow to ASEAN and the state of tourism market," *Tourism Sci.*, vol. 173, no. 4, pp. 89–99, 2019.

- [27] Q. Yao, Y. Shi, H. Li, J. Wen, J. Xi, and Q. Wang, "Understanding the tourists' spatio-temporal behavior using open GPS trajectory data: A case study of Yuanmingyuan park (Beijing, China)," *Sustainability*, vol. 13, no. 1, p. 94, Dec. 2020, doi: [10.3390/su13010094](https://doi.org/10.3390/su13010094).
- [28] J. A. Duro, "Seasonality of tourism: A new decomposition," *Tourism Econ.*, vol. 24, no. 5, pp. 615–621, Aug. 2018.
- [29] J. Xia, P. Zeepongsekul, and C. Arrowsmith, "Modelling spatio-temporal movement of tourists using finite Markov chains," *Math. Comput. Simul.*, vol. 79, no. 5, pp. 1544–1553, Jan. 2009.
- [30] Y. Asakura and T. Iryo, "Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument," *Transp. Res. A, Policy Pract.*, vol. 41, no. 7, pp. 684–690, Aug. 2007.
- [31] J. Xia, F. H. Evans, K. Spilbury, V. Ciesielski, C. Arrowsmith, and G. Wright, "Market segments based on the dominant movement patterns of tourists," *Tourism Manage.*, vol. 31, no. 4, pp. 464–469, Aug. 2010.
- [32] Z. Li et al., "Research on the evolution of urban tourism flow network structure based on online travel notes—A case study of Beijing," *Geograph. Sci.*, vol. 40, no. 2, pp. 298–307, 2020.
- [33] H. Seok, G. A. Barnett, and Y. Nam, "A social network analysis of international tourism flow," *Qual. Quantity*, vol. 55, no. 2, pp. 419–439, Apr. 2021.
- [34] M. Girvan and M. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [35] A. Java, X. Song, T. Finin, and B. Tseng, "Why we Twitter: Understanding microblogging usage and communities," in *Proc. 9th WebKDD 1st SNA-KDD Workshop Web Mining Social Netw. Anal.*, 2007, pp. 56–65.
- [36] M. Newman, "The structure of scientific collaboration networks," in *The Structure and Dynamics of Networks*, Oct. 2011, pp. 189–202.
- [37] Y. Cao, J. Liu, M. Shi, B. Cao, X. Zhang, and Y. Wang, "Relationship network augmented web services clustering," in *Proc. IEEE Int. Conf. Web Services (ICWS)*, Jul. 2019, pp. 247–254.
- [38] T. Nakanishi, M. Fujii, and A. Awazu, "Self-organization of diverse directional hierarchical networks in simple coupled maps with connection changes," 2021, *arXiv:2112.07212*.
- [39] D. Nan and X. Zhang, "Prediction of hot regions in protein-protein interactions based on complex network and community detection," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, Dec. 2013, pp. 17–23.
- [40] L. Becchetti, A. Clementi, P. Manurangsi, E. Natale, F. Pasquale, P. Raghavendra, and L. Trevisan, "Average whenever you meet: Opportunistic protocols for community detection," 2017, *arXiv:1703.05045*.
- [41] D. Rui et al., "Detecting the dynamics of urban network structure through the growth of rail networks, a case study of the greater Kuala Lumpur," *Phys. A, Stat. Mech. Appl.*, vol. 503, pp. 800–817, Jan. 2017.
- [42] F. Girardin, F. D. Fiore, C. Ratti, and J. Blat, "Leveraging explicitly disclosed location information to understand tourist dynamics: A case study," *J. Location Based Services*, vol. 2, no. 1, pp. 41–56, Mar. 2008.
- [43] N. Mou, Y. Zheng, T. Makkonen, T. Yang, J. Tang, and Y. Song, "Tourists' digital footprint: The spatial patterns of tourist flows in Qingdao, China," *Tourism Manage.*, vol. 81, Dec. 2020, Art. no. 104151.
- [44] K. S. Kung, K. Greco, S. Sobolevsky, and C. Ratti, "Exploring universal patterns in human home-work commuting from mobile phone data," *PLoS ONE*, vol. 9, no. 6, Jun. 2014, Art. no. e96180.
- [45] I. Önder, W. Koerbitz, and A. Hubmann-Haidvoge, "Tracing tourists by their digital footprints: The case of Austria," *J. Travel Res. Int. Assoc. Travel Res. Marketing Professionals*, vol. 55, no. 5, pp. 566–573, 2016.
- [46] Y. Gao, C. Ye, X. Zhong, L. Wu, and Y. Liu, "Extracting spatial patterns of intercity tourist movements from online travel blogs," *Sustainability*, vol. 11, no. 13, p. 3526, Jun. 2019.
- [47] J. M. S. Lam, H. Ismail, and S. Lee, "From desktop to destination: User-generated content platforms, co-created online experiences, destination image and satisfaction," *J. Destination Marketing Manage.*, vol. 18, Dec. 2020, Art. no. 100490.
- [48] N. Mou, R. Yuan, T. Yang, H. Zhang, J. Tang, and T. Makkonen, "Exploring spatio-temporal changes of city inbound tourism flow: The case of Shanghai, China," *Tourism Manage.*, vol. 76, Feb. 2020, Art. no. 103955.
- [49] K. Stock, "Mining location from social media: A systematic review," *Comput., Environ. Urban Syst.*, vol. 71, pp. 209–240, Sep. 2018.
- [50] Y. Zheng, N. Mou, L. Zhang, T. Makkonen, and T. Yang, "Chinese tourists in Nordic countries: An analysis of spatio-temporal behavior using geo-located travel blog data," *Comput., Environ. Urban Syst.*, vol. 85, Jan. 2021, Art. no. 101561.
- [51] M. E. J. Newman, "The structure of scientific collaboration networks," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 2, pp. 404–409, Jan. 2001.
- [52] T. Luthe and R. Wyss, "Resilience to climate change in a cross-scale tourism governance context: A combined quantitative-qualitative network analysis," *Ecol. Soc.*, vol. 21, no. 1, p. 27, 2016.
- [53] X. Ye, B. She, and S. Benya, "Exploring regionalization in the network urban space," *J. Geovisualization Spatial Anal.*, vol. 2, no. 1, pp. 1–11, Jun. 2018.
- [54] H. Shao, Y. Zhang, and W. Li, "Extraction and analysis of city's tourism districts based on social media data," *Comput., Environ. Urban Syst.*, vol. 65, pp. 66–78, Sep. 2017.
- [55] G. R. Salancik and R. S. Burt, "Structural holes: The social structure of competition," *Administ. Sci. Quart.*, vol. 40, no. 2, p. 345, 1995.
- [56] E. Lazega and R. S. Burt, "Structural holes: The social structure of competition," *Revue Française de Sociologie*, vol. 36, no. 4, p. 779, Oct. 1995.
- [57] A. Yaakop, "Tourist satisfaction in Malaysia," *Int. J. Bus. Social Sci.*, vol. 4, no. 5, pp. 221–226, May 2013.
- [58] L. P. Foo, M. Y. Chin, K. L. Tan, and K. T. Phuah, "The impact of COVID-19 on tourism industry in Malaysia," *Current Issues Tourism*, vol. 24, no. 19, pp. 1–5, 2020.
- [59] A. R. A. Arokiasamy, P. M. R. Smith, and T. Kijbumrung, "Conceptualizing post-COVID-19 Malaysia's tourism recovery: An auto-regressive neural network analysis," *Emerg. Sci. J.*, vol. 5, pp. 119–129, Sep. 2021.
- [60] N. Z. I. Hussin and A. Buchmann, "Understanding tourism development policies in Malaysia," *J. Policy Res. Tourism, Leisure Events*, vol. 11, no. 2, pp. 333–353, May 2019, doi: [10.1080/19407963.2018.1516091](https://doi.org/10.1080/19407963.2018.1516091).
- [61] F. Habibi, "The determinants of inbound tourism to Malaysia: A panel data analysis," *Current Issues Tourism*, vol. 20, no. 9, pp. 1–22, 2016.
- [62] R. Suppiah and D. P. Selvaratnam, "Tourism demand in Malaysia: A panel data analysis of 17 OECD countries," *J. Tourism, Hospitality Culinary Arts*, vol. 12, no. 3, pp. 52–64, 2021.



ZHENG CAO received the B.A. and M.A. degrees from the Zhengzhou University of Light Industry, China, in 2009 and 2012, respectively. She is currently pursuing the Ph.D. degree with the Management and Science University, Malaysia. Her research interest includes tourism management.



HENG XU received the bachelor's degree in management from the Zhengzhou University of Aeronautics, China, and the master's and Ph.D. degrees in management science and engineering from Northwestern Polytechnical University, China. He is currently serving as an Associate Professor with the Henan University of Technology, China, and the Co-Supervisor for Ph.D. program cum research fellow of Management and Science University, Malaysia. His research interests include text mining, sentiment analysis and its applications in E-commerce, online learning and public health, especially the behavior and psychological mechanism of consumers in the social network environment.



BRIAN TEO SHENG XIAN received the B.Sc. (Hons.) and Ph.D. degrees from Universiti Putra Malaysia (UPM), Malaysia, in 2008 and 2014, respectively. He is currently an Associate Professor at the Management and Science University (MSU) and heading the International Academic Affairs Department at MSU. His years of experience in multidisciplinary research in both business and scientific world have garnered unique areas of expertise. His research interests include a broad spectrum of areas, including data mining, data analytics, causal reasoning, scientific understanding of learning, data visualization, medical tourism, strategic management, complementary and alternative medicine, cosmeceuticals and skin care, and colloidal science.

• • •