Exploring tourism networks in the Guangxi mountainous area using mobility data from user generated content

LIU Yan-hua[1,3] https://orcid.org/0000-0002-5329-0377; e-mail: lyh_lucy@nnnu.edu.cn

CHENG Jian-quan[2,3]* https://orcid.org/0000-0001-9778-9009; e-mail: J.Cheng@mmu.ac.uk

LYU Yu-lan[4, 5] https://orcid.org/0000-0003-3242-349X; e-mail: yulanlv0729@163.com

**Correspondence author*

1 School of Natural Resources and Surveying, Nanning Normal University, Nanning 530001, China

2 Department of Natural Sciences, Manchester Metropolitan University, Manchester M1 5GD, UK

3 Joint Research Center for Urban Health and Security Intelligent Data Analytics, Nanning Normal University, Nanning 530001, China

4 School of Finance, Shandong University of Finance and Economics, Jinan 250014, China

5 Shandong Key Laboratory of Blockchain Finance, Jinan 250014, China

Citation: Liu YH, Cheng JQ, Lyu YL (2022) Exploring tourism networks in the Guangxi mountainous areas using mobility data from user generated content. Journal of Mountain Science 19(2). https://doi.org/10.1007/s11629-021-6883-3

Abstract: Tourism-led economic growth and tourism-driven urbanization have attracted increasing attention by provinces and regions in China with abundant tourism resources. Due to low data availability, the current tourism literature lacks empirical evidence of the tourism network in less-developed mountainous regions where the development of transport infrastructure is more variable. This paper aims to provide such evidence using Guangxi Zhuang Autonomous Region in China as a case study. Using User Generated Content (UGC) data, this study constructs a tourism network in Guangxi. By integrating social network analysis with spatial interaction modelling, we compared the impact of two different transport infrastructures, highway and high-speed railway, on tourist flows, particularly in less-developed mountainous regions. It was found that the product of node centrality and flow could best describe the significant pushing and pulling forces on the flow of tourists. The tourism by high-speed railway was sensitive to the position of trip destination on the whole tourism network but self-drive tourism was more sensitive to travelling time. The increase of high-speed railway density is crucial to promote local tourism-led economic development, however, large-scale karst landforms in the study area present a significant obstacle to the construction of high-speed railways.

Keywords: Tourism network; Mountainous region; User Generated Content; Social network analysis; Spatial interaction modelling; Guangxi

1 Introduction

Rapid urbanisation in China over the past four decades has enabled massive economic development. As a result China has been the second largest economy in the world since 2011. However, there has been an imbalance of economic development between provinces (Shen 2018). Some provinces, particularly those in the western China, have experienced economic-related inequality (Li et al. 2016) due to the slower processes of industrialization and urbanization than in eastern China. Tourism-led economic growth (Liu et al. 2017; Pratt 2015) and tourism-driven urbanization (Qian et al. 2012) have attracted increasing attention in provinces and regions with unique tourism resources. Guangxi Zhuang Autonomous Region (hereafter referred to as Guangxi), which is equivalent to a province in China, contains historical revolutionary sites, ethnic minority autonomous zones, international land border areas, and less-developed regions. It has abundant tourism resources, such as the well-known tourism city of Guilin (Hao et al. 2016), but it is characterised by regional inequality of economic development and faces the challenge of reducing poverty (Dai et al. 2018). In terms of the number of inbound overnight tourists, Guangxi was ranked fifth in 2015, accounting for 4.56% of the total inbound overnight tourists in China (NSB 2016). The impact of tourism as a driver of regional development in less developed regions, such as Guangxi, is rarely studied in the literature in China due to poor data availability. A deeper understanding of the tourism network in Guangxi will help to make better use of the tourism industry for the purposes of reducing inequality and alleviating poverty.

From the perspective of tourists, tourist destinations are regarded as geographic areas in which tourists can enjoy various tourism experiences (Bálint and Mátyás 2021). Destinations span various spatial scales, including specific places (Liu et al. 2018; Li, Ding and Wang 2016; Gao and Wu 2017), cities (Jin et al. 2018; Zhang et al. 2014), provinces (Liu et al. 2017), national (Wang et al. 2018; Liu and Nijkamp et al. 2017; Huang 2016) and global (Bendle 2018; Khadarooa and Seetanah 2008; Lozano and Gutiérrez 2018; Yahya 2003; Marrocu and Paci 2011). A tourism network is composed of attraction sites (nodes) and the transport connection (links) between these nodes. A tourism network can be viewed as a platform for tourism mobility. Flow of tourists on the tourism network is a quantitative measure of tourism mobility. The transportation system, which in this study includes highways and high-speed railways, allows tourists to flow on the network. In a large-scale tourism destination, such as a province or nation, tourism nodes include multiple cities, counties or attractions, and the flow of tourists between the nodes forms a network. Therefore, tourism destinations have been increasingly recognized as a network (Asero et al. 2016; Jin et al. 2018; Lozano and Gutiérrez 2018). To date tourism studies have largely focused on national and city scales and less on the provincial scale. There is a gap in the literature to understand tourism networks on a provincial scale (Liu et al. 2017) and its links to economic development and regional inequality. In this study the analysis of the tourism network in Guangxi provides systematic evidence on the spatial relationships between nodes, and reveals the spatial disparity of tourism development in a less-developed region with a unique Karst morphology.

The interdependent relationship between transport and tourism is well-established for multiple modes of transport including air transport (Spasojevic et al. 2018; Wu et al. 2012; Wu and Pan 2010), high-speed railways (Pagliara et al. 2017; Guirao et al. 2016; Wang et al. 2018; Wang and Chen et al. 2015; Wang et al. 2017; Wang and Niu et al. 2015), self-driving on motorways (Moyano et al. 2016; Jin et al. 2018), and cycling (Lumsdon 2000). Tourism-related transport contributes significantly to local economic development, although it also produces CO_2 emissions (Luo et al. 2018; Jin et al. 2018). As a result, sustainable tourism has become a main concern in tourism studies (Peeters et al. 2019). In China, the rapid development and large-scale coverage of the high-speed railway system has stimulated the tourism industry at a variety of scales. The impacts of the high-speed railway system have attracted growing concerns (Albalate and Fageda 2016; Wang et al. 2018; Wang et al. 2017; Masson and Petiot 2009). These studies have revealed a "time-space compression" effect of the high-speed railway system which enhances the spatial linkages between tourism cities and areas, shortens the psychological distance of tourists, reduces the time control risk, and promotes the marketing and spatial layout of tourism resources, and integration of tourism products. Self-driving tourism (Zhou and Huang 2016; Luo et al. 2018) has become an increasingly popular mode of tourism transport particularly for younger and middle-class tourists. Current literature on the relationship between transport and tourism has mainly focused on the network structure of tourist flows linked to a certain transport system. Comparative analysis studies of the effects of high-speed railway and highways on tourist flows within a tourism network are relatively rare. The accessibility of tourist attractions by highway and high-speed railway differs due to the speed, frequency and availability of the transport network (Wang et al. 2021; Moyano et al. 2016). The choice of travel mode by tourists is usually based on multiple factors including financial and time budgets. A comparison of the impacts of two transport systems on the tourism network would help explain how tourists use the different transport infrastructures and where improvements are needed spatially and temporally.

Data collection for tourism studies typically relies on questionnaire surveys (e.g., Liu et al. 2017; Yang et al. 2007) and interviews. However, the development of new technologies including sensors, internet and smart devices has stimulated the production of tourism big data (Li et al. 2018), which has high volume, high variability and high velocity (e.g., photographs and review texts). The development of User Generated Content (UGC) contributes to data collection as users share their original content through the Internet. YouTube, Flickr, and blogs are the main platforms for UGC. Travel strategies or online travel blogs are also a form of UGC for tourists. For example, Tourist A: 'day 1 at attraction site of Beihai Silver Beach, day 2 at the attraction site of Qingxiu mountain.... travel by private car on highway'; Tourist B: 'day 3 travelling from Guilin city to Yangshuo county by high-speed railway'. UGC data has been increasingly and extensively deployed in tourism studies (Jin et al. 2018). In particular, UGC data can be used to construct tourism networks and conduct social network analysis based on the tracked mobility of tourists. Social network analysis is a quantitative social space analysis method including several descriptive indicators, which can be used to measure, analyse and visualise the characteristics of the constructed social network (Zhao et al. 2021), but it cannot analyse the effects of transport mode. Gravity models of spatial interaction are able to link attraction sites and transport infrastructure such that tourist flows between attraction sites can be explained by the attributes of the sites and transport performance. The key advantage of a gravity model over social network analysis is the capacity to analyse and predict spatial interactions, which addresses the deficiencies of social network analysis methods for analysing the impact of transport modes on tourist flow.

To fill the research gaps described above, this study aims to model the effects of high-speed railway and highway systems on a tourism network using UGC data and integrating social network analysis into spatial interaction modelling, using Guangxi as a case study area. The contributions of this study are as follows: (1) Guangxi is a less-developed mountainous area, and tourism is a strategic pillar industry in this province. Research on tourism networks in Guangxi enriches current understanding of tourism-led economic growth in less-developed regions. (2) The application of UGC data provides a new method for the construction of tourism networks and outperforms the traditional data collection method of questionnaire surveys in tourism research. (3) Taking a province as the unit of spatial scale, and counties as the analytical unit, the characteristics and existing problems of the tourism networks are reflected on a more detailed basis than previously described which enriches the spatial perspective of tourism network studies. (4) Combining a gravity model and negative binomial regression, this study presents a comparative analysis of the impacts of highway and high-speed railway on the distribution of tourist flows over the tourism network. This will provide insight into tourism-related problems and provide quantitative evidence for local governments to improve and coordinate the tourism and transport sectors.

2 Materials and Methods

2.1 Study area

Guangxi Zhuang Autonomous Region is located in southern China. It extends between 20°54' and 26°23'N and 104°28' to 112°04' E (Fig. 1), occupying an area of 237,600 km². It is a major tourism province in China with abundant tourism resources. For example, the landscape in Guilin has been described as, "green hills, clear water, fantastic caves and spectacular rocks". Other resources include subtropical coastal vacation locations in Beihai Silver Beach and Weizhou Island; the customs of the Zhuang and Yao nationalities; China-Vietnam border customs and transnational tourism resources in Dongxing and Pingxiang counties; cultural relics and historical sites along the Lingqu Canal in Xing'an county and Huashan rock art in Ningming county; a national forest park by Mao'er Mountain, the first peak in Southern China; small town leisure tourism resources in Huangyao ancient town; leisure and health tourism resources in Bama county, the

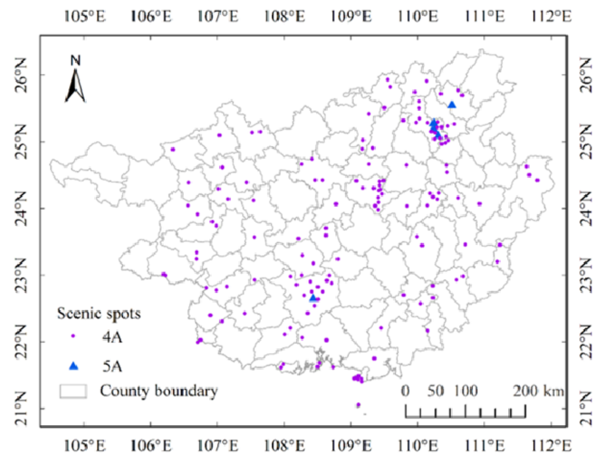

Fig. 1 Location of the study area and distribution of 4A, 5A scenic spots in Guangxi Province, China by 2017.

hometown of longevity; and communism cultural tourism resources in Baise city. In accordance with the Measures for the Administration of Quality Grades of Tourist Attractions (NTA 2012), all tourist attractions in China have been graded into five levels from 1A to 2A, 3A, 4A and 5A in ascending order. As of 2017, Guangxi had 173 4A and 5 5A tourism attractions at national level, including the Lijiang river scenic area, Solitary Beauty Peak-Prince City scenic area, Two Rivers and Four Lakes-Elephant hill scenic area in Guilin city, Lemandi leisure world in Xing'an county and Qingxiu mountain scenic area in Nanning city (Fig. 1).

2.2 Data collection

A web crawling technique was used to collect tourist flow data from the tourism web service called "Where to go?" (https://www.qunar.com/). A total of 2000 trip records between 2013-2016 with Guangxi listed as a destination were collected. The trip record included date of departure, place of departure, trip length, travel mode, and trip itinerary. After deleting incomplete or incorrect records, a total of 1664 valid trips were selected for this study. Trip destinations included more than 200 scenic spots distributed across 63 counties (cities) within Guangxi. A large volume of tourists visited Weizhou Island in Beihai city. The Island was treated as a county, creating a total of 64 spatial units for analysis. All 64 units were treated as nodes in the tourism network and the tourist flows between the units were treated as links. A 64×64 relation matrix of tourist flows was constructed based on the following rules (Fig. 2):

1) Flows (trips) between attractions (nodes) within a spatial unit (e.g., county) were not included in the matrix. Flows (trips) between attractions in different spatial units (e.g., attraction a in Unit A to attraction b in Unit B), were included as an element in the matrix and assigned a value of 1.

2) If there was a transfer or short stay in a spatial unit but no scenic spots were visited, this flow was treated as invalid.

3) When multiple trips occurred between attractions in different units, the trips were summarized by the value of the matrix element from Unit A to Unit B.

The resulting 64×64 matrix had 271 valid paths (with a value > 0) with a total of 2,920 tourist flows. The distribution of these flows between 63 counties and Weizhou Island was mapped using ArcGIS 10.5 (Fig. 3).

2.3 Analytical methods

The data set included tourism resources, tourist flows and mode of transport used. The resulting

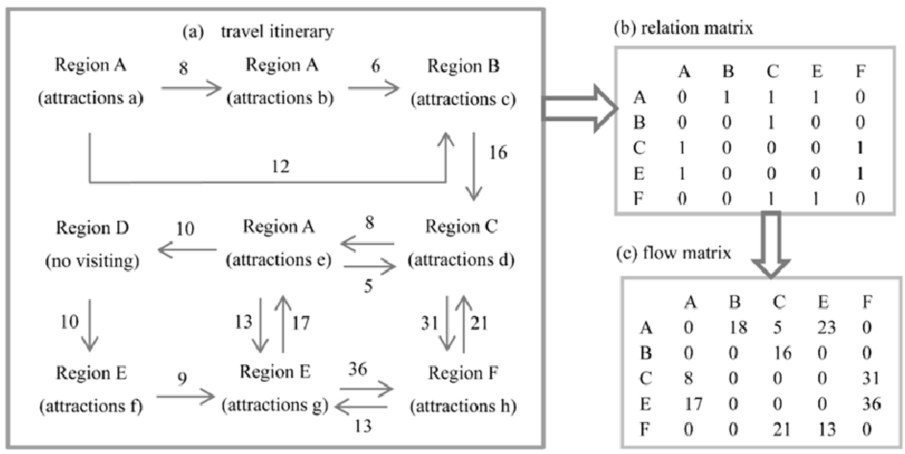

Fig. 2 A flow matrix from trip records. (a) travel itinerary, where a~h denotes tourist attractions, and A~F denote the administrative district where the attractions are located; (b) relation matrix of tourism nodes, 0 means no connection between the two nodes, 1 means there is a connection; (c) flow matrix of tourism nodes, numbers represent tourism flow.

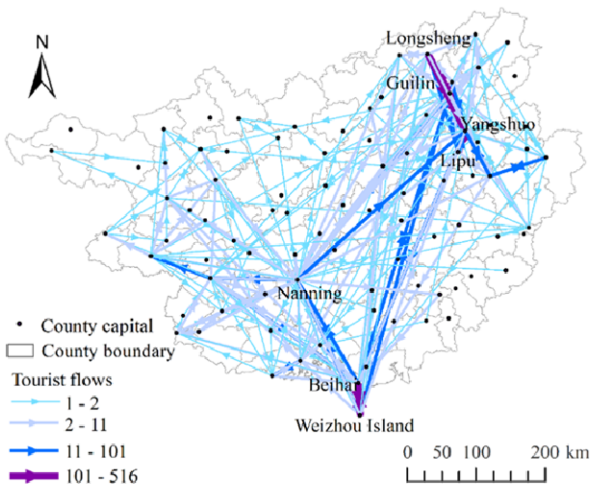

Fig. 3 Tourism networks in Guangxi Province, China. The tourism flow data is derived from the UGC released by tourists who traveled in Guangxi between 2013-2016 on "Where to go?" (https://www.qunar.com/).

tourism system and its interactions with the transport system was analysed and modelled using an assessment of tourism resources, social network analysis and spatial interaction modelling, which are explained below.

2.3.1 Integral assessment of tourism resources at county level

As mentioned in section 2.1, each scenic spot has a grade ranging from 1A to 5A, with higher scores indicating better resources. A weighting value from 1 to 5 was assigned to each category from 1A to 5A, and an aggregate tourism resource score for each county was calculated using a weighted sum as shown in Eq. (1).

$$A = \sum_{i=1}^{5} w_i \times n_i \quad (1)$$

where A is an aggregated score for a county, w_i ranges from 1 to 5 corresponding to the grade from 1A to 5A, and n_i is the total number of scenic spots in the category of grade i (1-5) within the county.

2.3.2 Social network analysis

Social network analysis aims to explore the relationship between (social) actors and their influence on a network (Adamic and Adar 2005). Existing studies show that this analysis is an effective method to study the flow structure of a tourism system (Jin et al. 2018; Lozano and Gutiérrez 2018). Processed UGC data can be used to construct flow matrices between the nodes in a tourism network, summarize the total flow of tourists between these nodes, and measure the overall size of a tourism network. In this study, network density was used to reflect the density distribution of flow, which is helpful to understand the overall distribution characteristics of the network. Network centrality was calculated to reflect the levels of connection and influence of nodes and links over the entire network, which is conducive to understanding the position of the nodes and links in the network. Network density and centrality are particularly useful for quantifying the concentration of the network and the position of nodes and links. These indicators measure the structural attributes of tourism networks from the perspective of tourist flow. In summary, network density, flows and centrality were used to analyze the structural characteristics of the constructed tourism network in Fig. 3. The definition and method of calculation of these indicators are listed in Table 1.

In Table 1, D represents the network density; E the total number of network paths; n the number of nodes in the network. F_{ini} and F_{outi} denote the inflow and outflow volumes at node i respectively. F_{ij} is the flow of trips from node i to node j; and F_{ji} the flow of trips from node j to node i. F_i represents the total flow of trips at node i. C_{ini} and C_{outi} denote the indegree and outdegree at node i respectively; c_{ij} is the number of connections from node i to node j, c_{ji} is the number of connections from node j to node i. C_i represents the centrality at node i. C_{mi} denotes the betweenness centrality of node i, and $g_{jk}(i)$ the number of short links passing through node i between nodes j and k. g_{ik} indicates the total number of shortcuts between nodes j and k, and $g_{jk}(i)/g_{ik}$ measures the probability that node i is located on the shortcut between nodes j and k. C_{rmi} is the relative betweenness centrality of node i, is the standardization of C_{mi} and reflects the influence of node i over other nodes on the network. L_{ma} indicates the betweenness centrality of link a, and l_{jk}(a) is the number of shortcuts bypassing link a between nodes j and k. l_{jk} denotes the total number of shortcuts between nodes j and k, and l_{jk}(a)/l_{jk} refers to the probability that link a is located on a shortcut between nodes j and k (Liu 2004).

2.3.3 Spatial interaction modelling

Social network analysis is useful for exploring network characteristics, however, it cannot explain the interaction between tourism and transport. To reveal their interaction, spatial interaction modelling has been deployed, combining the social network with regression analysis. A gravity model was used to

Table 1 Network indicators of tourism flow.

Indicator	Equation	Description
Network Density	$D=E/(n(n\text{-}1))$	Between 0~1, the larger the value, the higher the network density.
Inflow	$F_{ini} = \sum_{j=1}^{n} F_{ji}$	Tourist flow to node *i*
Outflow	$F_{outi} = \sum_{j=1}^{n} F_{ij}$	Tourist flow from node *i*
Total Flow	$F_i = F_{ini} + F_{outi}$	Tourists flow at node *i*
Indegree	$C_{ini} = \sum_{j=1}^{n} c_{ji}$	Number of connections to node *i*
Outdegree	$C_{outi} = \sum_{j=1}^{n} c_{ij}$	Number of connections from node *i*
Centrality	$C_i = C_{ini} + C_{outi}$	Number of connections at node *i*
Betweenness Centrality of Node	$C_{mi} = \sum_{j}^{n}\sum_{k}^{n} g_{jk}(i)/g_{jk}$	The probability that node *i* is on the shortcut of other node pairs in the network ($j \neq k \neq i$, and $j<k$)
Relative betweenness centrality of Node	$C_{rmi} = C_{mi}/(n * n - 3n + 2)$	Between 0~1, the larger the value, the stronger the ability to control other nodes.
Betweenness Centrality of Link	$L_{ma} = \sum_{j}^{n}\sum_{k}^{n} l_{jk}(\mathrm{a})/l_{jk}$	The probability that the link a is on the shortcut of other node pairs in the network ($j \neq k \neq a$, and $j<k$)

analyze and predict the spatial interactions within the network. This method has been widely applied in many mobility related studies such as transport, trade, population migration and tourism (Morley et al. 2014; Cheng et al. 2014). For example, Khadarooa and Seetanah (2008) have compared the roles of air, road and seaport transport infrastructure in international tourist flows using a calibrated gravity model. A simplified form of gravity model is typically constructed as follows:

$$G_{ij} = \frac{kP_iP_j}{f(d_{ij})} \tag{2}$$

where G_{ij} is the interaction intensity between i and j, P_i and P_j represent the push and pull forces at origin site i and destination site j respectively. d_{ij} denotes the distance between places i and j, and $f(d_{ij})$ is a distance decay function of d_{ij}. In practice, P_i and P_j are often indicated by the scaled variables at the origin and destination sites, such as population or GDP. It is argued that the distance decay function varies with spatial scale (Jin et al. 2018). For example, at an urban scale, the distance decay function $f(d_{ij})$ is often expressed as an inverse power function (Jin, Cheng and Xu 2018). At a regional scale (e.g., province), the function $f(d_{ij})$ is better expressed as a negative exponential function (e.g., $e^{-\beta d}$), where β is a spatial distance friction coefficient. A higher β value reflects greater sensitivity of flow to distance.

The volume of tourist flow results from the spatial interaction between nodes (as origin or destination sites). Similar to traffic assignment analysis in transport modelling, the distribution of these flow volumes is determined by the integral forces operating at the origin and destination sites. In this study, the flow volumes (including outflow and inflow) reflect the scale of tourism at a node, the centrality (including outdegree and indegree) represents the connection capacity of a node, and the product of flow volume and centrality, which are represented as F and C, reflects the combined forces at each node. Assuming that the outflow and outdegree have the same effect on the origin i, then P_i is represented by $F_{outi}C_{outi}$. Similarly, P_j is represented by $F_{inj}C_{inj}$. As there is no significant difference in the OD distance between the highway and high-speed railway but a significant difference in the OD travelling time between the modes of transport by highway and high-speed railway, so the distance in the distance decay function, $f(t_{ij})$, is represented by time, and a negative exponential function is used as the distance decay function.

Assuming the origin i and the destination j have different effects, the parameters a and b are used to reflect the intensity of thrust and attraction respectively. The gravity model in Eq. (2) is then modified as follows (Wilson 1970):

Table 2 The statistical characteristics of tourism network.

Threshold	Number of nodes	Number of links	Network density	Total flow	Average flow	Variance	Flow proportion
$F_{ij}\geq1$	64	271	0.067	2920	10.77	1888.63	100%
$F_{ij}\geq2$	36	135	0.107	2784	20.62	3598.01	95.3%
$F_{ij}\geq5$	18	59	0.193	2588	43.86	7272.52	88.6%
$F_{ij}\geq10$	14	34	0.187	2432	71.53	10812.01	83.3%
$F_{ij}\geq20$	10	22	0.244	2252	102.36	14013.41	77.1%
$F_{ij}\geq50$	7	12	0.286	1936	161.33	17983.72	66.3%

$$F_{ij} = k(F_{outi}C_{outi})^{a}(F_{inj}C_{inj})^{b}/e^{-\beta t_{ij}} \quad (3)$$

As flow (F_{ij}) is a discrete variable it is usually calibrated by Poisson regression analysis. This method assumes the sample's mean and variance are equal. However, in this study the mean (10.77) was much smaller than the variance (1888.63) as shown in Table 2. This is indicative of overdispersion. A negative binomial (NB) regression analysis was applied to calibrate Eq. (3) and, as shown in Eq. (4), it proved more suitable than the Poisson regression analysis (Cheng et al. 2014; Zhang et al. 2019).

Taking $X_{ij} = \{F_{outi}, C_{outi}, F_{inj}, C_{inj}, t_{ij}\}$, according to negative binomial regression (Cameron and Trivedi 2005), suppose the distribution F_{ij} of conditional on X_{ij}, v_{ij} is,

$$F_{ij}|X_{ij}, v_{ij} \sim P(u_{ij}v_{ij})$$

where P refers to the Poisson distribution with rate parameter (also its mean value) in the parenthesis. The random parameter v_{ij} indicates the unobserved heterogeneity. The parameter u_{ij} is a deterministic function of X_{ij} specified as follows,

$$ln(u_{ij}) = lnk + a \times ln(F_{outi}C_{outi}) + b \times ln(F_{inj}C_{inj}) + \beta \times t_{ij}$$

Assume v_{ij} is generated from a gamma distribution, then the marginal distribution of F_{ij}, conditional on X_{ij}, is derived by integrating out the term v_{ij},

$$F_{ij}|X_{ij} \sim NB\left(u_{ij}, f(u_{ij}, \alpha)\right) \quad (4)$$

where NB represents a negative binomial distribution with the mean and variance parameters in parentheses. To solve Eq. (4), we will assume $v_{ij} \sim Gamma(1/\alpha, \alpha)$, then adopt a maximum likelihood estimation to calculate the coefficients. It should be noted that the variance $f(u_{ij}, \alpha)$ is equal to $u_{ij} + \alpha u_{ij}$, as v_{ij} is assumed to have a gamma distribution.

In the above specification, α is a parameter to describe the variance of a negative binomial distribution, which can be used to verify the rationality of a negative binomial regression. If α is significantly different to 0, indicating that F_{ij} is over-dispersed, then it is more reasonable to use negative binomial regression.

The highway and high-speed railway networks in Guangxi extend in all directions. By 2015, the total length of road network across this province had reached 118,000 km, including 4,288 km of highway. The total length of the rail network had reached 5,086 km, including 1,811 km of high-speed railway network (GXLCCC 2016). The density of highway and high-speed railway were 1.80 km/100km² and 0.76 km/100 km² respectively. The majority of tourist trips within Guangxi are dominated by private car journeys on the highway or journeys on the public high-speed railway. These account for 95% of total tourist journeys. Therefore, these two transport infrastructures were selected as the target transport systems for tourism mobility analysis. In order to examine the effects of transport mode on tourism behaviour, Eq. (4) was split into Eq. (5) for high-speed railway, and Eq. (6) for highways:

$$F1_{ij}|X_{ij} \sim NB(u_{1ij}, f(u_{1ij}, \alpha)). ln(u_{1ij})$$
$$= lnk + a \times ln(F_{outi}C_{outi}) + b \times ln(F_{inj}C_{inj}) + \beta \times t_{1ij} \quad (5)$$
$$F2_{ij}|X_{ij} \sim NB(u_{2ij}, f(u_{2ij}, \alpha)). ln(u_{2ij})$$
$$= lnk + a \times ln(F_{outi}C_{outi}) + b \times ln(F_{inj}C_{inj}) + \beta \times t_{2ij} \quad (6)$$

where $F1_{ij}$ is the flow of tourists using high-speed railway and $F2_{ij}$ the flows of tourists using highways. t_{1ij} is the journey time by high-speed railway, collected from the official web-services of high-speed railway system (https://www.12306.cn/index/), and t_{2ij} is the journey time by highway between a pair of counties, calculated by network analysis in ArcGIS 10.5. The unit of travel time is minutes and the speed value was set at the maximum speed limit. As each node is not only the origin site of one flow but also the destination site of another flow, in this study, the subscripts In and Out were added to distinguish the two directions of tourist flow, $F_{outi}C_{outi}$ and $F_{ini}C_{ini}$. The thrust and pulling force of the nodes in the network are fixed and do not change with the transport mode chosen by tourists. The negative binomial regression was run using the Stata package.

3 Results

3.1 Link characteristics of the network

3.1.1 Flow of links

Mapping flows can be complicated by the number of overlapping lines. To better visualize the distribution of these flows, six separate maps were created showing the distributions of varied flows classified according to the magnitude of OD flows from 1 to 50 (a-f in Fig. 4): $F_{ij} \geq 1$, $F_{ij} \geq 2$, $F_{ij} \geq 5$, $F_{ij} \geq 10$, $F_{ij} \geq 20$, $F_{ij} \geq 50$. The basic measurements and their comparisons are listed in Table 2.

Table 2 and Fig. 4 show that when $F_{ij} \geq 1$, there were 271 valid links between nodes so the network density was only 0.067. When $F_{ij} \geq 2$, the number of network nodes decreased by almost 50% to 36, and there was a similar percentage reduction of valid links between the nodes. However, they accounted for 95.3% of the total number of flows. The network density increased by approximately 60%. When $F_{ij} \geq 5$, 18 network nodes and 59 links remained in the network, accounting for 88.6% of the total flow. The network density increased by 80%. When $F_{ij} \geq 50$, only 7 nodes and 12 links remained in in the network, but their flows still accounted for 66.3% of the total flow. Tourist flows tended to gather in two regions; one region including Guilin, Yangshuo, Longsheng and Lipu, had 4 nodes, 9 links, 1311 flows and a network density of 0.75, and a second region including Nanning, Beihai and Weizhou Island, had 3 nodes, 3 links, 625 flows and a network density of 0.5 (Fig. 4f).

Considering the data quality from the web service, the tourism network defined for $F_{ij} \geq 2$ was selected for further analysis. This network (Fig. 4b) had 36 nodes, 135 links and 2,784 flows, accounting for 95.3% of the total flow (Table 2). On this network, the express boat between Beihai and Weizhou Island was treated as equivalent to the high-speed railway, the slow boat was equivalent to the highway. The tourist flows by highway (with 36 nodes, 135 links and 1,933 flows) and high-speed railway (16 nodes, 48 links and 851 flows) are shown in Fig. 5.

Tourist flow by highway was mainly concentrated between Guilin (NodeID 34), Yangshuo (35), Longsheng (32), and, Lipu (31) nodes, as well as between Beihai (33) and Weizhou Island (36). The tourist flow by high-speed railway was mainly concentrated between Guilin and Yangshuo, and between Nanning (30) and Beihai.

3.1.2 Betweenness centrality of links

The betweenness centrality of each link measures the degree of control it has over the entire network. The results are displayed in Fig. 6, which demonstrates that the Beihai-Nanning link, Fangchenggang-Beihai link and Guilin-Nanning link were ranked as the most influential in the entire network.

3.2 Node characteristics of the network

Nodal indicator measures for all 36 nodes were calculated and are listed in Table 3.

3.2.1 Flow of nodes

The flow of nodes shown in Fig. 7 reveals that the top 4 outflows were Guilin (NodeID 34), Yangshuo (35), Beihai (33) and Weizhou Island (36) respectively, which collectively accounted for approximately 69% of the total outflows. The top 4 inflows were Yangshuo, Beihai, Weizhou Island and Guilin respectively, which collectively accounted for approximately 64% of the total. The top 4 total flows were Yangshuo, Guilin, Beihai and Weizhou Island respectively, together accounting for 67% of the total. The tourist flows by highway had a similar distribution trend to that of the whole tourism network. By contrast, the tourist flows by high-speed railway differed slightly from the whole network because some nodes had no high-speed railway stations, such as Longsheng (32), Lipu (31), Daxin (28), Zhaoping (29).

3.2.2 Centrality of nodes

The centrality of nodes shown in Fig. 8 indicates that the nodes with the four largest outdegree were Nanning (NodeID 30), Guilin (34), Yangshuo (35) and Longsheng (32) respectively. In terms of indegree, the top 4 nodes were Beihai, Nanning, Guilin and Yangshuo respectively, and in terms of centrality the top 4 nodes were Nanning, Guilin, Beihai and Yangshuo respectively. Nanning city, which, as the provincial capital, occupies a central position in the transport network, had the highest value of centrality. The centrality by highways had a similar distribution trend to the node centrality at most sites, whereas the centrality by high-speed railway only had a similar trend to node centrality at sites with high-speed railway stations.

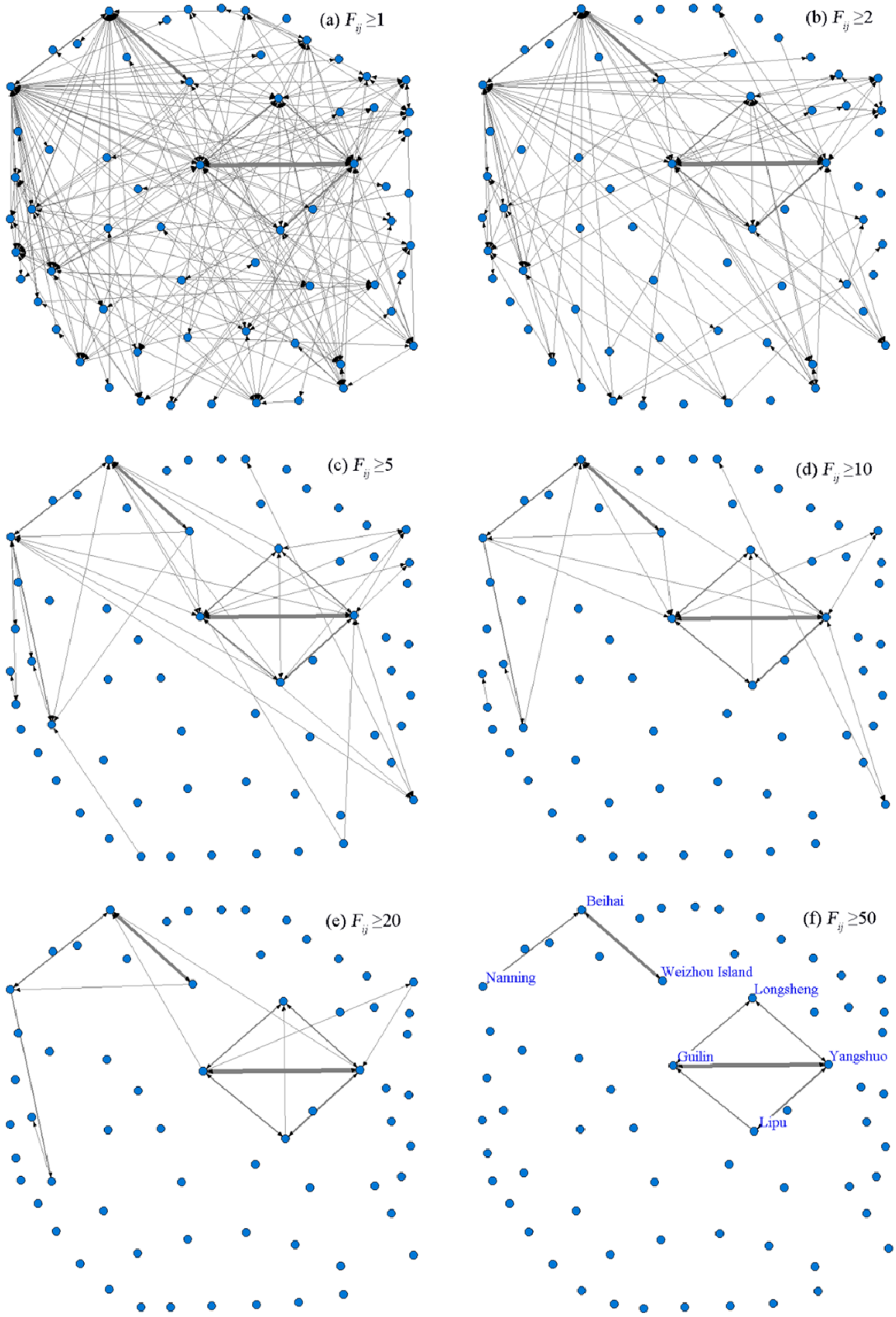

Fig. 4 Distribution of link in Guangxi tourism network under different thresholds. (a) $F_{ij} \geq 1$; (b) $F_{ij} \geq 2$; (c) $F_{ij} \geq 5$; (d) $F_{ij} \geq 10$; (e) $F_{ij} \geq 20$; (f) $F_{ij} \geq 50$. The thickness of the line in the figure indicates the amount of tourist flow, the thicker the line, the more tourist flow.

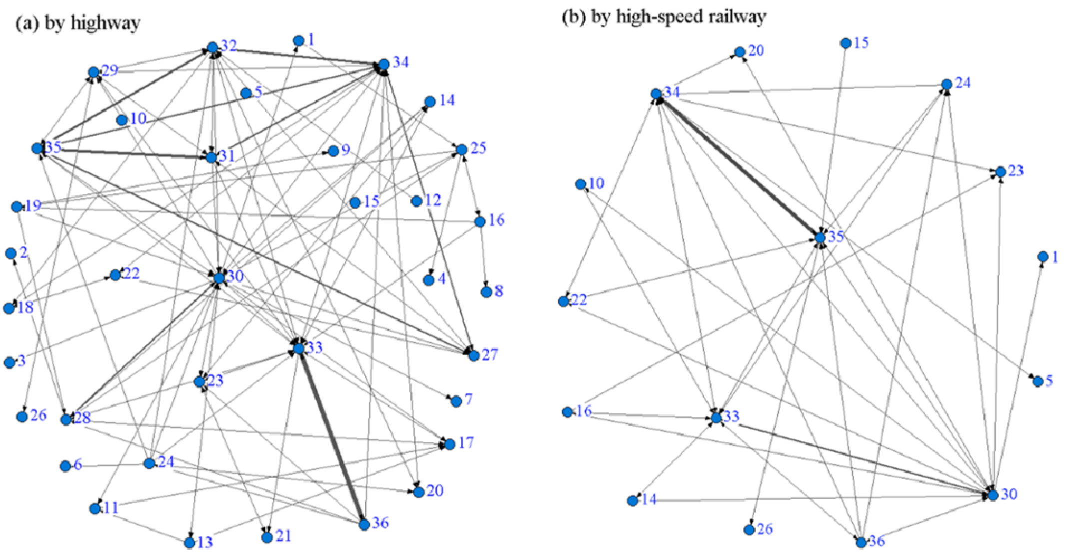

Fig. 5 Tourism network in Guangxi by highway and high-speed railway. (a) by highway; (b) by high-speed railway. The thickness of the line in the figure indicates the amount of tourist flow, the thicker the line, the more tourist flow. The number indicates the node number (NodeID), and the correspondence between NodeID and the name is mentioned in Appendix 1.

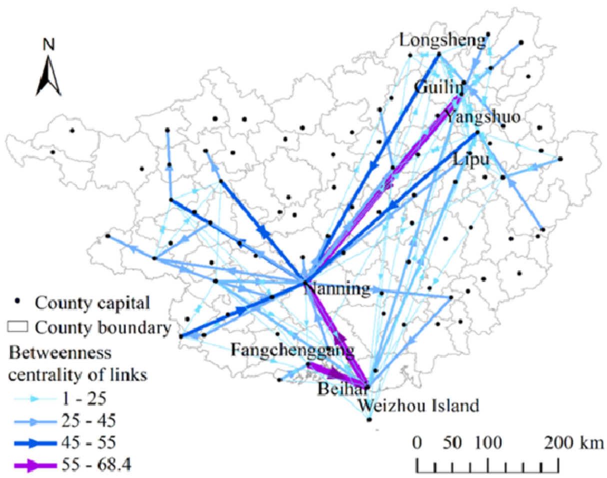

Fig. 6 Betweenness centrality of links in Guangxi tourism network based on UGC during 2013-2016.

3.2.3 Relative betweenness centrality of nodes

The relative betweenness centrality (abbreviated as rBetweenness) of nodes was used to measure the extent to which the interaction among other nodes was controlled by this node. If the rBetweenness of a node is 0, it does not control any other nodes and is at the margin of the network. If the rBetweenness of a node is 1, it controls other nodes completely and is at the core of network (Liu 2004). The rBetweenness of all nodes is shown in Fig. 9. The nodes with the four highest values of rBetweenness in the whole network were Nanning, Beihai, Guilin and Yangshuo respectively. The nodes with the four highest rBetweenness on the highway network were Nanning, Beihai, Guilin and Jingxi (NodeID 25) respectively, and on the high-speed rail network were Nanning, Beihai, Yangshuo and Guilin respectively. Regardless of the position, Nanning, Beihai, Guilin, and Yangshuo were all ranked in the top four nodes for centrality and rBetweenness. These results indicate that these four cities were core nodes in the tourism network, and they not only had a stronger connection capacity, but also a stronger controlling capacity over the other nodes. However, the degree of control of the core nodes on the Guangxi tourism network was very limited. Nanning, which had the highest level of control over the network, had an rBetweenness value of only 37.3%.

3.3 Gravity modelling by regression analyses

The empirical results of Eqs. (5) and (6) are presented in Table 4. Both models of tourist flow by either high-speed railway or highway were statistically significant. Firstly, the *alpha* parameters were all much greater than 0, which justified the use of a negative binomial regression for modelling. Secondly, all three coefficient parameters, a, b and β, which reflect pushing, pulling and spatial barrier factors, were significant at the 1% level. The pushing and pulling factors were both positive, and the spatial barrier coefficient was negative, which is consistent with the previous theoretical assumptions. Thirdly, the product of node flow and centrality showed the

Table 3 Nodal statistics of the network.

Node ID	Flow			Centrality			Betweenness Centrality
	Outflow	Inflow	Total flow	Outdegree	Indegree	Centrality	
1	2	2	4	1	1	2	0
2	0	2	2	0	1	1	0
3	0	2	2	0	1	1	0
4	0	2	2	0	1	1	0
5	0	2	2	0	1	1	0
6	0	3	3	0	1	1	0
7	0	2	2	0	1	1	0
8	0	2	2	0	1	1	0
9	0	2	2	0	1	1	0
10	2	2	4	1	1	2	0
11	3	5	8	1	2	3	0
12	2	0	2	1	0	1	0
13	6	3	9	3	1	4	0
14	7	5	12	3	2	5	0.333
15	3	0	3	1	0	1	0
16	10	5	15	5	2	7	34.583
17	10	12	22	3	4	7	34.333
18	4	10	14	2	3	5	0.4
19	9	10	19	3	3	6	50.4
20	5	10	15	2	4	6	1.25
21	4	20	24	1	2	3	0
22	11	22	33	3	5	8	6
23	19	23	42	2	7	9	37.2
24	29	9	38	7	3	10	29.417
25	16	36	52	4	4	8	60.867
26	15	20	35	1	2	3	0
27	54	63	117	5	5	10	3.65
28	52	73	125	6	6	12	70.1
29	45	53	98	6	6	12	41.333
30	214	93	307	21	12	33	443.717
31	148	216	364	7	5	12	11.1
32	201	284	485	9	8	17	64.1
33	424	366	790	8	14	22	200.017
34	770	306	1076	12	12	24	139.783
35	479	753	1232	10	11	21	110.25
36	240	366	606	7	2	9	3.167
Mean	77.33	77.33		3.75	3.75		
Standard deviation	163.5	155.0		4.32	3.61		
Variance	26734	24030		18.69	13.02		
Maximum	770	753		21	14		
Minimum	0	0		0	0		
Network density	0.107						

combined effect of push at the origin site and pull at the destination site. Comparatively, there was only a slight difference between the pushing and pulling effects on the flow of tourists by high-speed railway, but a much larger difference on flows using highway transport. The disparity shows that the pulling effect of a destination site has a larger influence on highway flows. Finally, although the impact of travel time on tourists' flows using the two transport systems was negative, the magnitude of the impact was very different. The coefficient of travel time in the highway model was almost twice as much as that in the high-speed railway model.

4 Discussion

The centrality and the aggregated tourism resource score of each node, calculated from Eq. (1), were mapped in Fig. 10. There was a large correlation (0.701) between centrality and the tourism resource grading. This finding justifies the reliability of UGC data, because the results are consistent with a similar study conducted at urban level (Jin et al. 2018).

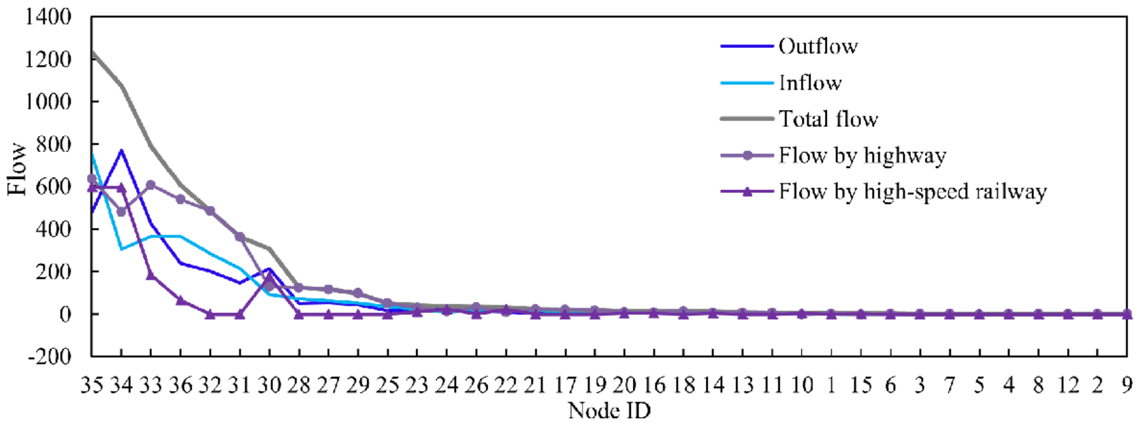

Fig. 7 Flow of nodes in Guangxi tourism network based on UGC during 2013-2016.

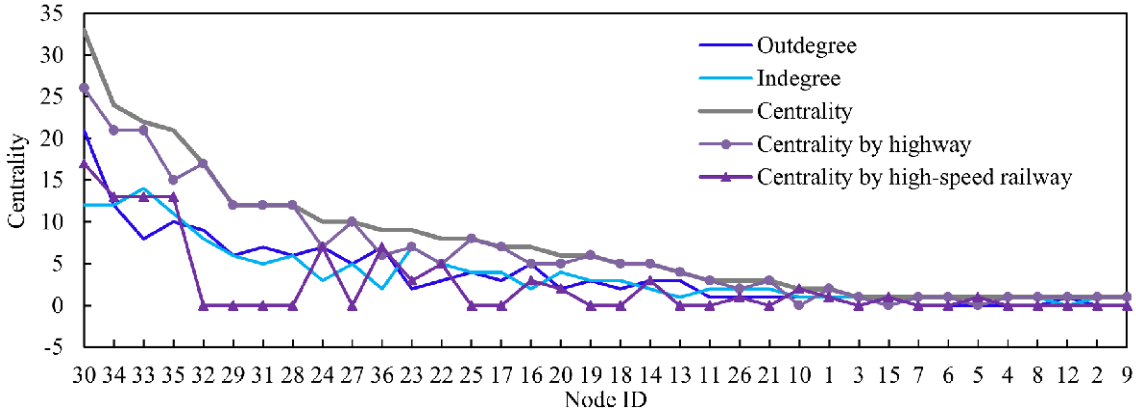

Fig. 8 Centrality of nodes in Guangxi tourism network based on UGC during 2013-2016.

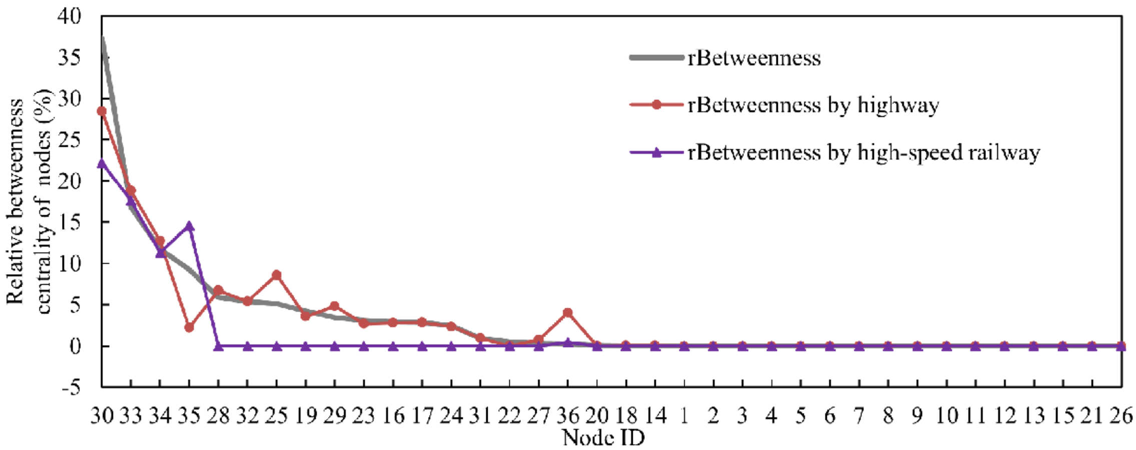

Fig. 9 Relative betweenness centrality of nodes in Guangxi tourism network based on UGC during 2013-2016. rBetweenness denotes the relative betweenness centrality of nodes.

The flow and betweenness centrality of links show that Nanning is in a very important position on the entire network. This is probably due to the geographical location of Nanning. The natural scenery tourism area in northern Guangxi and the coastal tourism area in southern Guangxi are two core tourism areas of Guangxi, with Nanning as the connection hub between them. In addition, the links that control the entire network are not necessarily the links with highest flows. High flow links mainly occurred between municipal-level cities (e.g., Guilin and Beihai) and their surrounding county-level nodes. The controlling links were between municipal tourism cities with high-speed railway stations, such as Nanning, Guilin, Beihai and

Table 4 Statistical results of regression analysis.

Coefficient parameter or variable	Highway	High-speed railway
a	0.295*** (5.56)	0.562*** (7.90)
b	0.389*** (6.96)	0.574*** (8.81)
β	-0.0088 *** (-8.05)	-0.0048*** (-3.86)
k	-0.95** (-1.96)	-5.91*** (-6.88)
n	135	48
Pseudo R^2	0.106	0.218
alpha	1.35	0.478

Note: ***p<0.01, **p<0.05. The numbers in parentheses are t statistic values.

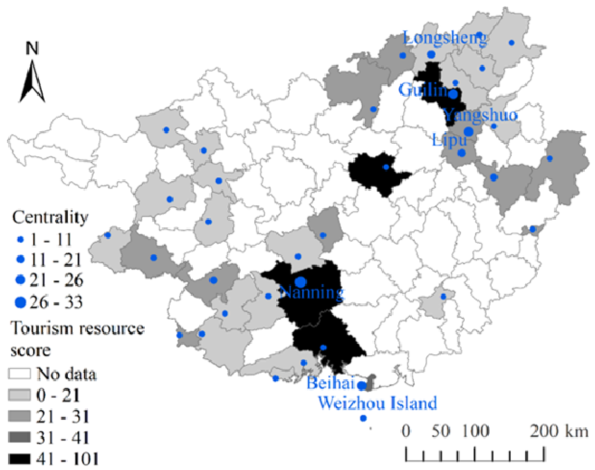

Fig. 10 Distribution of the node centrality and tourism resource score in Guangxi tourism network. The centrality is calculated by Table 1 based on UGC data during 2013-2016. The tourism resource score is calculated by Eq. (1), where the data of tourist attractions and grades are from the official website of the Department of Culture and Tourism of Guangxi Zhuang Autonomous Region (http://wlt.gxzf.gov.cn/).

Fangchenggang. Therefore, it might be expected that strengthening high-speed railway connectivity between nodes would improve the ability to control the entire tourism network.

The analytical results on flows and centrality of nodes on the tourist network (Fig. 11) indicated that Guilin, Yangshuo and Beihai had similar levels of flow and node centrality. There is a well-known Chinese proverb, "East or west, Guilin scenery is best", although the Yangshuo landscape is regarded as even better. Beihai Silver Beach is regarded as the best beach in China. Comparatively, Nanning, the capital of the province, had the highest centrality value but a much smaller number of tourist flows. This showed that the node flow did not match the node centrality

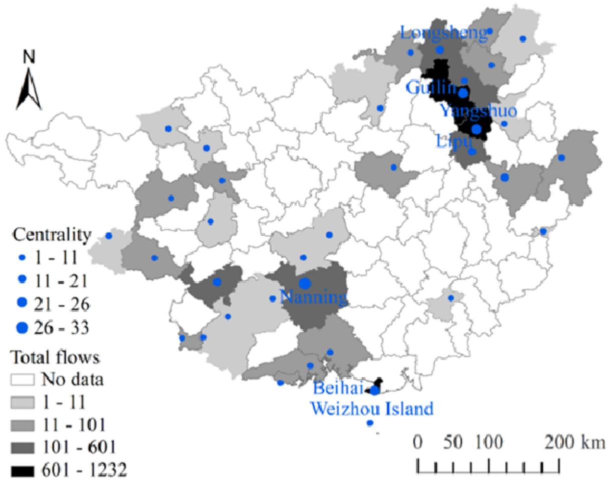

Fig. 11 Distribution of node flows and centrality in Guangxi tourism network. The flows and centrality are calculated by Table 1 based on UGC data during 2013-2016.

on the tourism network across the province. Nanning has only one 5A grade scenic spot, which was approved in 2014, so its national reputation and attraction is much lower than that of Guilin, Yangshuo or Beihai. The smaller number of flows to Nanning can be attributed to the low ranking of its scenic spots and its overall poor reputation as a tourism city. As a regional political and economic centre and provincial transport hub, Nanning has the highest centrality in the network. The above results also confirm that node flow and centrality were not individually sufficient to reflect the push and pull forces at each origin and destination site. Consequently, it is reasonable to use the product of node flow and centrality to represent both pushing and pulling forces as shown in Eqs. (3-6).

The distance decay effect, evaluated by either distance or time, has become a universal law of geographical mobility at a variety of scales (e.g., Cheng et al. 2014; Jin et al. 2018). For comparison, a more accurate indicator - travel time - was used in this study. A negative exponential function was shown to be an effective distance decay function for tourism mobility at the provincial scale. Travel time friction coefficients for the high-speed railway and highway transport systems were found to be different; the absolute value indicated that tourism by highway was more sensitive to time. Time is an important factor to consider when choosing a mode of transport. For medium and long distance travel, high-speed railway is more competitive due to increased travel comfort and lack of traffic jams. For short distances, self-drive tourism has the advantages of low travel costs, flexible timings, and high efficiency, especially for families, groups of four, younger and middle-class tourists (Moyano et al. 2016; Zhou and Huang 2016). Self-drive tours can depart directly from home at any time, without worrying about how to get to the station or how long it takes to arrive at the station in advance, and can also be adjusted to visit scenic spots along the route, time permitting. Consequently, self-drive tours are more popular for short distance trips. The results also showed that, taken separately, the friction coefficients of highway and high-speed railway were negative. This provides further confirmation of the law of distance decay in tourist mobility (Santeramo and Morelli 2016; Mehmet 2010; Jin et al. 2018). Moreover, they support Li et al.'s (2012) conclusions on tourism interaction, which showed that friction coefficients change with spatial scale. The larger the spatial scale, the greater the distance travelled by tourists, and the lower the friction coefficient. Li et al. (2012) found that when the average travel distance of mainland Chinese residents was 311 km, the corresponding friction coefficient was 0.00322. The friction coefficients under the four spatial scales of province, city, county and township were 0.00044, 0.0014, 0.0044 and 0.014 respectively (Li et al. 2012). The travel distance in this study included two spatial scales, city and county, and the friction coefficients were consistent with Li et al.'s (2012) results. Travel distance by highway was generally at the spatial scale of county whereas travel distance by high-speed railway was generally at the spatial scale of city due to the restrictions of high-speed railway stations and the extremely low distribution density, especially in Karst areas (Fig. 12). Therefore the friction coefficient for high-speed railway was smaller than that for highway.

5 Conclusions

Tourism mobility has received increasing attention in academia and policy practice in recent years. Focusing on county level tourism in a less-developed province in China, and using online UGC data, this study has provided new evidence of the spatial structure of the tourism network in Guangxi, and compared the impacts of two transport systems on tourism mobility. The comparative effects of high-speed railway and highways on the tourism network help address the challenges of spatial development in

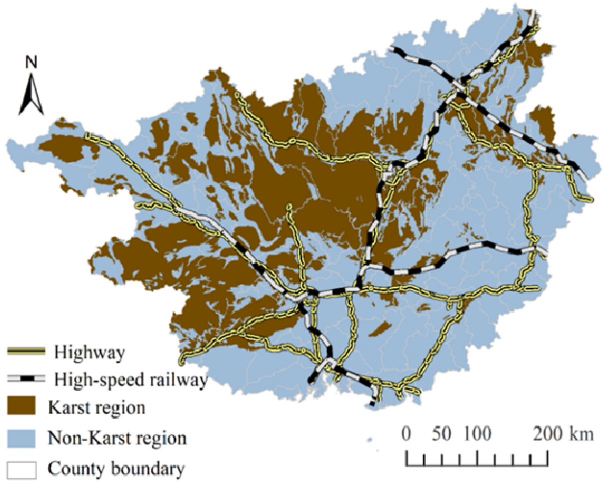

Fig. 12 The spatial distribution of highway, high-speed railway and karst areas in Guangxi Province by 2015.

this Karst region from a tourism perspective.

The tourism network in Guangxi has an extremely low density compared with other developed provinces in China (Zhou and Xu 2016; Fu et al. 2015; Ju et al. 2015). The development of tourism in Guangxi is extremely unbalanced. The density of the tourism network varied significantly among cities. Tourist flows were concentrated in a few cities, such as Guilin, Beihai, and in areas with rich tourism resources and convenient transportation. These areas have the type of landscapes preferred for sightseeing and seashore holiday tours, with the added advantage of leisure tours to nearby small towns. Other distinctive attractions include minority folk-customs, the China-Vietnam border and transnational villages, local cultural relics and historic areas, forest parks, health-care villages, and places associated with the revolution. However, these scenic spots had only a small number of tourist flows, and did not constitute large-scale tourism. The findings suggest that tourism resources in Guangxi have not yet formed an efficient integrated network.

As one of a few studies of tourism mobility at the provincial scale in China, this empirical study has validated the applicability of social network analysis for tourism networks. UGC data, a type of big data, outperforms the traditional questionnaire data collection method because web crawling is much cheaper, quicker and more objective. In this study, the details of tourism origin and destination sites and modes of transport within a specific period were captured to construct a flow matrix at the aggregate level of county. It is argued that the product of node flow and centrality is better for analysing the spatial patterns of a tourism network, and using a negative binomial regression model helps identify and distinguish the impacts of transport in a less-developed region. The distance decay effect of high-speed railway and highway transport on this scale follows the scale dependent universal law. The coverage of high-speed railway system should be increased in this region because tourism behaviour is less sensitive to travel time by high-speed railway than it is by highway. Self-drive tourism is less sustainable due to the higher consumption of energy and greater risks of health and safety. However, it remains a challenge is to develop the high-speed railway system in Guangxi, especially in the Karst mountainous areas where high-speed railway is underdeveloped.

This paper has some clear limitations which should be addressed in future research. In terms of social network analysis, the measure of centrality used in this study only considered spatial linkages between nodes and did not account for the temporal duration of trips. Longer stays at a node increase its control over the network. This information could be gleaned from review comments in UGC using text analytics. Another dimension of temporality is the heterogeneity between seasons, although Guangxi, as a sub-tropical region, has very mild weather and accordingly tourism is possible for the majority of the year. In terms of transport, in addition to the inclusion of air and seaport transport, local transport connecting railway stations and scenic spots should be considered in order to integrate tourism mobility on both regional and urban scales. From a tourism-led growth perspective, it is worth comparing the interactions between the tourism network and transport in other selected provinces using UGC data captured from tourism web services.

Acknowledgements

This study is funded by the Guangxi Natural Science Foundation (Grant No. 2020GXNSFAA159065), the Opening Fund of Key Laboratory of Environment Change and Resources Use in Beibu Gulf under Ministry of Education (Nanning Normal University) and Guangxi Key Laboratory of Earth Surface Processes and Intelligent Simulation (Nanning Normal University) (Grant No. GTEU-KLOP-K1701), the seventh batch of distinguished experts in Guangxi and National Natural Science Foundation of China (Grant No. 41867071).

Electronic supplementary material

Supplementary materials (Appendix 1) are available in the online version of this article at https://doi.org/10.1007/s11629-021-6883-3

References

Adamic L, Adar E (2005) How to search a social network. Soc Networks 27(3): 187-203.
Albalate D, Fageda X (2016) High speed rail and tourism: empirical evidence from Spain. Transport Res A-Pol 85: 174-185. https://doi.org/10.1016/j.tra.2016.01.009
Asero V, Gozzo S, Tomaselli V (2016) Building tourism networks through tourist mobility. J Travel Res 55(6): 751-763. https://doi.org/10.1177/0047287515569777
Bálint K, Mátyás G (2021) Tourism flows in large-scale destination systems. Ann Tourism Res 87: 103113. https://doi.org/10.1016/j.annals.2020.103113
Bendle LJ (2018) The structures and flows of a large tourist itinerancy network. Curr Issues Tour 21(1): 103-122. https://doi.org/10.1080/13683500.2015.1092948
Cameron AC, Trivedi PK (2005) Microeconometrics: methods and applications. Cambridge University Press, Cambridge, UK.
Cheng J, Young C, Zhang X, et al. (2014) Comparing inter-migration within the European Union and China: an initial exploration. Migr Stud 2(3): 340-368. https://doi.org/10.1093/migration/mnt029
Dai Q, Ye X, Wei Y, et al. (2018) Geography, ethnicity and regional inequality in Guangxi Zhuang Autonomous Region, China. Appl Spat Anal Polic 11: 557-580. https://doi.org/10.1007/s12061-017-9229-3
Fu QG, Liu DJ, Hu J, et al. (2015) Network structure and optimization research on the Hubei province tourist flow. Econ Geogr 35(3): 191-196. (In Chinese) https://doi.org/10.15957/j.cnki.jjdl.2015.03.028
Gao J, Wu B (2017) Revitalizing traditional villages through rural tourism: a case study of Yuanjia Village, Shaanxi Province, China. Tourism Manage 63: 223-233. https://doi.org/10.1016/j.tourman.2017.04.003
Guirao B, Campa JL, López ME (2016) The assessment of the HSR impacts on Spanish tourism: an approach based on multivariate panel data analysis. Transp Res Procedia 18: 197-204. https://doi.org/10.1016/j.trpro.2016.12.027
GXLCCC (Guangxi Local Chronicle Compilation Commit) (2016) Guangxi Yearbook 2016. Guangxi Yearbook Press, Nanning, China. (In Chinese)
Hao X, Wu B, Morrison AM, et al. (2016) Worth thousands of words? Visual content analysis and photo interpretation of an outdoor tourism spectacular performance in Yangshuo-Guilin, China. Anatolia 27(2): 201-213. https://doi.org/10.1080/13032917.2015.1082921
Huang T (2016) Evolution characteristics and factors affecting tourist flow potential in the Yangtze River Delta urban agglomerations. Resour Sci 38(2): 364-376. (In Chinese)
Jin C, Cheng J, Xu J (2018) Using user-generated content to explore the temporal heterogeneity in tourist mobility. J Travel Res 57(6): 779-791. https://doi.org/10.1177/0047287517714906
Jin C, Cheng J, Xu J, et al. (2018) Self-driving tourism induced carbon emission flows and its determinants in well-developed regions: a case

study of Jiangsu Province, China. J Clean Prod 186: 191-202. https://doi.org/10.1016/j.jclepro.2018.03.128
Ju SL, Tao ZM, Lai ZQ, et al. (2015) Research on network structure evolution of domestic tourist flow system in Zhejiang province. Geogr Geo-Inf Sci 31(2): 91-97. (In Chinese) https://doi.org/10.3969/j.issn.1672-0504.2015.02.019
Khadarooa J, Seetanah B (2008) The role of transport infrastructure in international tourism development: a gravity model approach. Tourism Manage 29(5): 831-840. https://doi.org/10.1016/j.tourman.2007.09.005
Li H, Chen J, Li G, et al. (2016) Tourism and regional income inequality: evidence from China. Ann Tourism Res 58: 81-99. https://doi.org/10.1016/j.annals.2016.02.001
Li J, Xu L, Tang L, et al. (2018) Big data in tourism research: a literature review. Tourism Manage 68: 301-323. https://doi.org/10.1016/j.tourman.2018.03.009
Li S, Wang Z, Zhong Z (2012) Gravity model for tourism spatial interaction: basic form, parameter estimation, and applications. Acta Geogr Sinica 67(4): 526-544. (In Chinese)
Li Y, Ding Y, Wang D (2016) A new approach for designing tourist routes by considering travel time constraints and spatial behavior characteristics of tourists. Tour Tribune 31: 50-60. (In Chinese) https://doi.org/10.3969/j.issn.1002-5006.2016.09.016
Liu B, Huang S, Fu H (2017) An application of network analysis on tourist attractions: the case of Xinjiang, China. Tourism Manage 58: 132-141. https://doi.org/10.1016/j.tourman.2016.10.009
Liu J (2004) An introduction to social network analysis. Social Sciences Academic Press, Beijing, China. (In Chinese)
Liu J, Nijkamp P, Lin D (2017) Urban-rural imbalance and tourism-led growth in China. Ann Tourism Res 64: 24-36. https://doi.org/10.1016/j.annals.2017.02.005
Liu J, Wang J, Wang S, et al. (2018) Analysis and simulation of the spatiotemporal evolution pattern of tourism lands at the Natural World Heritage Site Jiuzhaigou, China. Habitat Int 79: 74-88. https://doi.org/10.1016/j.habitatint.2018.07.005
Lozano S, Gutiérrez E (2018) A complex network analysis of global tourism flows. Int J Tour Res 20(5): 588-604. https://doi.org/10.1002/jtr.2208
Lumsdon L (2000) Transport and tourism: cycle tourism - a model for sustainable development? J Sustain Tour 8(5): 361-377. https://doi.org/10.1080/09669580008667373
Luo F, Becken S, Zhong Y (2018) Changing travel patterns in China and carbon footprint implications for a domestic tourist destination. Tourism Manage 65: 1-13. https://doi.org/10.1016/j.tourman.2017.09.012
Marrocu E, Paci R (2011) They arrive with new information. tourism flows and production efficiency in the European regions. Tourism Manage 32(4): 750-758. https://doi.org/10.1016/j.tourman.2010.06.010
Masson S, Petiot R (2009) Can the high speed rail reinforce tourism attractiveness? The case of the high speed rail between Perpignan (France) and Barcelona (Spain). Technovation 29(9): 611-617. https://doi.org/10.1016/j.technovation.2009.05.013
Mehmet E (2010) Factors affecting international tourism flows to Turkey: a gravity model approach. Tourism Econ 16 (3): 585-595. https://doi.org/10.5367/000000010792278374
Morley C, Rosselló J, Santana-Gallego M (2014) Gravity models for tourism demand: theory and use. Ann Tourism Res 48: 1-10. https://doi.org/10.1016/j.annals.2014.05.008
Moyano A, Coronado JM, Garmendia M (2016) How to choose the most efficient transport mode for weekend tourism journeys: an HSR and private vehicle comparison. Open Transport J 10(S1): 84-96. https://doi.org/10.2174/1874447801610010084
NSB (National Bureau of Statistics of the People's Republic of China) (2016) China Statistical Yearbook 2016. China Statistics Press, Beijing, China. (In Chinese)
NTA (China National Tourism Administration) (2012) Measures for the administration of quality grades of tourist attractions. http://zwgk.mct.gov.cn/zfxxgkml/zcfg/gfxwj/202012/t20201204_906214.html (Accessed on 11 May 2019) (In Chinese)
Pagliara F, Mauriello F, Garofalo A (2017) Exploring the interdependences between high speed rail systems and tourism: Some evidence from Italy. Transport Res A-Pol 106: 300-308. https://doi.org/10.1016/j.tra.2017.09.022
Peeters P, Higham J, Cohen S, et al. (2019) Desirable tourism transport futures. J Sustain Tour 27(2): 173-188. https://doi.org/10.1080/09669582.2018.1477785
Pratt S (2015) Potential economic contribution of regional tourism development in China: a comparative analysis. Int J Tour Res 17(3): 303-312. https://doi.org/10.1002/jtr.1990
Qian J, Feng D, Zhu H (2012) Tourism-driven urbanization in China's small town development: a case study of Zhapo Town, 1986-2003. Habitat Int 36(1): 152-160. https://doi.org/10.1016/j.habitatint.2011.06.012
Santeramo FG, Morelli M (2016) Modelling tourism flows through gravity models: a quantile regression approach. Curr Issues Tour 19(11): 1077-1083. https://doi.org/10.1080/13683500.2015.1051518
Shen J (2018) Urbanization, regional development and governance in China. Routledge Press, London, UK.
Spasojevic B, Lohmann G, Scott N (2018) Air transport and tourism-a systematic literature review (2000-2014). Curr Issues Tour 21(9): 975-997. https://doi.org/10.1080/13683500.2017.1334762
Wang D, Chen T, Lu L, et al. (2015) Mechanism and HSR effect of spatial structure of regional tourist flow: case study of Beijing-Shanghai HSR in China. Acta Geogr Sinica 70(2): 214-233. (In Chinese) http://doi.org/10.11821/dlxb201502004
Wang D, Niu Y, Qian J (2018) Evolution and optimization of China's urban tourism spatial structure: a high speed rail perspective. Tourism Manage 64: 218-232. https://doi.org/10.1016/j.tourman.2017.08.010
Wang D, Niu Y, Wang L (2015) Influence of high- speed rail on choices of tourist destination based on the gravity model: a case study of Beijing Shanghai high-speed rail in China. Geogr Res 34(9): 1770-1780. (In Chinese)
Wang D, Niu Y, Sun F, et al. (2017) Evolution and spatial characteristics of tourism field strength of cities linked by high-speed rail (HSR) network in China. J Geogr Sci 27(7): 835-856. https://doi.org/10.1007/s11442-017-1409-1
Wang Y, Wang M, Li K, et al. (2021) Analysis of the relationships between tourism efficiency and transport accessibility—a case study in Hubei province, China. Sustainability-Basel 13(15): 8649. https://doi.org/10.3390/su13158649
Wilson AG (1970) Entropy in urban and regional modeling. Pion Ltd, London, UK.
Wu J, Pan X (2010) Study on the relationship between inbound tourism flows network and aviation network. Tour Tribune 25(11): 39-43. (In Chinese)
Wu J, Ren R, Han L, et al. (2012) Chinese aviation international network and its impact on inbound tourism. Econ Geogr 32(5): 147-152. (In Chinese)
Yahya F (2003) Tourism flows between India and Singapore. Int J Tour Res 5(5): 347-367. https://doi.org/10.1002/jtr.443
Yang X, Gu C, Wang Q (2007) Urban tourism flow network structure construction in Nanjing. Acta Geogr Sinica 62(6): 609 - 620. (In Chinese)
Zhang L, Cheng J, Jin C (2019) Spatial interaction modeling of OD flow data: comparing geographically weighted negative binomial regression (GWNBR) and OLS (GWOLSR). Int J Geo-Inf 8(5): 220. https://doi.org/10.3390/ijgi8050220
Zhang Y, Li J, Yang M (2014) The tourism flow network structure of Xi'an based on tourism digital footprint. Hum Geogr 129(4): 111-118. (In Chinese)
Zhao J, Peng Z, Liu L, et al. (2021) Public space layout optimization in affordable housing based on social network analysis. Land-Basel 10(9): 955. https://doi.org/10.3390/land10090955
Zhou HL, Xu CX (2016) Study on spatial network structure of Hunan tourist flow based on travel arrangement. Econ Geogr 36(10): 201-206. (In Chinese)
Zhou Y, Huang X (2016) Aspects of lifestyle change in China: motivations for self-drive travel - a Chinese domestic market perspective. J Travel Tour Mark 33(6): 912-928. https://doi.org/10.1080/10548408.2015.1075933