Manchester Metropolitan University

Jin, C and Cheng, J and Xu, J (2017) *Using user-generated content to explore the temporal heterogeneity in tourist mobility.* Journal of Travel Research. ISSN 0047-2875 (In Press)

Using user-generated content to explore the temporal heterogeneity in tourist mobility

Cheng Jin [1,2,3] *
email: jincheng@njnu.edu.cn

Jianquan Cheng [4*]
email: J.cheng@mmu.ac.uk

Jing Xu [5]
email: xq514@163.com

[1] School of Geographical Science, Nanjing Normal University, Nanjing, Jiangsu 210023, China
[2] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, Jiangsu 210023, China
[3] Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing, Jiangsu 210023, China
[4] Division of Geography and Environmental Management, School of Science and the Environment, Manchester Metropolitan University (MMU), Chester Street, M1 5GD, UK,
[5] Tourism and Social Administration College, Nanjing Xiaozhuang University, Nanjing, Jiangsu 211171, China

*Correspondence author: jincheng@njnu.edu.cn, J.cheng@mmu.ac.uk

Using user-generated content to explore the temporal heterogeneity in tourist mobility

Abstract

In tourism studies, new means of data collection are opening up opportunities for disclosing hidden mobility patterns. This paper aims to analyze and model the tourist flow networks for different lengths of trip on urban scale, using user generated content (UGC) data collated from an open tourism web service. The textual UGC data, with high spatial and temporal resolution, is utilized to construct three tourist flow networks in response to length of trips. Social network analysis and a revised spatial interaction model are deployed for exploring the temporal heterogeneity in the tourist movements. This empirical study from Nanjing city has further confirmed the power law of distance decay in intra-urban tourist mobility. Furthermore, the research reveals temporal variations with length of trip. The paper highlights the role of time in the tourism study through incorporating a temporal dimension into the analyses and taking advantage of the availability of new data.

Key words: Tourist movement, user generated content, social network analysis, gravity model, power law, temporal heterogeneity.

1. Introduction

China, as the second largest economy in the world, is undergoing an economic transition from traditional manufacturing to modern services. In particular, rapid urbanization has stimulated the development of tourism. Domestic tourist numbers reached 3,262 million in 2013, following an average annual growth rate (AAGR) of 11.27% between 1990-2013. In 2013, China's domestic tourism revenue reached RMB 2,627.61 billion yuan, with an AAGR of 12.5% between 1995 and 2013 (CEIC China Database 2014). The challenges that this growth brings to China are subject to increasing research.

Human mobility is a key concept subject to multidisciplinary study (e.g. tourism, geography, transport and sociology), where research has focused on social mobility, geographic mobility, occupational mobility, smart mobility and even everyday mobility. Although there might be varied foci and interests between these areas, analyzing and modelling flows of people on different (spatial and temporal) scales is a major concern in the mobility literature. For example, Stillwell, Bell, Ueffing, Daras, Charles-Edwards, Kupiszewski and Kupiszewska (2016) explored the scaling effects of distance decay in the case of internal migration in an international context. Tourism, as a form of mobility, particularly as a sub-set of a vast and heterogeneous complex of global mobilities, is increasingly integrated into wider processes of economic and political development and even constitutive of everyday life (Hannam, Butler, and Paris 2014). Understanding how tourists move through time and space has important implications for infrastructure and transportation development, product development, destination planning and the planning of new attractions, as well as management of the social, environmental and cultural impacts of tourism (Lew and Mckercher 2006).

Understanding tourism mobility involves the scientific and empirical interpretations of spatial and temporal patterns of flows between destinations or attractions within a study area. The pre-requisite for reasonably analyzing and modelling tourist mobility is the collection of representative flow data of a high quality. Apart from secondary data, traditional methods of data collection are dominated by questionnaire surveys and interviews (e.g. Shoval and Raveh 2004; Asero, Gozzo, and Tomaselli 2016; Wong, Fong and Law 2016), which are time consuming, costly and labor intensive. Hannam, Butler, and Paris (2014) suggest that new technologies (e.g. sensor, social media and the internet), together with new means of high-resolution data collection, provide opportunities to develop sophisticated tools with which to explore mobility (Shoval and Ahas 2016), encompassing digital footprints (Önder, Koerbitz, and Hubmann-Haidvogel 2016), word of mouth (Ring, Tkaczynski, and Dolnicar 2016), user-generated content (UGC) (Lu and Stepchenkova 2015) and big data (Fuchs, Höpken, and Lexhagen 2014).

Most of these studies focus on the spatial dimension of mobility, however, Leiper (2004, p.129) suggests that ''different tourists tend to perceive the same destination in different ways and a very influential variable is the time available for spending in each place'', because time has been increasingly recognized as a scarce commodity and limited resource. The valuation of time as commodity or resource, which determines the distribution and order of attraction visits, is very much affected by time availability during a trip. Fuchs, Höpken, and Lexhagen (2014) stated that

"the evidence and understanding of the role played by time are incomplete". Using survey data, Shoval and Raveh (2004) revealed that both length of stay and number of previous visits to the city have a strong effect on the tourism consumption pattern, which is an initial exploration of the temporal dimension of tourism mobility. This paper, therefore, further explores the temporal heterogeneity in tourism mobility through the application of new means of data collection, to consider: how does the length of trip affect tourist mobility, which data sets can be utilized to analyze the influence, and what are the implications for tourism mobility theory?

The second section of the paper presents a literature review of tracking and analyzing tourist mobility, before introducing the study area of Nanjing City (China), data collection and processing, and analytical methods in section Three, followed by the presentations from the analyses and models in Section Four. Section Five proposes a conceptual model summarizing and comparing tourist flow patterns in response to different trip lengths. The paper concludes with a discussion of the potential contribution to the study of tourism mobility.

2. Literature review: tracking and analyzing tourist mobility

Tourism mobility involves the details of a person (who), location (where), time (when) and context (what or environment) (e.g. Kim and Fesenmaier 2015). The 4-W approach enables analysis of where and when the tourist is and what they are doing. Information can be recorded on a variety of spatial, temporal, and human scales, where: spatial scale considers resolution and extent; the temporal scale includes interval and duration or length; and the human scale consists of sample and population. This approach underpins both the tracking and analysis of tourist mobility.

2.1 Tracking

There has been increasing growth of the use of sensors for tracking purpose, e.g. GPS, cell-tower identification, Wi-Fi positioning, Bluetooth, RFID, and camera, to list but a few. In terms of citizen science (Goodchild 2007), the individual tourist acts as a kind of 'sensor' allowing the tracking of movement through their physical movements together with the personal data they volunteer and share online. By reviewing the progress of tracking technologies, from GPS to smart phone, Shoval and Ahas (2016) predict the eventual development of smartphones that will connect to various external sensors, which will become the main facility for rapidly collecting the spatial and temporal data of tourist mobility. For example, online social networks (e.g. Twitter, Flickr, and Foursquare), already provide geo-referenced messages and photos with high spatial and temporal resolution. These new data sets have motivated the spatial and temporal analyses of tourist movement (Shoval and Ahas 2016), who argue that new tracking technology is promoting data collection from networks to analyze high spatial, temporal, and human scales. The Internet, in particular, has changed from a "publishing-browsing-platform" to a "participation-interaction-platform" (Xiang, Wang, O'Leary, and Fesenmaier 2015). For instance, Zhou, Xu, and Kimmons (2015) developed a method for detecting tourism destinations using geotagged photos from social media. Liu, Sui, Kang, and Gao (2014) utilized social media check-in data set submitted by approximately half a million users to analyze inter-urban trip patterns. As social networks and online communities create the social DNA of the Network Society (Castells 2001), a range of platforms have emerged specifically within the tourist

sector which utilize UGC, e.g. virtual communities, consumer reviews, personal stand-alone blogs, blog aggregators, and microblogging platforms, social networks, media sharing tools, and wikis (Lu and Stepchenkova 2015). UGC, which is textual or visual (e.g. Zhang, Wu, and Mattila 2016), can improve data availability, reduce data cost, and promote speed and simplicity of data collection, when compared with other methods (e.g. GPS-tracking, travel diary or questionnaire survey). After completing a systematic review of empirical research using UGC as a data source to address issues in tourism and hospitality applications, Lu and Stepchenkova (2015) conclude that textual UGC and qualitative content analysis are the leading data type and research method employed, although there is an increasing trend in quantitative analysis (e.g. big data analytics) using both textual and visual data. However, most of these applications focus more on destinations or attractions rather than tourists flows.

2.2 Analyzing

Mckercher and Lau (2008) summarized 78 tourist movement patterns within an urban destination using qualitative techniques. Quantitatively analyzing the mobility patterns is very much dependent on these scales (spatial, temporal and human scales), complicating how to aggregate these data at individual level, spatially and temporally. One approach is to conduct analysis at the individual level. For example, Vu, Li, Law, and Ye (2015) modeled the travel behaviors of inbound tourists to Hong Kong using geotagged photos and the Markov Chain Model. Time geography focuses on the movement and interaction of tourists in time and space, and has been extensively applied to visualizing, analyzing and modelling patterns of tourist mobility (e.g. Grinberger, Shoval, and McKercher 2014), and time-space constraints (Shoval 2012). Modsching, Kramer, Gretzel, ten Hagen (2006) analyzed the tourist's spatial behavior, in particular the time spent for different activities at an urban destination using GPS tracking data. McKercher, Shoval, Ng, and Birenboim (2012) further explored the significant impacts of visitor numbers on their spatial patterns using the same kind of data. These studies highlighted the heterogeneous uses of time between activities during a trip.

There is an increasing trend involving the aggregation of mobility data at the individual level into networks, as the basis for the analysis of the topological structure of the attraction system (Smallworld, Beckley, and Moore 2012). This is because understanding the network characteristics of tourism attractions has practical implications for the competitiveness, management, and planning of tourism (Stienmetz and Fesenmaier 2015). Travel patterns can be viewed as network, and therefore, subject to network analysis (e.g. Shih 2006), which include several structural indicators, e.g. centrality. Based on a core-periphery analysis of attractions network, Zach and Gretzel (2011) argued that the network structure provides a strong and practical basis for dynamically bundling products that create value for tourists and the destination. Hwang, Gretzel and Fesenmaier (2006) concluded that (social) network analysis methods can provide insights into structural properties of trips and allow for testing a variety of structural differences, after their successful analyses of multicity tourism behavior across the U.S. More interestingly, using UGC data from overseas tourists (textual trip diaries) from 6 websites and data from social medias, Leung, Wang, Wu, Bai, Stahura, and Xie (2012) demonstrated the strength of social network analysis methods through the analysis of the patterns of intra-destination tourist

movement networks to examine the impacts of the 2008 Olympic Games. These case studies have proved that both UGC data and social network analysis can be effectively integrated into analyzing the spatial patterns of tourist flows. However, temporal patterns of tourist flows are not reported yet.

In the majority of tourism studies, distance plays a crucial role. Distance decay, sometimes referred to 'friction of distance', indicates declining interaction with an increase in distance. The distance decay effect is a universal phenomenon in geography for modelling spatial interaction between origin and destination sites. McKercher and Lew (2003) noted that distance decay was popular in tourism research in the 1960s and 1970s, when it was used as a proxy for forecasting. McKercher, Chan, and Lam (2008) summarized three types of distance decay curves: classic decay curve; plateauing decay curve; and a decay curve with a secondary peak to represent the global flow of tourists. The distance decay effect within tourism studies has focused on destination choices, such as outbound international travel (Hooper 2015). Actual decay curves have been found to exhibit a plateau, or a secondary peak (Smallwood, Beckley, and Moore 2012). However, few studies have explored the distance decay in tourist flow distributions.

Greer and Wall (1979) observed that the type of trip undertaken influences the decay rate—short duration trips experienced the steepest curve, while longer duration trips tended to display more of a plateauing curve. This means that the distance decay effect may demonstrate temporal variations with length of trip. Halás, Klapka and Kladivo (2014) have explored the heterogeneity of distance decay effect in commuting patterns. However, in network analysis case studies, temporal heterogeneity has been ignored. For example, tourists' spatial movement patterns during the fall might be different from the spring, summer, and winter seasons. Nicolau, Zach, and Tussyadiah (2016) confirmed a positive impact of both distance and first-time visitation on length of stay using data collected from 908 U.S. visitors to a tourism destination in the Atlantic Coast of the United States. This knowledge gap regarding the temporal heterogeneity in mobility analysis, therefore, forms a central aim of this paper.

3. Methodology

3.1 Study area

Nanjing City is the capital of Jiangsu Province, which is located in eastern China and at the middle and lower reaches of the Yangtze River (Figure 1). Nanjing City comprises eleven districts, with an area of 6597 km^2 and a total population of 8.22 million in 2014 (JSBS 2015). As a key node within the Yangtze River Delta, which is an emerging global city region, Nanjing has been a national center for education, research, transportation and tourism. Due to rapid urbanization, Nanjing has become a core mega city in this economically wealthy Yantze River Delta (Wu, Cheng, Chen, Hammel, and Wu 2014). In terms of tourism, this city, with 2500 years of urban history, has been the capital city of six different dynasties, and is now recognized as one of the four great ancient capitals of China (Yuan, Gao, and Wu 2016). With such a long historical culture and a unique natural landscape, Nanjing has been attracting domestic and overseas tourists. Nanjing hosted 5.66 million oversea tourists and 94.19 million domestic tourists in 2014 (JSBS

2015). With the total income of 150.48 billion yuan RMB in 2014, the tourism economy of Nanjing was ranked within the first few places and well recognized as a top tourism city (JSBS 2015). Being focused on the internal mobility of tourists within a city, the study area is composed of 47 sub-districts, located to the south of the Yangtze River and within the ring roads (Figure 1).
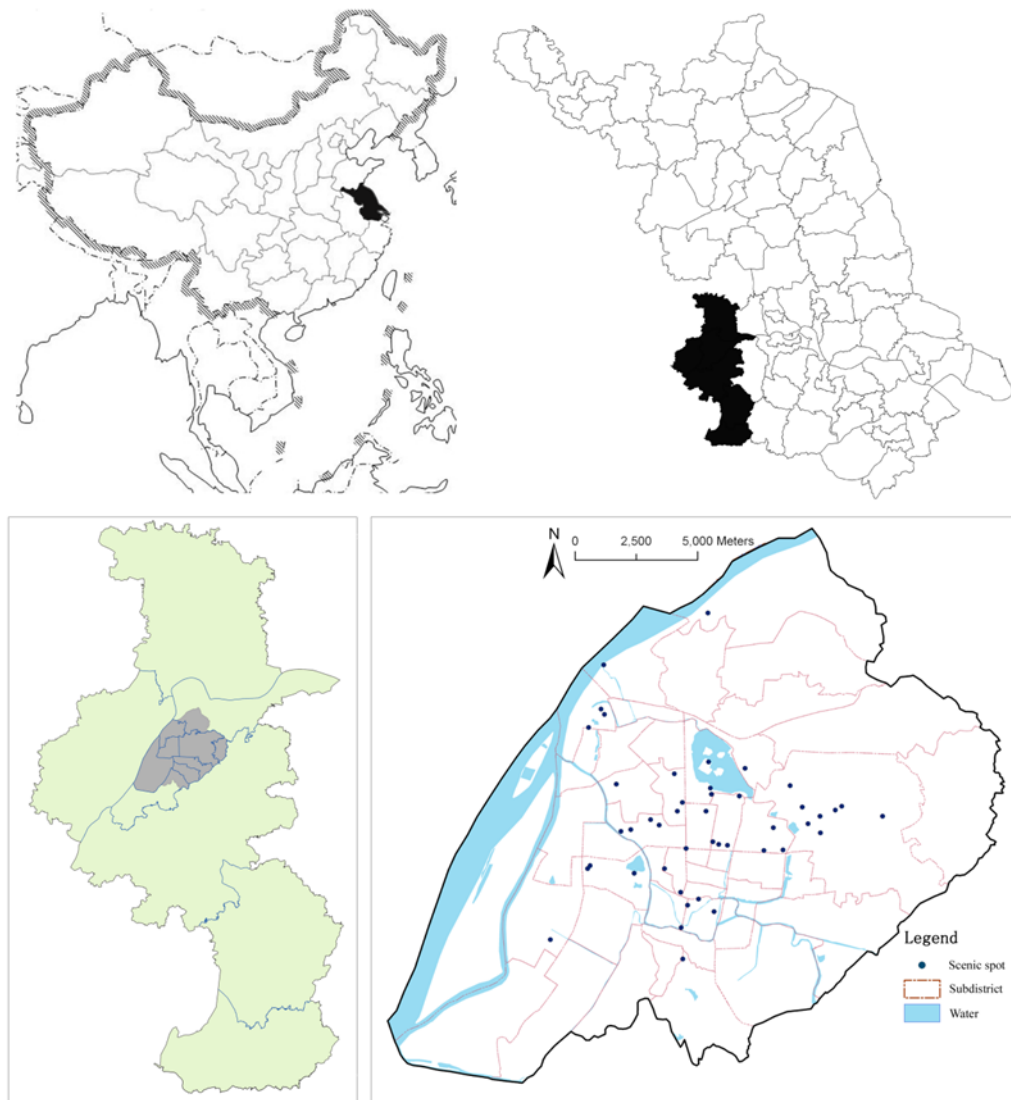

Figure 1. Location of the study area (Jiangsu province, Nanjing City and distribution of attractions)

3.2 Data collection and processing

The tourist flows dataset was compiled from a commercial web service called "Where to go?" （www.qunar.com）(another application example can be seen from the case study by Wu and Pearce 2016), which was launched in late 2010. This web service, an equivalent of Trip Advisor in the western world, provides a platform for the public or specifically tourists all over the world to upload and exchange trip related information (e.g. story, narrative and photo for attractions and routes) across the world. Such UGC data, although in text or photo format, has high spatial and

temporal resolution. Spatially, all of the attractions visited are detailed with a specific popular name together with easily recognizable photos. Each attraction can then be geo-referenced by using longitude and latitude coordinates extracted from an open Geographical Information System (GIS) e.g. Tianditu web service. A point layer with all of these attractions can be overlaid with other GIS layers (such as road networks, collected from secondary sources) for further spatial analysis, after being projected into the local coordinate system. Temporally, the visits to attractions are recorded in order on a daily basis (some even on an hourly basis) and therefore it is possible to recognize which attraction was visited and on which date. In total, 1424 valid samples of complete trips (those trips with only one attraction visited were removed from the sample), involving 45 attractions within Nanjing in 2012 (including traditional spots and also shopping streets, bars and pubs, and university campuses) (see Figure 1), have been recorded into a database.

In terms of the network, each attraction is represented as a node, the direct line between a pair of nodes is denoted as edge (*E*), the number of tourists on the line is represented as flow. All of these edges form a network of tourist flow (or mobility). Suppose that a complete trip is composed of a departure train station, hotel, arrival airport and eight attractions from *A* to *H* (but *E* is outside of the study area), then this trip can be semantically represented by a flow between the seven attractions excluding *E*. Further, this flow is mathematically represented by a 7x7 matrix in which value 1 indicates a move and 0 means no move between a pair of origin and destination spots.

Following this rule, each trip in the sample was transformed into a 45x45 relational matrix with only 1 or 0 values. For the whole set of samples, all of these 45x45 matrices are overlaid by a simple matrix addition operation. The final output – a flow network, is represented as a 45x45 matrix, in which the value for each matrix element *ij* is the total flow of tourists from attraction *i* to *j*. With the temporal information (the attractions visited each day) available from the web service, these trips can be classified into: one-day trip, two-day trip, three-and-more-day trip. Accordingly, the flow network can also be split into three sub-networks of flows, which are comprised of flows from only one-day or two-day or three-and-more-day trip respectively. All of these networks will be used to analyze and compare the mobility and behavior of tourists in central Nanjing. Statistically, this network, formed of 1424 trips in total is composed of 45 nodes (Figure 1), 590 edges, with a total flow of 5337 on the network. The statistical patterns of the three categories are summarized in Table 1. This table demonstrates a clear trend that the numbers of trips and flows decreased with the length of trip but the total number of attractions and average numbers of attractions and routes per trip increased with the trip length.

Table 1. Flow of tourists between attractions for different trip lengths

	One-day trip	Two-day trip	Three-and-more-day trip	Total
Number of trips	897	367	160	1424
Total Number of attraction visited	37	42	43	45
Average number of attractions visited	3.55	6.86	11.01	5.24
Number of edges valid	303	374	401	590
Number of flows	2203	1803	1331	5337

3.3 Methods

3.3.1 Analyzing mobility network between attractions

In social network analysis, the basic measures of centrality are calculated through a $n \times n$ matrix of attractions to evaluate the structural features of networks. The measures classify the links among attractions according to their nodal role along tourist routes and identify the most important network attractions. This measure is composed of several metrics including in-degree, out-degree and degree centrality. In-degree is the total number of tourists arriving at a node or attraction and conversely out-degree the total number of tourists departing from a node or attraction (Table 2). As the sum of in-degree and out-degree, degree centrality (Table 2) allows for recognition of central attraction in the tourism networks. This paper proposes, therefore, a new index vergence (between -1 and 1) to measure the difference between inflow and outflow at each node.

Table 2. Network metrics

Level	Metrics	Calculation	Equation
Node (attraction)	InDegree	$C_{ini} = \sum_{j=1}^{n} l_{ji}$	(1)
	OutDegree	$C_{outi} = \sum_{j=1}^{n} l_{ij}$	(2)
	Centrality	$C_i = C_{outi} + C_{ini}$	(3)
	Vergence	$S_i = (C_{outi} - C_{ini})/C_i \times 100\%$	(4)
Entire valid network	Network density	$D = E/(n(n-1))$	(5)
	Mean flow	$\bar{l} = \sum_{i=1}^{n}\sum_{j=1}^{n} l_{ij} / E$	(6)
	Mean distance weighted by flow	$\bar{d} = \sum_{i=1}^{n}\sum_{j=1}^{n} l_{ij}d_{ij} / \sum_{i=1}^{n}\sum_{j=1}^{n} l_{ij}$	(7)

Network density D, describing the global level of linkages among the nodes, is defined as division of the total number of edges or links E by the maximum number of links on the network. The more connected the nodes are, the denser the network. This global measure varies from 0 to 1. If the value of network density is equal to 1, it means that the nodes are tied to each other. In this study, seven network metrics are employed to analyze the mobility of tourists between the attractions, which are detailed in Table 2 and explained as follows:

C_{ini} and C_{outi} are the inflow and outflow of tourists arriving at and departing from attraction i respectively. l_{ji} is the number of flows or tourists from attraction j to i and l_{ij} from i to j. C_i is the degree centrality and S_i the vergence value at attraction i respectively. D is density of the network，

E is the total number of edges, C_D is the centrality of the network，$\bar{l}$ is a mean flow, d_{ij} is the

network distance between two attractions i and j calculated using ArcGIS, and n is the total number of attractions. In terms of vergence index, if it is greater than 0, it is called divergence，conversely convergence. When the vergence index is 1 (the maximum value), it means the first spot visited in the morning. When the vergence index is -1 (the minimum value), it means the last spot visited in the evening.

3.3.2 Modelling flows of tourists between attractions

Santeramo and Morelli (2016) concluded that distance is a main friction to tourist flows, but its effect varies with the intensity of tourist flows: the latter are less affected. A gravity model is commonly expressed as: $I_{ij}=k\,P_i\,P_j\,f(d_{ij})$, where I_{ij} is the intensity of interaction between i and j，P_i and P_j are the attraction force at locations i and j respectively, d_{ij} is the network distance between i and j, $f(d_{ij})$ is a function of distance d_{ij}. In practice, P_i and P_j are often measured by the economy and population in the two places, $f(d_{ij})$ is either an inverse power function d^β, or a negative exponential function $e^{-\beta d}$. β is a spatial barrier co-efficient or distance friction coefficient. The greater the parameter β is, the more sensitive to distance it is, then the higher speed of decay it is. Wilson has further developed the gravity model based on entropy maximization into flow model (Wilson, 1967), which is $I_{ij}=k\,O_i\,D_j\,/\,f(d_{ij})$, where d_{ij} and $f(d_{ij})$ remain the same, I_{ij} is the intensity of flow from location i to j. When moving to tourist flows, O_i is the outflow at location i and D_j the inflow at location j. In this study, l_{ij} is used to measure I_{ij}, the flows between two attractions, C_{outi} for O_i and C_{inj} for D_j，d_{ij} the real travelling distance between attractions.

$$l_{ij} = kC_{outi}C_{inj}/f(d_{ij}) \qquad （9）$$

In order to consider the variation in the effects of inflow and outflow at attractions, Equation 9 above needs to be revised into Equation 10 as follows (see another example in Patuelli, Mussoni, and Candela 2013):

$$l_{ij} = kC_{outi}^{a}C_{inj}^{b}/f(d_{ij}) \qquad （10）$$

Where two parameters a and b are added into the equation to reflect the intensity of impacts of inflow and outflow on flows. Parameters α and b can be viewed as the elasticity of attractions interactions. This gravity model has also been applied by Morley, Rosselló, and Santana-Gallego (2014) for modelling tourism demand. The interpretation is straightforward. For parameter α, a one-percent increase in inflow size is expected to produce a proportional increase in the number of local flows, and the same for parameter b.

To calibrate the parameters a, b and β, a natural log transformation is taken for equation 10, which results in equation 11 for the power function of $f(d_{ij})$ and equation 12 for the exponential function of $f(d_{ij})$. Then an ordinary least square (OLS) regression analysis is used to seek a best-fit equation with the real data set including flow of tourists, inflows, outflows and distance between attractions.

$$Ln(l_{ij}) = ln(k) + a * C_{outi} + b * C_{inj} - \beta * Ln(d_{ij}) \quad (11)$$

$$Ln(l_{ij}) = ln(k) + a * C_{outi} + b * C_{inj} - \beta * d_{ij} \quad (12)$$

4. Results

Using the three flow networks defined above, social network analysis has been conducted and flow models have been calibrated respectively. Their network and flow characteristics are calculated separately for the purpose of comparison using the metrics defined in Table 2 and models in equations 11-12.

4.1 Nodal position on the tourist flow networks

1）Nodal position on tourist flow networks

The centrality value at each attraction was calculated using equation 3 for each network separately (Figure 2). There is a general trend: the centrality value decreases with the length of trip. The average centrality value over a network has decreased from 97.91 (one-day trip) to 80.13 (two-day trip) and further down to 59.16 (three-and-more-day trip) and their corresponding standard deviation from 188.51 to 115.97 and further down to 64.27. This indicates that a one-day trip is focused on more selective well-known attractions, whereas a three-and-more-day trip on most attractions. These results reflect that the nodal position on the network of attractions varies with trip length. For simplicity and interpretation purposes, all attractions are classified into only two categories: primary and others, based on their centrality values. It is found that the top four attractions with the highest value of centrality are the same for each network, which include the Confucian Temple, Sun Yat-sen's Mausoleum, Xuanwu Lake and Presidential Palace. These four attractions are defined as primary spots and all the rest as others. However, each primary attraction has a varied value of centrality between the networks. On the one-day flow network, the centrality values of Confucian Temple, Sun Yat-sen's Mausoleum, and Xuanwu Lake are all greater than 500. On the two-day flow network, the centrality values of the four primary attractions range between 300 and 500, with the highest value of 476 taken by Confucian Temple. On the three-and-more-day flow network, the highest centrality value by Confucian Temple is 235 and lowest value by Xuanwu Lake 174. Therefore, as the core attractions in Nanjing, Confucian Temple and Sun Yat-sen's Mausoleum are ranked the first two for all networks. These two attractions, as the only two recognized as 5A level (highest) attractions, are the most well-known ones across Nanjing. Both Xuanwu Lake and Presidential Palace are classified as 4A-level tourist attractions. The general declining trend of centrality values indicates that the importance of the four primary attractions in tourists' travel behavior decreases with the length of trip. In some sense, this result confirms the role of time in ranking primary attractions, a theory of tourist attraction as highlighted by Bottia, Peypoch, and Solonandrasana (2008), which means the one-day trip is focused on discovery attractions, but the longer-length trip on escape attractions.
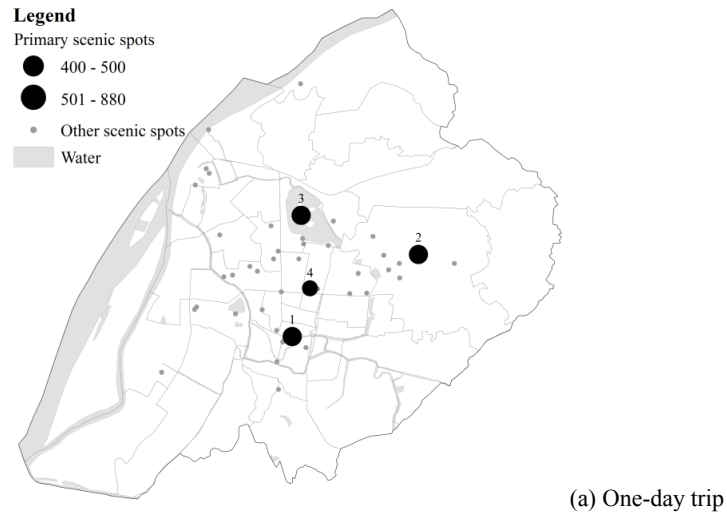
Primary scenic spots

● 400 - 500

● 501 - 880

· Other scenic spots

Water

(a) One-day trip

Legend
Primary scenic spots

● 300 - 400

● 401 - 500

· Other scenic spots

Water

(b) Two-day trip

Legend
Primary scenic spots

● 174 - 200

● 201 - 300

· Other scenic spots

Water

(c) Three-and-more-day trip

Figure 2. Mapping the centrality values of primary attractions across the study area

(Note：1 Confucian Temple, 2 Sun Yat-sen's Mausoleum, 3 Xuanwu Lake, 4 Presidential Palace)

2）The visitation rates to primary attractions

Visitation rates to the top four primary attractions are calculated for each length of trip, shown in Table 3, in which, the percentage of primary attraction k for trip t is calculated as the division of the number of samples involving attraction k for trip t over the total number of samples for trip t. It is clear to see that these visitation rates increase with the trip length. For short-length trips, only the selected primary attractions are visited quickly due to limited time availability. For longer-length trips, each of the most attractive primary sites, together with its surrounding minor sites will be visited each day due to higher time availability.

Table 3. The visitation rates to primary attractions between the three trips

| Primary attraction	One-day trip (%)	Two-day trip (%)	Three-and-more day trip (%)
Confucian Temple	50.06	64.85	73.44
Sun Yat-sen's Mausoleum	45.38	58.27	61.88
Xuanwu Lake	40.55	47.51	54.38
Presidential Palace	36.62	55.54	59.38

3）The vergence of attractions on flow networks

On each network, a vergence value is calculated for each attraction using Equation 4 and then all of the attractions are classified into three categories: divergence, balancing and convergence, according to the vergence values. A one-day trip shares the same set of convergence-type spots (Confucian Temple), divergence-type spots (Sun Yat-sen's Mausoleum) and balancing (the rest), with two-day trip. A three-and-more-day trip adds two more attractions (Xinjiekou Centre and Hulan street) to the convergence-type spots, and two other spots (Presidential Palace and Yangtze River Bridge) to the divergence-type spots. These variations indicate that tourists tend to start with a divergence-type attraction and end with a convergence-type attraction. It also implies that the flow network of a longer-length trip tends to be mixed with divergence-type and convergence-type spots, which might be determined by purchasing power, physical consumption and many other factors. These results demonstrate that temporality has great impact on intensity and direction of vergence.

4.2 Flow characteristics on the tourist flow networks

Using the values of tourist flows l_{ij}, the network density, mean flow and standard deviation are calculated using equations 5-7 for each network respectively. The network density is nearly the same between the three networks (0.216, 0.217 and 0.222 for one-, two- and three-and-more day trip respectively), which is very low. However, there is a large variation in mean flow and its standard deviation between the three networks, demonstrating a decreasing trend (7.27, 4.82 and 3.32 in mean flow and 15.17, 8.83 and 5.28 in standard deviation for one-, two- and three-and-more day trip respectively). The trip over three days is more spatially uniform. It is clear to see that the short-term trip is characterized by the large flows between primary attractions that

(a) One-day trip
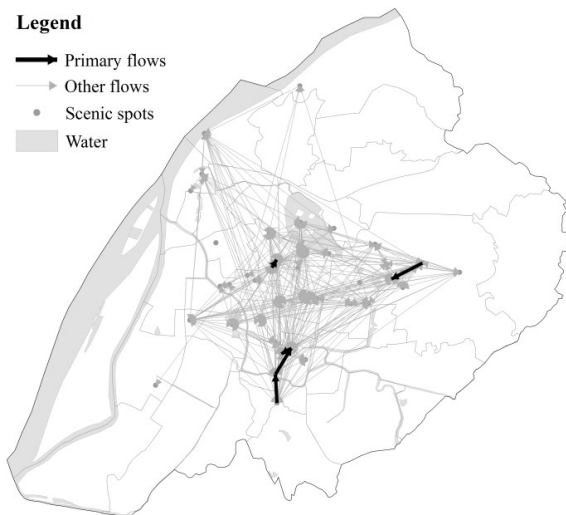
(b)Two-day trip

(c)Three-and-more-day trip

Figure 3. Tourist flow networks for different lengths of trip

have higher values of centrality and conversely the longer-term trip by a small number of flows between minor attractions. For example, in the one-day trip, the total number of tourists from the Presidential Palace to Confucian Temple is 108 (out of 2203). The shortest network distances between attractions are calculated using ArcGIS 13.1, based on the road transport network on Bing Maps. According to Equation 7, the weighted average distance d is calculated for each network. The one-day trip network has the longest distance (5.51 km), the three-and-more day trip network the shortest (3.45 km) and the two-day trip network between both (4.39 km). This index has also shown that the tourist flows on short-term trips tend to be long distance, but the flows of long-term trip tend to be short distance.

The top five flows over each network are defined as primary flows and the rest as others. The distributions of these primary flows are shown in Figure 3. Over the one-day trip network, the largest number of flows with a total value of 108 is from the Presidential Palace to Confucian Temple, followed by the flows with a value of 89 from Sun Yat-sen's Mausoleum to Confucian Temple. These flows are characterized by a long distance between primary attractions. Over the two-day trip network, the largest number of flows with a total value of 113 is from Sun Yat-sen's Mausoleum to Ming Xiaoling Mausoleum, followed by the flows with a value of 67 from the Presidential Palace to Confucian Temple. These flows are characterized by shorter distances between primary attractions than those over the one-day trip network. Over the three-and-more-day trip network, the largest number of flows with a total value of 67 is from Sun Yat-sen's Mausoleum to Ming Xiaoling Mausoleum. The primary flows are dominated by short distances between two primary attractions. The comparisons between the three networks indicate that there is strong association between primary attractions and primary flows. Primary flows are connected with at least one primary attraction. The primary flows from short-term trips tend to include long-distance mobility between primary attractions. However, the primary flows from long-term trips tend to be short-distance mobility between a primary attraction and its surrounding other attractions.

4.3 Flow models

For each trip network, the power and exponential functions and their revised versions, selected to represent distance decay effects, are calibrated by OLS regression analysis (Equations 11 and 12). Comparing the results of regression analysis (Table 4), the adjusted R^2 is the highest for the revised power functions, which are 0.724, 0.662 and 0.633 for one-day, two-day and a three-and-more-day trip respectively. Thereby, the model based on the revised power function is best fitted with the real tourist flows between attractions, so can better interpret the tourist flows. This has further confirmed that most flow networks of tourists have a degree distribution, which follows a power law for most part (Baggio, Scott, and Cooper 2010).

Table 4. Calibrated regression equations

	Function	Regression equation	Adjusted R^2
One-day trip	Power function	$T_{ij} = 0.0606 \times C_{outi}C_{inj} / d_{ij}^{0.373}$	0.455
	Revised power function	$T_{ij} = 0.1307 \times C_{outi}^{0.424}C_{inj}^{0.433} / d_{ij}^{0.248}$	0.724
	Exponential function	$T_{ij} = 0.0014 \times C_{outi}C_{inj} / e^{0.073d_{ij}}$	0.129
	Revised exponential function	$T_{ij} = 0.1208 \times C_{outi}^{0.425}C_{inj}^{0.426} / e^{0.045d_{ij}}$	0.611
Two-day trip	Power function	$T_{ij} = 0.0023 \times C_{outi}C_{inj} / d_{ij}^{0.558}$	0.429
	Revised power function	$T_{ij} = 02047 \times C_{outi}^{0.390}C_{inj}^{0.387} / d_{ij}^{0.370}$	0.662
	Exponential function	$T_{ij} = 0.0015 \times C_{outi}C_{inj} / e^{0.092d_{ij}}$	0.172
	Revised exponential function	$T_{ij} = 0.1537 \times C_{outi}^{0.369}C_{inj}^{0.391} / e^{0.046d_{ij}}$	0.517
Three-and-more-day trip	Power function	$T_{ij} = 0.0037 \times C_{outi}C_{inj} / d_{ij}^{0.747}$	0.529
	Revised Power function	$T_{ij} = 0.3848 \times C_{outi}^{0.308}C_{inj}^{0.330} / d_{ij}^{0.529}$	0.633
	Exponential function	$T_{ij} = 0.0032 \times C_{outi}C_{inj} / e^{0.173d_{ij}}$	0.164
	Revised exponential function	$T_{ij} = 0.3761 \times C_{outi}^{0.312}C_{inj}^{0.298} / e^{0.115d_{ij}}$	0.449

In the three calibrated models using the revised power function, parameters *a* and *b* are all statistically significant at the 1% level. However, parameters *a* and *b* vary with trip length, from 0.424 and 0.433 (one-day trip network) to 0.390 and 0.387 (two-day trip network) down to 0.308 and 0.330 (three-and-more-day trip network), respectively. Comparatively, the three-and-more-day trip network demonstrates a larger difference between *a* and *b*, which is 0.022, than the other networks. It indicates a roughly smaller effect from outflows than inflows on tourist flows between attractions.

The spatial barrier coefficients of flows between the attractions are all statistically significant at the 1% level, being 0.248, 0.370 and 0.529 for one-day, two-day and three-and-more-day trips respectively. In other studies of human mobility using mobile phone and taxi data sets, this parameter *β* is calibrated between 1 and 2 (Kang, Ma, Tong, and Liu 2012; Liu, Kang, Gao, Xiao,

and Tian 2012). Thereby, the parameter β in this study is beyond this range. In addition, in the study of long distance flows between cities in China using the data sets of passengers, its parameter β is also higher than this (Xiao, Wang, Liu, and Wang 2013). These variations indicate that the effect of distance on the flow of tourists within a city is smaller than that of long-distance travel and those of the daily travels of local residents. The travel from origin to destination cities needs to deal with a long distance, and with higher time and economic costs. However, the travel within a destination city only needs to deal with a short distance, as its friction coefficient is smaller than that from the long distance one. As a high transport cost has been invested, tourists often are concerned about flows rather than distance. In addition, distance friction is smaller than those of local peoples' daily activities.

The results from the regression analysis of the three types of trip exhibit that the influence of the attraction system on tourist flows decreases, but the distance decay effect increases with the increase of trip length. Or rather, for a short-period trip, a primary attraction has higher impacts on flows and the effects of distance between attractions decrease. Consequently, tourists tend to move to the primary attractions. For a long-term trip, the influence of attraction hierarchy on tourist flows decreases, but the distance decay effect increases, tourists tend to move to the other attractions within a shorter distance. This may imply that the flow network of tourists does not exhibit scale-free property as demonstrated by other tourism networks (see Baggio, Scott, and Cooper 2010).

5. Discussion

All of the attractions in Nanjing city can be classified into three attraction sub-systems: Sun Yat-sen's Mausoleum attraction sub-system, Confucian Temple attraction sub-system and city center attraction sub-system (Jin, Huang, Xu, and Gu 2014). Tourist flows can be classified into two categories: intra-system flows and inter-system flows. The primary flows within and out of the systems have large variations between the three networks. During a one-day trip, the top four flows are all inter-system mobility. This demonstrates that the one-day flow network is dominated by the long-distance inter-system mobility. During a two-day trip, the primary flows include two inter-system flows and three intra-system flows. The former mainly occurs between the core primary attractions. During a three-and-more-day trip, all the primary flows are intra-system moves, which reveals that this network is dominated by the short-distance mobility with limited smaller number of long-distance inter-system flows.

Based on this analysis, the patterns of tourist flows between attractions in Nanjing city are abstracted into a conceptual model (Figure 4). In the pattern of a one-day trip, primary flows are mostly distributed over the primary attractions between systems and few within the systems. These primary flows dominate the flows of the entire system, connecting different attraction sub-systems. Other flows are distributed within the system.

In the pattern of a two-day trip, primary flows are mostly distributed between the primary and other attractions within the system, and few between the primary attractions of different systems. The other flows distribute uniformly between systems and within a system. In the pattern of a three-and-more-day trip, the primary flows uniformly distribute between the primary and other

attractions within a system and other flows between the primary attractions within different systems, and also between the primary and other attractions within the system.

(a) One-day trip   (b)Two-day trip   (c)Three-and-more-day trip
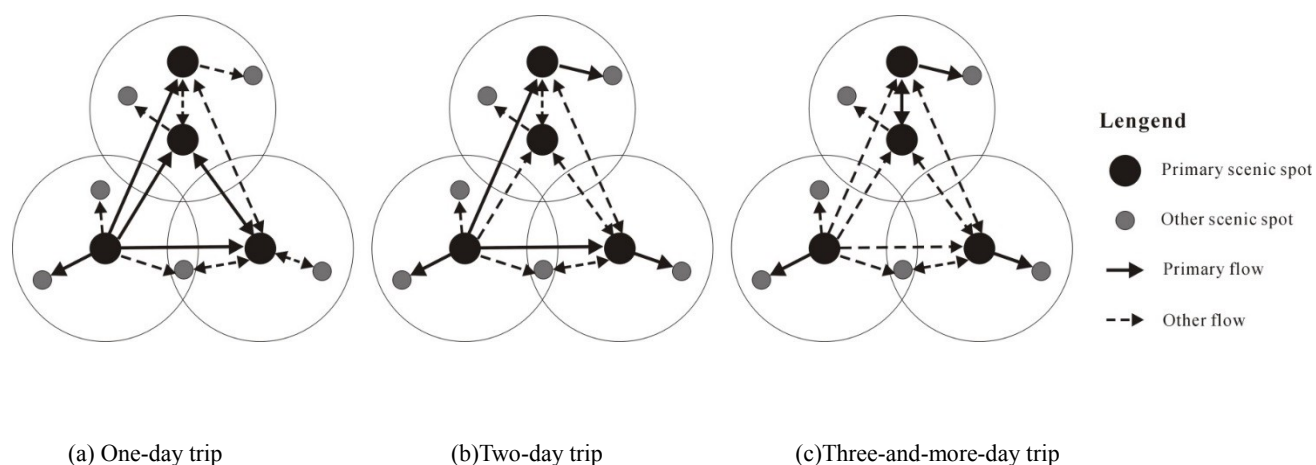
Figure 4. Patterns of tourist flows between attractions for three networks

There have been large variations in the patterns of tourist flow between the three networks. For the short-length trip, the flows between the primary attractions of different systems dominate the whole process of mobility. However, the long-length trip is characterized by intra-system flows. To understand the reasons, we have conducted interviews with the tourists from three networks, 10 from each one. It has been found that:

1) during the one-day trip, eight tourists decided to visit 2-3 well-known primary attractions within different systems (e.g. Sun Yat-sen's Mausoleum, Confucian Temple, Presidential Palace and Xuanwu Lake), but two tourists tended to concentrate on one system by visiting its primary attractions and the other attractions around it. Due to the long distance between the primary attractions within different systems, the one-day trip is dominated by the inter-system flows.

2) during the two-day trip, three tourists choose one system each day and have detailed visits of the primary attractions and its surrounding other attractions. Four tourists tend to visit all attractions within a system in one day, and only the primary attractions of the other two systems for the second day. A further three tourists did not have any specific preferences and focused more on primary attractions. As a result, the two-day trip network is characterized by long-distance inter-system and short-distance intra-system moves.

3）during the three-and-more-day trip, due to the availability of sufficient time, ten tourists all had detailed plans and tended to concentrate on one attraction sub-system per day so that they could explore all of the attractions in detail. Their travelling efficiency can be maximized due to the short distances between attractions within each system. Consequently, this trip network is characterized by short-distance moves.

6. Conclusions

In this paper, taking Nanjing City as a case study, user generated content (UGC) with high spatial and temporal resolution has been successfully utilized for constructing three flow networks corresponding to different lengths of trip, analysing the temporal heterogeneity of tourist mobility, allowing significant conclusions to be drawn, as follows:

The position of each attraction, highly correlated with the centrality value, in the local attraction systems, declines with the length of trip. The intensity and direction of flow vergence also demonstrate significant temporality. The distance decay effect in intra-urban tourist movements, following power law, declines with the length of trip. Their spatial barrier coefficients demonstrate that these spatial flows are less sensitive to distance, when compared with inter-city (Xiao, Wang, Liu, and Wang 2013) and the daily human movements (Liu Kang, Gao, Xiao, and Tian 2012). It also indicates the scale-dependent property of tourist flows.

With the increase of trip length, the influence of attraction systems on flows is weakening, but the influence of distance is continuously enhancing. Tourists tend to change from long-distance inter-system to short-distance intra-system movements. These tourist behaviours might be explained by the theories of "process" oriented tourists (Lew and Mckercher 2006) and 'Escape/discovery attractions' (Bottia, Peypoch, and Solonandrasana 2008). To achieve this, more contextual data from the web service needs to be extracted and analysed with text analytics, such as opinions and comments on these attractions. These temporal (decaying) effects mentioned above indicate the roles of time in perceiving, ranking and visiting primary attractions, which provide empirical evidence to the theories regarding time in tourism studies (Leiper 2004; Fuchs, Höpkenb, and Lexhagena 2014).

This case study has confirmed that UGC data, a form of big data and digital footprint, can be utilized for exploring temporal effects in tourist mobility through a static network analysis and spatial interaction model at the aggregate level. However, social network analysis and spatial interaction models focusing on pattern analysis are not able to model spatial and temporal processes. The UGC data used in this paper is a small sample in terms of its temporal scale (only 1 year). With a large size sample in the future, these UGC data can be further used to model the dynamics of tourists' mobility (e.g. Yang, Fik, and Zhang 2013) through developing Markov Chain models (e.g. Vu, Li, Law, and Ye 2015). As one form of big data, many other elements of digital footprint, such as photos and narratives are not explored yet, which contain information on tourists' experience, altitude and demographics, and can be used to analyse the tourists' behaviour and emotions (e.g. sentiment analysis). The study is also subject to sampling bias as only a small proportion of tourists have recorded their trips online, and in particular, those who might be deprived of digital access are not sufficiently represented in the samples. This reflects the drawbacks of UGC though it has also provided evidence for the linkage between citizen science (Goodchild, 2007) and tourism studies (e.g. Brosnan, Sebastian, and Rock 2015). It is rewarding to see how citizen science can potentially be incorporated into travel studies. The limitations of the web service (where to go?) include no attribute data for tourists (e.g. age, gender, origin), number of visits (first-time or repeat visit), types of tourism (e.g. pleasure), where tourists' stayed, and

how they travelled to the destination. These data, which are recommended for the web service to add in the future, will not only help to explore more spatial and temporal patterns and behaviour (e.g. different consumption patterns between first time and repeat visitors, and influence of hotel location on behavior patterns) but also interpret the mechanisms behind the disclosed spatial and temporal heterogeneity from the perspective of tourists' behaviour. The incorporation of the temporal dimension into tourist mobility study will enhance understanding of the valuation of time in tourists' behaviour.

Theoretically, this case study reflects the tourist mobility in a rapidly developing city, whose tourism studies remains conceptually ill-equipped, given most of its theories have been generated from Western contexts (Winter 2009). This paper, therefore, begins to address this significant gap in the field through exploring the temporal heterogeneity of tourism mobility in a non-western context.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Author Biographies

Cheng Jin (jincheng@njnu.edu.cn) is an associate professor of tourism geography at Nanjing Normal University, whose research interests include regional tourism development and GIS applications. Dr Jin gained his doctoral degree from Nanjing Normal University, China.

Jianquan Cheng (J.cheng@mmu.ac.uk) is a reader of urban studies at MMU, whose research interests include big data for urban applications and the urbanisation in China. Dr Cheng gained his doctoral degree from Utrecht University, Netherlands.

Jing Xu (xq514@163.com) is a lecturer of tourism geography at Nanjing Xiaozhuang Univeristy. Her main research interests are internet and tourism, and traveler consumption behavior. Dr Xu gained her doctoral degree from Nanjing Normal University, China.

References

Asero, Y., S. Gozzo, and V. Tomaselli. (2016). "Building tourism networks through tourist mobility." *Journal of Travel Research*, 55 (6): 751-63.

Baggio, R., N. Scott, and C. Cooper. (2010). "Network science a review focused on tourism." *Annals of Tourism Research*, 37(3): 802-27.

Bottia, L., N. Peypoch, and B. Solonandrasana. (2008). "Time and tourism attraction." *Tourism Management*, 29(3): 594-96.

Brosnan, T., F. Sebastian, and J. Rock. (2015). "Exploring synergies: Hopeful tourism and citizen Science." *Annals of Tourism Research*, 53: 96-98.

Castells, M. (2001). The Internet Galaxy: Reflections on the Internet, Business, and Society. Oxford: Oxford University Press.

CEIC China Database. (2014). Retrieved from http://www.ceicdata.com (Authorized access).

Fuchs, M., W. Höpken, and M. Lexhagen (2014)," Big data analytics for knowledge generation in tourism destinations − A case from Sweden." *Journal of Destination Marketing & Management*, 3(4):198–209.

Goodchild, M. F. (2007). "Citizens as voluntary sensors: Spatial data infrastructure in the world of web 2.0." *International Journal of Spatial Data Infrastructures Research*, 2: 24–32.

Greer, T., and G. Wall (1979). "Recreational Hinterlands: A Theoretical and Empirical Analysis." In Recreational Land Use in Southern Ontario (Department of Geography Publication Series No. 14), edited by G. Wall. Waterloo, Canada: Waterloo University, pp. 227-46.

Grinberger, A. Y., N. Shoval, and B. McKercher. (2014) "Typologies of tourists' time–space consumption: a new approach using GPS data and GIS tools." *Tourism Geographies*, 16(1): 105-23.

Halás, M., P. Klapka, and P. Kladivo. (2014). "Distance-decay functions for daily travel-to-work flows." *Journal of Transport Geography*, 35:107-19.

Hannam, K., G. Butler, and C. M. Paris. (2014). "Developments and key issues in tourism mobilities." *Annals of Tourism Research*, 44(1): 171-85.

Hooper, J. (2015). "A destination too far? Modelling destination accessibility and distance decay in tourism." GeoJournal, 80(1): 33-46.

Hwang, Y. H., U. Gretzel, and D. R. Fesenmaier, (2006). "Multicity trip patterns: tourists to the united states." *Annals of Tourism Research*, 33(4): 1057-1078.

Jiangsu Bureau of Statistics (JSBS). (2015). *Jiangsu Statistical Yearbook 2015*. Beijing: China Statistical Press.

Jin, C., Z. Huang, J. Xu, and Q. Gu. (2013). "Evaluation of travel convenience for residents in the Yangtze River Delta, China." *Asia Pacific Journal of Tourism Research*, 18(8): 865-79.

Kim, J. and D. R. Fesenmaier, (2015). "Measuring emotions in real time: Implications for tourism experience design." *Journal of Travel Research*, 54(4): 419-29.

Leiper, N. (2004). Tourism management (3rd ed). Australia: Pearson Education.

Leung, X., F. Wang, B. Wu, B. Bai, K. A. Stahura, and Z. Xie. (2012). "A Social Network Analysis of Overseas Tourist Movement Patterns in Beijing: the Impact of the Olympic Games." *International Journal of Tourism Research*, 14(5): 469-84.

Lew, A., and B. Mckercher. (2006). "Modeling tourist movements: A local destination analysis." *Annals of Tourism Research,* 33(2): 403-23.

Liu, Y., C. Kang, S. Gao, Y. Xiao, and Y. Tian. (2012). "Understanding intra-urban trip patterns from taxi trajectory data". *Journal of Geographical Systems*, 14(4): 463-83.

Liu, Y., Z. Sui, C. Kang, and Y. Gao. (2014). "Uncovering patterns of inter-urban trip and spatial interaction from social media check-In data." *PLoS ONE*, 9(1): e86026.

Lu, W., and S. Stepchenkova. (2015). "User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software." *Journal of Hospitality Marketing & Management*, 24(2): 119-54.

McKercher, B., and A. Lew. (2003). "Distance decay and the impact of effective tourism exclusion zones on international travel flows." *Journal of Travel Research*, 42(2), 159-65.

McKercher, B., A. Chan, and C. Lam. (2008). "The Impact of distance on international tourist movements." *Journal of Travel Research*, 47(2): 208-24.

Mckercher, B., and G. Lau. (2008). "Movement patterns of tourists within a destination." *Tourism Geographies*, 10(3): 355-74.

McKercher, B., N. Shoval, E. Ng, and A. Birenboim. (2012). "First and repeat visitor behaviour: GPS tracking and GIS analysis in Hong Kong." *Tourism Geographies*, 14(1): 147-61.

Modsching M., R. Kramer, U. Gretzel and K. ten Hagen. (2006). "Capturing the beaten paths: a novel method for analysing tourists' spatial behaviour at an urban destination". In: Hitz M., Sigala M., Murphy J. (eds) *Information and Communication Technologies in Tourism*, 75-86, Springer, Vienna.

Morley, C., J. Rosselló, and M. Santana-Gallego. (2014). "Gravity models for tourism demand: theory and use." *Annals of Tourism Research*, 48: 1-10.

Nicolau, J.L., F. J. Zach, and I. P. Tussyadiah. (2016). "Effects of distance and first-time visitation on tourists' length of stay." *Journal of Hospitality & Tourism Research*, (in press) doi:10.1177/1096348016654972.

Önder, I., W. Koerbitz, and A. Hubmann-Haidvogel. (2016). "Tracing tourists by their digital footprints: the case of Austria", *Journal of Travel Research*, 55: 566-73.

Patuelli, R., M. Mussoni, and G. Candela. (2013). "The effects of world heritage sites on domestic tourism: a spatial interaction model for Italy." Journal of Geographical System, 15(3):369-402.

Ring, A., A. Tkaczynski, and S. Dolnicar. (2016). "Word-of-Mouth Segments: Online, offline, visual or verbal?" *Journal of Travel Research*, 55(4): 481-92.

Santeramo, F. G. and M. Morelli. (2016). "Modeling tourism flows through gravity models: a quantile regression approach." *Current Issues in Tourism*, 19(11):1077-83.

Shih, H. Y. (2006). "Network characteristics of drive tourism destinations: an application of network analysis in tourism." *Tourism Management*, 27 (5): 1029-39.

Shoval, N. (2012). "Time geography and tourism". In J. Wilson (Ed.), *The Routledge Handbook of Tourism Geographies* (pp. 174–180). London: Routledge.

Shoval, N., and A. Raveh. (2004). "Categorization of tourist attractions and the modeling of tourist cities: based on the co-plot method of multivariate analysis." *Tourism Management*, 25(6): 741-50.

Shoval, N., and R. Ahas. (2016). "The use of tracking technologies in tourism research: the first decade." *Tourism Geographies*, 18(5): 587-606.

Smallwood, C. B., L. E. Beckley, and S. A. Moore. (2012). "An analysis of visitor movement patterns using travel networks in a large marine park, north-western Australia." *Tourism Management*, 33: 517-28.

Stienmetz, J. L., and D. R. Fesenmaier. (2015). "Estimating value in Baltimore, Maryland: An attractions network analysis." *Tourism Management*, 50: 238-52.

Stillwell, J., M. Bell, P. Ueffing, K. Daras, E. Charles-Edwards, M. Kupiszewski, and D.

Kupiszewska. (2016) "Internal migration around the world: comparing distance travelled and its frictional effect." *Environment and Planning A*, 48(8):1657-75.

Vu, H., G. Li, , R. Law, and B. Ye. (2015). "Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos." *Tourism Management*, 46: 222-32.

Wilson, A. G. (1967). "A statistical theory of spatial distribution models." *Transportation Research*, 1(3): 253-69.

Winter, T. (2009). "Asian tourism and the retreat of anglo-western centrism in tourism theory." *Current Issues in Tourism*, 12(1): 21-31.

Wong, I. A., L. H. N. Fong, and R. Law. (2016). "A longitudinal multilevel model of tourist outbound travel behavior and the dual-cycle model." *Journal of Travel Research*, 55(7): 957-70.

Wu, M., and P. L. Pearce. (2016). "Tourism blogging motivations: why do Chinese tourists create little "lonely planets"?" *Journal of Travel Research*, 55(4): 537-49.

Wu, Q., J. Cheng, G. Chen, D. J. Hammel, and X. Wu. (2014). "Socio-spatial differentiation and residential segregation in the Chinese city based on the 2000 community-level census data: A case study of the inner city of Nanjing." *Cities*, 39: 109-19.

Xiao, Y., F. Wang, Y. Liu, and J. Wang. (2013). "Reconstructing gravitational attractions of major cities in China from air passenger flow data 2001-2008: a particle swarm optimization approach." *The Professional Geographer*, 65(2): 265-82.

Xiang, Z., D. Wang, J. T. O'Leary, and D. R. Fesenmaier. (2015). "Adapting to the internet: trends in travelers' use of the web for trip planning." *Journal of Travel Research*, 54(4): 511-27.

Yang, Y., T. Fik, and J. Zhang, (2013). "Modeling sequential tourist flows: where is the next destination?" *Annals of Tourism Research*, 43(7): 297-320.

Yuan, F., J. Gao, and J. Wu. (2016). "Nanjing-an ancient city rising in transitional China." *Cities*, 50: 82-92.

Zach, F. and U. Gretzel. (2011). "Tourist-activated networks: implications for dynamic bundling and EN Route Recommendations." *Information Technology & Tourism*, 13(3), 229-238.

Zhang, L., L. Wu, and A. S. Mattila. (2016). "Online reviews: The role of information load and peripheral factors." *Journal of Travel Research*, 55(3): 299-310.

Zhou, X., C. Xu, and B. Kimmons. (2015). "Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform." *Computers, Environment and Urban Systems*, 54: 144-53.