Tourism Management 82 (2021) 104195

Contents lists available at ScienceDirect

Tourism Management

journal homepage: http://www.elsevier.com/locate/tourman

Characterizing destination networks through mobility traces of international tourists — A case study using a nationwide mobile positioning dataset

Yang Xu [a], Jingyan Li [a], Alexander Belyi [b,c], Sangwon Park [d,*]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[b] Singapore-MIT Alliance for Research and Technology, 1 Create Way, Singapore
[c] Faculty of Applied Mathematics and Computer Science, Belarusian State University, Minsk, Belarus
[d] Hospitality and Tourism Research Centre (HTRC), School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ARTICLE INFO

Keywords:
Tourist mobility
Mobile positioning
Network science
Community detection
Tourism big data

ABSTRACT

This article demonstrates how large-scale tourist mobility data can be linked with network science approaches to better understand tourism destinations and their interactions. By analyzing a mobile positioning dataset that captures the nationality and movement patterns of foreign tourists to South Korea, we employ a few metrics to quantify the network properties of tourism destinations, aiming to reveal the collective dynamics of tourist movements and key differences across nationalities. According to the results, the number of inbound tourists to destinations follows a log-normal distribution, which indicates a notable heterogeneity of destination attractiveness. Although this finding holds across different nationalities, we find that tourists from different countries tended to visit different places in South Korea. A community detection algorithm partitions South Korea into several tourism regions, each covering a set of destinations that are closely connected by tourist flows. The implications for transportation development and regional tourism planning are discussed.

1. Introduction

Tourism researchers have substantially contributed to understanding travel behaviors at the destination and estimating tourism demand. The relevant literature has suggested the existence of inequality in tourism income and demand across different travel regions (Li, Chen, Li, & Goh, 2016). This can be fundamentally attributed to the disintegrated nature of tourism destinations, which consist of diverse stakeholders (i.e., destination marketing organizations: DMOs) with various business goals (Wang & Xiang, 2007). Accordingly, the destination marketing and promotion carried out by individual DMOs can be challenged to form holistic destination image and is hard to make it successful in a sustainable manner (Wang, 2008). Recognizing competitive marketplace in tourism, the collaborative marketing that involves collective efforts from different DMOs can create competitive advantage for a destination (Fyall & Garrod, 2005). In this sense, this study explores the network structure of tourism destinations, which are the foundation of collaborative strategy in tourism marketing based on an understanding of tourist mobility. Much of the knowledge from existing research has been built upon augmented capabilities to track tourist mobility in time and space (Shoval & Ahas, 2016). Apart from the continuous adoption of surveys, more and more studies are benefiting from the use of technologies such as Global Positioning System (GPS), mobile positioning, Bluetooth tracking, and geocoded social media (Li, Xu, Tang, Wang, & Li, 2018).

Studies that link destinations and tourist mobility can generally be categorized into two groups — one focusing on tourist dynamics within destinations, and the other mainly investigating inter-destination movements. Among studies in the latter group, some have employed network science approaches (Asero, Gozzo, & Tomaselli, 2016; Miguéns & Mendes, 2008; Shih, 2006; Stienmetz & Fesenmaier, 2019). These studies treat tourism destinations as complex networks, of which the topological and structural properties can be derived and analyzed through observations of tourist movements. From a quantitative point of view, network science provides a rich set of tools and metrics for tourism research (Baggio, 2017; Baggio, Scott, & Cooper, 2010), and many of

* Corresponding author.
E-mail addresses: yang.ls.xu@polyu.edu.hk (Y. Xu), jing-yan.li@connect.polyu.hk (J. Li), alex.bely@smart.mit.edu (A. Belyi), sangwon.park@polyu.edu.hk (S. Park).

https://doi.org/10.1016/j.tourman.2020.104195
Received 4 February 2020; Received in revised form 19 May 2020; Accepted 6 July 2020

them can be used to characterize tourism destinations and their interactions (e.g., degree or strength of node, network density, community structure).

Despite the increasing availability of tourism big data, there has been a scarcity of research that investigates destination networks based on mobility of foreign tourists. Although there have been studies for understanding tourism networks (Asero et al., 2016; Miguéns & Mendes, 2008; Shih, 2006; Stienmetz & Fesenmaier, 2019), the focus has been more on characterizing destinations (e.g., degree centrality and distribution) and the structural properties of the networks (e.g., network density). Little effort has been devoted to quantifying the spatial interactions of tourism destinations through large-scale mobility observations. Based on the idea that tourism destination is a complex network system (Baggio, 2008), understanding of spatial interactions formulating clustered destinations guides developing strategic planning to accelerate destination collaboration. From the data perspective, survey and GPS data collection are costly and time-consuming (Shoval & Ahas, 2016). These data usually cover small sample sizes and are mostly used to understand tourist mobility within destinations (e.g., a city) or around attractions (e.g., a park). Social media data (e.g., geo-tagged photos) allow for tourism analysis at broader scales and they could reveal rich contextual information about travelers. However, such data can be sparse and irregular in time and space (Lo, McKercher, Lo, Cheung, & Law, 2011). Mobile positioning data could capture location footprints of large populations. At the downside, such data are more difficult to acquire and usually fall short of collecting sociodemographic attributes of travelers (Ahas, Aasa, Roose, Mark, & Silm, 2008). These issues have hindered a systematic understanding of: (1) the network properties of destinations and their spatial interactions in relation to mobility of international tourists, and (2) the collective behavioral difference among travelers from different foreign countries.

This study aims to address these gaps by linking tourist mobility data with network science approaches. By analyzing a large-scale mobile positioning dataset that captures the nationality and movement patterns of foreign tourists to South Korea, we employ a few metrics to quantify the network properties of tourism destinations, aiming to reveal the collective dynamics of tourist movements and key differences across nationalities. First, we extract inter-destination tourist movements from the mobile positioning dataset to build destination networks. Two types of networks (directed and undirected) are formed to reflect the properties of destinations and their interactions. The directed networks are primarily used to quantify destination attractiveness, and the undirected networks — which are generated based on the topology of directed networks — are used to describe aggregate tourist flows between destinations. Then, by analyzing the distributions of node strength in the directed networks, we explore the attractiveness of destinations to the overall tourist population, followed by a segmentation analysis that unveils the mobility preference of different nationalities. Finally, we apply a community detection algorithm over the undirected networks, with the purpose to reveal groups of destinations that are closely connected by tourist flows. The main research questions we aim to address are as follows:

- Which destinations in South Korea were traveled more by international tourists? Is there a large heterogeneity of attractiveness among the destinations?
- Did travelers from different foreign countries tend to visit the same set of destinations? If not, what are the key differences?
- Which destinations were closely connected by tourist flows? Do networks derived from different nationalities show diverging community structures? What are their implications for regional tourism planning?

This research provides important insights contributing to knowledge of tourism destination networks and destination collaboration. The findings of this paper discover non-linear and dynamic structures of tourism network as well as heterogeneity of destinations' attractiveness through the lens of travel inflow, suggesting complex tourism systems (Baggio, 2008). As opposed to previous tourism studies that use social network analysis mainly describing the shape of network (Casanueva, Gallego, & García-Sánchez, 2016), this study applies community detection algorithm as a type of network science analytics. The insights derived from community detection method suggests important knowledge in the development of strategic destination collaboration and cooperation as a way to enhance destination competitiveness (Fyall, Garrod, & Wang, 2012). The analytical framework (e.g., community detection algorithm) can also be applied to comparable datasets in other countries and regions to facilitate destination management such as transportation efficiency and tourism planning. Furthermore, considering tourism literature largely focusing on travel decision-making process (e.g., perceived intention) (Jönsson & Devonish, 2008), this research identifies different travel flow between travelers from different origin countries. As a result, this finding implies an innovative clustering method based on the travel movement patterns.

2. Literature review

2.1. Network science

Network science is a study of network models based on mathematical theory, which investigates, analyzes and characterizes networks' behavior (Javed, Younis, Latif, Qadir, & Baig, 2018). The study of networks has observed a significant advancement in understanding and evaluating structural and dynamic properties of large scale networks, whereby researchers utilize a set of tools and techniques to assess topological properties of a network and its influence on behavior and evolution (Newman, 2003). In general, networks are represented by graphs composing a group of nodes (vertices) with links between them (edges), which denotes the graph theory. Network theory develops on the basis of the assumption that a cause, effect, or association between objects (or aspects) involves something that can be conceptualized as a network (Brandes, Robins, McCranie, & Wasserman, 2013). That is, a network system can be modeled as an ensemble of connected elements (Baggio, 2017). Basically, network science involves identifying the unifying principles that illustrate generic patterns/rules of dynamic behaviors and explicate the structural features being uncovered. Newman, Barabási, and Watts (Newman, Barabási, & Watts, 2006) have shown that network science focuses on the properties of real-world networks derived from empirical and theoretical questions and understands networks not only as topological objects, but also as a framework of dynamical systems.

A network system is not a simple aggregation of consistent elements, but it is patterned. The random graph theory proposed by Erdős and Rényi (Erdős & Rényi, 1960) shows the invariants of graphs, signifying that the links in a network are placed randomly between pairs of nodes. Indeed, the degree distribution (i.e., number of connections to nodes) follows the average degree of the network. Importantly, however, recent scholars have identified the existence of dependence among ties. That is, many ties centered on a popular actor attract the presence of more ties to the same actor: the rich get richer. The distribution of connections is largely uneven; some nodes play as largely connected hubs with a large number of ties, whereas most nodes have a small number of links (Milo et al., 2002). This feature calls for scale-free networks exhibiting a power-law distribution. The tie dependence, or power-law distribution, implies that the network structure evolves and self-organizes, which refers to a notion of complex networked systems.

Along with a sound theoretical framework, a number of network analytical methods have been developed to compute, draw, and analyze the patterns of connections between the elements (or actors) in a network system. Network analysis assists researchers in modeling an ensemble of distinct elements (the nodes or vertices of the network) and identifying the relationships (the links or edges) that include linkage weights and directions. Social network analysis is a quantitative method that is most widely used in social science, which examines the arrangement of coordination and integration between actors in a network structure, such as individuals, groups and organizations (Leung et al., 2012; Wang & Xiang, 2007). Network analysis has mainly been applied in tourism in order to (1) understand the evolution of business networks and assess inter-organizational relationships (Pavlovich, 2003; Scott, Baggio, & Cooper, 2011) and (2) estimate the relationships between public and private sectors and the structure of tourism destination with involvement of manifold stakeholders (Baggio & Cooper, 2010; Erkuş-Öztürk & Eraydın, 2010). The results of social network analysis imply that tourism organizations act through interactive networks, where value is generated by nurturing collaboration (Williams Hristovet al., 2018).

2.2. Tourism destinations as a complex network system

"Network" denotes a set of components that interact with each other, generally in a nonlinear manner. Complex systems involve the evolution of self-organization, appearing as neither completely regular nor fully random patterns, in order to facilitate development of emergent behavior (Sayama, 2015). A tourism destination has been regarded as a complex dynamic system composed of interrelated components, including not only a large amount of natural, cultural, artistic and economic resources, but also institutional actors who provide tourism services (Baggio et al., 2010; Dâ€™Agata, Gozzo, & Tomaselli, 2013; Miguéns & Mendes, 2008). These elements of the system share interdependent relationships, and the system evolves to respond to the influences of external and internal factors (Baggio, 2008). The actors, such as local firms, organizations, associations, and people, represent "nodes". A multi-actor structure that is closely connected in creating the tourism products and services in the destination network represents "links".

McKercher (McKercher, 1999) has suggested that tourism functions are non-linear and behave in a dynamical manner in which it is difficult to see direct cause and effect between actions. The network reflecting the dynamic relationships of organizations (or stakeholders) in a tourism destination assists in enumerating the "real world" in tourism. Building upon the implications of McKercher's study, a number of tourism researchers have exhibited the complex structure of tourism systems by applying complex theory and the existence of complex attributes, such as ontological realism, non-linear and complex relationships, difficulty to predict, fragility of the system to abrupt changes of direction, and emergence of modular structures (Baggio et al., 2010; Byrne & Callaghan, 2013; Farrell & Twining-Ward, 2004; Newman et al., 2006; Paget, Dimanche, & Mounet, 2010; Zahra & Ryan, 2007). Sainaghi and Baggio (Sainaghi & Baggio, 2017) have empirically demonstrated the complex structure of tourism destinations and confirmed its structural evolution over time. Williams and Hristov (Williams Hristovet al., 2018) examined the role and effect of DMO (destination marketing organizations) on interactions of destination networks, such as communication between members, based on the assumption that the network structure has an influence on the efficiency of communication and knowledge sharing activities in destination networks.

The range of network science is broad and covers extensive facets of tourism fields. The initial application has been explored to identify the structural features of tourism destination with multiple stakeholders (Hede & Stokes, 2009), the fundamental properties of inter-organizational destination networks (Scott, Cooper, & Baggio, 2008) and global airport networks (Guimera, Mossa, Turtschi, & Amaral, 2005). Another stream applying network science focuses on tourism supply chain emphasizing collaboration in tourism destinations (Baggio, 2011; Erkuş-Öztürk & Eraydın, 2010), destination value systems (Stienmetz & Fesenmaier, 2019), knowledge transfer (Baggio & Cooper, 2010), and business networks considering hotel performance (Sainaghi & Baggio, 2014). These studies have concluded that the insights obtained from network science and understanding of network evolution make it possible to create and replenish competitive advantage for destinations and to develop a strategy for attracting the interest of new targets (Leiper, 2000; Sainaghi & Baggio, 2017).

Tourism researchers have suggested that understanding tourist mobility has become one of the most important issues in the development of tourism planning (Grinberger & Shoval, 2019). The notion of network science has been applied to better understand tourism flows from both global (Lozano & Gutiérrez, 2018) and local perspectives (González-Díaz, Gómez, & Molina, 2015). For example, a primary study by Hwang, Gretzel, and Fesenmaier (Hwang, Gretzel, & Fesenmaier, 2006) has explored international tourists' multicity trip patterns within the United States. They considered multidestination trip patterns as network structures that imply the representation of locations as nodes and travel movement between cities as links, which allows for the formulation of the structural properties of a multicity trip. Three sets of network-related concepts, including node centrality, betweenness centrality, and closeness centrality, have been taken into account across travelers from different origin destinations. Likewise, Shih (Shih, 2006) has used network analysis to investigate the structural configuration of destinations in Nantou County, Taiwan, by measuring degree centrality, closeness centrality, betweenness centrality, and structural holes. As a result, they identified key places to understand travel movement patterns better. In addition, network analysis has been used to assess the configuration of tourism demand. González-Díaz et al. (González-Díaz et al., 2015) explored the structural dynamic of the regional tourism network based on changes in tourist flow over time in both hotel and non-hotel accommodations and identified the evolutionary trend of tourism in Spain. With increasing accessibility to various massive tourism big data sources, such as digital footprints, usage of network analysis associated with network theory has been accelerated by tourism researchers (Lozano & Gutiérrez, 2018), for example, the structure of visitor flow exploring geotagged data from Flickr associates with economic values at the tourism destinations (Stienmetz & Fesenmaier, 2019). A tourism destination is regarded as a system consisting of divergent supply-side actors interdependent on each other through various activities and sharing resources. Stienmetz and Fesenmaier's study (Stienmetz & Fesenmaier, 2019) demonstrated the mechanism by which the network structure of aggregate demand generates the value created within a destination. Network science guides tourism researchers to illustrate the spatial distribution of tourism mobility and develop destination clustering based on structural equivalence in a

Table 1
Example of an individual's mobile phone records in the dataset.

User ID	Nationality	Date	Starting Time	Ending Time	Longitude	Latitude
123***	***	2018-08-01	00:14:00	08:57:00	126.***	37.***
123***	***	2018-08-01	09:47:00	10:41:00	127.***	37.***
123***	***	2018-08-01	11:35:00	12:29:00	127.***	37.***
…	…	…	…	…	…	…
123***	***	2018-08-04	19:11:00	20:59:00	128.***	38.***
123***	***	2018-08-04	21:53:00	23:25:00	128.***	38.***

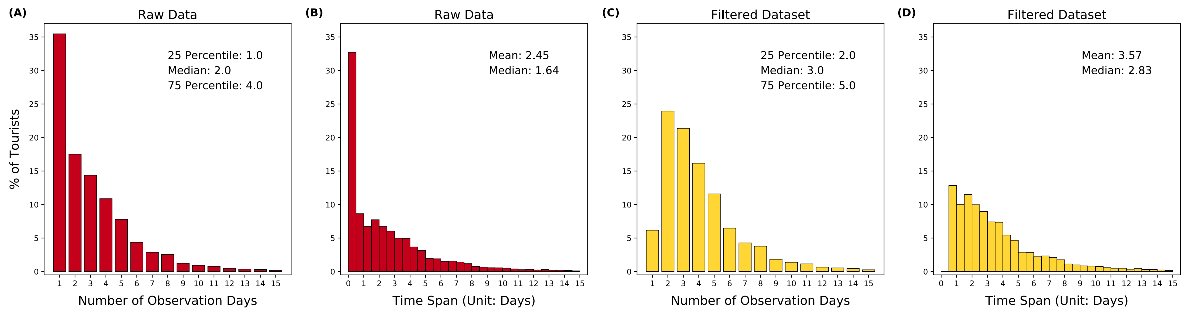


Fig. 1. Distribution of number of observation days and time span of users in: (A–B) the raw data and (C–D) the filtered dataset.

network (Dâ€™Agata et al., 2013). As a result, network science helps tourism researchers identify the structure of complex real-world destination networks and determine communities within the tourism networks so as to understand the entire phenomenon and its complexities in tourism (Farrell & Twining-Ward, 2004; Newman, 2003).

3. Research design

3.1. Mobile positioning dataset

This study uses a large-scale mobile positioning dataset collected in South Korea. The anonymized dataset tracks the location footprints of 192,302 users during a period of 15 days (August 1st to 15th, 2018). Different from Call Detail Records (CDRs, see (Gonzalez, Hidalgo, & Barabási, 2008; Xu, Belyi, Bojic, & Ratti, 2018)) or mobile signaling data (Yan, Wang, Zhang, & Xie, 2018) — which capture individuals' sightings at discrete time points — this dataset consists of records that document the location and dwelling time of users during their stays in South Korea. The locations in the dataset were tracked at the level of cellphone towers. The dataset was preprocessed by the data provider to generate estimates of dwelling time at the tower level.

Table 1 shows an example of an individual's phone records. Each record tracks the unique ID and the nationality of the user, the location (lng/lat of cellphone tower) he or she stayed, as well as the date, starting time and ending time that define the corresponding stay period. In other words, each user's diary consists of records that document the stay activities, and the time periods between consecutive records indicate movements among locations. For example, the first two rows in Table 1 indicate that the user stayed at two different locations between [00:14:00–08:57:00] and [09:47:00–10:41:00] respectively, and a trip was possibly conducted by the user in between (i.e., [08:57:00–09:47:00]).

The densities of cellphone towers in space reflect the spatial granularity of the dataset. To obtain a good understanding of their spatial arrangement, we measure, for each cellphone tower, its distance to the nearest peer. According to the distribution, the 25^{th} percentile, median, and 75^{th} percentile of the values are 139.7, 264.6 and 632.2 m, respectively. Overall, the dataset provides a fine-grained view of tourist mobility in time and space.

3.2. Filter users with a brief stopover

For the 192,302 mobile phone users, their duration of stay in South Korea could vary from one person to the other. Here, we compute two values for each user — the *number of observation days* and the *time span* — to better understand the characteristics of the dataset. The number of observation days is defined as the total number of days with records, while the time span is measured as the duration between the starting time of the first record and the ending time of the last record. These two values, which are highly correlated, reflect how long the users tended to stay in South Korea. As shown in Fig. 1A, a substantial amount of users stayed in South Korea for only a few days. This observation is further confirmed by the distribution of time span (Fig. 1B), where more than 30% of users have a time span less than 12 h (i.e., half of a day). This suggests that many users had a brief stopover in the country and some of them could be transfer passengers. To tackle this issue and target primarily at tourists, this study focuses on users who have a time span greater than 12 h. This results into a subset of 129,332 users. The median number of observation days changes from 2.0 in the raw data (Figure 1A) to 3.0 in the filtered dataset (Fig. 1C), and the average time span changes from 2.45 days (Figure 1B) to 3.57 days (Fig. 1D).

3.3. Extract inter-city tourist movements

The administrative divisions of South Korea follow a general hierarchy. On the top there are seventeen provincial-level divisions. These first-tier divisions are subdivided into municipal-level units, which consist of cities, counties, and districts that are defined based on the population.[1] The municipal-level divisions are used in this study to generate destination networks. For simplicity, we refer to all of them as "cities" in the remaining of the article. There are a total of 250 cities and they are represented as nodes in the destination networks (Fig. 2).

To generate city-level tourism networks, the initial task is to convert individual trajectories to the compatible spatial scale. Note that an individual's mobile phone records can be represented as a list of tuples $T = \{(l_1, t_1^s, t_1^e), (l_2, t_2^s, t_2^e), \ldots, (l_n, t_n^s, t_n^e)\}$, where l_i denotes the location of the i^{th} record, and t_i^s and t_i^e denote the starting and ending time of the stay activity. For each location l_i, we first map it to the corresponding city. For records in T, we iteratively group them into a segment if their locations fall within the same city. Otherwise, a new segment is created. This results into a city-level trajectory $T' = \{(l'_1, t'^s_1, t'^e_1), (l'_2, t'^s_2, t'^e_2), \ldots, (l'_m, t'^s_m, t'^e_m)\}$, where each tuple (l'_j, t'^s_j, t'^e_j) corresponds to a segment that is defined by one or more records in T. Here, l'_j denotes the j^{th} city visited by the individual, and t'^s_j and t'^e_j refer to the starting time of the first record and ending time of the last record in T that defines the visit to l'_j. Note that an individual could visit the same city more than once. Therefore, for some individuals, the value m will be larger than the number of unique cities visited (i.e., $|set(l'_1, l'_2, \ldots, l'_m)|$. Fig. 3 shows an example of how an individual's city-level trajectory is derived from the raw data.

3.4. Derive destination networks and their properties

In this study, two types of networks (directed and undirected) are formed to reflect the properties of destinations and their interactions. The directed networks are used to quantify the attractiveness of cities and also to distinguish inbound and outbound tourists. The undirected networks are generated based on the topology of directed networks and

[1] Cities usually have a population of at least 150,000, while counties generally have a population below 150,000. Cities with a population greater than 500,000 are further divided into districts.

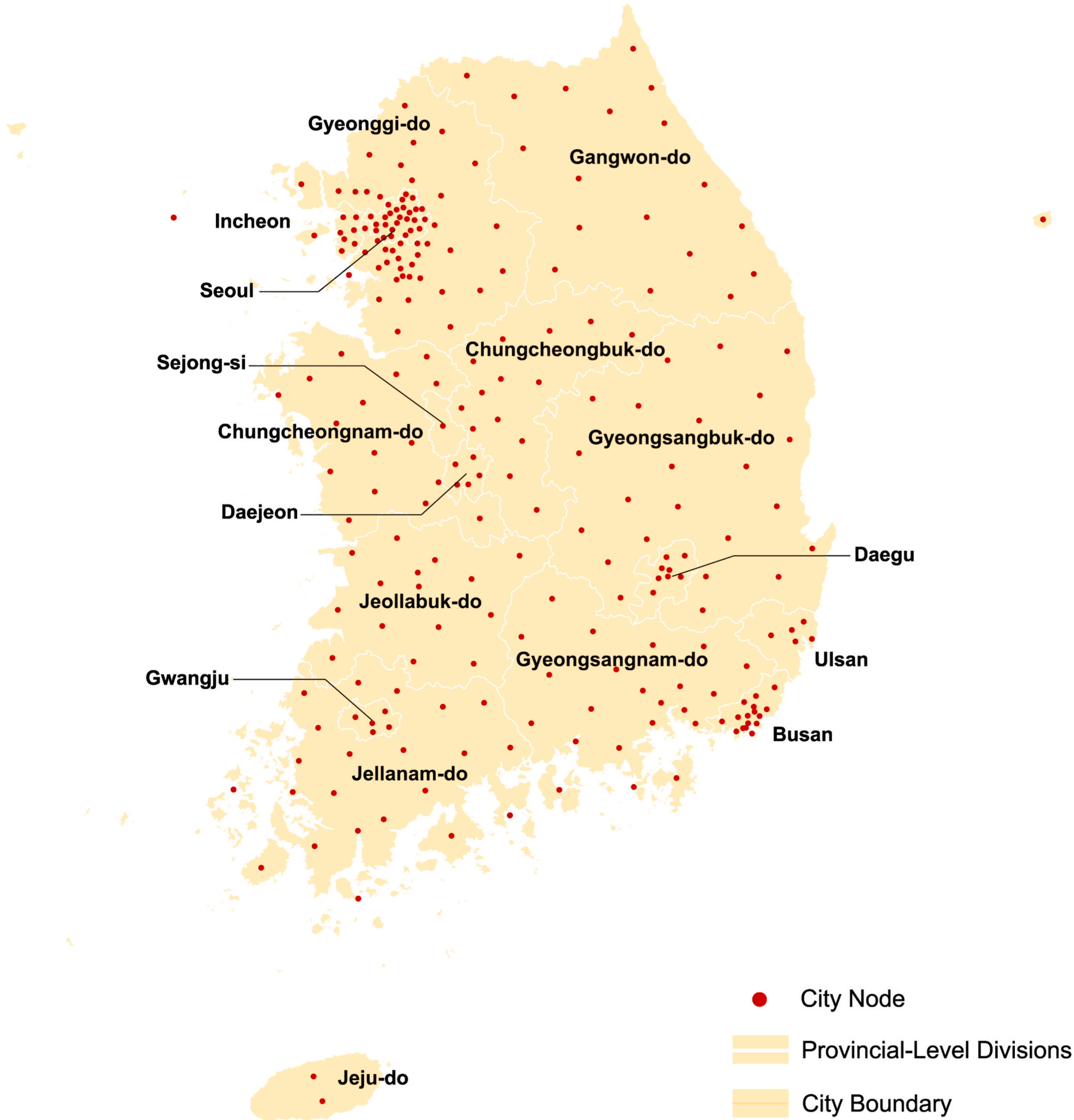


Fig. 2. Provincial-level divisions and municipal-level units of South Korea.

will be used to quantify the interactions among cities.

The information extracted from city-level trajectories can be aggregated to form a directed network $G = (V, E)$, where V denotes a set of vertices and E refers to a set of edges. Each vertex in the network $v_i \in V$ corresponds to a particular city, while the edge of the network $e_{i,j} \in E$ represents tourist flows from one city (v_i) to another (v_j). Note that $e_{i,j}$ and $e_{j,i}$ are two different edges in the network. For an edge $e_{i,j}$, its weight $w(e_{i,j})$ carries the information of the volume of tourist flows. The network G is formed in a simple way. For each tourist, we extract all the inter-city movements from the city-level trajectory. Suppose a tourist's trajectory T' traverses through three cities in a sequential order $v_i \rightarrow v_j \rightarrow v_k$, then two inter-city movements will be derived, and both edge weights $w(e_{i,j})$ and $w(e_{j,k})$ will be incremented by one unit. One issue worth mentioning is that since the dataset captures international travelers, for each trajectory T', the inbound tourist to the initial city — in the aforementioned example, v_i — will not be counted. To tackle this issue, we add a virtual node v_X to G, and when processing each T', we append this node v_X to the beginning of T'. In the previous example, $v_i \rightarrow v_j \rightarrow v_k$ becomes $v_X \rightarrow v_i \rightarrow v_j \rightarrow v_k$. In this way, we intuitively represent the fact that "the tourist came from somewhere (e.g., his or her home

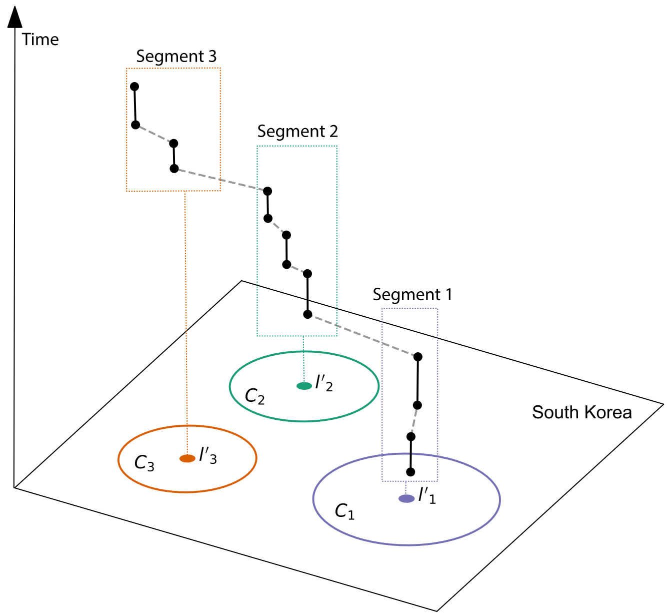


Fig. 3. Transforming individual phone trace (T) into city-level trajectory (T'). Black lines denote phone records in the raw data. C_1, C_2 and C_3 denote individual cities. l'_1, l'_2 and l'_3 refer to the centroids of cities (i.e., nodes in destination networks).

country) to visit city v_i in South Korea". Thus, the weight of a virtual edge $w(e_{X,i})$ will also be incremented by one unit. Adding this virtual node is essential because otherwise we would underestimate the vertices' *instrength*, which will be used in this study to quantify the destinations' attractiveness. To fully grow network G, we update the weight of edges until all trajectories are processed.

A few metrics, namely node *degree* and *strength*, are employed to quantify the properties of tourist destinations. Given a vertex $v_i \in V$, we use its indegree or outdegree to describe the number of cities that is directly linked to it. The *indegree*, denoted as $deg^-(v_i)$, measures the number of edges directed to v_i, while the *outdegree*, $deg^+(v_i)$, refers to the number of edges that v_i is directed to.

Strength is introduced as an extension of nodal degree in a weighted network (Barrat, Barthelemy, Pastor-Satorras, & Vespignani, 2004). Given a vertex $v_i \in V$ and a set of edges directed to v_i (denoted as E_i^-), the *instrength* is defined as:

$$s^-(v_i) = \sum_{e_{j,i} \in E_i^-} w(e_{j,i}) \tag{1}$$

$s^-(v_i)$ quantifies the strength of the node in relation to the weight of the edges in E_i^-. Intuitively, it represents the total number of inbound tourists. Similarly, the *outstrength* of v_i represents the total number of outbound tourists:

$$s^+(v_i) = \sum_{e_{i,j} \in E_i^+} w(e_{i,j}) \tag{2}$$

Here, E_i^+ denotes the set of edges that v_i is directed to.

The distributions of node strength reflect important characteristics of G. For example, given the instrength $s^-(v_i)$ of all nodes, we can measure the statistical probability $p(s^-)$, i.e., the fraction of vertices having an instrength s^-. The empirical distribution $p(s^-)$ can then be examined or fitted to describe the heterogeneity of destination attractiveness. Note that when building network G, we repeat the same procedure using: (1) the full dataset and (2) tourist trajectories of each nationality. This results into a set of G which quantify the properties of the overall network as well as the variations across nationalities.

We further derive an undirected network G' to understand the interactions among destinations without considering the directionality of tourist flows. The network G' can simply be derived from the topology of the directed network (G). Given two vertices v'_i and v'_j in G', an edge $e_{(i,j)}$ is added if its weight $w'(e_{i,j})$ — the total number of tourists traveling between the two cities — is greater than zero:

$$w'(e_{i,j}) = w(e_{i,j}) + w(e_{j,i}) \tag{3}$$

Here $w(e_{ij})$ and $w(e_{j,i})$ denote respectively the weight of $e_{i,j}$ and $e_{j,i}$ in network G. Fig. 4 illustrates the key concepts in the two types of networks using three hypothetical cities (nodes) as an example. In section 3.5, we introduce how the undirected network G' and a community detection algorithm will be used to quantify the interactions among cities.

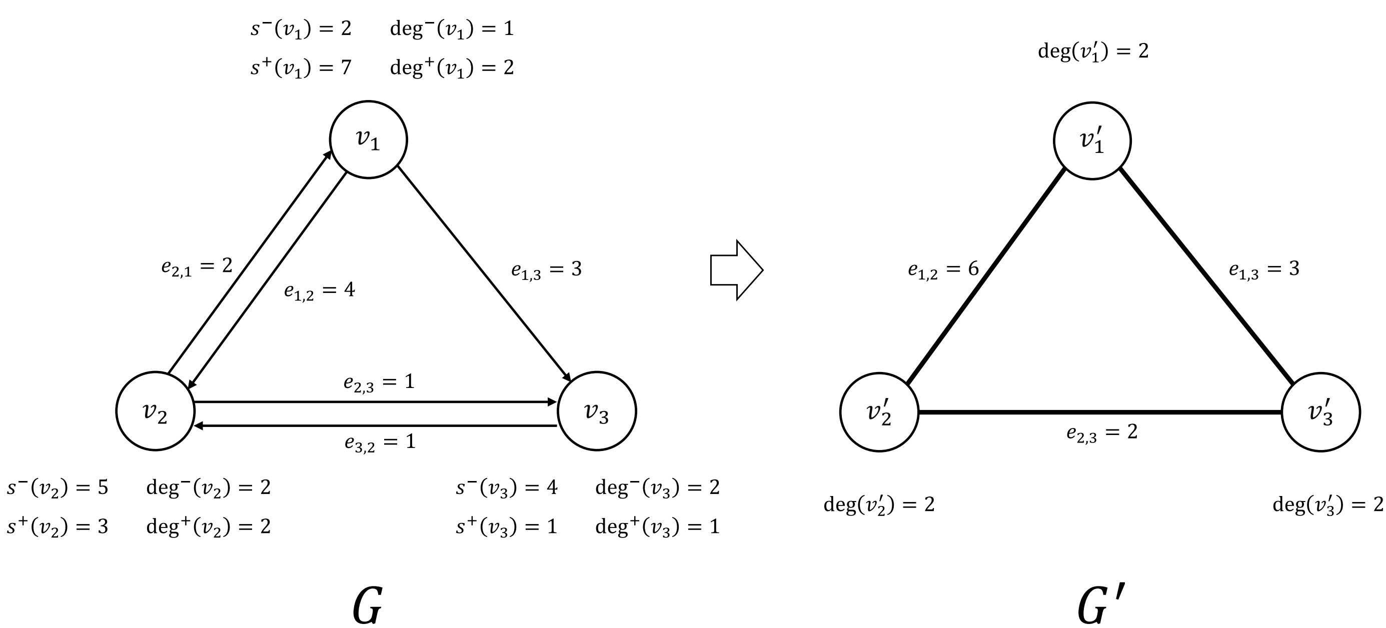


Fig. 4. An example of destination networks with three hypothetical cities (nodes). The undirected network (G') is derived from the topology of the directed network (G).

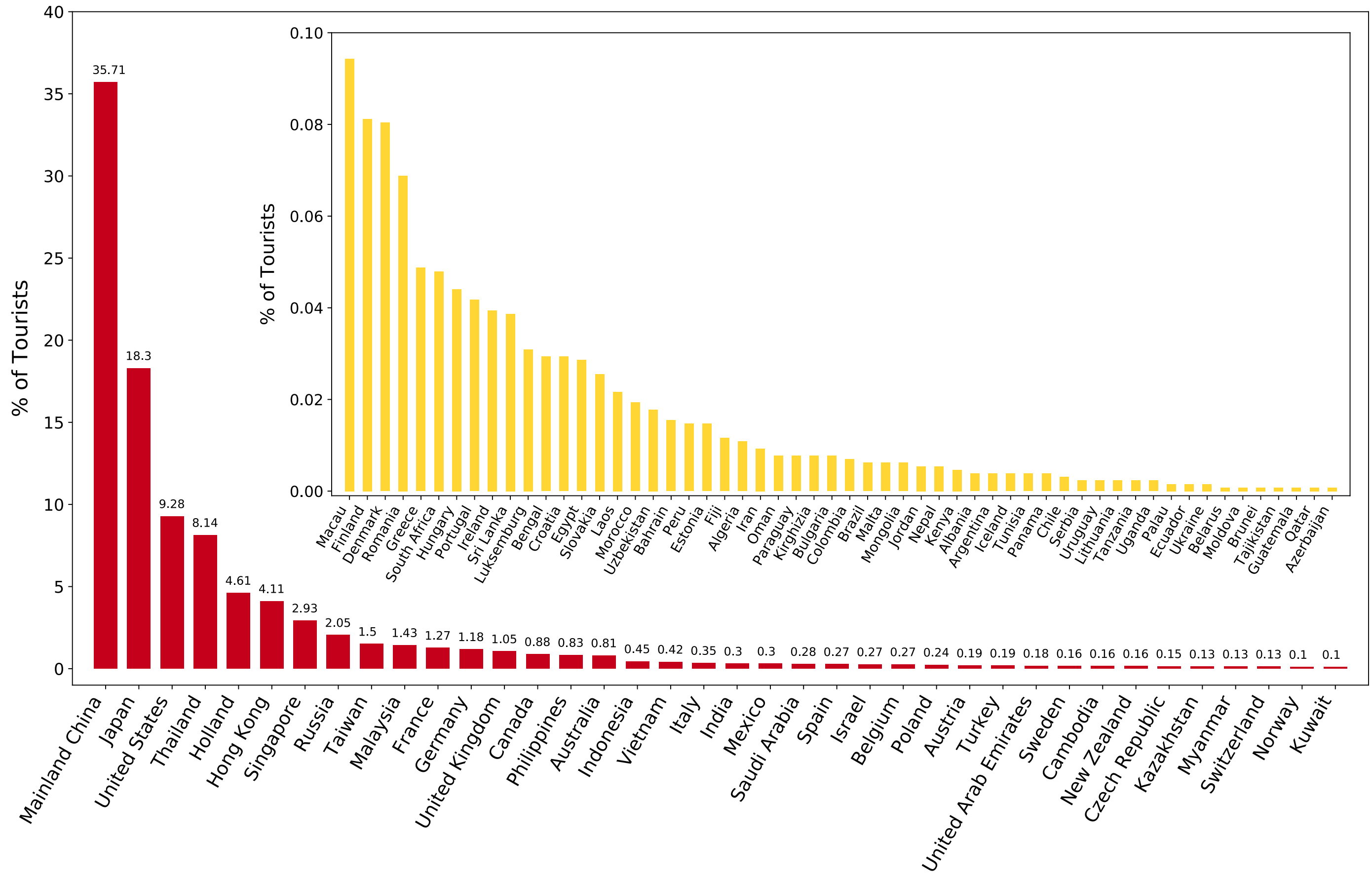


Fig. 5. Percentage of tourist arrival by country or region. In this dataset, information of tourists from Mainland China, Hong Kong, Macau and Taiwan is provided separately.

3.5. *Community detection over undirected networks*

One of the important properties of real-world networks is their community structure. Intuitively, it reflects partition of the network into groups of nodes that are closely tight together, i.e. there are many strong links between nodes of the same group, but nodes from different groups have much weaker connections. There are many ways to formalize this intuitive definition and then determine the community structure of a network (Fortunato, 2010). Here we employ one that is widely used in literature and is based on optimization of modularity function. Given an undirected graph *G*' modularity is defined as follows:

$$Q = \frac{1}{2W}\sum_{e_{i,j}}\left(w'(e_{i,j}) - \frac{s(v_i)\cdot s(v_j)}{2W}\right)\delta(c(v_i), c(v_j)) \quad (4)$$

where $W = \sum_{i<j} w'(e_{i,j})$ represents sum of weights of all links in the network, $s(v_i) = \sum_j w'(e_{i,j})$ is the strength of node *i*, $c(v_i)$ is a group number of node *i*, $\delta(c_i, c_j)$ is Kronecker delta function that equals to 1 when $c_i = c_j$ and 0 otherwise. This way, we add to the total score only terms that correspond to edges between nodes from the same group. For a given network and its partition modularity score characterizes how good this partition is, the higher the score, the better the partition. Then finding the best partition becomes equivalent to finding the one with the best modularity score. In last decade many methods to maximize modularity were proposed in literature. We picked an algorithm called Combo (Sobolevsky, Campari, Belyi, & Ratti, 2014), implementation of which is freely available. It was shown to produce high modularity scores and it runs fast for networks of tens of thousands of nodes.

4. Analysis results

4.1. *Descriptive statistics*

A large variation of tourist arrival is observed among countries. The top ten countries or regions account for more than 88.1 percent of all tourists, suggesting their importance to the tourism market of South Korea (Fig. 5). There are also many countries with a small arrival number. Countries including Indonesia and the ones that follow are observed with less than 1000 tourists during the 15-day period, and many of them account for less than 0.1 percent of the total visitors (inset of Fig. 5). By further exploring the time span, we find that tourists from Europe and North America tend to spend more time in South Korea than visitors from Asian countries. They also visited more cities in general than the Asian tourists (Fig. 6).

4.2. *Distribution of node strength*

In this section, we investigate the distribution of node strength in the destination networks. As mentioned in section 3.4, the instrength and outstrength of a node represents, respectively, the total number of inbound and outbound tourists of a city. In Fig. 7, we plot the cumulative probability distribution of node instrength for the overall network and the ones derived from each nationality. Note that in this analysis we focus on the top 16 countries or regions by tourist arrival — from Mainland China to Australia as shown in Fig. 5. Since there is a large variation of tourist arrival by nationality, to make results comparable, for each network *G*, we normalize the node instrength by the total instrength of the network. The meaning of normalized instrength is intuitive. Given a node *i*, its normalized instrength $s^-_{nor}(v_i)$ denotes the

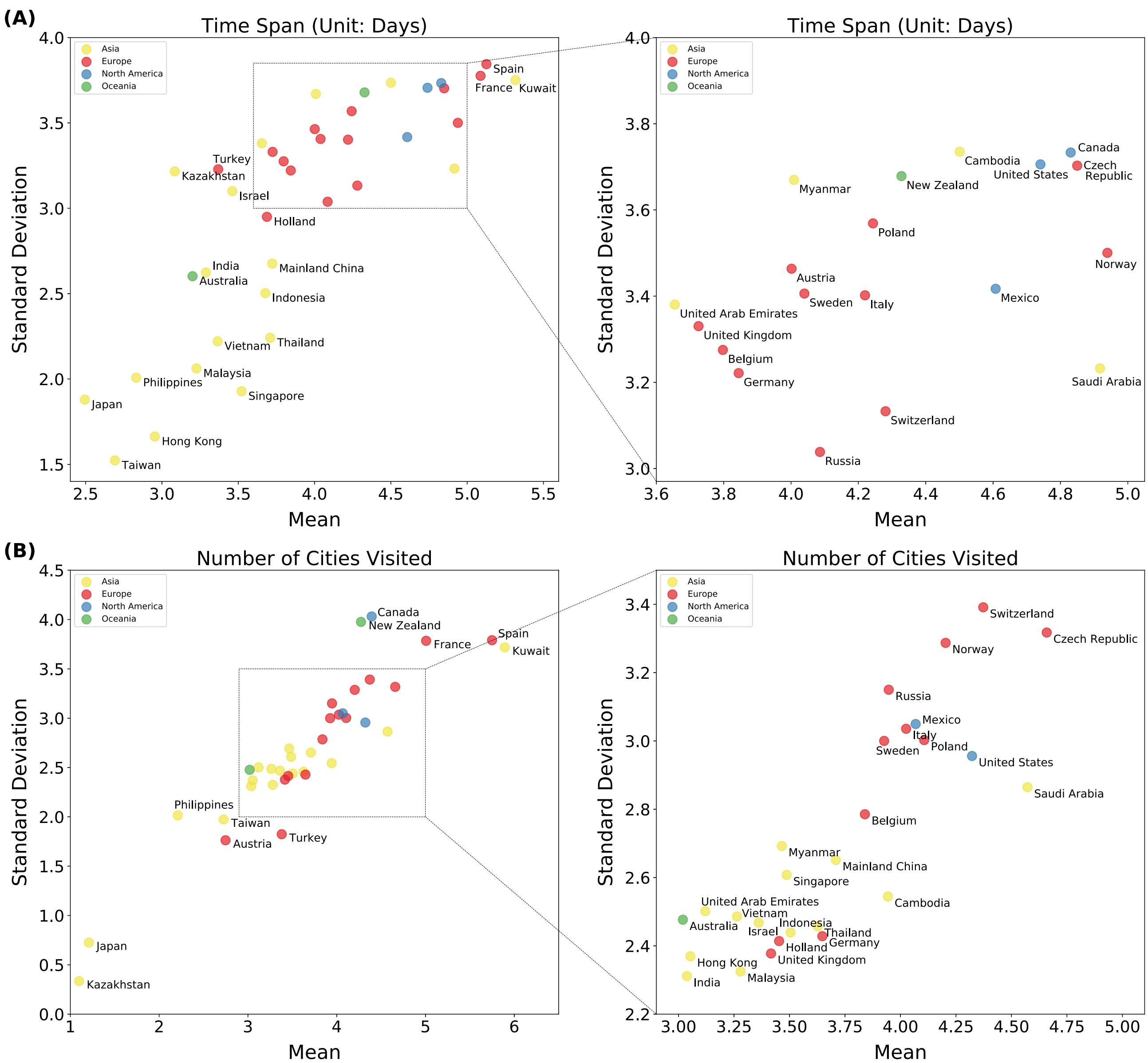


Fig. 6. Mean and standard deviation of: (A) time span and (B) number of cities visited by tourists from each foreign country or region. Here only countries or regions with over 1% of total tourist arrival are analyzed.

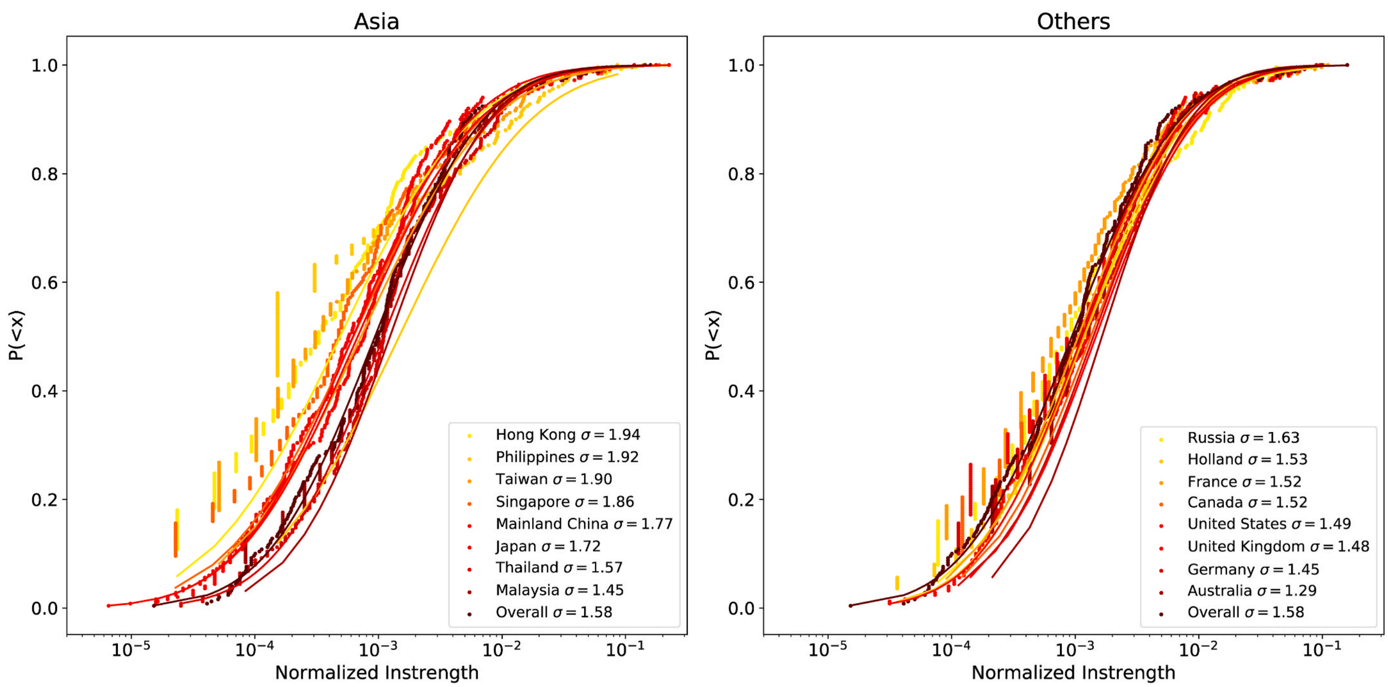


Fig. 7. Cumulative probability distribution of normalized node instrength. Lines show fitted log-normal distribution.

proportion of tourist visits that goes to city *i*.

According to the result, the node instrength in most of the networks can be well approximated by a log-normal distribution $\ln(s^{-}_{nor})\sim \mathcal{N}(\mu, \sigma^2)$. This indicates a heterogeneity of destinations' attractiveness, i.e., few cities attract a large number of visits while many cities attract few tourists. This applies not only to the overall network, but also to the networks derived from different nationalities. By further exploring the value of σ from the fitted distributions — the standard deviation of the natural logarithm of node instrength — we find that countries or regions in Asia tend to have higher σ (Fig. 7A) than other countries (Fig. 7B). The western countries and Australia have a σ lower than that of the overall network ($\sigma = 1.59$). This dichotomy suggests that tourist visits from areas in Asia, as compared to the visits from western countries, are more concentrated in a few destinations in South Korea. Note that we also observe similar distribution patterns for outstrength — given that a very high correlation is observed between node instrength and outstrength in all the networks. Readers could refer to Figure A1 and Table B1 in Appendix for more details.

4.3. Similarity between countries and regions by tourist visitation patterns

The comparison of node instrength distributions suggests that for travelers from most of the countries, the top few cities in South Korea tend to attract a large fraction of their visits. An intriguing question is whether these top performers refer to the same set of destinations, i.e., whether tourists from different foreign countries prefer to visit the same set of places in South Korea. To answer this question, we perform an agglomerative hierarchical clustering algorithm over different nationalities, using the normalized instrength of nodes in the corresponding network as the input feature. Given tourists from a given country or region, the normalized node instrength of cities in South Korea can be represented as a feature vector:

$$X = \left[s^{-}_{nor}(v_1), s^{-}_{nor}(v_2), \ldots, s^{-}_{nor}(v_{250})\right] \tag{5}$$

Note that:

$$\sum_{i=1}^{250} s^{-}_{nor}(v_i) = 1 \tag{6}$$

The hierarchical clustering algorithm works as follows. Initially, each cluster includes one feature vector *X* (i.e., one country or region). At each step, two clusters with the smallest distance are merged into a new cluster. In this analysis, the distance between two feature vectors *X* and *X*' is defined based on the Euclidean distance measure:

$$I = \sqrt{\sum_{i=1}^{250} \left(s^{-}_{nor}(v_i) - s^{-}_{nor}(v_i')\right)^2} \tag{7}$$

The distance between any two clusters is calculated based on the Ward's method (Ward, 1963). Note that we choose agglomerative hierarchical clustering over other alternatives (e.g., k-means) for two reasons. First, it embraces a "bottom-up" approach and the algorithm does not require the total number of clusters to be predefined. Second, it allows for examining how individual elements are grouped at different steps of the algorithm (through a dendrogram), which provides a comprehensive picture of the similarities and differences across nationalities.

Fig. 8 demonstrates the clustering result. Each column in the figure represents one feature vector that summarizes the destination attractiveness (i.e., normalized node instrength) to a given country or region. Darker color represents higher destination attractiveness. For ease of comparison, we also include the feature vector of the overall network into the clustering process. The destination cities are organized in descending order of their attractiveness in the overall network (the first column in Fig. 8).

As can be seen from the dendrogram, the clustering algorithm divides the 16 countries and regions into three distinct groups. Strikingly, areas in Asia are completely separated from Western countries and Australia, with Philippines as an "outlier" that is quite different from the rest of others. For the Asian cluster, the very top destination — *Jung-gu* in the Seoul metropolitan area — attracts a large fraction of tourist visits. Although *Jung-gu* is also the top attraction for Western countries, the city accounts for a much larger fraction of visits in the Asian cluster. Another key difference between the Asian and Western clusters is that after excluding *Jung-gu*, the most popular attractions for Asian travelers are not always the top choices for western countries. It is also worth noting that a large concentration of tourist visits is observed in the top one or

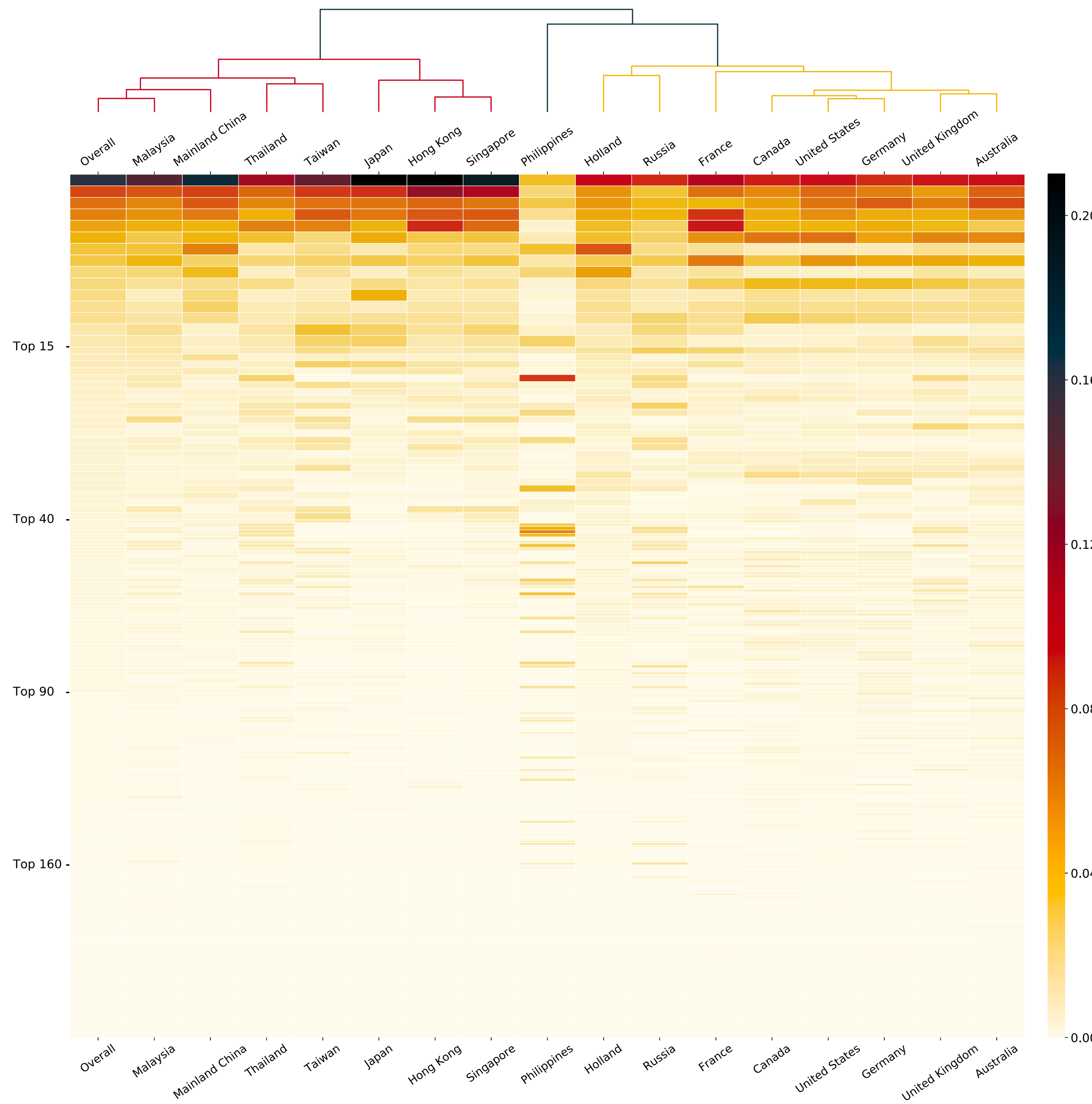


Fig. 8. Clustering result of tourist visitation patterns by country or region. Each column represents a feature vector that summarizes the attractiveness (i.e., normalized node instrength) of 250 destination cities to tourists from a given country or region. Darker color represents higher destination attractiveness. Destination cities are organized in descending order of their attractiveness to the overall tourist population (first column). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

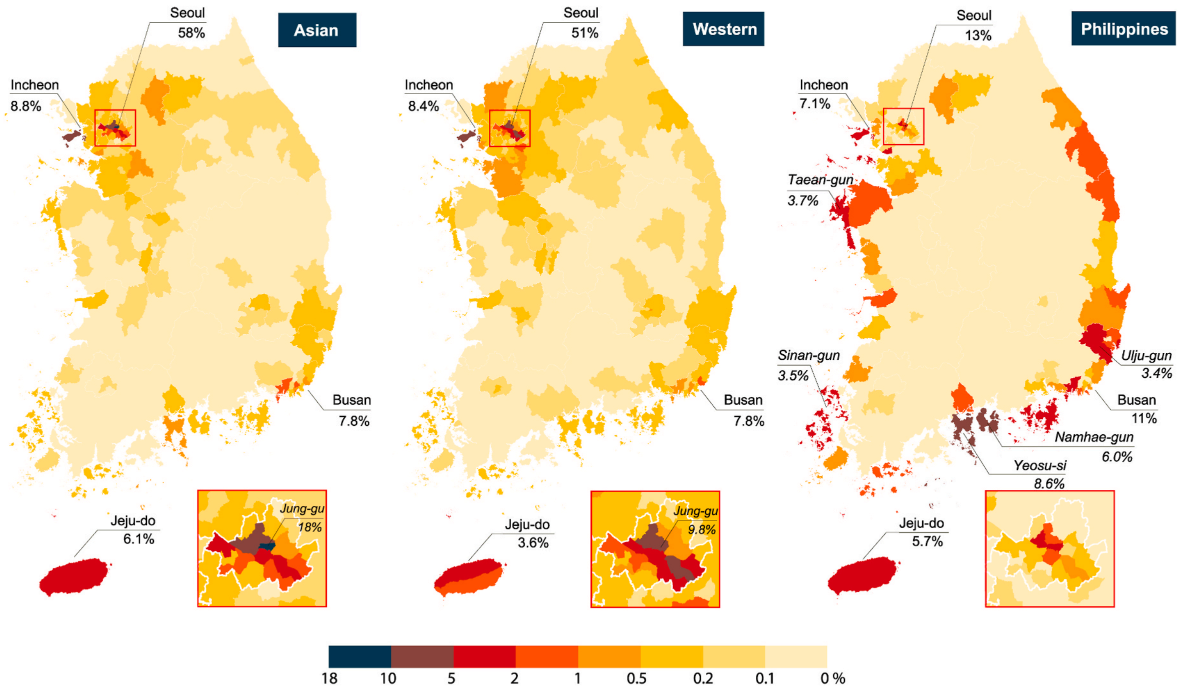


Fig. 9. Destination attractiveness to travelers from the Asian cluster, the Western cluster and Philippines.

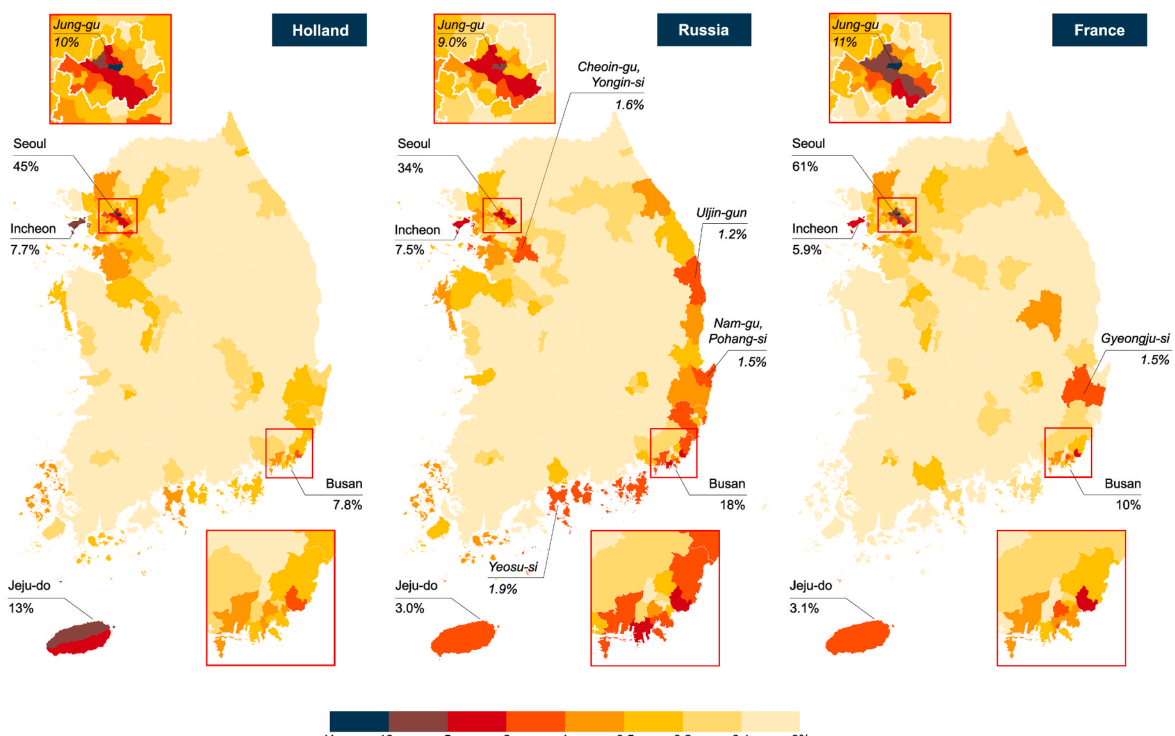


Fig. 10. Destination attractiveness to travelers from Holland, Russia and France.

two destinations in the Asian cluster, but not in the Western cluster. This further explains the finding from Fig. 7 that Asian countries and regions are observed with higher values of σ — higher heterogeneity of attractiveness among the destinations. Tourists from Philippines have unique travel patterns. A lot of their visits fall within destinations that are not popular to travelers from other countries.

To better understand the key difference between the Asian cluster, Philippines and the Western cluster, we aggregate countries and regions in each group, and recalculate their feature vectors. Fig. 9 shows the spatial patterns of destination attractiveness to the three groups. In general, Asian and Western countries show similar spatial patterns, and a large fraction of their visits fall within the Seoul metropolitan area (Asian: 58%; Western: 51%). The key difference, as discussed previously, is that travelers from Asia paid more visits to Jung-gu (18%, compared to 9.8% for the Western). Both groups distribute 7.8% of their visits to Busan, while the Asian group paid more visits to the Jeju island.

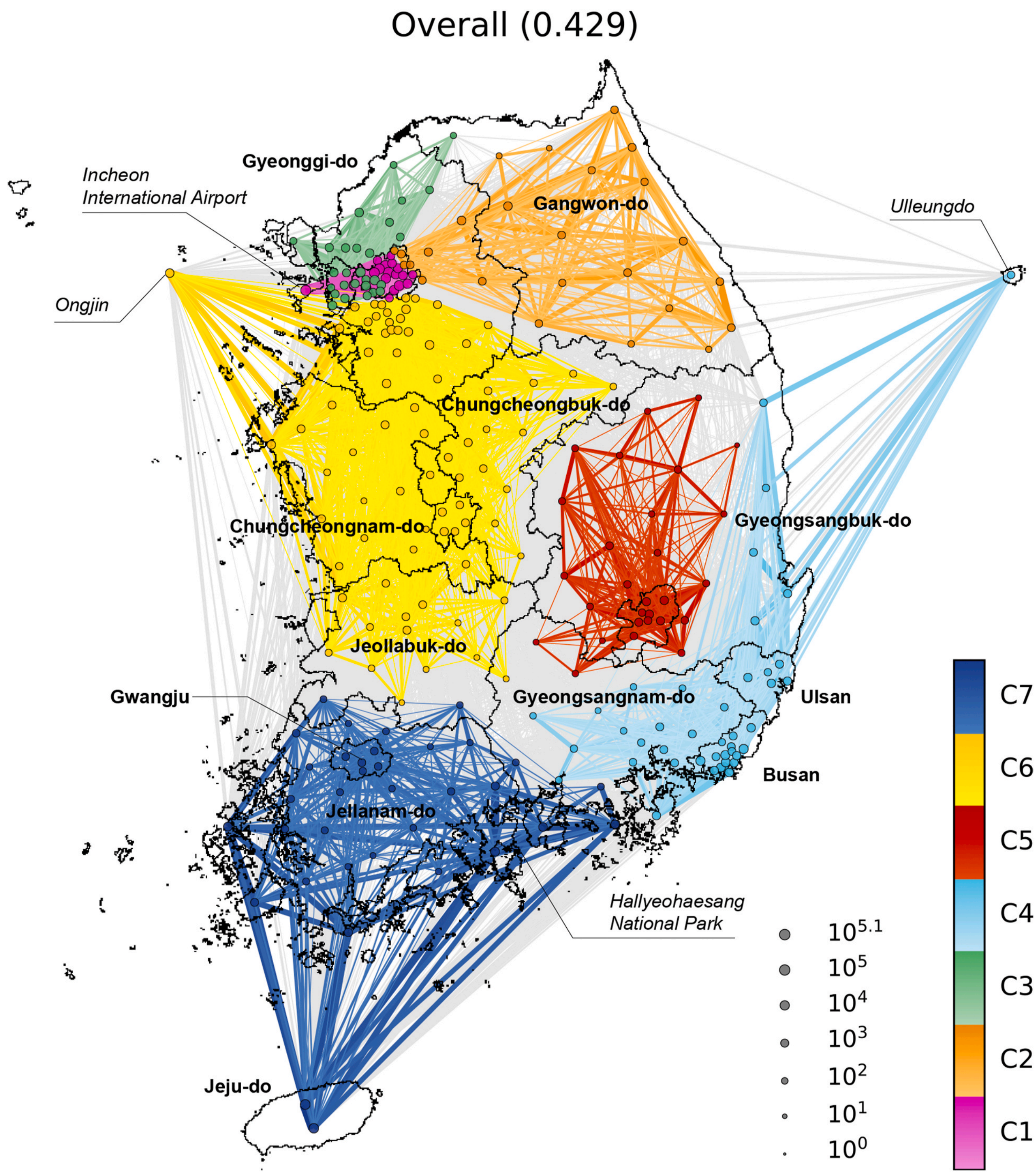


Fig. 11. Community detection for the overall network. Size of the nodes represents destination attractiveness and color denotes the communities they belong to. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Modularity and number of communities derived from the destination networks.

Country or Region	Modularity	No. Of Clusters
Philippines	0.648	9
Russia	0.591	10
Thailand	0.564	8
United Kingdom	0.511	11
Taiwan	0.499	9
Malaysia	0.496	12
Holland	0.496	7
Germany	0.468	11
Australia	0.468	9
Canada	0.454	10
United States	0.423	7
Singapore	0.408	7
Mainland China	0.407	6
Japan	0.380	7
France	0.367	8
Hong Kong	0.326	7

For Philippines, only 13% of their visits went to the Seoul metropolitan area. Interestingly, the popular attractions for them mainly refer to the islands and coastal areas, such as Busan (11.0%), Yeosu-si (8.6%), Namhae-gun (6.0%), and Taean-gin (3.7%).

We also want to discuss a few countries in the western group. As shown in Fig. 8, although Holland, Russia and France are grouped with other western countries, the height of their leaves in the dendrogram suggests that they are relatively unique in this cluster. Fig. 10 shows the destination attractiveness to these three countries. Compared to the overall visitation patterns of the western group, the Hollander distribute more visits to the Jeju island (13%) while a lot of visits of the French fall within the Seoul metropolitan area (61%). Some areas along the east coast are popular attractions for the Russian, but not for visitors from other western countries.

Note that the results on instrength distribution (Fig. 7) and hierarchical clustering (Fig. 8) are generated from the full dataset without considering visitors' duration of stay in each city. In other words, even a short stay of a traveler in a city would be counted as one visit. Although it is challenging to identify tourist stays due to reasons such as flight connections or transfer of ground transportation, in this study, we evaluate the sensitivity of our analysis by filtering tourist stays in cities below a certain threshold. In particular, we tested two different thresholds — 1 h and 3 h. For each parameter setting, only tourist stays in a city that are beyond the threshold are counted when calculating the destination attractiveness. This would allow us to mitigate the impact of transfer passengers on the analysis results. We report the instrength distribution and clustering result in Appendix. In general, we find that our findings on the heterogeneity of destination attractiveness still hold even when brief stopovers are filtered (Figure C2 and Figure C3). The clustering result based on the threshold of 1 h shown in Figure C4 is also highly similar to that of the full dataset (Fig. 8). Some changes are observed in the clustering result when the threshold changes to 3 h (Figure C5). Specifically, countries that show more unique characteristics in the western cluster in the original result (Fig. 8) are now grouped with other Asian countries. However, other western and eastern countries are still separated relatively well. On the one hand, it suggests that our findings are relatively robust even when we filter tourists who stayed in a city for a very short period of time. On the other hand, it indicates that considering the length of stay may provide additional insights into destination attractiveness. For instance, some cities might be more attractive to same-day visitors while other cities might be more appealing to over-night stayers (Rodriguez, Martinez-Roget, & Gonzalez-Murias, 2018). Incorporating duration of stay as an explicit measure of destination attractiveness is a future direction of this research.

4.4. Community structure of destination networks

In this section, we report the results of community detection. Fig. 11 demonstrates a visualization of communities extracted from the overall destination network. The color of a node represents the community it belongs to while the size describes destination attractiveness (log scale). The links represent the interactions between destinations and their widths are proportional to the logarithm (log_{10}) of the volume of tourist flow. If links connect nodes within the same community, they will be rendered using the same color of the nodes, while links that cross communities will be rendered in grey.

The community detection algorithm partitions the network into seven communities, yielding a modularity score of 0.429. The modularity score is a clear evidence of community structure in the network — tourist movements are not randomly distributed between destinations. In general, most of the communities consist of areas that are geographically cohesive, indicating that tourist movements are more likely to occur between destinations in closer proximity.

The first community (C1 - pink) mainly consists of destinations in the Seoul metropolitan area — a top tourism attraction in South Korea. Interestingly, we find that these destinations are grouped together with the Incheon International Airport. Given Seoul as the most visited region in South Korea and Incheon Airport as the major point of entry ("2018 international visit, 2018), this strong connection is likely to be a joint outcome of the two-way interactions between these two areas. First, it could imply that many tourists, once landing in South Korea, tended to visit areas in Seoul rather than exploring places that are nearby. Second, those areas in Seoul can also be the final destinations for many travelers before they departed from the airport.

Community two (C2 - orange) captures the whole area of Gangwon-do along with a few destinations in Seoul and Gyeonggi-do. Most of the destinations in Gangwon-do are well connected, indicating diverse travel patterns of tourists once visiting this region. The northern part of Gyeonggi-do forms community three (C3 - turquoise), while southern Gyeonggi-do along with the western part of South Korea are grouped as community six (C6 - yellow). Community four (C4 - light blue) mainly consists of attractions along the east coast. Within this community, strong connections are observed between the Ulleungdo island and eastern Gyeongsangbuk-do. The remaining part of Gyeongsangbuk-do forms the community five (C5 - red). Community seven (C7 - dark blue) consists of areas in Gwangju, Jellanam-do and Jeju island. Tourist flows between the destinations in this region are generally high, knitting them into a closely connected tourism community.

We then repeat the community detection analysis over networks derived from different nationalities. As shown in Table 2, all the networks achieve modularity scores between 0.3 and 0.7, which again, indicates that travelers' movements do not occur randomly across cities. For countries such as Philippines, Russia, Thailand, UK and Malaysia, the derived networks have a very high modularity score. Meanwhile, the number of communities in these networks are relatively high. That means for travelers from these countries, their movements tended to be concentrated in many dedicated areas.

We visualize the community structures of these networks in Fig. 12. There are notable variations among the networks, suggesting that travelers from different countries have different travel patterns. For networks of Mainland China, Japan and Malaysia, the Incheon International Airport is grouped together with the core part of Seoul, which is consistent with the finding from the overall network. However, for some other networks (e.g., United States, Thailand, Taiwan, Germany, UK, Canada, Philippines and Australia), the Incheon Airport is also grouped together with adjacent destinations. That means some travelers from these countries or regions, once landing at the Airport, would explore cities that are immediately nearby.

For countries such as Singapore, Russia, Malaysia, France and Germany, we find that Gangseo-gu (Seoul) and Jeju-do — two areas that are far away from each other — are grouped into the same community. They

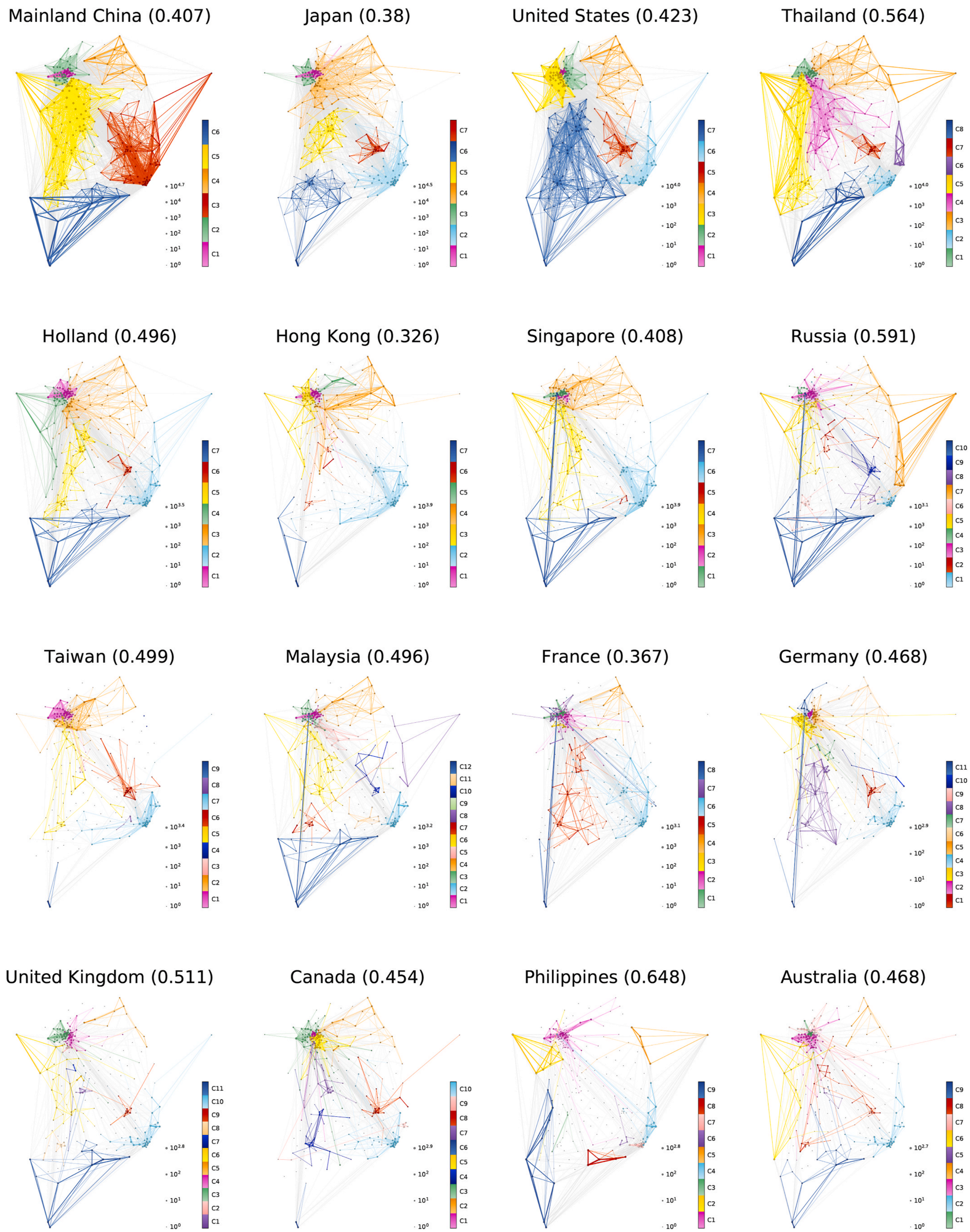


Fig. 12. Community detection for networks derived from different nationalities. Size of the nodes represents destination attractiveness and color denotes the communities they belong to. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

also refer to the areas where Gimpo Airport and Jeju International Airport are located, respectively. The result indicates frequent travel patterns between Gangseo-gu (Seoul) and Jeju-do for selected nationalities, revealing a hidden linkage between these two areas.

Apart from these key differences among nationalities, we find that in many of the networks, two regions — namely the Gangwon-do area and the lower "triangular area" that links Jeju island and coastal areas in Jeollanam-do and Gyeongsangnam-do — are relatively well defined. In other words, the destination cities in these two areas are strongly connected, and this finding holds across many nationalities.

5. Discussion and conclusion

This study investigates the network characteristics of tourism destinations through mobility traces of international travelers to South Korea. It demonstrates the great potential of linking mobility data with network science approaches for tourism research. By extracting tourist movements from a large-scale mobile positioning dataset, we reconstruct a series of destination networks, from which the collective dynamics of tourists and key differences across nationalities are revealed. This research implies, as a result, the network science approach enables tourism researchers to identify better insights of tourist flow and to explicate the complex structure of interplay between tourism destinations (Baggio et al., 2010).

The findings of this research provide important academic implications. Previous tourism literature discusses tourism systems as a set of inter-connected parts including market, travel, destination and marketing (Mill & Morrison, 2002). This means that tourism is too complex to be assessed and understood by a deterministic model. In spite of the complex nature of tourism, most of extant studies have focused narrowly on selected destination elements or a linear relationship between markets and destinations (McKercher, 1999). This current study applied a notion of network science in understanding the structure of tourism destinations, which reveals a complex, non-linear and dynamic systems of tourism. The destination attractiveness, measured as the number of inbound tourists to a city, generally follows a log-normal distribution. That is, the sophisticated level of data analytic emerges recognizable patterns of travel flow and enumerates the "real world" in tourism destination as a whole. The result indicates a notable heterogeneity of destination popularity, and this finding continues to hold when tourist visits are separated by nationality. One previous study using social media data has found that country attractiveness to international travelers follows a log-normal distribution (Belyi et al., 2017). Our analysis suggests that strong preferences of tourists also exist when they pick destinations within a country.

Despite the universality of mobility preference, we find that tourists from different countries tended to visit different places in South Korea. The clustering analysis divides countries and regions into three distinct groups, with Asian countries completely separated from the western ones. Besides this clear evidence of continental difference, the countries within each group (e.g., Asian cluster, Western cluster) show similar visitation patterns. A number of tourism studies have attempted to identify different perceptions and travel decision-making process between international travelers from different originality (i.e., behavioral intentions) (Jönsson & Devonish, 2008). Importantly, however, this current study demonstrated heterogeneous travel mobility among travelers originating from different countries. The implications are twofold. On the one hand, tourism planning and services in a destination country (e.g., destination branding, location recommendation) should be tailored to accommodate the varying needs of international travelers. On the other hand, cultural, religious and institutional contexts of originating countries — which are to some extent related to geographic proximity — could play a role in shaping tourist mobility patterns.

The community detection algorithm partitions the overall network into seven communities. Each community covers a set of destinations that are densely connected by tourist flows. This finding contributes to literature on destination collaboration. A study by Fyall, Garrod and Wang (Fyall et al., 2012) suggested that destination collaboration is shaped by chaotic, non-linear and non-deterministic processes. This paper empirically demonstrates the proposition of complex destination system by presenting the dynamic formations of destination community through the lens of travel flow. The spatial layout of these communities (Fig. 11) could serve as a useful guide for regional tourism planning and suggest actions of collaborative management in destinations. For example, local destinations with clustered cities should develop collaborative destination marketing and fulfill knowledge sharing between the clustered destinations, which potentially improves destination competitiveness. Furthermore, the transportation efficiency is of the utmost importance to facilitate travel flow within and across destination community (Lew & McKercher, 2006). This study suggests, for instance, inter-city transportation services in each community can be strengthened to meet intensive travel needs among the destinations. Some hidden linkages (e.g., Incheon International Airport – Seoul metropolitan area; Gangseo-gu of Seoul – Jeju island), as revealed by the algorithm, suggest the need for maintaining or improving transportation connectivity between specific areas.

The visualization of community structure, which demonstrates both the attractiveness and spatial interactions of destinations, can be useful in revealing their interplay in the tourism context. According to the result, communities that correspond to the Gangwon-do area and the lower "triangular area" are relatively well defined. Destinations in these two communities are densely connected, suggesting a symbiotic relationship among them. The methods used in this study can thus be applied over longitudinal datasets to better understand and advise long-term cooperation among destinations (Baggio et al., 2010).

With regard to methodological contributions, this study suggests innovative tourism big data analytics for examining mobile sensor data, such as community detection algorithm and clustering method. Advancement of mobile technology facilitates tourism researchers in obtaining real-time movement information of travelers. This research suggests a set of mobile big data approaches to identifying the network structure of destinations and to classifying them based upon tourist movement patterns.

We want to point out a few limitations of this research. First, the destination networks are extracted based on movement patterns of international travelers who subscribed to the cellular operator's service when they visited South Korea. Despite the ubiquitous use of mobile phones worldwide, our dataset did not capture tourists who relied on other types of services (e.g., used WIFI only) during their journeys. Second, since the dataset covers a 15-day period, our nationality segmentation analysis only focuses on the top 16 countries or regions with an adequate number of travelers. Extending our methods to long-term observation datasets would provide a more comprehensive picture of the behavioral difference among various nationalities. It will also allow us to examine the temporal evolution of destination networks, for example, whether there exist seasonal variations of tourist mobility that alter the network characteristics. These are possible directions for future research. Nevertheless, this study contributes to the emerging field of *tourism big data* (Li et al., 2018) by integrating network science approaches with large-scale tourist mobility data. The framework can be applied in other countries or regions to provide data-driven insights for destination management.

Impact statement

Along with the advancement of big data in general and tourism sector in particular, this study attempts to integrate mobility big data and network science approaches to understand tourism destination system. This research generates significant impact to tourism literature on tourism destination networks and destination collaboration. The findings of this paper discover non-linear and dynamic structures of tourism network as well as heterogeneity of destinations' attractiveness through the lens of travel inflow, suggesting complex tourism systems. The insights derived from community detection method suggests important knowledge in the development of strategic destination collaboration and cooperation as a way to enhance destination competitiveness. The findings of this study suggests, for instance, inter-city transportation services in each community can be strengthened to meet intensive travel needs among the destinations. Furthermore, considering tourism literature largely focusing on travel decision-making process (e.g., perceived intention), this research identifies different travel flow between travelers from different origin countries. The implications are twofold. On the one hand, tourism planning and services in a destination country (e.g., destination branding, location recommendation) should be tailored to accommodate the varying needs of international travelers. On the other hand, cultural, religious and institutional contexts of originating countries which are to some extent related to geographic proximity could play a role in shaping tourist mobility patterns.

Declaration of competing interest

None.

CRediT authorship contribution statement

Yang Xu: Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **Jingyan Li:** Formal analysis, Writing - original draft. **Alexander Belyi:** Formal analysis, Writing - original draft. **Sangwon Park:** Conceptualization, Data curation, Writing - original draft.

Acknowledgement

The authors would like to express our great appreciation to Korea Tourism Organization (particularly to Team of Tourism Big Data) for their full support in initiating the project and collecting the big data set as well as sharing their insights on this project. This paper and research project (Project Account Code: 5-ZJLW) are jointly funded by Research Grant of Hospitality and Tourism Research Centre (HTRC Grant) of the School of Hotel and Tourism Management, The Hong Kong Polytechnic University, the Hong Kong Polytechnic University PTeC Grant (Grant no. CA19-00204), and the Hong Kong Polytechnic University Start-Up Grant (Grant no. 1-BE0J).

Appendices.

A Distribution of node outstrength in the destination networks

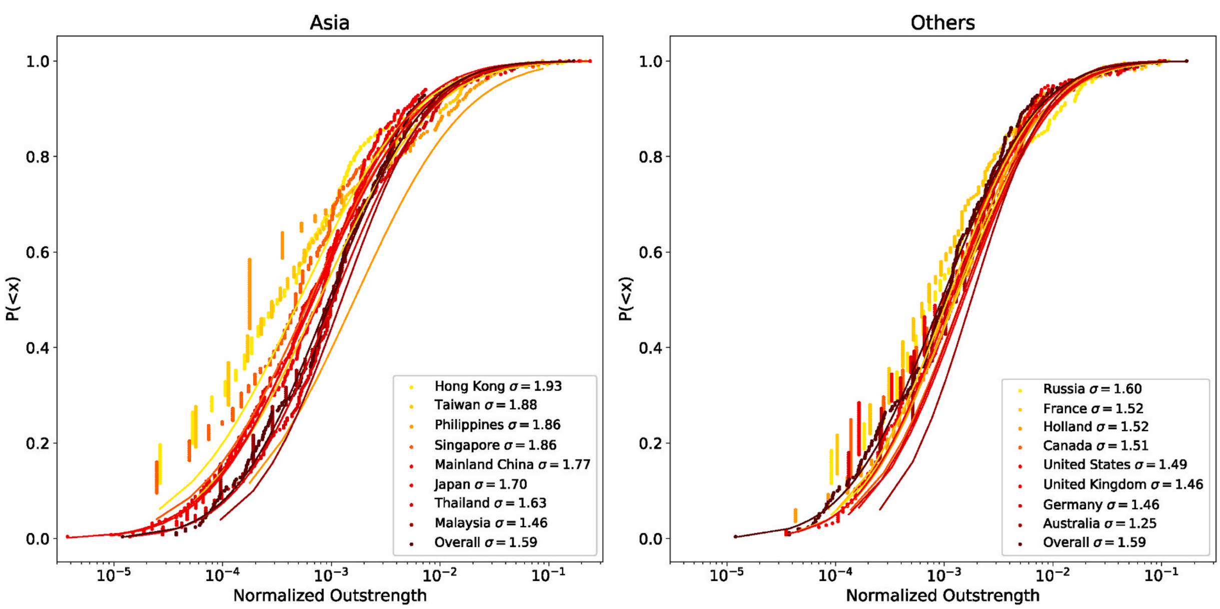


Fig. A.1. Cumulative probability distribution of normalized node outstrength. Lines show fitted log-normal distribution.

B Correlation between node instrength and outstrength in the destination networks

Table B.1
Correlation between node instrength and outstrength

Network	Pearson's r	Spearman's r
Overall	0.984	0.999
Mainland China	0.983	0.999
Japan	0.981	0.999
Thailand	0.985	0.997
Malaysia	0.986	0.991
Hong Kong	0.988	0.997

(*continued on next page*)

Table B.1 (*continued*)

Network	Pearson's r	Spearman's r
Taiwan	0.990	0.996
Singapore	0.989	0.998
Philippines	0.993	0.993
Russia	0.991	0.993
France	0.993	0.993
Holland	0.985	0.998
Canada	0.987	0.994
United States	0.983	0.999
United Kingdom	0.968	0.996
Germany	0.977	0.994
Australia	0.968	0.992

C Instrength distribution and clustering results by filtering brief stopovers of tourists in cities

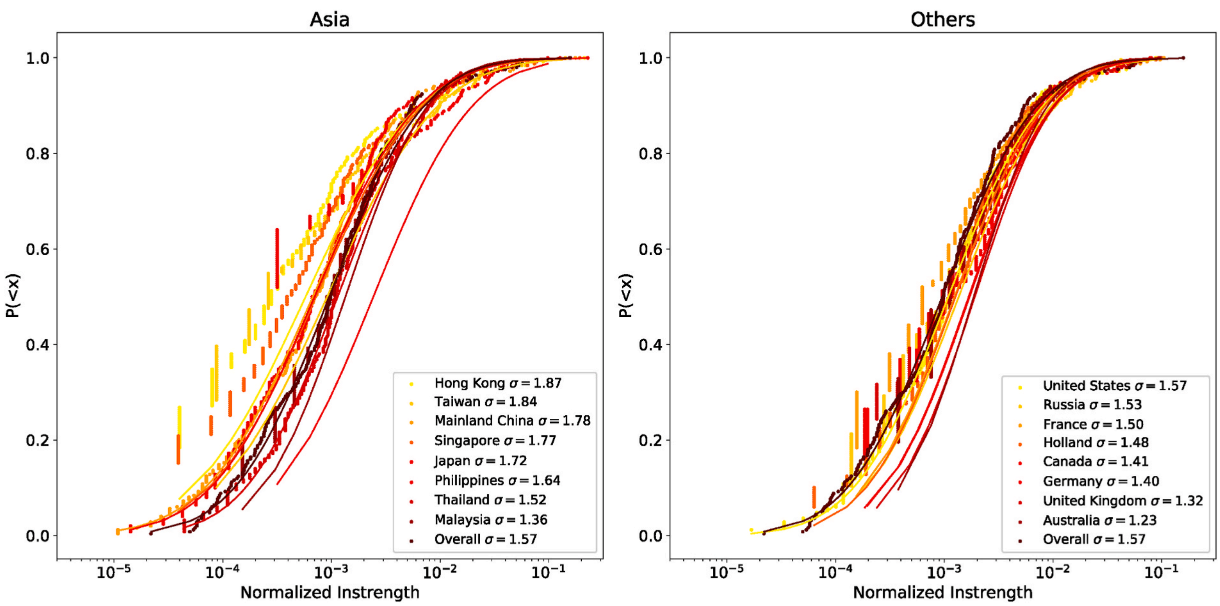


Fig. C.2. Cumulative probability distribution of normalized node instrength using threshold of 1 h.

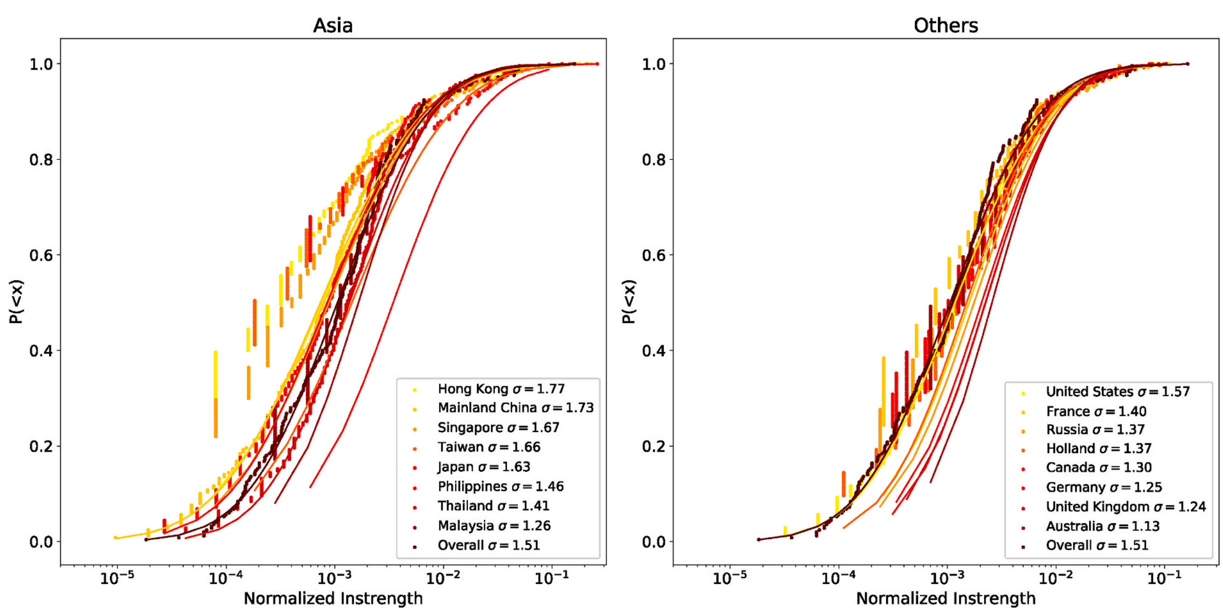


Fig. C.3. Cumulative probability distribution of normalized node instrength using threshold of 3 h.

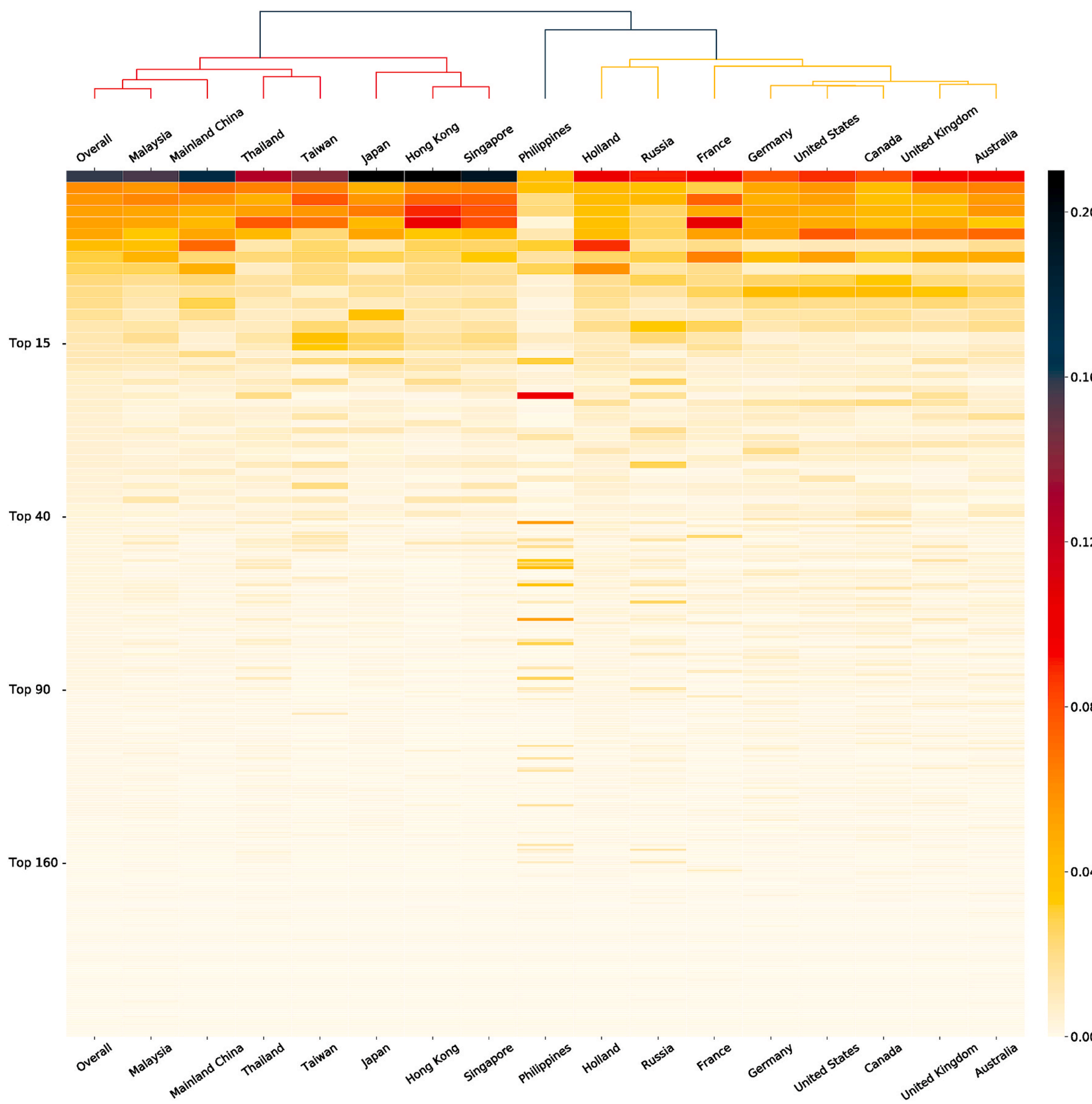


Fig. C.4. Clustering result of tourist visitation patterns by country or region based on threshold of 1 h.

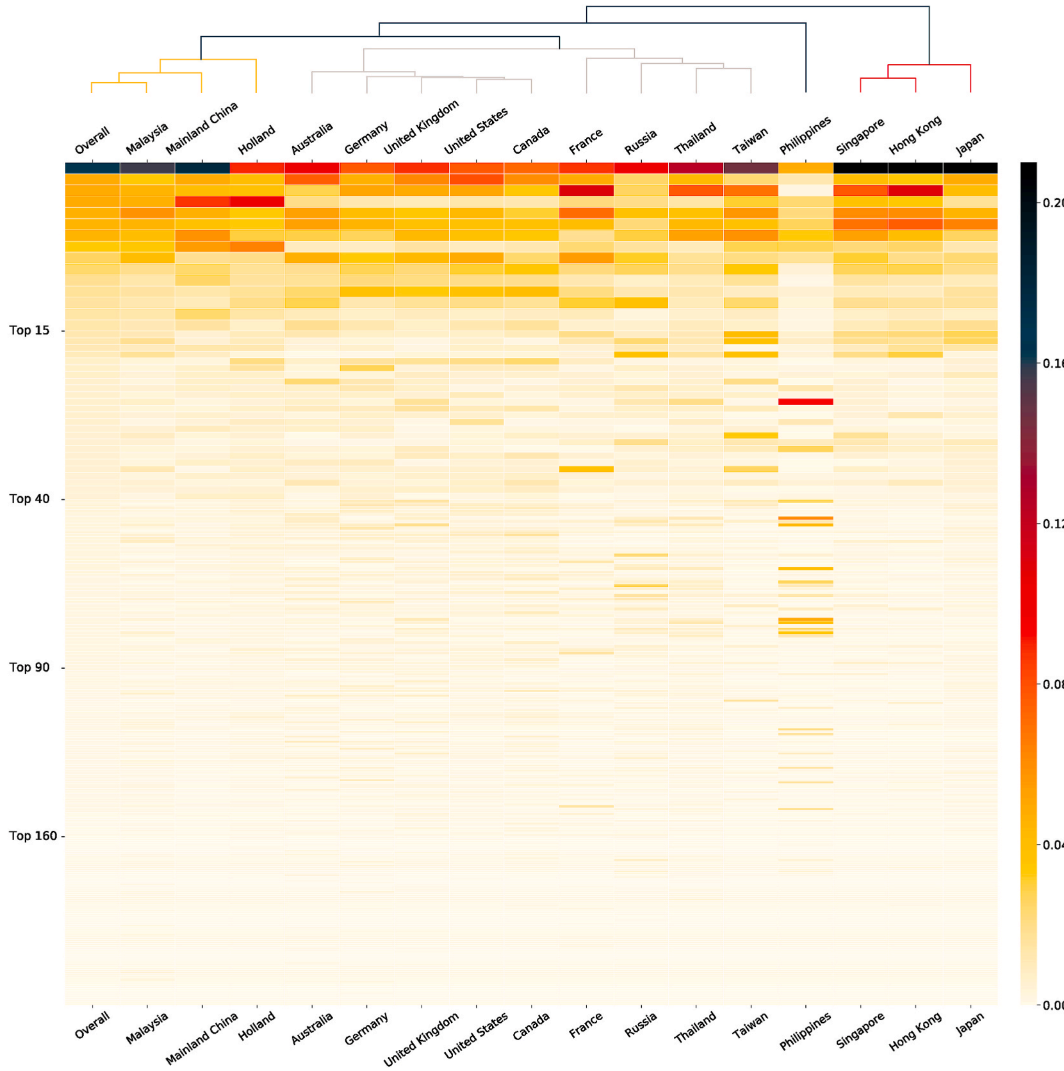


Fig. C.5. Clustering result of tourist visitation patterns by country or region based on threshold of 3 h.

References

2018 international visitor survey. (2018). *tech. rep., Ministry of Culture, Sports and Tourism*. the Republic of Korea.

Ahas, R., Aasa, A., Roose, A., Mark, Ü., & Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management, 29*(3), 469–486.

Asero, V., Gozzo, S., & Tomaselli, V. (2016). Building tourism networks through tourist mobility. *Journal of Travel Research, 55*(6), 751–763.

Baggio, R. (2008). Symptoms of complexity in a tourism system. *Tourism Analysis, 13*(1), 1–20.

Baggio, R. (2011). Collaboration and cooperation in a tourism destination: A network science approach. *Current Issues in Tourism, 14*(2), 183–189.

Baggio, R. (2017). "Network science and tourism – the state of the art. *Tourism Review, 72*(1), 120–131.

Baggio, R., & Cooper, C. (2010). Knowledge transfer in a tourism destination: The effects of a network structure. *Service Industries Journal, 30*(10), 1757–1771.

Baggio, R., Scott, N., & Cooper, C. (2010). Network science: A review focused on tourism. *Annals of Tourism Research, 37*(3), 802–827.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences, 101*(11), 3747–3752.

Belyi, A., Bojic, I., Sobolevsky, S., Sitko, I., Hawelka, B., Rudikova, L., et al. (2017). Global multi-layer network of human mobility. *International Journal of Geographical Information Science, 31*(7), 1381–1402.

Brandes, U., Robins, G., McCranie, A., & Wasserman, S. (2013). What is network science? *Network science, 1*(1), 1–15.

Byrne, D., & Callaghan, G. (2013). *Complexity theory and the social sciences: The state of the art*. Routledge.

Casanueva, C., Gallego, Á., & García-Sánchez, M.-R. (2016). Social network analysis in tourism. *Current Issues in Tourism, 19*(12), 1190–1209.

Dâ€™Agata, R., Gozzo, S., & Tomaselli, V. (2013). Network analysis approach to map tourism mobility. *Quality and Quantity, 47*(6), 3167–3184.

Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci, 5*(1), 17–60.

Erkuş-Öztürk, H., & Eraydın, A. (2010). Environmental governance for sustainable tourism development: Collaborative networks and organisation building in the antalya tourism region. *Tourism Management, 31*(1), 113–124.

Farrell, B. H., & Twining-Ward, L. (2004). Reconceptualizing tourism. *Annals of Tourism Research, 31*(2), 274–295.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports, 486*(3), 75–174.

Fyall, A., & Garrod, B. (2005). *Tourism marketing: A collaborative approach* (Vol. 18). Channel View Publications.

Fyall, A., Garrod, B., & Wang, Y. (2012). Destination collaboration: A critical review of theoretical approaches to a multi-dimensional phenomenon. *Journal of Destination Marketing & Management, 1*(1–2), 10–26.

González-Díaz, B., Gómez, M., & Molina, A. (2015). Configuration of the hotel and non-hotel accommodations: An empirical approach using network analysis. *International Journal of Hospitality Management, 48*, 39–51.

Gonzalez, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779.

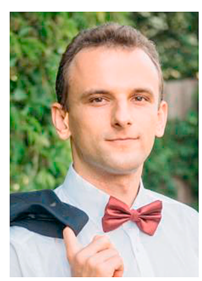
Grinberger, A. Y., & Shoval, N. (2019). "Spatiotemporal contingencies in touristsâ€™ intradiurnal mobility patterns. *Journal of Travel Research, 58*(3), 512–530.

Guimera, R., Mossa, S., Turtschi, A., & Amaral, L. N. (2005). "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences, 102*(22), 7794–7799.

Hede, A.-M., & Stokes, R. (2009). Network analysis of tourism events: An approach to improve marketing practices for sustainable tourism. *Journal of Travel & Tourism Marketing, 26*(7), 656–669.
Hwang, Y.-H., Gretzel, U., & Fesenmaier, D. R. (2006). Multicity trip patterns: Tourists to the United States. *Annals of Tourism Research, 33*(4), 1057–1078.
Javed, M. A., Younis, M. S., Latif, S., Qadir, J., & Baig, A. (2018). Community detection in networks: A multidisciplinary review. *Journal of Network and Computer Applications, 108*, 87–111.
Jönsson, C., & Devonish, D. (2008). Does nationality, gender, and age affect travel motivation? A case of visitors to the caribbean island of Barbados. *Journal of Travel & Tourism Marketing, 25*(3–4), 398–408.
Leiper, N. (2000). "Are destinations' the heart of tourism'? The advantages of an alternative description. *Current Issues in Tourism, 3*(4), 364–368.
Leung, X. Y., Wang, F., Wu, B., Bai, B., Stahura, K. A., & Xie, Z. (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), 469–484.
Lew, A., & McKercher, B. (2006). Modeling tourist movements: A local destination analysis. *Annals of Tourism Research, 33*(2), 403–423.
Li, H., Chen, J. L., Li, G., & Goh, C. (2016). Tourism and regional income inequality: Evidence from China. *Annals of Tourism Research, 58*, 81–99.
Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management, 68*, 301–323.
Lo, I. S., McKercher, B., Lo, A., Cheung, C., & Law, R. (2011). Tourism and online photography. *Tourism Management, 32*(4), 725–731.
Lozano, S., & Gutiérrez, E. (2018). A complex network analysis of global tourism flows. *International Journal of Tourism Research, 20*(5), 588–604.
McKercher, B. (1999). A chaos approach to tourism. *Tourism Management, 20*(4), 425–434.
Miguéns, J., & Mendes, J. (2008). Travel and tourism: Into a complex network. *Physica A: Statistical Mechanics and Its Applications, 387*(12), 2963–2971.
Mill, R. C., & Morrison, A. M. (2002). *The tourism system*. Kendall Hunt.
Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science, 298*(5594), 824–827.
Newman, M. E. (2003). The structure and function of complex networks. *SIAM Review, 45*(2), 167–256.
Newman, M. E., Barabási, A.-L. E., & Watts, D. J. (2006). *The structure and dynamics of networks*. Princeton university press.
Paget, E., Dimanche, F., & Mounet, J.-P. (2010). A tourism innovation case: An actor-network approach. *Annals of Tourism Research, 37*(3), 828–847.
Pavlovich, K. (2003). The evolution and transformation of a tourism destination network: The waitomo caves, New Zealand. *Tourism Management, 24*(2), 203–216.
Rodriguez, X. A., Martinez-Roget, F., & Gonzalez-Murias, P. (2018). Length of stay: Evidence from santiago de compostela. *Annals of Tourism Research, 68*, 9–19.
Sainaghi, R., & Baggio, R. (2014). Structural social capital and hotel performance: Is there a link? *International Journal of Hospitality Management, 37*, 99–110.
Sainaghi, R., & Baggio, R. (2017). Complexity traits and dynamics of tourism destinations. *Tourism Management, 63*, 368–382.
Sayama, H. (2015). *Introduction to the modeling and analysis of complex systems*. Open SUNY Textbooks.
Scott, N., Baggio, R., & Cooper, C. (2011). Network analysis methods for modeling tourism inter-organizational systems. In *Tourism sensemaking: Strategies to give meaning to experience* (pp. 177–221). Emerald Group Publishing Limited.
Scott, N., Cooper, C., & Baggio, R. (2008). Destination networks: Four australian cases. *Annals of Tourism Research, 35*(1), 169–188.
Shih, H.-Y. (2006). Network characteristics of drive tourism destinations: An application of network analysis in tourism. *Tourism Management, 27*(5), 1029–1039.
Shoval, N., & Ahas, R. (2016). The use of tracking technologies in tourism research: The first decade. *Tourism Geographies, 18*(5), 587–606.
Sobolevsky, S., Campari, R., Belyi, A., & Ratti, C. (Jul 2014). General optimization technique for high-quality community detection in complex networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics, 90*, Article 012811.
Stienmetz, J. L., & Fesenmaier, D. R. (2019). Destination value systems: Modeling visitor flow structure and economic impact. *Journal of Travel Research, 58*(8), 1249–1261.
Wang, Y. (2008). Collaborative destination marketing: Understanding the dynamic process. *Journal of Travel Research, 47*(2), 151–166.
Wang, Y., & Xiang, Z. (2007). Toward a theoretical framework of collaborative destination marketing. *Journal of Travel Research, 46*(1), 75–85.
Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236–244.
Williams, N. L., Hristov, D., et al. (2018). An examination of dmo network identity using exponential random graph models. *Tourism Management, 68*, 177–186.
Xu, Y., Belyi, A., Bojic, I., & Ratti, C. (2018). Human mobility and socioeconomic status: Analysis of Singapore and Boston. *Computers, Environment and Urban Systems, 72*, 51–67.
Yan, L., Wang, D., Zhang, S., & Xie, D. (2018). *"Evaluating the multi-scale patterns of jobs-residence balance and commuting time–cost using cellular signaling data: A case study in shanghai," transportation* (pp. 1–16).
Zahra, A., & Ryan, C. (2007). "From chaos to cohesionâ€"complexity in tourism structures: An analysis of New Zealand's regional tourism organizations. *Tourism Management, 28*(3), 854–862.

Yang Xu is an Assistant Professor in the Department of Land Surveying and Geo-Informatics at the Hong Kong Polytechnic University. His research interests include Geographic Information Science (GIScience), Urban Informatics and Trajectory Data Mining.

Jingyan Li is a research assistant in the Department of Land Surveying and Geo-Informatics (LSGI) at the Hong Kong Polytechnic University. She received her bachelor's degree in Geomatics in LSGI. Her research interests include Geographic Information Science (GIScience), Urban Informatics and Mobility Science.



Alexander Belyi is a Software Engineer at Singapore-MIT Alliance for Research and Technology (SMART) Centre, he is a candidate for PhD in the Faculty of Applied Mathematics and Computer Science at the Belarusian State University. Before joining SMART, he spent a year at MIT as a visiting PhD student in Senseable City Lab. He received his bachelor's degree in Applied Mathematic from the Belarusian State University and master's degree in Applied Mathematics and Informatics from the Researchers Training Institute of the National Academy of Sciences of Belarus. His research interests include algorithms for networks and data analysis, optimization, big data and quantitative research methods for urban studies.

Sangwon Park is Associate Professor, School of Hotel and Tourism Management, The Hong Kong Polytechnic University. His research interests include tourism big data, information technology and labour productivity in tourism and hospitality.