Application of Graph Theory to Mining the Similarity of Travel Trajectories

1. Introduction

A considerable amount of research has focused on characterizing travel movements (Mckercher & Lau, 2008; Vu, Li, Law, & Ye, 2015). A comprehensive understanding of travel movements within a destination provides fundamental insights into tourist behavior, and such insights are directly applicable to destination management activities such as planning, product development, transport management, attraction planning, and accommodation development (Mckercher & Lau, 2008). Big data representing human mobility that is mostly generated from mobile technologies and social media platforms offers unprecedented opportunities to tourism researchers as an innovative source of information for smart tourism. Thanks to the advancement of information technology, tourism researchers and marketers have identified travelers' movement patterns (Gao et al., 2019; Mckercher & Lau, 2008; Rong et al., 2012; Huang & Wu, 2012) through such measures as predicting next travel movement behavior (Zheng et al., 2017), exploring spatial interactions/destination value systems (Park et al., 2020; Stienmetz & Fesenmaier, 2019), and assessing activity preferences (Vu, Li, Law, & Zhang, 2019).

Although a number of tourism studies have explored travel movement patterns (Lue et al., 1993), little research has focused on measuring the similarities between travel movement behaviors. Assessing the similarities in patterns of travel mobility is crucial for developing destination recommendation systems, which are key constituents for achieving smart destinations (Park et al., 2020). Trajectory similarity measures play an important role in trajectory mining to determine travel mobility (Xia et al., 2011). Assuming that the locations visited by travelers imply their interests and preferences (Li et al., 2008), trajectory similarity can identify travelers who exhibit similar movement patterns and, in turn, classify individuals who have similar interests and preferences based on homogenous groups of travel movement behaviors. This insight can be an underpinning of a real-time destination recommendation system.

Most extant studies of travel movement have focused on exploring the spatial features of travel trajectories that represent a sequence of places visited by travelers—that is, although the relevant literature has explored the spatial process, an approach to considering the temporal (time) dimension is limited. Travelers use their time to conduct diverse activities at their destinations, and such usage of time is highly heterogeneous (Xu et al., 2020). This notion implies that considering the spatial process to understand travel movement patterns provides a partial understanding of travel behavior. Indeed, the current research fills this research gap by considering the spatial and temporal aspects in estimating trajectory similarity.

The current research applies graph theory, which is the foundation of network science, to elucidate travel movement patterns (Hu et al., 2019). Since tourism is a complex system in nature, quantitatively describing intrinsic knowledge from tourism big data analytics and multifaceted travel flows is a critical issue (Baggio et al., 2010; McKercher, 1999). Graph theory facilitates the illustration of complex travel mobility via vertices (places visited by travelers) and edges (directions of travel flow) through a mathematical theory. Travel patterns are highly divergent—e.g., there are substantial variations in the number of visited places and the directional sequences of places—and thus, graph-based spatiotemporal analytics should function as a guide for discovering insights by considering different mobility contexts.

Thus, the purpose of current research analyzing mobile sensor big data is to (1) propose the trajectory similarity of travel mobility by considering the spatial and temporal dimensions of travel flow and (2) adopt graph theory to gain important knowledge of in-transit travel behavior.

This work provides important contributions to tourism knowledge. In particular, along with extant studies that identify key travel patterns, this study suggests an innovative approach to assessing trajectory similarity that reflects the potential of data-driven clustering methods. Indeed, as opposed to the extant literature on identifying collective travel movement patterns, this trajectory similarity study performs a pairwise comparison of comprehensive travel movement patterns at the individual level not only to identify key behavioral patterns but also to classify travelers who show similar behaviors. In addition, the graph theory approach guides the characterization of complex destination systems through the lens of travel flow. This research demonstrates the applicability and existence of a scaling law of network science in travel mobility as demonstrated by heavy-tailed distribution and spatial heterogeneity (Jiang, 2018).

2. Literature Review

2.1. Tourist Mobility

By definition, travel includes spatial movements for any purpose and duration from the place where a person lives to a single or multiple destination. Moving from one attraction to another within a destination is also regarded as travel (Hwang et al., 2006; Leiper, 1979). This definition implies that travel behavior is fundamentally composed of three dimensions: the “what” (e.g., activity), the “when” (temporal aspect), and the “where” (spatial aspect) (Shoval & Isaacson, 2007). Travel behavior is closely related to the fundamental problem in the economics of tourism attractions and experiences that will be used to identify the current market condition and predict the future flow of travelers. Therefore, tourist movement or mobility should be one of the critical aspects of understanding the various purposes of destination marketing organizations (DMOs), i.e., planning, management, and destination marketing (McKercher & Lau, 2008; Park et al., 2020).

Previous research has shown that travelers typically move through several attractions within a given time and space that are the basic building blocks of travelers' experiences (Stienmetz et al., 2020; Zach & Gretzel, 2011). Scholars have demonstrated the effectiveness and usefulness of applying knowledge regarding tourist behavior, particularly tourist mobility, to managing, planning, and marketing tourism destinations. For example, Stienmetz and Fesenmaier (2019) emphasized that aggregated travelers' activities reflect value cocreation processes between the supply and demand sides of a destination, leading to a "network orchestrator" approach in destination management. Zhang, Li, and Su (2017) suggested effective theme park attributes and spatial design by examining visitors' movements in a theme park. Vu, Li, Law, and Zhang (2018) provided marketing implications for international travels by discovering sequential travel patterns and strong associations among tourism destinations from geotagged photos via sequential rule mining. Park et al. (2020) identified several popular tourist attractions and spatial interactions across hot spots in travel destinations that will be used as the foundation for tourism design and planning.

The inter- and intra-destination movement patterns of tourists have been studied at different scales: regional, national, and global. Lue et al. (1993) conceptualized the role and structure of multi-destination trips and identified five distinctive spatial movement patterns of pleasure travelers (i.e., single destination, in transit, base camp, regional tour, and trip chaining). Lew and McKercher (2006) proposed four types of territorial models that reflect the effects of distance and intervening opportunities (e.g., transportation and time budget) and three types of linear path models showing tourists' movements and spatial patterns based primarily on the geography of a destination. Although some variations exist, follow-up research in tourism studies has repeatedly confirmed the findings of earlier studies. For example, McKercher and Lau (2008) identified 11 unique movement styles of Hong Kong visitors that are largely influenced by territoriality and intensity of consumption (e.g., number of stops and activity participation). Their study supported the territorial models proposed by Lew and McKercher (2006); however, linearity was not observed in the obtained data. Park et al. (2020) presented evidence of circular loop patterns (or the base camp model), a linear point-to-point chaining pattern, and radiating hubs by applying sequential pattern-mining analytics to the mobile big data of international tourists.

The advancement of information technology allows researchers to utilize more diverse data sources than direct methods (e.g., observation, survey, diary). Recent studies have applied advanced tracking technologies (e.g., GPS, Wi-Fi, and Beacon) and social media posts (e.g., geotagged photos) (Wong et al., 2017). Advanced tracking methods are less burdensome to participants and relatively less interrupted by external situations, allowing researchers to avoid any potential biases during data collection (Raun et al., 2016; Shoval & Isaacson, 2007). However, social media data (e.g., geotagged information) exhibits several limitations in delineating travel movements within a destination, i.e., data sparsity, selection bias, scarcity of sources, and incorrect information (Girardin et al., 2008; Wong et al., 2017). Mobile phone data, on the other hand, helps researchers to overcome such limitations of social media data in terms of (1) greatly improved spatial and temporal accuracy, (2) population-like sample, (3) extensive sampling/tracking period, (4) relative lack of selection bias, and (5) ease of data collection (Park et al., 2020; Raun et al., 2016). In other words, unlike geo-tagged information posted on social media, which is solely reliant on extremely few records for only selected travelers, mobile phone data often contains almost the entire population as a sampling frame along with its spatial and temporal movement information. However, mobile phone data generally includes anonymized user IDs, spatial and temporal information, or some additional information depending on the purpose of the study (i.e., nationality in the current study), indicating that no other personal information is available due to the privacy issue (Li, Xu, Tang, Wang, & Li, 2018).

Through these devices and data, tourism researchers can effectively obtain individuals' spatial and temporal positions while traveling and then investigate travelers' movements at a destination with considerable detail and accuracy. Methodologically, data obtained from mobile location–based systems provides new opportunities for modeling the spatiotemporal behavior of travelers at a destination. In particular, the mobile sensor data is suitable for measuring trajectory similarity because numerical information for a trajectory is continuous as opposed to that collected from social media (e.g., Flickr, TripAdvisor) (Yuan & Raubal, 2014). Specifically, social media data cannot accurately delimitate movement trajectories and similarities between trajectories due to data sparsity issues (e.g., few records for individuals and slightly different time frames among individuals). However, mobile phone data, which contains spatial and temporal information in a continuous form with the same spatial and temporal resolution, enables researchers to identify movement trajectories of individuals throughout their trips and then quantify the similarities between users' movement trajectories.

With these advantages of mobile phone data, several studies have already utilized mobile phone data in accurately measuring tourist flow or tourism destination (e.g., Raun et al., 2016; Shoval & Isaacson, 2007). In earlier studies, Baggio and Scaglione (2018) showed the value of mobile phone data in monitoring and identifying strategic tourist flow, which can improve the governance of local DMOs. In recent years, Park and his colleagues utilized a mobile positioning dataset to detect the spatial structures of tourism destinations using trajectory mining (Park et al., 2020), the time use of tourists at the destination (Xu et al., 2020), and the heterogeneity and collective dynamics of inbound tourist movements (Xu et al., 2021). These recent studies built upon previous empirical works, which focused more on the movement patterns, by utilizing tourist movement patterns from mobile positioning data to detect various characteristics of tourist movements and destinations. Thus, the present study focuses primarily on the similarity of travel trajectories within the destination by analyzing mobile positioning data in terms of graph theory.

2.2. Graph Theory

Graph theory is the study of graphs based on a mathematical structure that is represented by a set of nodes (vertices) and links (edges) that connect pairs of nodes (Hayes, 2000; Riaz & Ali, 2011). A graph or network in graph theory represents the relationship between nodes (vertices) and their connections (links or edges). Studies in tourism and geography generally consider points in space (locations, areas, or regions) as nodes and pathways (fluxes, exchanges, or spatial interactions) and as edges (Phillips et al., 2015). A graph-based approach has been used in many disciplines (e.g., computer science, geography, and transportation) to quantify structural and functional connectivity within a network system based on nodes and edges. In such applications, nodes frequently contain information about qualitative and quantitative characteristics, and edges exhibit unique properties, i.e., weights and directions (Dale & Fortin, 2010).

The advancement and widespread application of graph theory have inspired the field of network science. Graph theory specifically describes a well-structured graph, i.e., one with a common shape and relatively small size, whereas network science focuses on structurally complex and unpredictable graphs (or networks) (Derrible & Kennedy, 2011). Consequently, studies in network science emphasize the dynamic properties of real-world networks and consider networks a dynamic system (Newman et al., 2006). A network or system often involves nonlinear interactions among elements, and each element will react based on its historical behavioral pattern and locally available information (Baggio et al., 2010).

In recent years, tourism researchers have utilized graph theory and network science to quantify functions and relationships among entities or components within the tourism and hospitality industries, including structural complexity, diversity, dynamic complexity, and interactions (Strogatz, 2001). Tourism functions are frequently described in a nonlinear and dynamic manner, making them suitable for network theory (McKercher, 1999). By applying graph theory and network science, tourism destinations have been described as a complex network system consisting of various interrelated components, e.g., natural resources, man-made attractions, and service providers (Baggio et al., 2010; Miguéns & Mendes, 2008; Xu et al., 2020). A network approach has further modified the conceptualization of tourism destinations based on the notion of "tourist-activated networks" and a market-oriented system wherein tourists' behavior and flow at a destination activate the supply-and-demand system of tourism destinations (Stienmetz & Fesenmaier, 2019; Zach & Gretzel, 2011).

A core feature of networks in the real world (i.e., scale-free networks) perfectly fits tourist mobility within a tourism destination, allowing researchers to determine the structural properties and characteristics of tourism destinations on the basis of tourists' flows. Real networks are neither completely random nor completely regular, indicating that links between nodes are not created randomly (Wang & Chen, 2003; Watts & Strogatz, 1998). This characteristic implies that tourist movements between places may reflect the popularity and necessity of such routes depending on tourists' motives or locations' functions within tourism destinations. Scholars have recently observed the uneven distributions of connections in many large-scale complex networks in the real world. For example, if a certain tourism attraction or place (i.e., an actor or node in the network theory) attracts more tourists or transportation connections with other places (i.e., ties and connections in the network theory), then that particular place can easily become a well-connected hub with a large number of ties (Milo et al., 2002). In the real world, the connections or nodes of each destination (or the connectivity of each node) follow a power law distribution and are independent of network scale (i.e., the size of the entire network). Consequently, network science (i.e., scale-free networks) helps tourism researchers understand the spatial structure and complexities of tourism destinations (e.g., Baggio et al., 2010; Miguéns & Mendes, 2008). These characteristics can explain the evolution of a tourism destination, which consists of several popular tourist spots and changes through the lens of travel flow.

Despite their widespread applications, many earlier studies that focus on tourist mobility using graph theory have critical limitations with regard to (1) weights and directions (i.e., edge contains weights and directions) and (2) sampling (i.e., extremely sparse geotagged or survey data) (Girardin et al., 2008; Park et al., 2020; Wong et al., 2017). First, many earlier studies using graph theory often neglected the importance of the directions of tourist flow, possibly due to the data availability. However, the direction of tourist flow might be an important issue in managing the tourism destination or providing the proper transportation services at the destination. Indeed, links and edges have unique properties (i.e., weights and directions) for understanding tourist movement patterns within a destination (Dale & Fortin, 2010). Such information helps researchers reconstruct accurate tourist mobility patterns, including movements between locations and their sequence and, more importantly, the similarities between travel trajectories. Second, social media data, which is widely used these days, does not always cover the entire population, and it includes a subset of users who agree to share their locations, thus presenting a potential bias of the population. Such practices indeed bring about a data sparsity issue which hinders the accurate estimation or identification of the entire tourist flow at the destination (extremely little coordinate information for individuals). To resolve the aforementioned challenges, the current study utilizes mobile phone sensor data as another type of data source for identifying the sequential behavioral patterns of tourists within a destination. The mobile sensor data collected from a massive sample size records the comprehensive movement of international travelers in an automatic way rather than relying on information that travelers have spontaneously shared. Baggio and Scaglione (2018) demonstrated that network and graph theory work well with large amounts of data such as mobile phone data while depicting visitor flow, and the authors emphasized that visitor flow identified by mobile phone data can facilitate development of innovative tourism products and services.

3. Methodology

3.1. Data Collection

The researchers collaborated with one of the largest mobile communication companies in the Republic of Korea (hereafter Korea) in conducting a tourism big data project and were able to access the mobile sensor data of international travelers who visited Jeonju City in Korea in the last 12 months (i.e., from August 2017 to July 2018). The international travelers were defined as those travelers who used a mobile roaming service while visiting the city. Jeonju is located in the southwestern part of Korea, covering an area of 206.22 km^2. Jeonju is one of the most popular and attractive cities for tourists due to its traditional culture and distinctive food as a “Creative City of Gastronomy” and part of the “Creative Cities Network” of the United Nations Educational, Scientific, and Cultural Organization.

The mobile data was collected from 18,584 mobile users, generating 202,255 mobile sensor records. The dataset represents a tracking trajectory that comprises four elements: user ID, nationality, spatial (coordinate) information, and temporal information (time of arrival at and departure from a certain location associated with the corresponding coordinates). Coordinate information refers to the locations of cellphone towers, with 782 cellphone towers covering a radius of 181 m on median nearest distance. Table 1 provides an example of a user’s mobile sensor data.

[Please insert Table 1 about here]

For example, User ID 1 visited a place on 127.*** longitude and 35.*** latitude at 15:43 on November 25, 2017, and stayed there for 10 min. Then, the traveler moved to another location described as 127.*** longitude and 36.*** latitude at 17:06. This information also means that the traveler took 73 min to move from the first place (coordinate of cell towers: 127.*** of longitude, 35.*** of latitude) to the second place (coordinate of cell towers: 127.*** of longitude, 36.*** of latitude). If a traveler stays in a same location, the timestamp keeps recording the duration of stay with constant coordinate information. The fine-grained mobile sensor data that reflects the discrete spatiotemporal movement records allows researchers to comprehensively discover the trajectory information of individual travelers at a destination (e.g., Jeonju). Notably, no information that can potentially enable researchers to identify individuals, such as gender, name, and phone number, was collected.

3.2. Data Analysis

This study implements a series of mobile big data analytics to explore trajectory similarity measurements, as shown in Figure 1.

[Please insert Figure 1 about here]

3.2.1. Trajectory synthetization

Trajectory synthetization identifies sequential locations (coordinates) from individual trajectories and extracts the corresponding arrival and departure times for each location. The extracted individual trajectories are essential for calculating travel time between two locations (coordinates) in the full trajectory. Suppose that two people are traveling. Traveler 1 visited a, b, c, d, and e places, and Traveler 2 visited, a, b, c, d, e, and f places. Two trajectory extractions can be represented as follows:

$$\text{traj}_1 = \langle a \xrightarrow{\Delta t_1} b \xrightarrow{\Delta t_2} c \xrightarrow{\Delta t_3} d \xrightarrow{\Delta t_4} e \rangle,$$

$$\text{traj}_2 = \langle a \xrightarrow{\Delta t_1{}'} b \xrightarrow{\Delta t_2{}'} f \xrightarrow{\Delta t_3{}'} c \xrightarrow{\Delta t_4{}'} d \xrightarrow{\Delta t_5{}'} e \xrightarrow{\Delta t_6{}'} f \rangle,$$

where traj_1 and traj_2 refer to the trajectories of Users ID1 and ID 2, respectively.

The number of place names (e.g., a, b, c, and so on) indicates the number of places visited successively by a user in the corresponding locations of cellphone towers. Given the temporal information, such as the arrival and departure times at each cellphone tower for each traveler, the researchers can estimate the time interval between the locations of two cellphone towers (or two places visited in such sequence). That is, Δt_i and $\Delta t_j{}'$ denote the *i*th or *j*th transfer times of $traj_1$ and $traj_2$ between two locations.

3.2.2. Trajectory matching

Similar trajectory: Similar trajectory matching detects pairs of trajectories ($traj_1$ and $traj_2$) as similar if and only if they satisfy certain conditions such as spatial and temporal constraints.

Spatial constraint: Spatial path (trajectory) is one of the primary spatial movement parameters for spatial similarity measures and is an ordered list of actually recorded locations (e.g., coordinates of cellphone tower). A directed straight line can connect each two consecutive recorded locations in a trajectory. In a similar trajectory, more than one (at least two) location with the same coordinates should exist in a pair of trajectories, and the visiting order in this pair of trajectories must be similar (Li et al., 2008).

Given two trajectories, $traj_1$ is the trajectory of User ID 1 with length m, and $traj_2$ is the trajectory of User ID 2 with length n.

$$traj_1 = < a_1 \xrightarrow{\Delta t_1} a_2 \xrightarrow{\Delta t_2} \ldots \xrightarrow{\Delta t_{m-1}} a_m >,$$

$$traj_2 = < b_1 \xrightarrow{\Delta t_1{}'} b_2 \xrightarrow{\Delta t_2{}'} \ldots \xrightarrow{\Delta t_{n-1}{}'} b_n >.$$

$$a_i = (Lon_i, Lat_i)$$
$$b_j = (Lon_j, Lat_j)$$
$$a_i = b_j = (Lon_i = Lon_j, Lat_i = Lat_j)$$

Where a_i denotes the *i*th location (coordinate of cellphone tower) for User ID 1 in Jeonju, b_j denotes the *j*th location (coordinate of cellphone tower) for User ID 2 in Jeonju, and Lon and Lat correspond to the longitude and latitude of a cellphone tower, respectively.

Temporal constraint: Temporal duration (e.g., transition time) is one of the important derived movement parameters for temporal similarity measures. It is obtained from the corresponding arrival and departure times for each location in a trajectory. A quantitative measure is the difference between two temporal durations (Ranacher & Tzavella, 2014). The absolute value of the difference between two users' transition times should be within a predefined threshold, denoting that the two users have similar movement times between the same regions (Li et al., 2008).

$$\forall 1 \leq \mathrm{i} \leq \mathrm{m}, \forall 1 \leq \mathrm{j} \leq \mathrm{n}, \left|\Delta \mathrm{t_i} - \Delta \mathrm{t_j}'\right| \leq \mathrm{t_{th}},$$

where ∀ indicates all the coordinates in a trajectory; $\Delta \mathrm{t_i}$ is the *i*th transfer time of $traj_1$ between two locations; $\Delta t_j'$ is the *j*th transfer time of $traj_2$ between two locations; $\mathrm{t_{th}}$ refers to the time threshold; and m and n are the lengths of $traj_1$ and $traj_2$, respectively.

For example, the two preceding trajectories (e.g., $traj_1$ and $traj_2$) can be represented as follows:

$$traj_1 = <a \xrightarrow{56\ \mathrm{min}} b \xrightarrow{136\ \mathrm{min}} c \xrightarrow{78\ \mathrm{min}} d \xrightarrow{14\ \mathrm{min}} e>,$$

$$traj_2 = <a \xrightarrow{34\ \mathrm{min}} b \xrightarrow{67\ \mathrm{min}} f \xrightarrow{31\ \mathrm{min}} c \xrightarrow{86\ \mathrm{min}} d \xrightarrow{3\ \mathrm{min}} e \xrightarrow{8\ \mathrm{min}} d \xrightarrow{39\ \mathrm{min}} f>,$$

where a, b, and so on represent the accurate coordinates of the cellphone towers in $traj_1$ and $traj_2$. The order of the nodes represents the visiting order of users in their trajectories.

In terms of spatial constraints, finding the same pair of sequential trajectories is required. In this example, <a→b>, which satisfies the spatial constraint. Once the spatial condition is satisfied, the temporal constraint is then considered with 30 min as the threshold (Park et al., 2020). Based on local knowledge, there are basically three ways for travelers to move within the city: bus, taxi, and walking. The travel time between one of the most famous travel attractions located in the north area (i.e., Deokjin Park) and another famous travel attraction (e.g., Hanok Village) in the south area by public transportation (i.e., bus) is around 25–30 minutes. When taking a taxi, it takes around 15 minutes, and when walking, it takes about 1 hour 20 minutes. So, even considering the farthest distance between two popular travel attractions, the difference of time transition when using public transportation (e.g., taxi or bus) is less then 30 minutes. This means that when travelers use public transportation to move within Jeonju city, the transition time is very likely to be within the threshold, which meets the temporal constraint. In addition, the threshold (30 min) has been decided following the guidance of mobile big data analysis developed by Korea Tourism Organization and relevant literature (Bifulco et al., 2010; Jongno City Office, 2019; Park et al., 2020). They considered the 30-min threshold as an indication to discern whether a traveler is engaged in a travel activity at the cell.

Considering the preceding example, the transition time from “a” to “b” for $traj_1$ is 56 min and for $traj_2$ it is 34 min, resulting in a 22-min difference (56 min – 34 min). Thus, the similar trajectory satisfies the temporal constraint, which detects a two-length (a→b) similar trajectory. From the perspective of graph theory, a directed network from vertex a to vertex b is shown.

3.2.3. M-length similar sequence

The length of a similar trajectory indicates the number of locations (vertices in graph theory) visited in a similar sequence (edges in graph theory) by a pair of travelers (Li et al., 2008). M refers to the numbers of similar sequences that satisfy the two constraints discussed above, labeled as a m-length similar sequence. Two rules—namely, trajectory extension and pruning—must be followed when calculating a m-length similar sequence.

Trajectory extension: M-length similar trajectory refers to the longest common trajectory in a pair of trajectories (Li et al., 2008). Thus, if the common trajectories identified in a pair of trajectories are shorter than the longest common trajectory, then they will be *skipped*.

Trajectory pruning: The previous longest common trajectory will be *replaced* by the newer longest common trajectory (Li et al., 2008).

In other words, trajectory extension is performed to detect all the elements (locations with transition times) which satisfy spatiotemporal constraints but do not need to be drawn from the contiguous positions in pairwise trajectories, whereas trajectory pruning is done to detect the maximum-length (m-length) similar sequences.

In the following successive sequences ($traj_1$ and $traj_2$), in addition to trajectory <a→ b>, another pair of trajectories (<b→ c>) satisfies the spatial constraint. The temporal constraint is also satisfied on the basis of the 18-min difference in the time interval between $traj_1$ (136 min) and $traj_2$ (67 min + 20 min remaining at “f” + 31 min). By combining sequential trajectories <a→ b> and <b→ c> that satisfy the spatial and temporal conditions, a three-length similar sequence <a→b→ c> is detected, replacing the two-length similar sequence. The three-length similar sequence can be represented as follows:

$traj_1 = \langle a \xrightarrow{\text{56 min}} b \xrightarrow{\text{136 min}} c \xrightarrow{\text{78 min}} d \xrightarrow{\text{14 min}} e \rangle$,

$$traj_2 = \langle a \xrightarrow{34\ min} \underbrace{b \xrightarrow{67\ min} f \xrightarrow{31\ min} c}_{(20\ min)} \xrightarrow{86\ min} d \xrightarrow{3\ min} e \xrightarrow{8\ min} d \xrightarrow{39\ min} f \rangle.$$

The calculation continues until the two (spatial and temporal) constraints are not satisfied. In accordance with the rule of "trajectory pruning," a five-length similar trajectory between $traj_1$ and $traj_2$ is detected, and its composition is <a → b → c → d → e>.

3.2.4. Algorithm for assessing similar trajectory matching: Longest common subsequence (LCSS) algorithm

To calculate m-length similar trajectory matching, this research applied an innovative algorithm-labeling method, the LCSS algorithm (Toohey & Duckham, 2015). This algorithm aims to discover the longest common subsequence in a set of full sequences (or trajectories). The fundamental idea is to match the common subsequences appearing in both full sequences to eliminate outliers (i.e., some elements remain unmatched) (Bergroth et al., 2000).

There are a couple of benefits in adopting the LCSS algorithm. First, the LCSS model supports trajectories of different/variable lengths and is good for processing with noisy trajectory data. Most travelers undertake multi-destination trips (i.e., they visit more than one place) (Lue et al., 1993; Tideswell & Faulkner, 1999). It is very common to have multi-destination trajectories with different lengths and many outliers (unmatching/noise points). LCSS is almost immune to outliers (noise points) because the algorithm is designed to match two trajectories by allowing them to stretch without rearranging the order of the elements (locations with transition times).

Second, the LCSS model can analyze detailed trajectories without many transformations applied to the original trajectories. To reduce the data size, plenty of previous research has applied the clustering methods to movement data (e.g., Andrade et al., 2020; Li et al., 2008). However, clustering of the trajectory will result in significant loss of information. In this research, each matching trajectory consists of sequential locations (coordinates of cellphone towers) at every change in direction or velocity (transition time). The matching trajectories are nearly the original trajectories. The LCSS algorithm facilitates for generating good performance for uncovering the travelers' trajectories with multiple elements such as locations and transition times. However, other trajectory similarity measures (e.g., Euclidean distance measures, dynamic time warping–based measures [e.g., DTW, PDTW], edit distance–based measures [e.g., EDR, ERP], etc.) are sensitive to the detailed travelers' trajectories with multiple elements (Wang et al., 2013).

More specifically, $Z = \langle z_1, z_2, \ldots, z_l \rangle$ in all the common subsequences of the two given trajectories, $traj_1 = \langle a_1, a_2, \ldots, a_m \rangle$ and $traj_2 = \langle b_1, b_2, \ldots, b_n \rangle$ ($\forall 0 \leq i \leq m, \forall 0 \leq j \leq n, \forall 0 \leq l \leq k$). Dynamic programming is generally applied to calculate LCSS (Cormen, 2013):

$$A = \langle a_1 \rightarrow a_2 \rightarrow \ldots \rightarrow a_m \rangle,$$

$$B = \langle b_1 \rightarrow b_2 \rightarrow \ldots \rightarrow b_n \rangle,$$

$$\text{LCSS}(A, B) = Z = \langle z_1 \rightarrow z_2 \rightarrow \ldots \rightarrow z_k \rangle,$$

where a_i, b_j, and z_l are the *i*th, *j*th, and *l*th places visited in sequence A, sequence B, and the longest common subsequence Z, respectively; m and n are the lengths of sequences A and B, respectively; and z is the length of the longest common subsequence (Z) between sequences A and B.

From the recursion principle of the LCSS algorithm, if sequence A or B is empty, then the longest common subsequence of sequences A and B is 0. If the corresponding elements at the ends of sequences A and B are the same, i.e., $a_m = b_n$, then the longest common subsequence of sequences A and B is the longest common subsequence of sequences A_{m-1} and B_{n-1} plus 1. If the corresponding elements at the end of sequences A and B are different, i.e., $a_m \neq b_n$, then the longest common subsequence of sequences A and B is the maximum of the longest common subsequence of sequences A_m and B_{n-1} and the longest common subsequence of sequences A_{m-1} and B_n.

$$\text{LCSS}(A_m, B_n) = \begin{cases} 0, if\ m = 0 \text{ or n} = 0 \\ 1 + A_{m-1}, B_{n-1}, if\ m, n > 0\ and\ A_m = B_n \\ \max(A_{m-1}, B_n, A_m, B_{n-1}), if\ m, n > 0\ and\ A_m \neq B_n \end{cases} \quad (3)$$

This recursive algorithm continuously performs until it finds the longest common sequence between a pair of sequential trajectories in the entire sample. Consequently, over 345 million combinations of trajectory comparisons were conducted in this research. The LCSS algorithm has been used in similar trajectory matching areas such as human mobility (Vlachos et al., 2005), transportation vehicle mobility (Kim & Mahmassani, 2015), and animal mobility (Cleasby et al., 2019).

3.2.5. Post-normalized trajectory similarity (PNTS)

After estimating LCSS, which determines m-length similar trajectory, this section calculates the similarity scores and uses them to classify travelers from different nationalities. In particular, the similarity score indicates the levels of homogenous travel flow between pairs of travel trajectories.

To estimate trajectory similarity, this study calculated PNTS, which is represented as follows:

The PNTS between Trajectories A and B was proposed by Zhao, Yuan, Peng, and Wang (2002).

$$\text{Trajectory A} = <a_1 \rightarrow a_2 \rightarrow \ldots \rightarrow a_m>,$$

$$\text{Trajectory B} = <b_1 \rightarrow b_2 \rightarrow \ldots \rightarrow b_n>,$$

$$\text{LCSS (A, B)} = Z = <z_1 \rightarrow z_2 \rightarrow \ldots \rightarrow z_k>,$$

$$\text{PNTS}_{\text{p}} (A_m, B_n) = \frac{\text{LCSS (A,B)}}{\max\ (m,n)} \in [0, 1],$$

where a_1, a_2, …, a_m are the locations in Trajectory A; b_1, b_2, …, b_n are the locations in Trajectory B; m and n are the lengths of Trajectories A and B, respectively; LCSS (A, B) is the length of the longest common subsequence of Trajectories A and B; and max (m, n) refers to the maximum trajectory length between Trajectories A and B.

The PNTS aims to compare the similarity between trajectories of different lengths based on similarity scores obtained from LCSS. In this sense, the normalization from the m-length similar trajectory is required to make it comparable. Specifically, the longer the similar trajectory is, the higher the similarity score (PNTS) the similar trajectory can obtain, while the higher the similarity score (PNTS) is, the more similar the pairwise trajectories are.

The PNTS scores of each nationality were used to perform (1) heat map analysis, which presents similar trajectory matching between a pair of nationalities, and (2) agglomerative hierarchical clustering, which groups nationalities on the basis of similarity scores (Xu et al., 2020). Cluster analysis divides a universal set into subsets (Hastie et al., 2009), and it provides widely used unsupervised market segment techniques (D'urso et al., 2013). Notably, we chose agglomerative hierarchical clustering over other alternatives (e.g., *k*-means) for two reasons. First, it adopts a bottom-up approach, and the algorithm does not require predefining of the total number of clusters. Second, agglomerative hierarchical clustering enables the examination of how individual elements are grouped at different steps of the algorithm (through a dendrogram), providing a comprehensive picture of similarities and differences across nationalities.

4. Results

4.1. Temporal travel demand

Figure 2 presents the number of international travelers who used mobile roaming services and visited Jeonju. April and July are top two months when the largest numbers of inbound tourists visited the city. By contrast, January and February are the months when the lowest numbers of international travelers visited Jeonju. This finding shows the existence of tourist demands across seasonal levels. Summer is apparently the peak season, and winter is the shoulder season.

[Please insert Figure 2 about here]

With regard to the days of the week, international travelers are likely to visit Jeonju on Friday (n = 20,834), followed by Thursday (n = 20,469) and Saturday (n = 20,375). Interestingly, the result shows that Sunday is the day with the lowest number of international travelers visiting the city.

4.2. Number of places visited / transition time by travelers

Figure 3 presents numbers of places (cellphone towers) travelers have visited in daily movement. Approximately, 38% of travelers tend to visit a place covered by a single cellphone tower (about 300m radius). About 60% of travelers are likely to visit places covered by up to two cellphone towers (about 600m radius). The frequency rapidly drops as the number of cellphone towers travelers visited increases.

[Please insert Figure 3 about here]

Checking transition time between each cellphone tower, it shows approximately 29 minutes of 46,771 daily travel trajectory on average with 48 minutes in standard deviation.

4.3. Length of Stay at the destination

Next, this study analyzed length of stay by calculating the differences between the arrival date (the first time the digital footprint was observed at the city) and the departure date (the last time the digital footprint was observed at the city) of international travelers. Most travelers (98.7%) stay in Jeonju for less than 15 days (Figure 3). The majority of travelers stay in Jeonju for 1 days, and consistently, the average length of stay in the city is 2.7 days.

[Please insert Figure 4 about here]

4.4. Travel Mobility Models Based on Network Science Analytics

From the data of total 782 cellphone towers, the weighted network of the destination presents a heavy-tailed distribution on the same line of network science (Baggio et al., 2010). Figure 5a presents the cumulative distribution function of vertex (or node) degree consisting of numbers of edges (spatial links by travel flow) at the *X*-axis and the cumulative probability of frequency at the *Y*-axis. Figure 5b shows the cumulative distribution function of weighted edges (or travel flow) consisting of weighted flows (the number of edges between cellphone towers taking into account the number of travelers exhibiting travel flow) at the *X*-axis and the cumulative probability of frequency as edges at the *Y*-axis. Understanding a log–log plots (Figure 5a and 5b), the relationships are likely to be linear in a visual assessment. Furthermore, the statistical estimation demonstrates that the distributions follow the lognormal distributions. This means that travel flow exhibits a scaling law of complex network (power-law/lognormal distributions), as indicated by network science. In other words, few areas (the coverage of cellphone towers) attract a large number of visits while many areas attract few travelers, demonstrating the existence of spatial heterogeneity (Jiang, 2018).

[Please insert Figures 5a and 5b about here]

4.5. Similar Trajectory

Table 2 presents the summary results of the m-length similar trajectory analysis by using the LCSS algorithm. Note that m-length denotes numbers of nodes in graph theory. The trajectory similarity has been applied to identify similar travel flow according to different numbers of nodes (i.e., identical places visited between two travelers). Notably, although the results of over nine-length similar trajectory exist, their frequency is insignificant to show the results. In particular, the number of similar trajectories with two to nine lengths accounts for 98% of the number of categories and international travelers (Appendix I). Thus, this study focuses on m-length similar trajectories ranging from two to nine. Table 1 provides two measurements in accordance with the lengths of similar trajectories. For example, in the case of a two-length similar trajectory, 515,048 trajectories are identified, including 2,438 different types of trajectories. Overall, when the length of a similar trajectory is longer, less similar trajectories are identified. Based on the nature of complex destination networks, the results reveal a variety of categories of similar trajectory. The following section discusses the findings of the graph-based spatiotemporal analytics for semantic understanding of travel behaviors according to different M-lengths (i.e., vertices in graph theory).

[Please insert Table 2 about here]

4.5.1 Two-length similar trajectory

The results of the two-length similar trajectory are presented in Figure 6a. The color changing from light gray to Mandarin black indicates the intensity level of the trajectories. Three key zones of high density are evident, namely, Jeonju Hanok Village, Deokjin Park (or Chonbuk National University), and the train station. For further details, Figure 6b presents the top 20% two-length similar trajectories, which appear at a popular area of Jeonju Hanok Village. On the basis of the semantic meaning, travelers tend to use a train to visit Jeonju City. When exploring the most frequent two-length similar trajectories, the associations represent the flow between a traditional restaurant and the entrance to Jeonju Hanok Village or between a traditional restaurant and a popular travel attraction (e.g., Jeondong Catholic Church).

[Please insert Figures 6a and 6b about here]

4.5.2 Three-length similar trajectory

Total 2,950 types of three-length similar trajectories are identified, not only representing identical orders of places visited but also taking similar travel time in the daily travel trajectory. Figure 7a presents the top three three-length similar trajectories. In particular, Figures 7b, 7c, and 7d show each selective trajectory from Figure 7a. Compared with the patterns in two-length similar trajectory, the movement patterns in three-length similar trajectory do not increase the number of places visited by travelers. Instead, they exhibit a circular (or loop) movement pattern. That is, three-length similar trajectory essentially presents loop patterns leaving the starting point and returning to the origin point. On the basis of local knowledge, most trajectories are observed at Jeonju Hanok Village along with a traditional restaurant, a traditional market, and an area that includes a number of accommodation facilities.

[Please insert Figures 7a, 7b, 7c, and 7d about here]

4.5.3 Four-length similar trajectory

There are total 2,349 types of four-length similar trajectory. From the features of the visited places (Figure 8), the patterns in four-length similar trajectory have an additional point at Jeonju train station compared with three-length similar trajectory. Moreover, four-length similar trajectory represents a travel chain model in which travelers visit places in sequence, similar to a chain, without returning (Lue et al., 1993). Three key patterns exhibit different directions of the chain model comprising Jeonju train station and Jeonju Hanok Village.

[Please insert Figures 8a, 8b, 8c, and 8d about here]

4.5.4 Five-length similar trajectory

The patterns of five-length similar trajectory apparently have two distinctive routes. A pattern similar to three-length similar trajectory starting from Jeonju train station is observed. Another route shows the travel movement starting from the western part of the city and stopping at its center (i.e., the area of Jeonju city hall). Essentially, five-length trajectory has circular and chain pattern models (Park et al., 2020): (1) arriving at the city by train and traveling to a key travel attraction (Hanok Village) and leaving the city by train, i.e., labeled as a day trip (Figure 9a); and

(2) departing from an accommodation and traveling to the city center, referred to as a stay-over trip that involves extensively exploring the city (Figure 9b).

[Please insert Figures 9a, 9b, and 9c about here]

4.5.5 Six-length similar trajectory

Two key patterns are observed in six-length similar trajectory (Figures 10a, 10b and 10c). Six-length similar trajectory appears as a chain model traveling extensively at a part of Hanok Village from north to south. Compared with five-length similar trajectory, six-length similar trajectory has an uplinked mutual dyad model wherein the mutual dyad (i.e., repeated visitations) is observed ahead of the chain model (Figure 10a) (Lew & McKercher, 2006).

[Please insert Figures 10a, 10b and 10c about here]

4.5.6 Seven-length similar trajectory

The results show three types of linear chain models in similar trajectories. In terms of semantic meaning, the patterns in seven-length similar trajectory involve travelers who like to eat authentic food (Figure 11b) and explore the city by visiting multiple places/attractions. The three key trajectories have different starting points, such as the central, northern, and western parts of Hanok Village. Interestingly, however, two other patterns (Figures 11c and 11d) reflect a centralized mutual dyad wherein the mutual dyad is located at the middle of the chain. This mutual dyad is observed at the area of Hanok Village. Travelers demonstrating the third pattern (Figure 11d) may exhibit high preference for experiencing traditional culture at a destination given that the trajectory includes Hanok Village at its middle and the Hanok living experience center at its end.

[Please insert Figures 11a, 11b, 11c, and 11d about here]

4.5.7 Eight-length similar trajectory

Travelers that present eight-length similar trajectories are likely to visit another key travel attraction, such as Deokjin Park/Chonbuk National University, apart from Hanok Village (Figure 12a). A couple of selective patterns (Figures 12b and 12c) represents a linear chain model. That is, travelers tend to stop over at popular places during the entire journey, leaving from a starting point (i.e., accommodation) to visit Deokjin Park/Chonbuk National University (northern part of the city) and finishing their trips by visiting Hanok Village.

[Please insert Figures 12a, 12b, and 12c about here]

4.5.8 Nine-length similar trajectory

Travelers in this category are likely to extensively explore Jeonju City. Instead of spending most of their time at key travel attractions, these travelers tend to visit a variety of places and search for diverse activities in the city. Their movement pattern presents an alternative version of the base camp pattern (Lue et al., 1993) and/or a radiating hub (Lew & McKercher, 2006) where people make several trips from the hub. This trajectory includes multiple patterns, such as point-to-point and circular loops.

[Please insert Figure 13 about here]

4.6 Trajectory Similarity-based Clustering analysis

An intriguing question is whether travelers from close countries show similar trajectory patterns, i.e., whether tourists from close origin countries visit the destination in a similar way. To answer this question, we perform the clustering analysis based on the score of trajectory similarity. Figure 14 presents the result of the heat map analysis, which assesses the pair-wise comparisons of similar trajectory matching among different nationalities. Notably, the clustering result shows the top 16 countries/regions, explaining 91% of the total sample.

[Please insert Figure 14 about here]

The results show that travelers from the same countries/continents do not necessarily demonstrate similar travel behavior. In particular, the trajectory similarity of Chinese travelers does not exhibit high correlation with those from the same country. Instead, their similarity patterns tend to be more correlated with travelers from Austria and Hong Kong. Similarly, the travel trajectories of travelers from the USA are more likely to correlate with those from Hong Kong and Holland than with individuals from their own country. Meanwhile, travelers from Japan present high correlation with those from Austria, Hong Kong, and France. These findings imply that travel trajectories are more likely to be quite dynamic and individual behaviors rather than a collective understanding (Miller, 2004)

Furthermore, HCA is conducted to identify segments with different similarity scores in accordance with various nationalities. The results consistently demonstrate that not all the near things are more related than distant things (Miller, 2004). In conclusion, six clusters are formed. Group 1 includes Austria; Group 2 includes Russia; Group 3 includes Hong Kong; Group 4 includes Malaysia, France, the USA, and China; Group 5 includes Singapore, Canada, Japan, and

Holland; and Group 6 includes Australia, the UK, Taiwan, Thailand, and Germany (see Figure 15). That is, travelers from the same countries/continents do not necessarily show the similar travel patterns of movement. Their behaviors are dynamic and individual, which suggest to consider it as important standard to cluster travelers.

[Please insert Figure 15 about here]

5. Discussion

Understanding travel mobility has elicited considerable attention among tourism scholars. The literature has extensively discussed movement patterns (Vu et al., 2018) and destination interactions from the perspective of travel flow (Xu et al., 2020). Along with the advancement of information technology, such as social media and mobile technologies, studies on travel mobility have remarkably evolved. Advanced technologies enable tourism researchers to obtain fine-grained and large sets of data, encouraging the use of innovative analytics and another class to explain the social phenomenon. This study, which analyzes mobile sensor big data, suggests the method for identifying similar travel trajectories and illustrates such trajectories using graph-based spatiotemporal analytics to gain insights from tourism big data analytics. Thus, this work provides important findings regarding trajectory similarity that consider the spatial and temporal dimensions in contrast with the analysis of movement patterns that focuses only on a sequence of spatial locations (e.g., Park et al., 2020). These insights, in recognizing travelers who follow similar movement patterns, should guide DMOs to execute data-driven (or flow-based) segmentation and suggest the foundation of destination recommendation systems, which are essential for smart tourism destinations.

In terms of theoretical implications, the existing tourism literature has substantially investigated the patterns of travel movements in a conceptual (Lew & McKercher, 2006; Lue et al., 1993) and empirical manner (Vu et al., 2018; Zhao et al., 2018). These existing studies have focused on discovering collective movement patterns. However, the current research suggests an approach to measuring the degree and types of similar trajectory patterns by comparing a pairwise trajectory of individual travelers. This approach contributes to the literature on market segmentation by proposing a data-driven clustering method. A number of studies have suggested criteria for classifying travelers such as demographic characteristics and travel features. On the basis of the assumption that travel locations imply the interests and preferences of travelers (Li et al., 2008), this study demonstrates a potential for considering travel trajectory as a type of clustering method based on the similarity of travel movement.

The extant tourism literature has largely discussed the sequential patterns of places visited by travelers (e.g., Mckercher & Lau, 2008; Park et al., 2020). However, the temporal dimension, which denotes the transition time of visiting travel attractions, is lacking when estimating travel mobility. The current research considers both spatial and temporal dimensions to identify key travel patterns and estimate pattern similarity. Because travelers use time differently while visiting places, considering time (or temporal) information is critical to better understanding travel behavior. The algorithm used in this research provides tourism researchers with a method of simultaneously reflecting spatial and temporal information.

Travel destination is a complex system; it is dynamic and not linear (Zahra & Ryan, 2007). In this regard, this study adopts graph theory as the foundation of network science, which facilitates the investigation of complexity. This research denotes the places visited by travelers as vertices and travel flows as directional edges. Trajectory similarity is developed in accordance with the different numbers of places visited (variations of vertices) and directions (directed networks). In addition to estimating mobility similarity, consistent with the rationale of network science, this study demonstrates the heavy-tailed distribution of a destination network, which suggests spatial heterogeneity (Jiang, 2018). This implies that a few areas (the coverage of cellphone towers) attract a large number of visits while many areas attract few travelers. As a result, the application of graph-based spatiotemporal analytics sheds light on the complex destination system in general and apprehends similar trajectories according to different vertices in particular.

Most tourism studies have emphasized the first law of geography, called Tobler's law, which states that "everything is related to everything else, but near things are more related than distant things" (Taleb, 2007). A number of tourism scholars have demonstrated the suitability of this law in explaining the association between travel distance and demands, particularly distance decay (Mckercher & Lew, 2003; Zhang, Li, Muskat, & Law, 2020). Importantly, however, the current study suggests that Tobler's law is not necessarily applicable to understanding travel movement patterns by presenting variant similarity scores within nationalities denoting travel distance. In the consistent vein, Miller (2004) suggests that the advancement of transportation and information technology make the world much smaller and more accessible and have an influence on the understanding of spatial nearness. Thus, the application of a complex adaptive systems theory that emphasizes the importance of local interactions among entities is proposed to explain complex global behavior that is not completely predictable or controllable.

With regard to practical implications, DMOs can obtain immense benefits from developing destination management activities from the perspectives of destination marketing and planning. The approach to measuring trajectory similarity can be applied to creating destination recommendation systems. This finding supports the notion of micro-marketing in tourism destinations which aims to provide a better touristic experience through highly personalized and customized services (Fesenmaier et al., 2016). Given that travelers who exhibit similar patterns have similar preferences, trajectory similarity can be considered a data-driven clustering method. With the real-time monitoring system, DMOs can provide customized information and services to individuals who demonstrate high similarity in their travel behaviors. Tourism companies having larger properties (e.g., integrated resorts or theme parks) could adopt a similar marketing strategy (e.g., location-based advertisement) using diverse data sources (e.g., Wi-Fi data, mobile app log data). This strategy is fundamental to creating smart destinations that manage tourism systems on the basis of big data analytics and IT (Gretzel et al., 2015).

Destination planning (e.g., transportation, place design, crowdedness control) could be another important area in which to apply the dynamic algorithm for calculating movement similarity and identifying travelers who exhibit similar patterns. For example, DMOs can create new tour products and transportation systems, such as tour buses. In particular, in accordance with the notion that graph theory considers the numbers of places visited (i.e., vertices), directions (i.e., edges), number of tourist flows (i.e., weighted edges/flows), and temporal information (i.e., length of time of moving and staying), this study suggests the number of tour buses and their routes and optimizes the time intervals of the bus services to be provided. Additionally, such information could help DMOs and tourism enterprises identify several necessary locations that could be used for visitor centers, bulletin boards, or any necessary tourist-support facilities, thereby improving the quality of tourism offerings at the destination.

Last but not least, considering the current situation in which the tourism industry has suffered substantially due to the COVID-19 pandemic, trajectory similarity is particularly useful for DMOs in making decisions to control the crowdedness within destinations to effectively control the spread of the disease. DMOs can identify places where travelers are likely to visit, which movement patterns are demonstrated, and who is presenting similar patterns; thus, they can plan a proactive strategy for preventing the spread of infection in the future, including tracking possibly contagious tourists, targeted cleaning and disinfecting, and announcing urgent information. This approach demonstrates the advantages of smart tourism destinations in managing the safety of travelers and the residents at a destination.

While this research has made important contributions, there are several limitations. First, considering the nature of mobile data, it is difficult to identify the exact locations that travelers visited and the activities they engaged in at those places. Thus, the researchers suggest an approach to data integration that combines information about behavioral and psychological (perceptions) aspects so that comprehensive travel behaviors and experiences can be identified. Second, the data analyzed in this study reflects travel mobility in a certain city. In order to test the generalizability of the findings and suitability of the pattern recognition algorithm, it is suggested that future researchers assess the travel mobility patterns in various contexts of destinations and diverse nationalities.

References

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

Bergroth, L., Hakonen, H., & Raita, T. (2000). *A survey of longest common subsequence algorithms.* Paper presented at the Proceedings Seventh International Symposium on String Processing and Information Retrieval. SPIRE 2000.

Bifulco, G. N., Cartenì, A., & Papola, A. (2010). An activity-based approach for complex travel behaviour modelling. *European Transport Research Review*, *2*(4), 209-221

Cleasby, I. R., Wakefield, E. D., Morrissey, B. J., Bodey, T. W., Votier, S. C., Bearhop, S., & Hamer, K. C. (2019). Using time-series similarity measures to compare animal movement trajectories in ecology. *Behavioral Ecology and Sociobiology*, 73(11), 151.

Cormen, T. H. (2013). *Algorithms Unlocked* (The MIT Press). Cambridge: MIT Press.

Dale, M. R. T., & Fortin, M. J. (2010). From Graphs to Spatial Graphs. *Annual Review of Ecology, Evolution, and Systematics, 41*(1), 21-38. doi:10.1146/annurev-ecolsys-102209-144718

Derrible, S., & Kennedy, C. (2011). Applications of graph theory and network science to transit network design. *Transport reviews, 31*(4), 495-519.

D'Urso, P., De Giovanni, L., Disegna, M., & Massari, R. (2013). Bagged Clustering and its application to tourism market segmentation. *Expert Systems with Applications*, 40(12), 4944-4956.

Fesenmaier, D. R., Bharadwaj, N., Stienmetz, J. L., & Xaing, Z. (2016). Micro-marketing and Big Data analytics: An information system for destination marketing management. *Management Science in Hospitality and Tourism: Theory, Practice, and Applications*, 67.

Gao, Y., Ye, C., Zhong, X., Wu, L., & Liu, Y. (2019). Extracting Spatial Patterns of Intercity Tourist Movements from Online Travel Blogs. *Sustainability*, *11*(13), 3526.

Girardin, F., Calabrese, F., Dal Fiore, F., Ratti, C., & Blat, J. (2008). Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive computing, 7*(4), 36-43.

Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: foundations and developments. *Electronic Markets, 25*(3), 179-188. doi:10.1007/s12525-015-0196-8

Hastie, Trevor, Tibshirani, Robert, & Friedman, Jerome. (2009). *Elements of Statistical Learning* (Springer series in statistics). New York: Springer.

Hayes, B. (2000). Computing Science: Graph Theory in Practice: Part II. *American Scientist, 88*(2), 104-109.

Hu, F., Li, Z., Yang, C., & Jiang, Y. (2019). A graph-based approach to detecting tourist movement patterns using social media data. *Cartography and Geographic Information Science, 46*(4), 368-382.

Hwang, Y. H., Gretzel, U., & Fesenmaier, D. R. (2006). Multicity trip patterns. *Annals of Tourism Research, 33*(4), 1057-1078. doi:10.1016/j.annals.2006.04.004

Jiang, B. (2018). Geospatial analysis requires a different way of thinking: The problem of spatial heterogeneity. *Trends in Spatial Analysis and Modelling*, 23-40.

Jongno City Office (2019). Jongno Tourism Master Plan. Accessed from http://www.prism.go.kr/homepage/researchCommon/downloadResearchAttachFile.do;jsessionid=302CA26C9BE8A40153924F51A12A0816.node02?work_key=001&file_type=CPR&seq_no=001&pdf_conv_yn=N&research_id=3000000- 201900004.

Kim, J., & Mahmassani, H. S. (2015). Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories. *Transportation Research Procedia, 9*, 164-184.

Leiper, N. (1979). The framework of tourism: Towards a definition of tourism, tourist, and the tourist industry. *Annals of Tourism Research, 6*(4), 390-407.

Lew, A., & McKercher, B. (2006). Modeling tourist movements: A local destination analysis. *Annals of tourism research, 33*(2), 403-423.

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management, 68*, 301-323.

Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W., & Ma, W. Y. (2008). *Mining user similarity based on location history.* Paper presented at the Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems.

Lue, C. C., Crompton, J. L., & Fesenmaier, D. R. (1993). Conceptualization of multi-destination pleasure trips. *Annals of tourism research, 20*(2), 289-301.

McKercher, B. (1999). A chaos approach to tourism. *Tourism management, 20*(4), 425-434.

Mckercher, B., & Lau, G. (2008). Movement patterns of tourists within a destination. *Tourism geographies, 10*(3), 355-374.

Mckercher, B., & Lew, A. A. (2003). Distance decay and the impact of effective tourism exclusion zones on international travel flows. *Journal of Travel research, 42*(2), 159-165.

Miguéns, J., & Mendes, J. (2008). Travel and tourism: Into a complex network. *Physica A: Statistical Mechanics and its Applications, 387*(12), 2963-2971.

Miller, H. J. (2004). Tobler's first law and spatial analysis. *Annals of the Association of American Geographers, 94*(2), 284-289.

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., Barabási, A. L. E., & Watts, D. J. (2006). *The structure and dynamics of networks*: Princeton university press.

Park, S., Xu, Y., Jiang, L., Chen, Z., & Huang, S. (2020). Spatial structures of tourism destinations: A trajectory data mining approach leveraging mobile big data. *Annals of tourism research, 84*, 102973.

Phillips, J. D., Schwanghart, W., & Heckmann, T. (2015). Graph theory in the geosciences. *Earth-Science Reviews, 143*, 147-160. doi:10.1016/j.earscirev.2015.02.002

Ranacher, P., & Tzavella, K. (2014). How to compare movement? A review of physical movement similarity measures in geographic information science and beyond. *Cartography and geographic information science, 41*(3), 286-307.

Raun, J., Ahas, R., & Tiru, M. (2016). Measuring tourism destinations using mobile tracking data. *Tourism Management, 57*, 202-212.

Riaz, F., & Ali, K. M. (2011). *Applications of graph theory in computer science.* Paper presented at the 2011 Third International Conference on Computational Intelligence, Communication Systems and Networks.

Rong, J., Vu, H. Q., Law, R., & Li, G. (2012). A behavioral analysis of web sharers and browsers in Hong Kong using targeted association rule mining. *Tourism management, 33*(4), 731-740.

Shoval, N., & Isaacson, M. (2007). Tracking tourists in the digital age. *Annals of Tourism Research, 34*(1), 141-159. doi:10.1016/j.annals.2006.07.007

Shoval, N., McKercher, B., Ng, E., & Birenboim, A. (2011). Hotel location and tourist activity in cities. *Annals of tourism research, 38*(4), 1594-1612.
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.
Stienmetz, J. L., Kim, J., Xiang, Z., & Fesenmaier, D. R. (2020). Managing the structure of tourism experiences: Foundations for tourism design. *Journal of Destination Marketing & Management*. doi:10.1016/j.jdmm.2019.100408
Strogatz, S. H. (2001). Exploring complex networks. *nature, 410*(6825), 268-276.
Taleb, N. N. (2007). *The black swan: The impact of the highly improbable* (Vol. 2): Random house.
Toohey, K., & Duckham, M. (2015). Trajectory similarity measures. *Sigspatial Special, 7*(1), 43-50.
Vlachos, M., Kollios, G., & Gunopulos, D. (2005). Elastic translation invariant matching of trajectories. *Machine learning, 58*(2-3), 301-334.
Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism management, 46*, 222-232.
Vu, H. Q., Li, G., Law, R., & Zhang, Y. (2018). Travel diaries analysis by sequential rule mining. *Journal of Travel research, 57*(3), 399-413.
Vu, H. Q., Li, G., Law, R., & Zhang, Y. (2019). Exploring tourist dining preferences based on restaurant reviews. *Journal of Travel research, 58*(1), 149-167.
Wang, X. F., & Chen, G. (2003). Complex networks: small-world, scale-free and beyond. *IEEE circuits and systems magazine, 3*(1), 6-20.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature, 393*(6684), 440-442.
Wong, E., Law, R., & Li, G. (2017). Reviewing geotagging research in tourism. In *Information and Communication Technologies in Tourism 2017* (pp. 43-58): Springer.
Xia, Y., Wang, G. Y., Zhang, X., Kim, G. B., & Bae, H. Y. (2011). Spatio-temporal similarity measure for network constrained trajectory data. *International Journal of Computational Intelligence Systems, 4*(5), 1070-1079.
Huang, X., & Wu, B. (2012). Intra-attraction tourist spatial-temporal behaviour patterns. *Tourism geographies, 14*(4), 625-645.
Xu, Y., Li, J., Belyi, A., & Park, S. (2020). Characterizing destination networks through mobility traces of international tourists—A case study using a nationwide mobile positioning dataset. *Tourism management, 82*, 104195.
Xu, Y., Li, J., Xue, J., Park, S., & Li, Q. (2020). Tourism Geography through the Lens of Time Use: A Computational Framework Using Fine-Grained Mobile Phone Data. *Annals of the American Association of Geographers*, 1-25.
Yuan, Y., & Raubal, M. (2014). Measuring similarity of mobile phone user trajectories–a Spatio-temporal Edit Distance method. *International Journal of Geographical Information Science, 28*(3), 496-520.
Zach, F., & Gretzel, U. (2011). Tourist-Activated Networks: Implications for Dynamic Bundling and EN Route Recommendations. *Information Technology & Tourism, 13*(3), 229-238. doi:10.3727/109830512x13283928066959
Zahra, A., & Ryan, C. (2007). From chaos to cohesion—Complexity in tourism structures: An analysis of New Zealand's regional tourism organizations. *Tourism management, 28*(3), 854-862.

Zhang, Y., Li, G., Muskat, B., Law, R., & Yang, Y. (2020). Group pooling for deep tourism demand forecasting. *Annals of tourism research, 82*, 102899.

Zhang, Y., Li, X. R., & Su, Q. (2017). Does spatial layout matter to theme park tourism carrying capacity? *Tourism Management, 61*, 82-95.

Zhao, L., Yuan, S. S., Peng, S., & Wang, L. T. (2002). *A new efficient data cleansing method.* Paper presented at the International Conference on Database and Expert Systems Applications.

Zhao, X., Lu, X., Liu, Y., Lin, J., & An, J. (2018). Tourist movement patterns understanding from the perspective of travel party size using mobile tracking data: A case study of Xi'an, China. *Tourism management, 69*, 368-383.

Zheng, W., Huang, X., & Li, Y. (2017). Understanding the tourist mobility using GPS: Where is the next place? *Tourism management, 59*, 267-280.

Table 1. Example of a user's mobile data (User ID 1)

ID	Date	Longitude	Latitude	Arrival time	Leaving time
1	20171125	127.***	35.***	15:43:00	15:53:00
1	20171125	127.***	36.***	17:06:00	17:17:00
1	20171125	128.***	35.***	18:09:00	19:01:00
1	20171125	127.***	35.***	19:07:00	19:34:00

Table 2. The total number of m-length similar trajectory

The length of similar trajectory (m)	Total number of categories of m-length similar trajectory	Total number of m-length similar trajectory
2-length similar trajectory	2,438	515,048
3-length similar trajectory	2,950	56,962
4-length similar trajectory	2,349	7,785
5-length similar trajectory	988	1,426
6-length similar trajectory	366	417
7-length similar trajectory	180	187
8-length similar trajectory	95	97
9-length similar trajectory	86	87

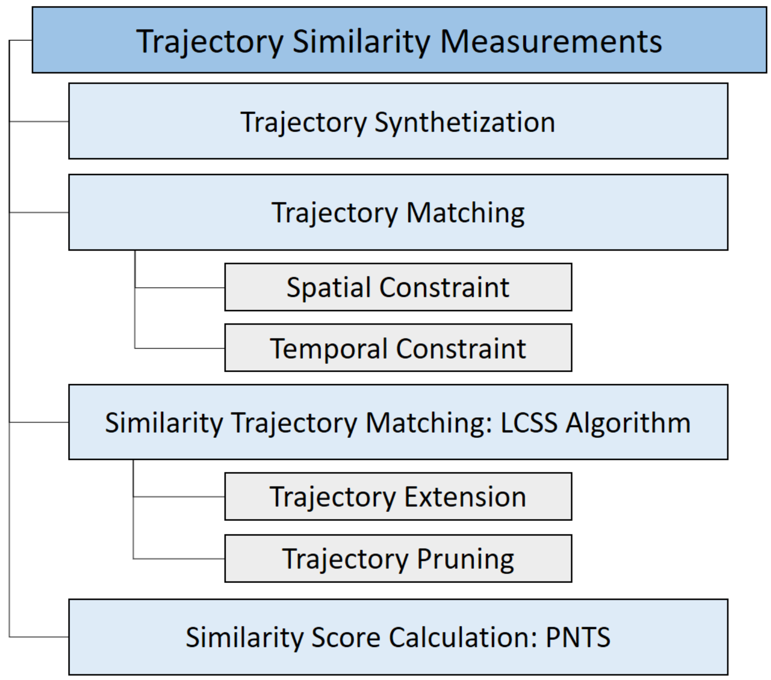

Figure 1. Flow chart of data analysis

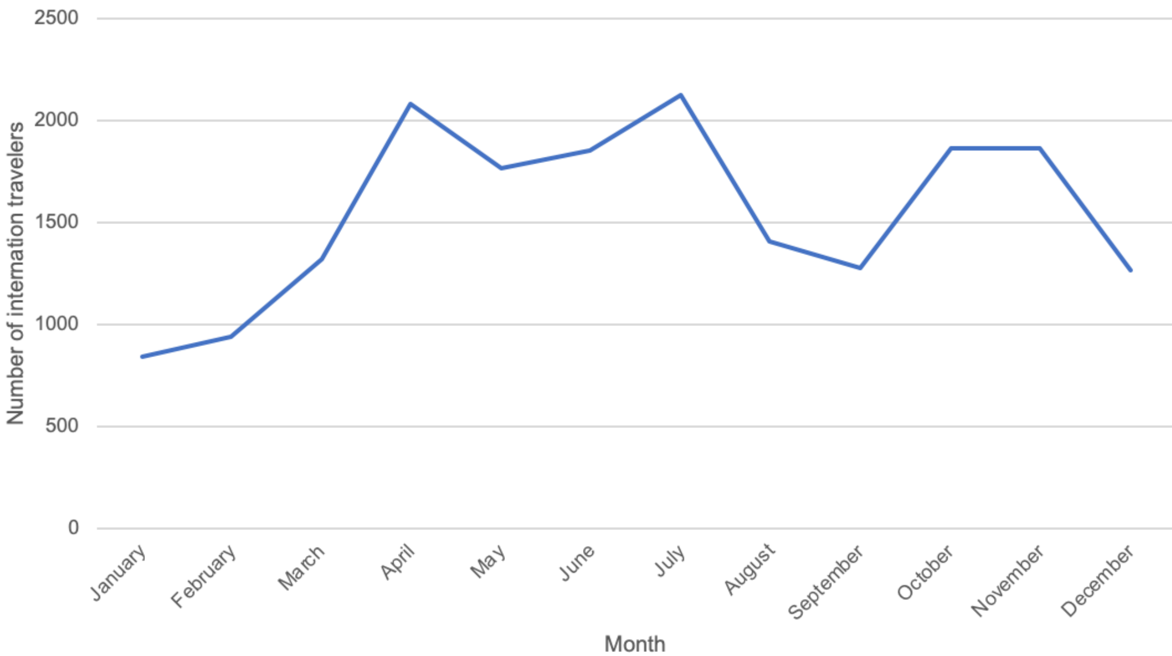

Figure 2. The distribution of the number of International travelers in a year

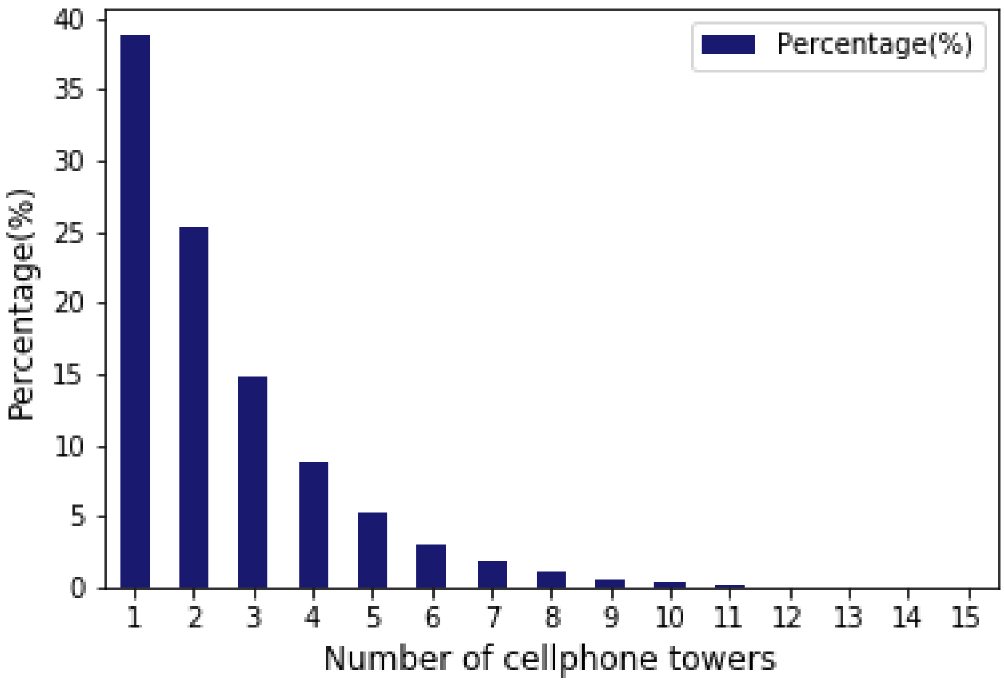

Figure 3. The number of cellphone towers travelers visited in daily trajectory

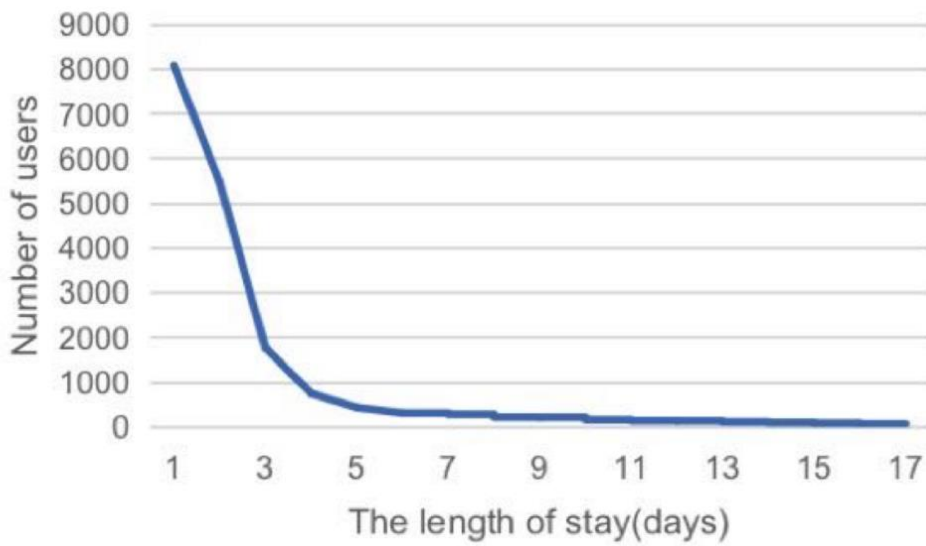

Figure 4. International travelers' length of stay in Jeonju

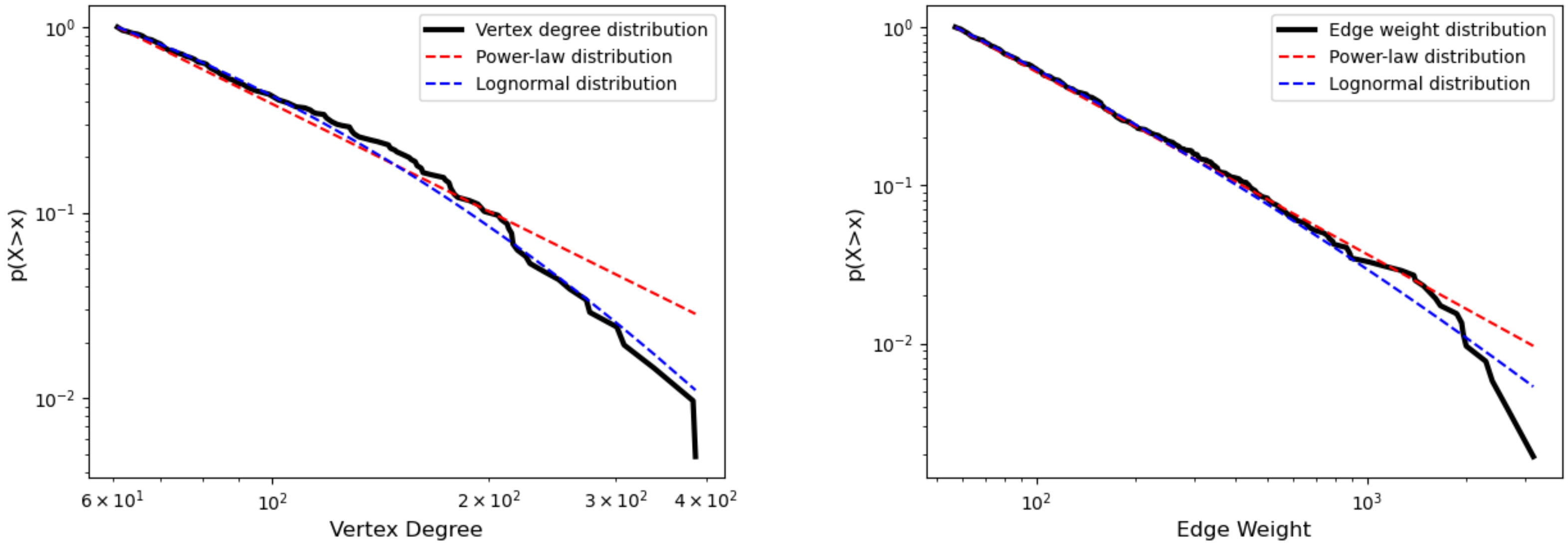

Figure 5. Cumulative distribution function of vertex degree (A) and weighted edge (B)

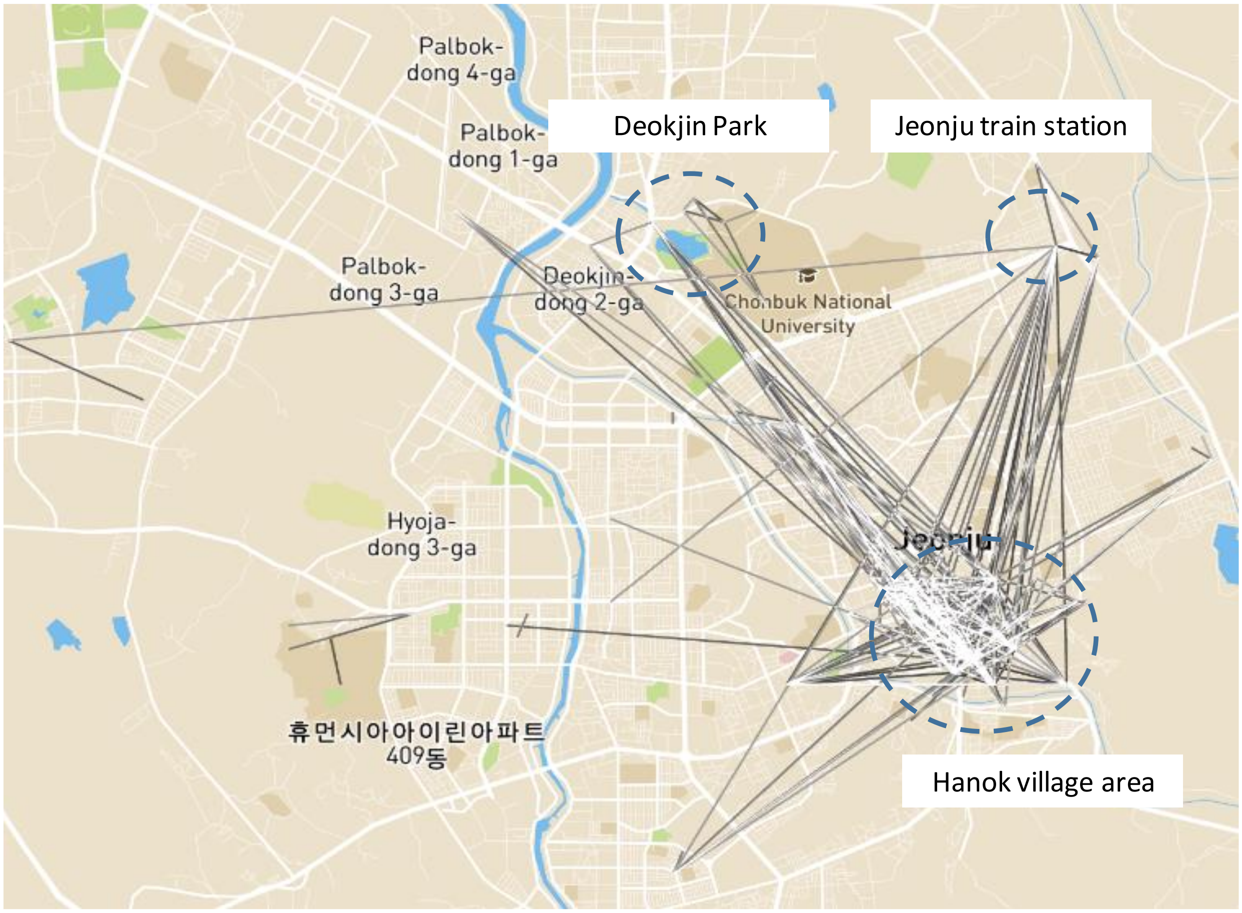

Figure 6a. Results of 2-length similar trajectory

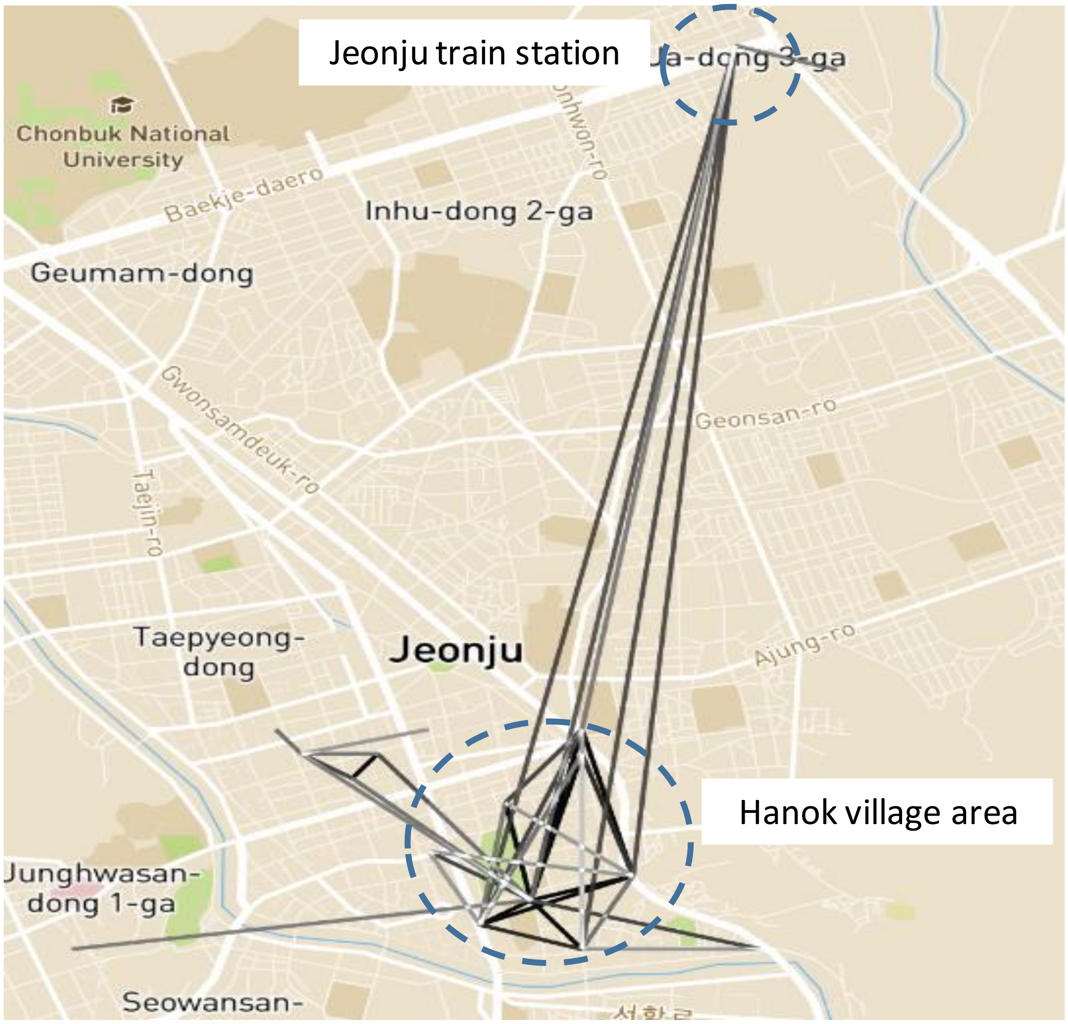

Figure 6b. Results of top 20% of 2-length similar trajectory

Figure 7a. Results of 3-length similar trajectory

Figure 7b. Type A of 3-length similar trajectory

Figure 7c. Type B of 3-length similar trajectory

Figure 7d. Type C of 3-length similar trajectory

Figure 8a. Results of 4-length similar trajectory

Figure 8b. Type A of 4-length similar trajectory

Figure 8c. Type B of 4-length similar trajectory

Figure 8d. Type C of 4-length similar trajectory

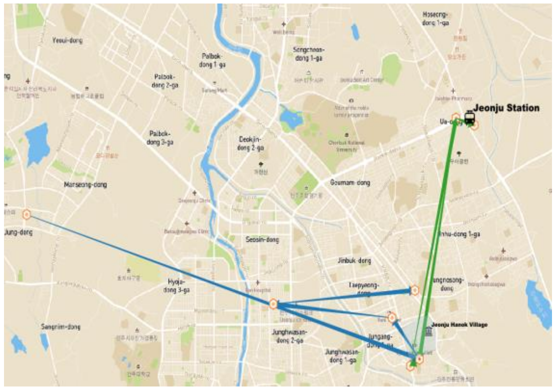

Figure 9a. Results of 5-length similar trajectory

Figure 9b. Type A of 5-length similar trajectory

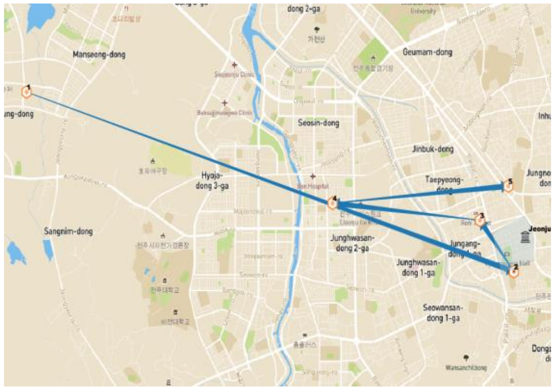

Figure 9c. Type B of 5-length similar trajectory

Figure 10a. Results of 6-length similar trajectory

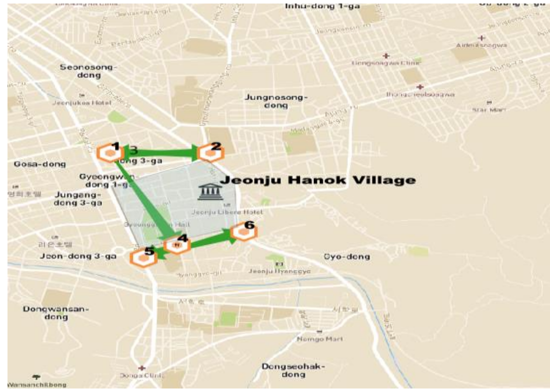

Figure 10b. Type A of 6-length similar trajectory

Figure 10c. Type B of 6-length similar trajectory

Figure 11a. Results of 7-length similar trajectory

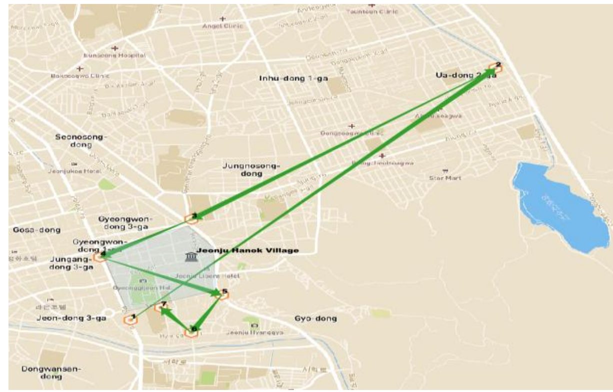

Figure 11b. Type A of 7-length similar trajectory

Figure 11c. Type B of 7-length similar trajectory

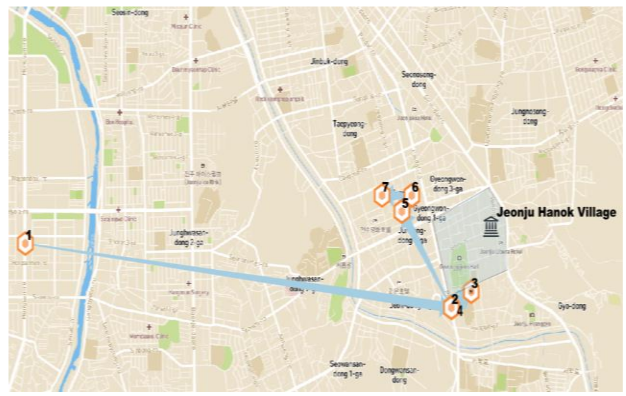

Figure 11d. Type C of 7-length similar trajectory

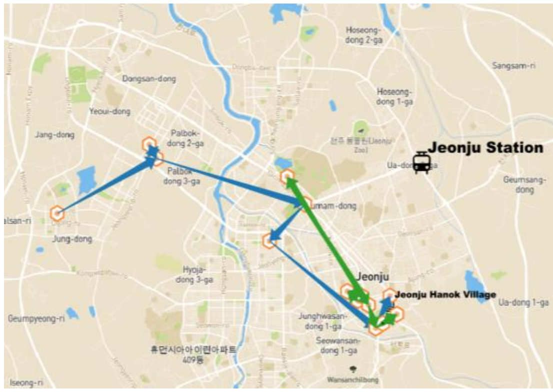

Figure 12a. Results of 8-length similar trajectory

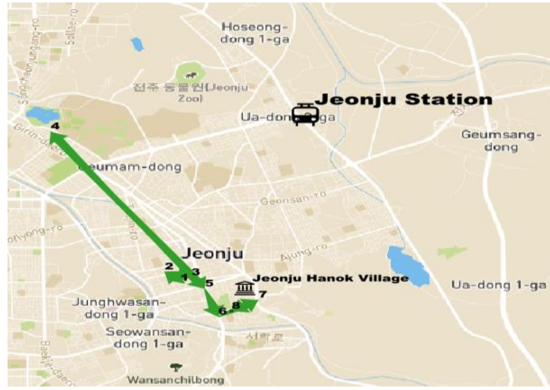

Figure 12b. Type A of 8-length similar trajectory

Figure 12c. Type B of 8-length similar trajectory

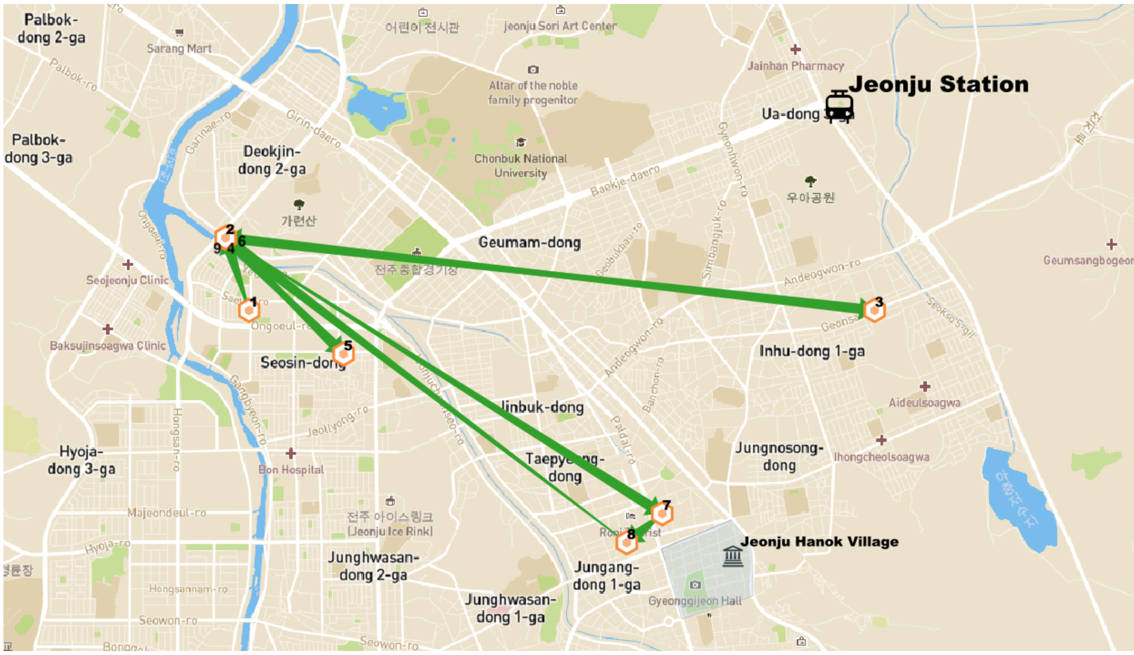

Figure 13. Results of 9-length similar trajectory

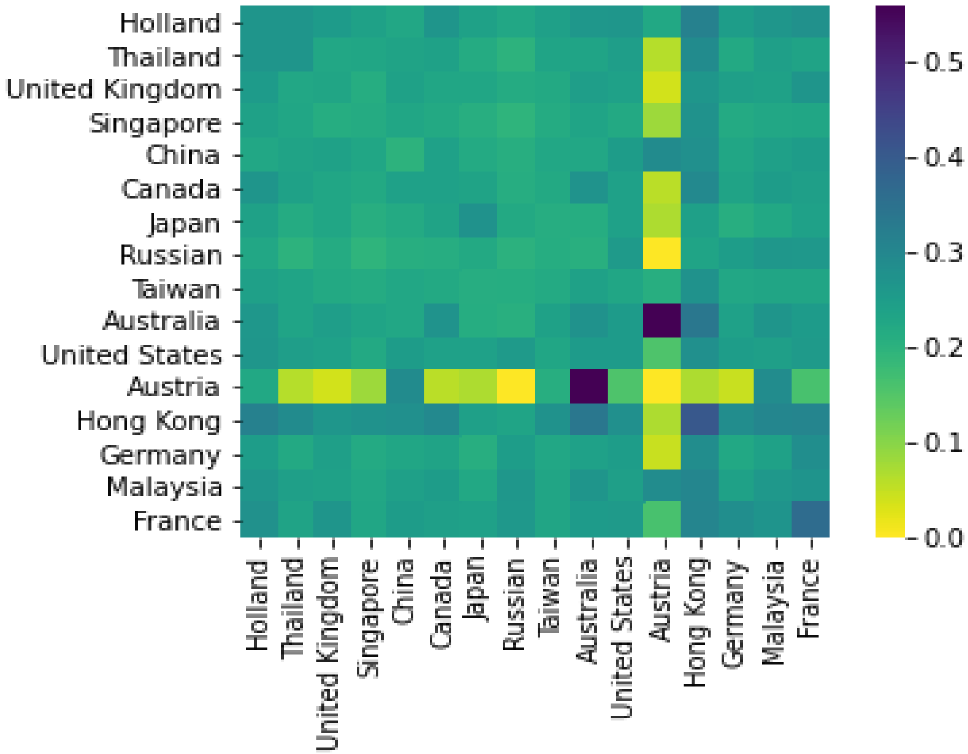

Figure 14. Heatmap analysis of trajectory similarity matching between a pair of nationality

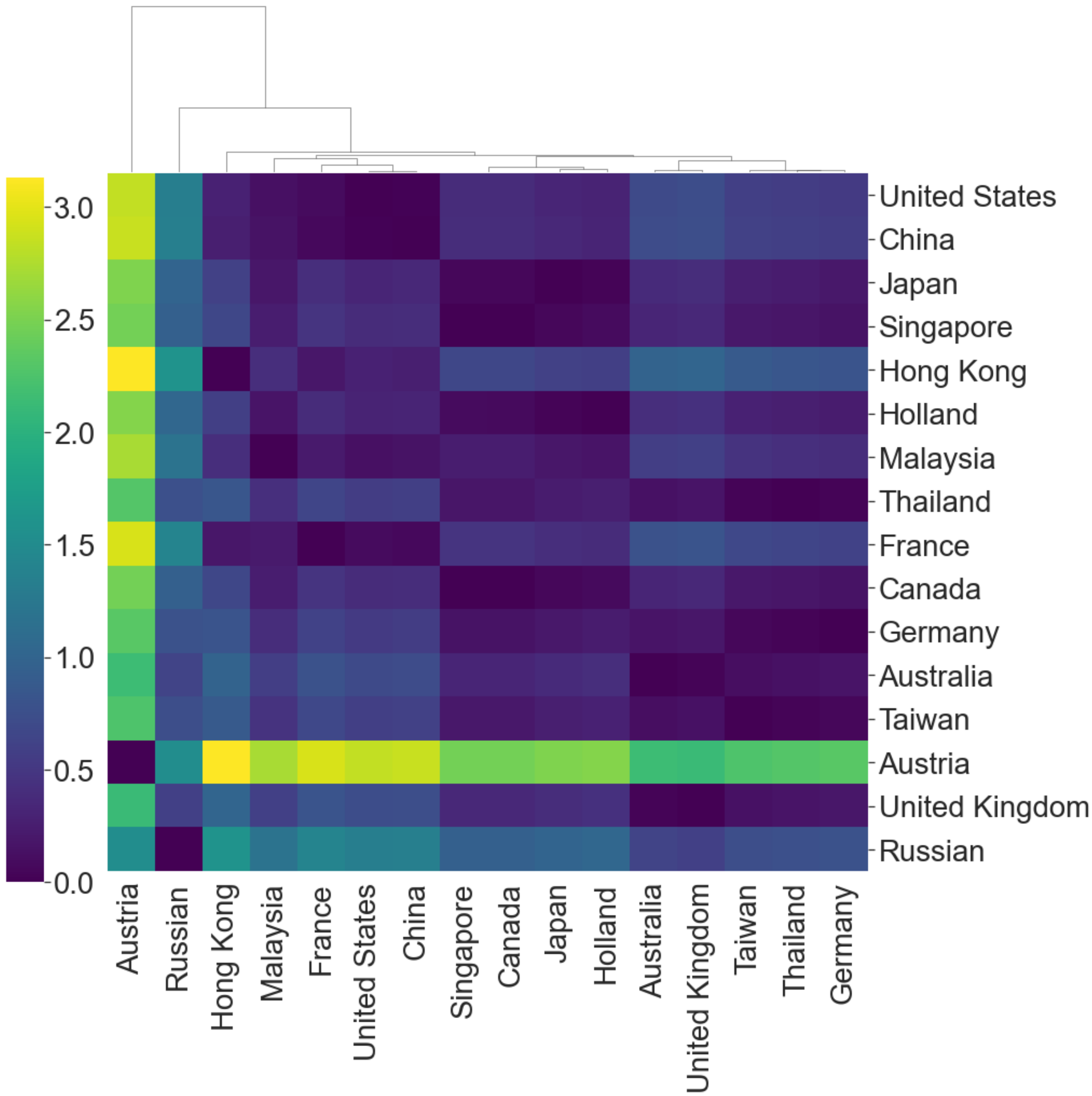

Figure 15. Clustering result of trajectory similarity matching by country