Tourism geography through the lens of time use — A computational framework using fine-grained mobile phone data

Yang Xu*[1], Jingyan Li[1], Jiaying Xue[1], Sangwon Park[2], and Qingquan Li[3]

[1]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University
[2]School of Hotel and Tourism Management, The Hong Kong Polytechnic University
[3]Guangdong Key Laboratory of Urban Informatics, School of Architecture and Urban Planning, Shenzhen University

Abstract

Location-aware technologies and big data are transforming the ways we capture and analyze human activities. This has particularly impacted tourism geography, which aims to study tourist activities within the context of space and places. In this study, we argue that the tourism geography of cities can be better understood through the time use of tourists captured by fine-grained human mobility observations. By using a large-scale mobile phone dataset collected in three cities in South Korea (Gangneung, Jeonju and Chuncheon), we develop a computational framework to enable accurate quantification of tourist time use, the visualization of their spatio-temporal activity patterns, and systematic comparisons across cities. The framework consists of several approaches for the extraction and semantic labelling of tourist activities, visual-analytic tools (time use diagram, time-activity diagram) for examining their time use, as well as quantitative measures that facilitate day-to-day comparisons. The feasibility of the framework is demonstrated by performing a comparative analysis in three cities during representative days, when tourists tended to show more regular patterns. The framework is also employed to examine tourist time use under special events, using Gangneung during the 2018 Winter Olympics (WO) as an example. The findings are validated by comparing the spatio-temporal patterns with the calendar events of WO. The study provides a new perspective that connects time geography and tourism through the usage of spatio-temporal big data. The computational framework can be applied to compatible datasets to advance time geography, tourism, and urban mobility research.

Key words: time use; human mobility; time geography; tourism geography; mobile phone data

1 Introduction

Geography and tourism are naturally connected. The inherent linkage between the two has forged a research area that aims to study and understand tourist activities within the context of space and places. The field of tourism geography, in a global and interconnected world, is becoming more important given the enhanced mobilities of people across locations, regions and countries. According to the *The Travel & Tourism Competitiveness*

*Electronic address: yang.ls.xu@polyu.edu.hk

Report 2019 (Calderwood and Soshkin 2019) by the World Economic Forum, the number of international arrivals worldwide has reached 1.4 billion in 2018, and is predicted to reach 1.8 billion by 2030 (pp.7). The increasing connections between travelers and tourism destinations, as foreseen by Franklin and Crang (2001), will generate "new ways of living, new ties of space, (and) new places" (pp.8) that inspire future research of tourism geography.

Through the years, numerous discussions have taken place to define, review and anticipate the development of the field (Williams 1998; Gibson 2008; Hall 2013; Hall and Page 2014). Despite the diverse views from the scholars, the key questions concern the *where* (do tourism activities develop), *who* (are the tourists), *how* (is tourism developed) and *what* (is the impact upon the physical, cultural and socioeconomic systems) aspects of tourism (Williams 1998). To tackle these questions, geography has played a key role, by providing observations, theories, methodologies, and tools to facilitate the understanding of tourism activities in the context of space and time. Although the above questions continue to be central to tourism geography, there is no doubt that the field is rapidly evolving. This is largely due to the emergence of disruptive technologies — such as location-aware technologies and big data — that transform the ways tourists travel, and the ways we capture and analyze them (Birenboim and Shoval 2016; Shoval and Ahas 2016; Li et al. 2018).

One example is the impact of tracking technologies. As claimed by Shoval and Ahas (2016), increasing studies are benefiting from the use of technologies such as Global Positioning System (GPS), mobile positioning, Bluetooth tracking and geocoded social media. The adoption of these tracking technologies has produced a variety of tourism big data. Such data have been used to study important research topics, such as the spatio-temporal behavioral patterns of tourists (Tchetchik et al. 2009; Shoval et al. 2011; Xiao-Ting and Bi-Hu 2012; Edwards and Griffin 2013; Bauder and Freytag 2015; De Cantis et al. 2016; Zheng et al. 2012; Wood et al. 2013; Vu et al. 2015; Kim et al. 2019; Ahas et al. 2007), tourism activities during special events (Pettersson and Zillinger 2011; De Cantis et al. 2016), tourist route choice and planning (Lu et al. 2010; East et al. 2017), behavior of repeated visitors (Tiru et al. 2010; Kuusik et al. 2011; McKercher et al. 2012), among others.

Despite these fruitful research outcomes, one aspect of tourist activities that has been under-studied is their time use, which is an important topic tied to the theory of time geography (Hägerstrand 1970). To date, limited research has been conducted to link time geography concepts with tourist travel behavior (Pultar and Raubal 2010). As tourism is often considered as an escape from the daily routines (Hall 2005a;b), it usually involves additional efforts and expenditure of travelers. Their time allocation during travels will be affected by their financial capacity, time budget and other constraints. It is therefore important to obtain an improved understanding of how tourists allocate their time across space and places. Such efforts will generate new knowledge of tourism geography, new hypotheses of tourist activities, and valuable information that is beneficial to the development of tourism destinations.

As mentioned above, studies on the time use of tourists are scarce. One study by Birenboim et al. (2013) uses GPS data to analyze the time use of visitors at a theme park in Catalonia, Spain. The sample covers a few hundreds of individuals. Similar studies that leverage other tracking technologies (e.g., geocoded social media, positioning) are almost absent. There are a few reasons that hinder the adoption of these tracking technologies for detailed time use analysis. GPS data are able to capture travelers' movements at a fine spatio-temporal resolution. However, as data collection is costly and time-consuming, most of the GPS-based studies aim at small geographic scales and limited sample sizes (Tchetchik et al. 2009; Shoval et al. 2011; Pettersson and Zillinger 2011; Birenboim et al. 2013; De Cantis et al. 2016). Social media data enable tourism analysis at broader scales, and they have clear advantages for cross-site comparisons and even global-level analysis (Wood et al. 2013; Keeler et al. 2015; Belyi et al. 2017). The sample sizes are also much larger. However, individual observations can be very sparse (Lo et al. 2011) and certain demographic tiers tend to be underrepresented (Salas-Olmedo et al. 2018). Such issues make social media data unfitted for accurate quantification of time use. Mobile tracking technologies and the data produced are suitable for large-scale mobility analysis. For instance, one typical type of mobile phone data — namely call detail records (CDRs) — has been widely used to study human mobility patterns from different perspectives (Gonzalez et al. 2008; Song et al. 2010; Becker et al. 2013; Silm and Ahas 2014; Xu et al. 2015; Yuan and Raubal 2016; Xu et al. 2017; 2018; 2019). However, CDRs are regarded as "passive mobile phone data" (Blondel et al. 2015; Chen et al. 2016) as locations of phone users are passively generated during certain types of phone usage activities (e.g., call and text message). In other words, observations are still sampled irregularly over time, and data points can be very sparse for inactive phone users. Therefore, CDRs could generate biased estimates of time use characteristics.

But are we facing a dead end here? Probably not. In recent years, our capabilities to capture detailed observations on human activities have been further augmented. For instance, many studies have adopted mobile signaling data (MSD) for human mobility analysis (Li et al. 2019; Yan et al. 2019). Different from CDRs, MSD could track phone users' locations in a more continuous manner through various signaling events (e.g., data usage, cellular handover, periodic update) triggered by the telecommunication systems (Janecek et al. 2015). This has greatly improved the data granularity, especially from the temporal perspective. Moreover, it is very likely that other types of datasets (e.g., Wi-Fi data, CCTV) will become more fine-grained in their spatio-temporal resolutions, and more accessible due to the emergence of new technologies such as 5G and Internet of Things (IoT).

Inspired by these technological advancements, we ask the following questions — How can we better understand tourist time use in cities as mobility observations of travelers become both massive and granular? What new knowledge will this bring to the geography of tourism? To answer these questions, we analyze a large-scale mobile phone dataset collected in three cities in South Korea (Gangneung, Jeonju and Chuncheon) during a period of one year. The dataset provides a fine-grained view of tourist mobility by capturing travelers' stay (dwelling) activities at a high spatial resolution. A computational framework is developed to enable accurate quantification of tourist time use, the visualization of their spatio-temporal activity patterns, and systematic comparisons across cities. The major contributions of this research are as follows:

- We first develop an anchor point extraction approach that converts each individual's cellphone trajectory into a sequence of stays at meaningful activity locations. A semantic labelling process is then introduced to annotate each individuals' stays or movements into different activity categories. The semantics derived in this step can indicate travelers dwelling places in a city, when they tended to travel, where their activities occurred, and when they were outside of a city. The labelling process provides an effective solution to enriching the semantics of travelers' activities when contextual information is absent in big data (e.g., mobile phone data).
- We introduce an approach to derive individuals' daily time use diagram, which de-

scribes a traveler's time allocation across different types of activities in a day. These individual time use diagrams are further aggregated by each observation day to form the daily time use diagrams of a city. A workflow based on: (1) a coarse segmentation of individuals based on travel styles, and (2) a hierarchical clustering algorithm based on a time-use similarity metrics is introduced to determine how individuals are organized in the aggregated diagrams. The resulting diagrams could convey meaningful patterns that reveal the heterogeneity of tourist time use as well as their collective dynamics that shape the signature of a city.

- A day-to-day similarity measure is introduced to identify the so-called representative days in a city when tourists tended to show more regular patterns, compared to days when they exhibited unique characteristics.

- We apply the computational framework over cellphone trajectories collected in Gangneung, Jeonju and Chuncheon. A comparative analysis is performed to understand tourist time use in the three cities during representative days. We also introduce another visualization layout — the time-activity diagram — along with spatial mapping of tourist activities, to complement the findings derived from the time use diagrams.

- The framework is also employed to examine tourist time use under special events. We use Gangneung during the 2018 Winter Olympics (WO) as an example, and demonstrate the capability of the framework in capturing the transitions of tourist activity patterns before, during and after this mega event. A close examination of tourist activities on Feb 25, 2018 (last day of 2018 WO) is performed and the results are validated by comparing the spatio-temporal patterns with the calendar events of WO.

We believe the study provides a new perspective that connects time geography and tourism through the usage of spatio-temporal big data. The computational framework enables a new way for analyzing large-scale and fine-grained human mobility observations. The framework can be leveraged to generate meaningful insights of travelers, which are beneficial to spatial planning of cities, management of tourism destinations, and may inspire new hypotheses of tourist travel behaviors.

2 Study Area and Dataset

2.1 Mobile phone dataset

This study uses a mobile phone dataset collected in South Korea. The anonymized dataset tracks the location footprints of 116,807 international travelers[1] who visited any of the three following cities — Gangneung, Jeonju and Chuncheon — between August 1st, 2017 and July 31st, 2018. The one-year dataset was collected by one of the major cellular operators in South Korea. It is part of a tourism big data project with the Korea Tourism Organization (KTO). The authors have received full permission to use the dataset for research purposes.

Previous studies have mainly relied on two types of mobile phone data, namely Call Detail Records (CDRs, see Gonzalez et al. (2008); Xu et al. (2018)) and mobile signaling data (Yan et al. 2019). These datasets document mobile phone users' geolocations at discrete time points. Different from CDRs or mobile signaling data, the dataset used in this study tracks the sequence of locations where an individual tended to stay and for how long. Table 1 shows an example of an individual's phone records. Each record tracks the unique ID of the user, the location (lng/lat) he or she stayed, as well as the date, starting time and ending time that define the corresponding stay period. Each row in the table corresponds to one stay activity and the time periods in between indicate movements among locations. For example, the first two rows in Table 1 indicate that the user stayed at two different locations between [00:07:00 - 09:22:00] and [09:48:00 - 12:48:00] respectively, and a trip was possibly conducted by the user in between (i.e., [09:22:00 - 09:48:00]).

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

User ID	**Date**	**Starting Time**	**Ending Time**	**Longitude**	**Latitude**
123 ∗ ∗∗	2018-05-13	00:07:00	09:22:00	127. ∗ ∗∗	35. ∗ ∗∗
123 ∗ ∗∗	2018-05-13	09:48:00	12:48:00	127. ∗ ∗∗	35. ∗ ∗∗
123 ∗ ∗∗	2018-05-13	18:07:00	18:43:00	126. ∗ ∗∗	37. ∗ ∗∗
...	...	...	...	...	...
123 ∗ ∗∗	2018-05-15	15:32:00	15:53:00	126. ∗ ∗∗	37. ∗ ∗∗
123 ∗ ∗∗	2018-05-15	15:59:00	23:25:00	126. ∗ ∗∗	37. ∗ ∗∗

The locations were tracked at the level of mobile base stations (MBS) and their densities in space reflect the spatial granularity of the dataset. To better understand their spatial arrangement, for each of the three cities, we measure the distance from each MBS to its nearest peer. The summary statistics of the three cities are shown in Table 2. It can be seen that the spacing gaps between MBS in all three cities are relatively small. Overall, the dataset provides a fine-grained view of tourist mobility in both time and space. Such information allows for accurate quantification of travelers' time use characteristics.

Table 2: Summary statistics of mobile base stations (MBS) in each city

City	**Total number of MBS**	**Average nearest distance**	**Median nearest distance**
Gangneung	704	420 meters	344 meters
Jeonju	782	250 meters	181 meters
Chuncheon	861	443 meters	283 meters

2.2 General background of the three cities

Gangneung, Jeonju and Chuncheon are three cities in South Korea (Figure 1A). Gangneung is a city in the province of Gangwon-do with a population of 213,658 (as of 2017) and a total area of $1040km^2$. The city has two expressways, one national highway and a high-speed train line that connect Gangneung to other major Korean cities, such as Seoul, Incheon and Busan. Gangneung is also one of the cities that hosted the 2018 Winter Olympics during February 9-25, 2018. Indoor ice events were held in the Gangneung Olympic Park.

Jeonju is the capital of North Jeolla Province and the 16th largest city in South Korea. It covers an area of $206km^2$ and a population of 652,392 (as of 2017). The city is renowned for traditional Korean food and historical sites. It was chosen as a creative city for Gastronomy as part of UNESCOs creative Cities Network. It is accessible from Seoul by Jeolla Train Line.

Chuncheon is the capital city of Gangwon Province. It has a total population of 281,725 (as of 2012) and a total area of $1116km^2$. It is the terminal city of Cyeongchun Line, which connects Chuncheon to Seoul, and a series of train stations on the Cyeongchun Line are located in Chuncheon suburban area. Chuncheon is also the northern terminus of Jungang Expressway and the essential destination of Seoul-Yangyang Expressway that connects to Seoul and east coastal region. Major cities in mainland South Korea are accessible through Chuncheon Bus Terminal.

According to the mobile positioning dataset, a total of 33,219 phone users visited Gangneung during the data collection period, compared to 18,625 for Jeonju, and 66,646 for Chuncheon. Figure 1B to Figure 1D illustrate the daily number of visitors[2] in the three cities. Some fluctuations are observed throughout the year. Notably, there is a significant spike in Gangneung that matches well with the period of the 2018 Winter Olympics. Therefore, the dataset offers a unique opportunity for examining tourist time use both during regular days and under special events.

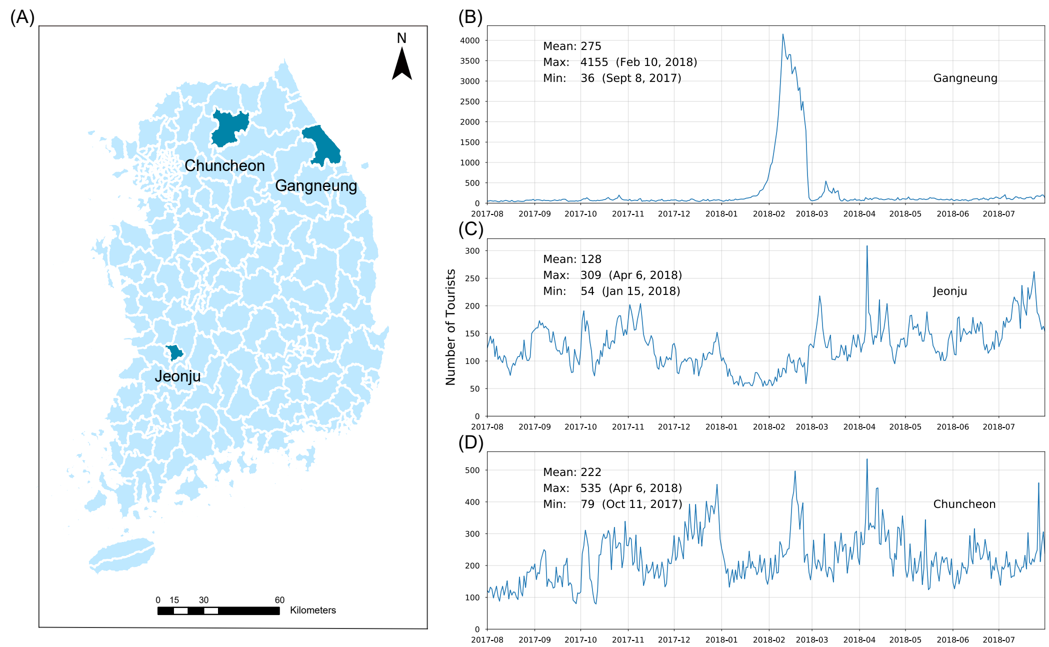

Figure 1: Locations of the three cities in South Korea and daily variations of visitors.

3 Methodology

3.1 Derive activity anchor points from individual cellphone trajectory

In order to extract the time use of travelers in a city, an initial and important step is to identify and represent individuals' meaningful activity locations. One challenge of identifying such locations from mobile phone data is that an individual's observations could switch among adjacent cellphone towers or mobile base stations due to cellphone load balancing or signal strength variation (Csáji et al. 2013; Isaacman et al. 2012). To tackle this issue, we adopt the concept of *activity anchor points*, which were frequently used in the geography literature to denote a person's activity locations (Dijst 1999; Schönfelder and Axhausen 2003; Xu et al. 2016). In this article, we define an activity anchor point as a set of MBS that are close to each other and where an individual has spent a certain amount of time.

Given an individual's cellphone trace $T = \{(l_1, t_1^s, t_1^e), (l_2, t_2^s, t_2^e), ..., (l_n, t_n^s, t_n^e)\}$, where l_i denotes the MBS location of the i^{th} record, and t_i^s and t_i^e denote the starting and ending time of the stay, the extraction of activity anchor points works as follows. First, we compute the total time that the individual stayed at each MBS and sort them in descending order. We then select the MBS with the longest duration of stay, and group all other MBS within a roaming distance (Δd) of the selected MBS into a cluster. Among the MBS that have not been assigned to any cluster, we select the next one with the longest duration of stay and perform the same grouping process. The process is repeated until all the MBS in T are processed. We refer to these clusters as *activity anchor points*. Regarding the choice of Δd, given that the mean and median nearest distance between MBS in three cities are all below 500 meters (Table 2), we set Δd as 500 meters in this study.

In doing so, we can label each record in T using the corresponding activity anchor point. Figure 2 demonstrates the process of anchor point extraction and the labeling of activity locations. As shown in the Figure, an individual's cellphone trace T traverses through five distinct MBS in chronological order: $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$. To group them into clusters, the MBS with the highest total duration of stay — A in this example — is selected. Since E is within 500 meters of A, it is grouped with A to form the first cluster r_1. Then, the next MBS with the longest duration of stay (C) is selected and it alone forms the second cluster r_2, leaving B and D to form the third cluster r_3. Thus, the individual's cellphone trace can be represented as a sequence of activity anchor points: $r_1 \rightarrow r_3 \rightarrow r_2 \rightarrow r_3 \rightarrow r_1$. Such sequences will be used as input to further derive travelers' time use characteristics.

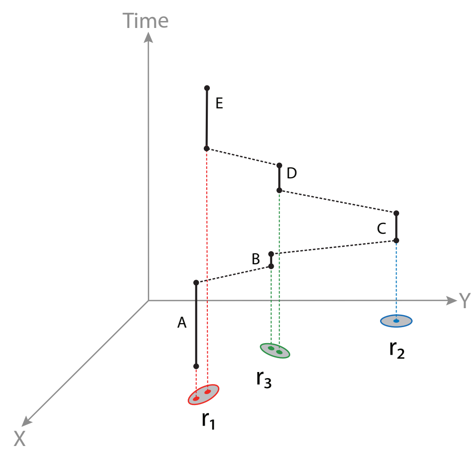

Figure 2: Extraction of activity anchor points from an individual's cellphone trace T. Each vertical segment corresponds to one mobile phone record in the raw data (e.g., rows shown in Table 1). A to E denote five distinct mobile base stations (MBS) traversed by T. r_1, r_2 and r_3 denote the extracted activity anchor points. After the anchor point extraction, the individual's cellphone trace can be represented as a sequence of stays at these activity locations $r_1 \rightarrow r_3 \rightarrow r_2 \rightarrow r_3 \rightarrow r_1$.

3.2 Generate daily time use diagram of individuals

In this study, we use a single *24-hour day* as the basic unit to derive and analyze the time use of travelers. It is a typical representation of *cyclical time*, which allows us to examine the cyclic process of travelers' time use characteristics and how they collectively portray the pulses and rhythm of a city. Since time use is often associated with specific activity types or purposes (Ellegård 1999), which are not directly available from the mobile phone dataset, we introduce an approach to infer and describe the semantics of individual stays and movements as they explore a destination city.

We first define $T' = \{(r_1, t_1^s, t_1^e), (r_2, t_2^s, t_2^e), ..., (r_n, t_n^s, t_n^e)\}$ as an individual's *daily activity sequence*, where r_i denotes the activity anchor point of the i^{th} record, and t_i^s and t_i^e denote the starting and ending time of the stay. The proposed approach aims to label the individual's stays or movements into one of the following five categories:

- **Night-Time Anchor Point (NAP)**: To label the activity anchor point where an individual tended to spend most of the time between midnight and 7am. The NAP can be used as a reasonable estimate of one's dwelling place (e.g., a hotel that an individual stayed during that day). Since an individual could stay at multiple locations between midnight and 7am for a small amount of time, we only label the most-stayed activity anchor point as NAP if the total duration of stay is above 3 hours. Otherwise we label them as *Other Activity Locations* (OTHERS). Note that the choice of the threshold (e.g., 3 hours in this study) will not affect estimation of time use at different locations, but just the representation and visualization of the time use patterns. This parameter can be adjusted based on specific research purposes, or validated when ground truth data of travelers' dwelling places is available (e.g., through questionnaire). One special case of labelling NAP is that if an individual entered a city in the middle of a day and stayed overnight, we first detect the individual's NAP for the second day. If a NAP is detected, the proportion of time the individual was observed at this location (NAP) in the previous day is also labelled as NAP. For instance, if a visitor checked in a hotel at 21:00 in the first day and stayed in the hotel till 8:00 in the next day, the activities during both [21:00:00 - 24:00:00 (day1)] and [00:00:00 - 08:00:00 (day 2)] are labelled as NAP.

- **New Night-Time Anchor Point (NNAP)**: By exploring the mobile phone dataset, we find that the night-time anchor points of an individual could switch from day to day, which indicates a possible change of his/her dwelling place (e.g, change of hotel location). For example, an individual might check out from a hotel in the morning and check in another hotel in the evening in the same day. In this case, the first hotel location will be labelled as the NAP and the second one will be labelled as *New Night-Time Anchor Point* (NNAP). The identification of such cases requires evaluating an individual's activity sequences for two consecutive days (T' and T'_{next}).

- **Other Activity Locations (OTHERS)**: To label activity anchor points that do not belong to the above two categories. They are used to reflect an individual's activity locations beyond night-time anchor points (e.g., activities at scenic spots).

- **Travel (TRAVEL)**: To label individual movements or trips between consecutive stays. More precisely, the category of TRAVEL aims to describe individual movements within a city (i.e., intra-urban movements).

- **Out of the City (OUT)**: Note that the time use of individuals is derived from the perspective of a specific destination city. Thus, if an individual is not observed within the city being analyzed, we label this portion of time as OUT. This category can effectively reveal when tourists enter or leave a city, which is critical information for tourism sectors and local stakeholders.

Figure 3 shows examples on the time use patterns of four individuals in the city of Jeonju. Following the concept of space-time path (Hägerstrand 1970), we develop a space-time plot to visualize an individual's activities and movements over time. We use cylinders to denote activity anchor points and dashed lines to denote travels. Five different colors are used to denote the semantic category of activities or movements. The first individual visited the city for only one day (Figure 3A). The starting time of the first record and the ending time of the last record he was observed in the city are 09:08:00 and 17:23:00, respectively. Thus, the time periods of [00:00:00 - 09:08:00] and [17:23:00 - 24:00:00] are labelled as OUT (black). The first stay activity of the individual was observed during [09:08:00 - 09:46:00], and this activity is labelled as OTHERS (yellow). Similarly, three other stay activities were observed during [10:54:00 - 13:05:00], [13:35:00 - 15:40:00] and [16:07:00 - 17:23:00], respectively. Since we do not observe any activity anchor points with a total duration of stay greater than 3 hours between midnight and 7am, no activity is labelled as NAP or NNAP. The fraction of time occurred between consecutive stay activities was labelled as TRAVEL (dash lines in blue). At the right side of the space-time plot, a bar plot is created to visualize the time use of the individual in chronological order, along with the associated semantic information. We refer to this as the *daily time use diagram* of the individual.

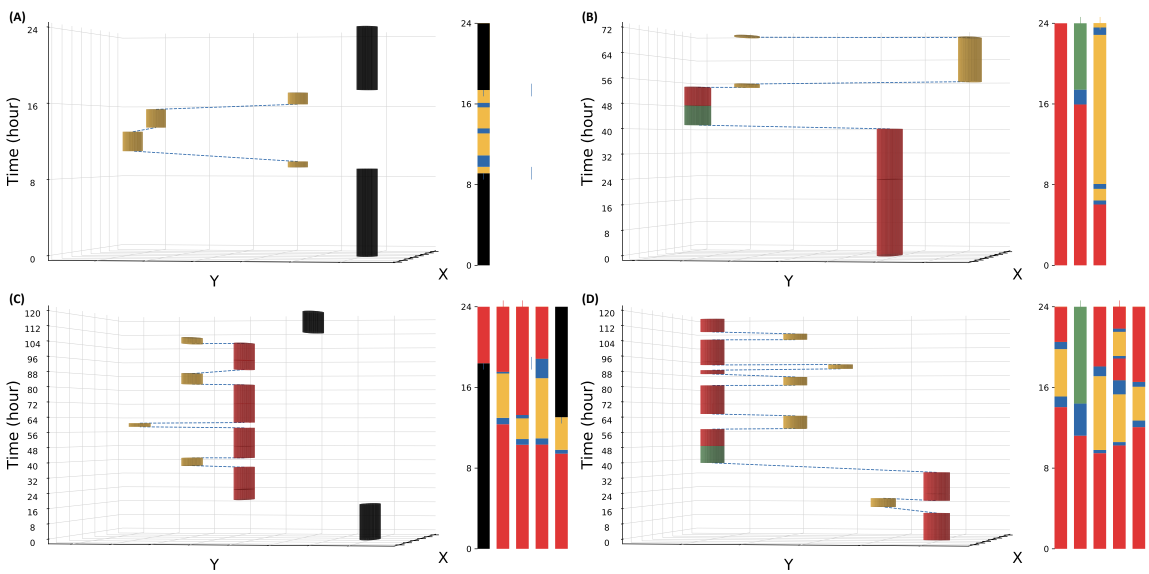

Figure 3: Examples on the time use patterns of four selected individuals in the city of Jeonju. Each example consists of a space-time plot (left side) that shows the sequence of activities and movements, along with the corresponding daily time use diagrams of the individual (right side). Five different colors are used to denote the activity categories: NAP (red), NNAP (green), OTHERS (yellow), TRAVEL (blue), and OUT (black).

The second individual stayed in Jeonju city for three whole days (Figure 3B). The individual spent the first whole day at a single activity anchor point. Thus, the activity is labelled as NAP (red). In the second day, The individual continued to stay at the same location during [00:00:00 - 15:57:00 (day 2)]. Later in that day, the individual was observed at another location during [17:25:00 - 24:00:00 (day 2)]. Interestingly, this location happened to be the *Night-Time Anchor Point* of the third day, where the individual stayed during [00:00:00 - 06:00:02 (day 3)]. Therefore, we label the activity during [17:25:00 - 24:00:00 (day 2)] as NNAP (green), which indicates a change of possible dwelling place of the individual during the second day. Three other activities were observed during the third day. As these activities did not occur at the *Night-Time Anchor Point* (of the third day), they are labelled as OTHERS. Since this individual has three days of observations, three different *daily time use diagrams* were derived. Note that although this traveler has NNAP (green color) identified in the second day, the associated time segment in the next day (i.e., the third day) is still recognized as NAP (red color). Similarly, the space-time plots and time use diagrams of the other two individuals are shown in Figure 3C and Figure 3D.

3.3 Generate daily time use diagrams of a city

Once the daily time use diagrams of individuals are derived, it is easy to organize them by each observation day, forming what we call the *daily time use diagrams of a city*. Such diagrams serve as the signatures of a city that reflect the pulses and rhythm of tourist activities. As an example, Figure 4A shows the daily time use diagram of Gangneung city on January 31, 2018. Individual time use diagrams are organized from left to right, with the vertical axis showing their time use characteristics. Since individuals are organized randomly, the diagram does not convey meaningful information about the city. Thus, we propose two strategies to reduce visual clutter and further improve the readability of the diagram.

First, we categorize individuals into eleven classes based on the types of their first and last activity observed within a specified day. Individuals within the same class are then organized together in the aggregated diagram. We define these classes based on the first and last activity because their combinations enable a coarse segmentation of individuals by revealing different travel styles. The definition of the eleven classes and some typical scenarios are summarized in Table 3.

Table 3: Coarse segmentation of individuals based on different travel styles in a city.

Class	First and Last Activity	Scenario
C1	NAP → NAP	Started and ended the day at the same dwelling place (e.g., hotel)
C2	NAP → NNAP	Started the day at one dwelling place and switched to a new dwelling place at the end of the day
C3	NAP → OTHERS	Started the day at the dwelling place and was observed at a different activity location at the end of the day
C4	OTHERS → NAP	Started the day at a place and ended the day at the dwelling place
C5	OTHERS → NNAP	Started the day at a place, returned to his dwelling place (e.g., a hotel) later, and switched to another dwelling place at the end of the day
C6	OTHERS → OTHERS	Started and ended the day at locations other than the dwelling place
C7	NAP → OUT	Started the day at the dwelling place, and left the city before the end of the day
C8	OTHERS → OUT	Started the day somewhere and returned to dwelling place later, and left the city before the end of the day
C9	OUT → NAP	Started the day somewhere and returned to dwelling place at the end of the day
C10	OUT → OTHERS	Out of the city at the beginning of the day, but was observed at a location other than the dwelling place at the end of the day
C11	OUT → OUT	Out of the city at the beginning of the day, and then came to visit the city and then left

Figure 4B illustrates the diagram after grouping individuals by the eleven classes. The diagram starts to reflect some interesting patterns of the city. However, individuals within the same class are still organized randomly. For example, some individuals belong to C1

(NAP $\rightarrow$ NAP) stayed in the city for the whole day, while some others could be out of the city in the middle of the day (i.e., individual diagrams with black segments). These individuals cannot be easily distinguished in the diagram.

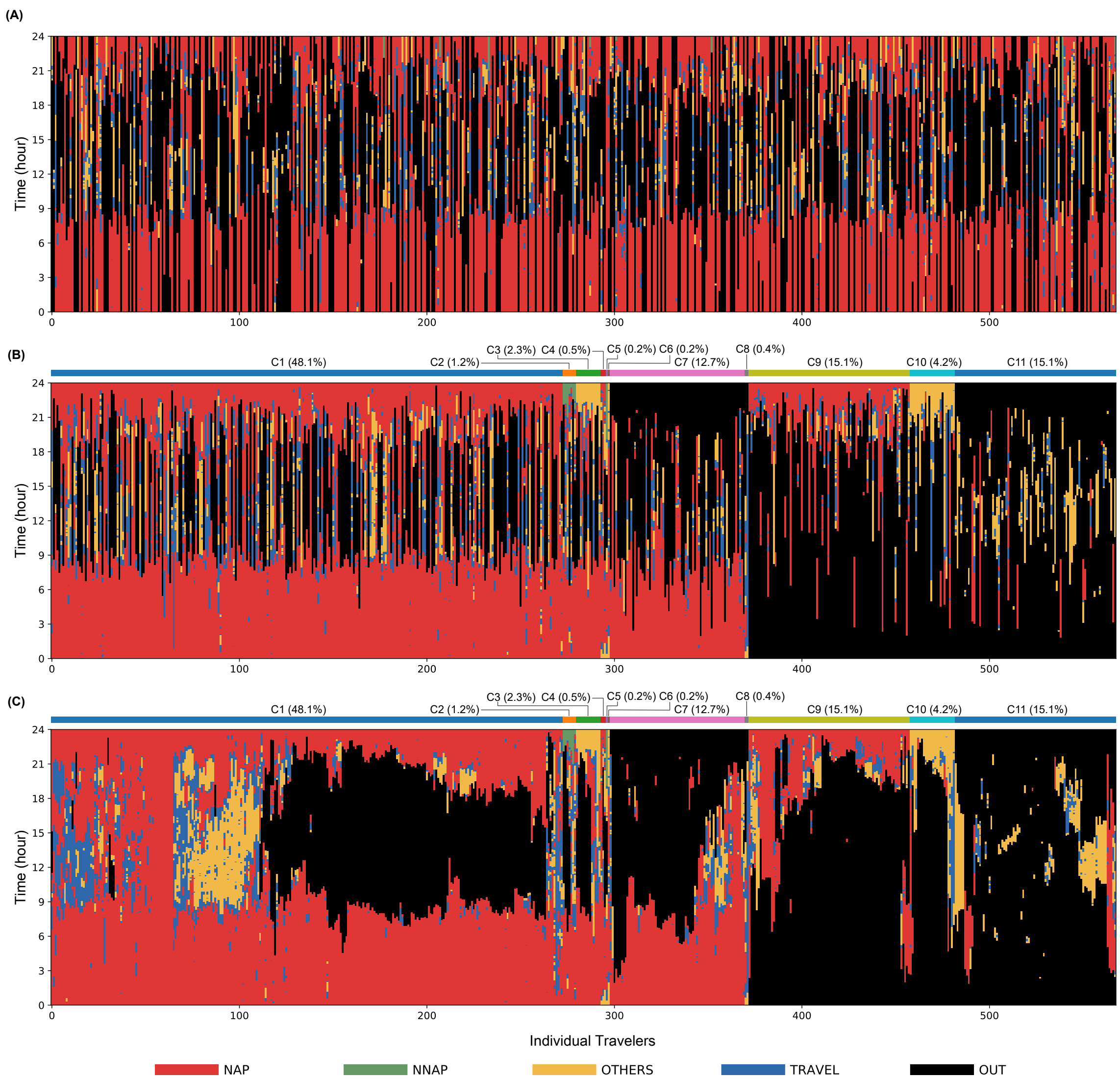

Figure 4: The time use diagram of Gangneung city on January 31, 2018 by: (A) organizing individuals randomly, (B) grouping individuals into eleven predefined classes, and (C) further performing a hierarchical clustering algorithm over individuals within each class based on a time-use similarity metric.

Thus, in the second step, we further perform an agglomerative hierarchical clustering algorithm over individuals within the same class. Given each individual's daily time use diagram, we sample the activity semantics (i.e., types) once every 10 minutes. This give us a feature vector X with 144 elements:

$$X = [x_1, x_2, ..., x_{144}] \tag{1}$$

where x_j denotes the activity type of j^{th} element, with $x_j \in \{$"NAP", "NNAP", "OTHERS", "TRAVEL", "OUT"}. When performing the hierarchical clustering algorithm, the distance between two feature vectors X and X' is defined based on the Hamming distance measure:

$$\Delta(X, X') = |\{j : x_j \neq x'_j\}|, j \in [1, 144] \tag{2}$$

The value of $\Delta(X, X')$ ranges between 0 and 144. For instance, two individuals with exactly the same daily time use patterns will have a Hamming distance of zero, while two people with complete different time use patterns will have a Hamming distance of 144. Figure 4C shows the time use diagram of the city by performing the hierarchical clustering algorithm over individuals in each of the eleven classes (using the average linkage method). The diagram now can effectively distinguish individuals with different time use patterns, both within and across the classes. For instance, four types of individuals can be quickly spotted out within C1: (1) individuals who spent a notable proportion of time on TRAVEL, but with the rest of time mostly at night-time anchor points; (2) individuals who stayed at night-time anchor points almost for the whole day; (3) individuals who spent a significant proportion of time at other activity locations (OTHERS); (4) individuals who visited other cities in the middle of the day. It is also clear to see that most of the travelers in C1 departed from their NAP around 8 or 9 AM in the morning, an important rhythm of their travel patterns. Note that the diagram can also reflect the percentage of individuals belonging to each class, which can be used to facilitate day-to-day comparisons. To sum up, the daily time use diagram we develop in this research could serve as a powerful tool for visualizing and analyzing the time use of individual travelers, the interpersonal variations, as well as their collective patterns that form the signature of a city.

4 Day-to-day similarity of a city's daily time use diagrams

Given the methodology introduced in section 3.3, we process the full dataset to generate each city's daily time use diagrams through the data collection period (365 days), from which we can further measure the day-to-day similarities. Such information would allow us to identify "representative" days with more regular time use patterns of travelers as opposed to days with more unique characteristics that can be caused by special events.

Given a city on a specified day, we summarize the characteristics of its daily time use diagram as a feature vector V:

$$V = [c_1, c_2, ..., c_{11}] \tag{3}$$

where c_1 to c_{11} denote the proportion of individuals within each of the eleven classes defined in Table 3. Note that $\sum_{i=1}^{11} c_i = 1$. The Euclidean distance measure is then used to quantify the similarity between the feature vectors of two daily time use diagrams:

$$I(V, V') = 1 - \sqrt{\sum_{i=1}^{11} (c_i - c'_i)^2} \tag{4}$$

The value ranges between 0 and 1, with larger values indicating higher similarities between the two days.

It's worth noting that the similarity measure defined here is based on the proportion of individuals with different travel styles (i.e., eleven classes). Although the measure does not consider the detailed time use arrangement of individuals — such as how much time they spent on each type of activity — the measure enables an efficient way to compare the aggregate behavioral characteristics between any two observation days. In the next section, we demonstrate how the time use diagrams and the similarity measure can be used to uncover the rhythm of cities and the underlying space-time geographies of tourist activities.

5 Analysis Results

5.1 Discover and understand representative days

When tourists explore a destination city, their time use can be affected by a variety of factors, such as the spatial arrangement of tourist attractions, the deployment of transportation services, how their visits to the city relate to prior and future travel plans, among others. Their collective behavioral patterns form the rhythm and signature of a city. The dataset used in this study allows us to examine and compare tourists' time use across Gangneung, Jeonju and Chuncheon. To accomplish this, an essential step is to identify those *representative days* when tourists tended to show more regular patterns compared to days when tourists exhibited unique characteristics.

Using the similarity measure defined in section 4, we compute the day-to-day similarity of the time use diagrams of each city, which is presented as a similarity matrix in Figure 5. Grids with lighter colors represent higher similarities between the corresponding day pairs. Given a matrix, if a certain row (or column since the matrix is symmetric) is dominated by light colors, it means that the observation day is similar to the majority of other days in terms of the distribution of travelers by the eleven classes. However, if a row or column is dominated by dark colors, it suggests that travelers show more unique travel patterns, which could be caused by special events or other latent factors. For instance, in the matrix of Gangneung, there are a few notable dark stripes that match with the period of the 2018 Winter Olympics (Figure 5A). Although the matrices look different across the three cities, one thing they share in common is that the majority of the rows or columns are dominated by light colors, which indicates the existence of *representative days* in these cities.

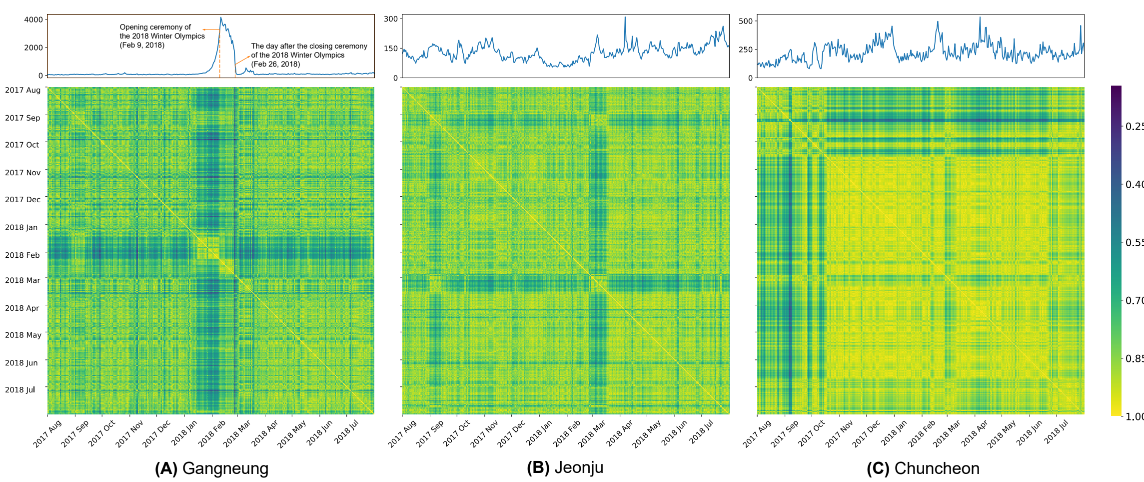

Figure 5: The day-to-day similarity matrix of the three cities. The daily number of visitors in each city is also shown as a reference.

To identify these representative days, for each observation day k in a city, we compute its average similarity to other observation days:

$$E_k = \frac{1}{N}\sum_{j=1}^{N} I(V_k, V_j) \tag{5}$$

where N refers to the total number of days in the dataset ($N = 365$).

In Figure 6, we present the frequency distribution of days by average similarity score (E) and overlay it with the cumulative probability distribution. It can be seen that the majority of days in the three cities have a large E. By further examining the cumulative probability distributions, we find that, for all three cities, the average similarity score of the top 75% of the days are above 0.75 (i.e., 0.77 for Gangneung, 0.82 for Jeonju, and 0.86 for Chuncheon). This suggests that for all three cities, the majority of observation days are similar to each other. Therefore, we use the 75 percentile as a benchmark to filter days that show more distinctive patterns. More specifically, for each city, we define the top 75% of days by average similarity score as *representative days*.

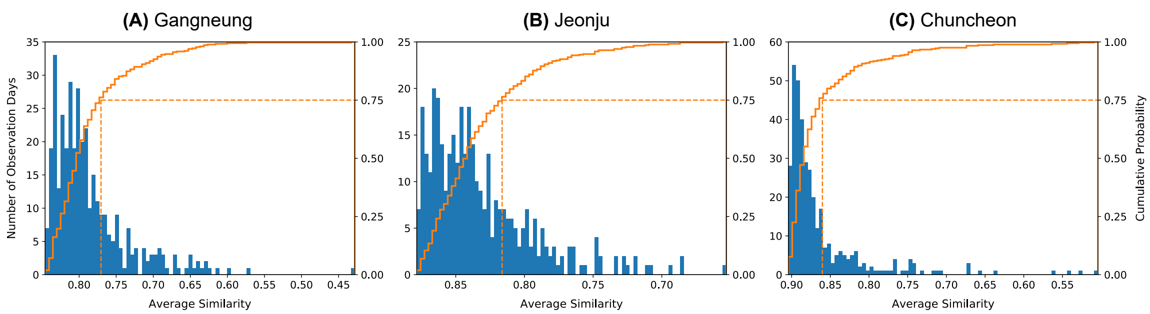

Figure 6: Distribution of average similarity score (E) of observation days in the three cities.

Following this definition, we compute the average proportion of individuals in the eleven classes during the representative days. This would reveal the general split of tourists with different travel styles. As shown in Table 4, tourists in the three cities are dominated by a few classes such as C1, C7, C9 and C11. These classes cover some typical types of travelers, such as those who were at NAP in the beginning and end of the day (C1), who left the city from NAP during the day (C7), the new comers (C9), and visitors with a brief stay (C11). Another interesting finding is that the percentages of travelers in C2 and C5 are very low in the three cities, which may suggest that overnight visitors tended not to switch their dwelling places.

Table 4: Average proportion of travelers in the eleven classes during representative days.

	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**
Gangneung	17.7	0.91	1.16	0.36	0.12	0.40	18.1	2.17	15.7	5.87	37.5
Jeonju	23.3	1.22	1.91	0.55	0.20	0.35	21.2	1.99	17.9	5.81	25.6
Chuncheon	6.50	0.37	0.46	0.17	0.05	0.18	5.33	0.62	4.61	1.82	79.9

Substantial differences are also observed across cities. For instance, in Chuncheon, the percentage of travelers in C1 (6.5%) is much lower than that of Gangnueng (17.7%) and Jeonju (23.3%). However, the proportion of travelers with a brief stay (C11) is significantly higher (79.9%). This reflects a unique characteristic of Chuncheon. It also suggests that cities can be very different in the ways they attract visitors and the spatio-temporal organizations of tourist activities.

To provide more contexts into the time allocation of tourists, we select one of the representative days (May 15, 2018) and visualize the time use diagram of the three cities. From a visual perspective, the diagrams of Gangneung (Figure 7A) and Jeonju (Figure 7B) look similar. There are many travelers (in C1) who stayed in the cities through the whole day. Some visitors left the cities during the day and mostly in the morning (C7). May 15, 2018 also marks the first day of many travelers to the two cities. Interestingly, some of them headed to the night-time anchor points upon arrival (C9), while others were still at other places or travelling around in the late night (C10). The comparison suggests that the diagrams can be a powerful tool for hypothesis formulation and testing. For instance, tourism sectors and decision makers could leverage the diagrams to further explore the places visited by travelers in C9 and C10. This might shed light on tourists' preferences in hotel location selection (C9), or reveal night-time attractions to visitors (C10).

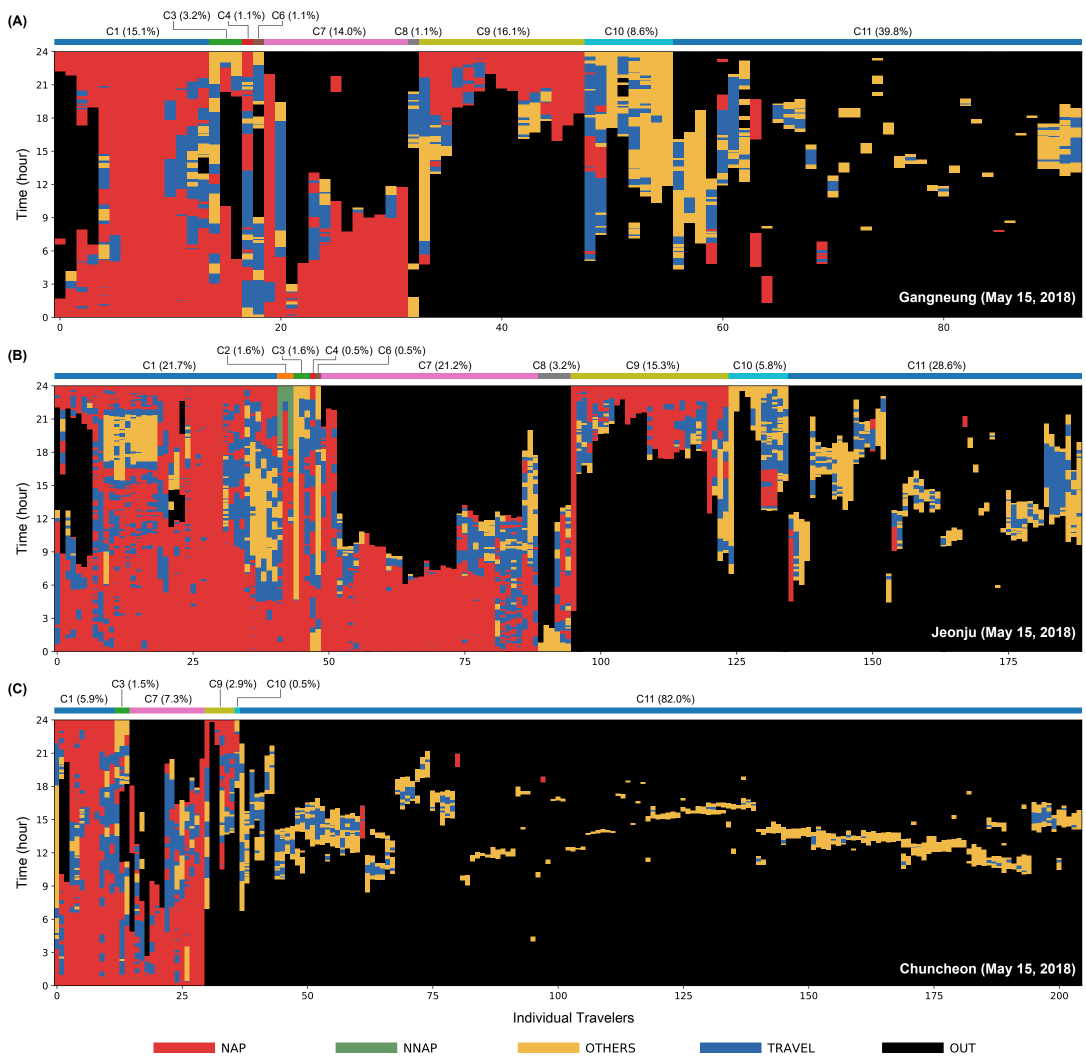

Figure 7: The daily time use diagram of: (A) Jeonju, (B) Gangneung and (C) Chuncheon on May 15, 2018.

As discussed previously, Chuncheon is dominated by temporary visitors (e.g., C11). This finding is reaffirmed by looking at the time use diagram in Figure 7C. Besides illustrating the large share of such visitors (C11), the diagram also reveals that most of them stayed in the city for a very short period of time, usually a few hours, or even shorter. Again, it suggests that the time use diagram is a useful tool that enables the investigation of tourist time use in cities, both visually and quantitatively.

5.2 Transform the time use diagram — From individual to time oriented perspective

The time use diagram enables an individual-oriented perspective of cities by revealing travelers with different time use characteristics. In the meantime, we are also interested in understanding the proportion of individuals participating in different activities at any given time point. The time use diagrams can be easily adjusted to such a time-oriented perspective. Given each representative day in a city, we calculate the percentage of individuals in each activity category once every five minutes. We then compute the average proportion of individuals by time and activity category across all representative days. The results — what we call the *time-activity diagrams* shown in Figure 8 — illustrate the average proportions of individuals by activity category and how they change over time. Since NNAP are rare cases in this study, we combine this category with NAP when visualizing the results (renamed as (N)NAP)

As shown in Figure 8A, the proportion of individuals belonging to OUT is stable over time in Gangneung (around 0.6). It tells the fact that on an average representative day — if we use the travelers who were observed at least once in Gangneung as a base — the proportion of travelers who were outside of the city tended to remain constant over time. That means on a representative day, the visitors who entered or left the city tended to balance with each other, a finding that seems not too surprising. However, if compared with Jeonju (Figure 8B), we can see that the overall proportion of such individuals is around 0.5, which is lower than that of Gangneung (0.6). The comparison indicates that Gangneung has a higher "metabolic rate" — the pace a city absorbs and emits travelers. An equivalent expression is that more tourists tended to stay overnight in Jeonju than in Gangneung.

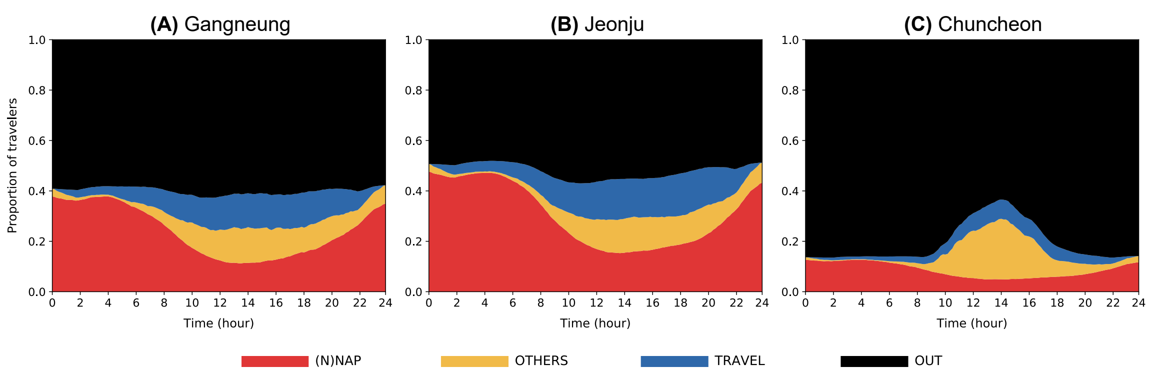

Figure 8: The time-activity diagrams of: (A) Jeonju, (B) Gangneung and (C) Chuncheonon on an average representative day.

The other three categories ((N)NAP, OTHERS, TRAVEL) reflect intra-urban activities.

As shown in Figure 8A, the proportion of individuals belonging to (N)NAP started to drop after 6:00 and gradually increased in the evening. Daytime and early evening mark the periods when tourists actively engaged in travels and activities. Similar patterns are observed in Jeonju (Figure 8B). Different from the time use diagram, the time-activity diagram is powerful in revealing the time periods when people were mostly observed at their dwelling places, travelling, or at other locations such as scenic spots.

Given prior knowledge of Chuncheon, we would expect a special time-activity diagram. As shown in Figure 8C, the diagram is mainly black, which suggests a very high "metabolic rate" of the city. Few people stayed overnight while many visitors had a brief stopover. Most of their activities occurred during the daytime, with the peak around 14:00. Again, the results in this section suggest that cities can be different in terms of the rhythms of tourist activities. The two types of diagrams could complement each other in portraying cities from both individual-oriented (time-use diagram) and time-oriented perspectives (time-activity diagram).

5.3 Time allocation of travelers and spatial patterns of activities

While cities during representative days have been clearly portrayed through the two types of diagrams, the time allocation of individual travelers has not been measured quantitatively. In this section, we examine how travelers allocate their time to different types of activities. Since travelers in the three cities are dominated by a few classes such as C1, C7, C9 and C11, in this analysis we only focus on these four types of travelers. Note that for C1, we have further split travelers who stayed in a city for the whole day and those who spent part of their time in a city. We name these two subgroups as C1-full and C1-partial. For each of these five groups (C1-full, C1-partial, C7, C9, C11), we compute individuals' time allocation to the following activity categories — NAP, OTHERS, TRAVEL and OUT — and examine their distribution patterns.

In Figure 9, we present the results for the five groups and three cities. Each row maps one city and each column corresponds to one traveler group. The error plots illustrate the mean and standard deviation of time that an individual spent on the activity, and the violin plots show the underlying distributions. By comparing the same group across cities, we find that they show very similar distribution patterns, despite of large interpersonal variations. For instance, for travelers who stayed in a city for a whole day (C1-full), they on average spent 16∼18 hours at night-time anchor points and 4∼5 hours travelling or moving around. The results suggest that tourists spent the majority of their time at NAP, locations that are highly likely to be their dwelling places. Regarding the time spent on TRAVEL, it can be a combination of time spent on regular trips (e.g., heading somewhere through bus, taxi or other transportation modes) and exploration of the city (e.g., walking tours). That means tourists tended to experience more around where they live or sense a city "in motion". Very limited time — usually one or two hours — was spent by travelers on other activity locations (OTHERS), although they could be tourist attractions in cities. For C1-partial, C7 and C9, tourists' time allocation at NAP and intra-urban movements (TRAVEL) were partially replaced by time spent out of the city. Since C7 and C9 are very likely to be last-day and first-day visitors to a city, it is reasonable that they spent even less time on other activity locations (OTHERS) than C1-full and C1-partial. On average, travelers in C11 — those who had a brief stopover in a city — spent most time on OTHERS than other groups. It is likely that these travelers had clear purposes visiting particular places when coming to cities, and most of these travelers would stay in a city for less than 4 hours.

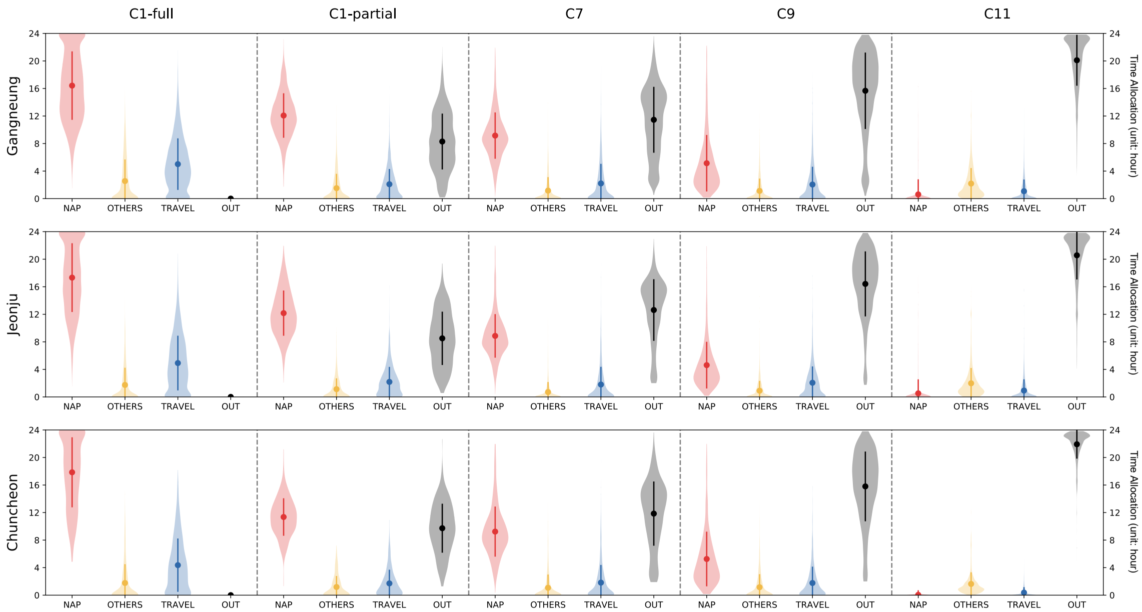

Figure 9: Distributions of individual time allocation to different activities for the five traveler groups in the three cities.

We next visualize the spatial patterns of these activities for selected traveler groups. Figure 10A shows the likelihood that each MBS location is used as NAP for travelers who stayed in Gangneung for a whole day (C1-full). The probability is calculated by counting the number of times each location is identified as NAP during the representative days, normalized by the total frequency of all locations. The map shows that the central part of the city and certain areas along the beach are likely to be chosen by travelers as dwelling places[3]. When looking at travelers' activities beyond night-time anchor points (OTHERS), we find that these activities are also pronounced in central Gangneung, but tended to cover almost all parts of the city (Figure 10B). For full-day visitors in Jeonju, popular areas of dwelling places (Figure 10C) and other activities (Figure 10D) generally point to the same set of places. These places cover some notable tourism attractions such as the Jeonju Hanok Village, the two universities, and the Deokjin Park. Similar to what we find in Gangneung, the spatial choices of OTHERS are more widely spread.

For Chuncheon, the spatial choices of NAP and OTHERS for full-day travelers are mainly distributed in the city center and the Elysian Ski Resort (Figure 10E and Figure 10F). Since travelers who paid a brief stopover (C11) represent the majority of the city's visitors, we also visualize the spatial distribution of their activities (Figure 10G). It turns out that the spatial choices of their activities are very different from those of full-day travelers. In particular, the Jade Garden, Elysian Ski Report and Nami Island accounts for a large proportion of total visits.

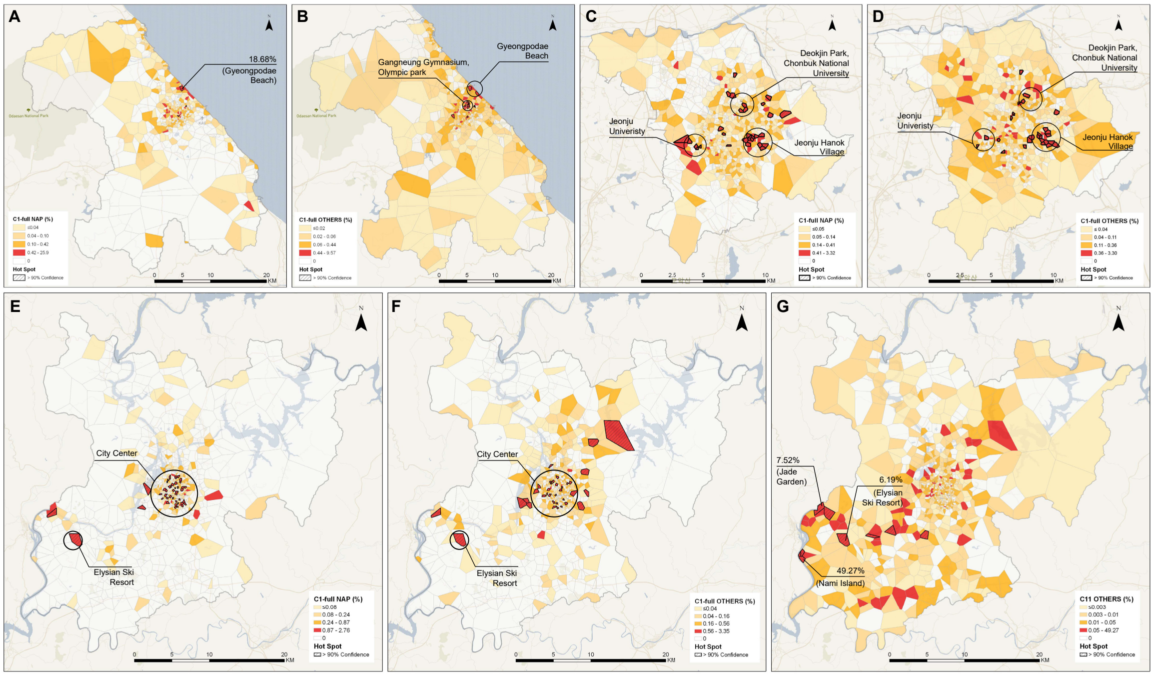

Figure 10: The intensity of activities for selected traveler groups: (A-B) NAP and OTHERS of C1-full in Gangneung; (C-D) NAP and OTHERS of C1-full in Jeonju; (E-F) NAP and OTHERS of C1-full in Chuncheon; (G) OTHERS of C11 in Chuncheon. The Thiessen polygons are generated based on the locations of MBS as approximations of their service areas.

5.4 Tourist dynamics during special events — An example of Winter Olympics in Gangneung

The results presented so far have demonstrated the effectiveness of the computational framework in revealing tourist dynamics during representative days. In this section, we employ the framework to further explore tourist time use under special events, using Gangneung during the Winter Olympics (WO) as an example. Gangneung is one of the cities in South Korea that hosted the WO between February 9-25, 2018. The event has resulted in a surge of visits to the city during this period (Figure 1B). By examining the day-to-day similarity matrix of the city (Figure 5A), we find three distinctive periods around the WO:

- January 16, 2018 to February 8, 2018 (Before WO)
- February 9, 2018 to February 25, 2018 (During WO)
- February 26, 2018 (The day after the closing ceremony of WO)

The first two correspond to periods before and during the WO. According to the similarity matrix in Figure 5A, the observation days within these two periods are much more similar to each other than to other days throughout the year. Regarding February 26, 2018, the dark stripe associated with the date (Figure 5A) suggests that tourists show unique activity patterns right after the WO. The result indicates notable variations of tourist behavioral patterns across the three periods. It also motivates us to further examine tourist time use during the three stages of this mega event.

Table 5 summarizes the average proportion of travelers in the eleven classes during the three periods. Similar to representative days, travelers are still dominated by the following four classes: C1, C7, C9, C11. However, substantial differences are observed across the periods. Before the WO, the city was observed with a large proportion of overnight stayers (C1: 39%) and incoming visitors (C9: 17.1%). During the WO, the shares of C1 and C9 dropped, but there were more outgoing (C7: 18.5%) and temporary visitors (C11: 28.5%). On February 26, 2018, the day right after the closing ceremony of WO, a large proportion of visitors left the city (C7: 62.9%; C8: 10.7%). The results suggest notable transitions of tourist activity patterns across the three periods.

Table 5: Average proportion of travelers in the eleven classes in Gangneung before (January 16 to February 8, 2018), during (February 9 to 25, 2018) and after the Winter Olympics (February 26, 2018).

	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**
Before	39.3	0.85	2.47	0.82	0.24	0.34	12.3	0.96	17.1	5.11	20.4
During	21.8	0.60	6.68	1.27	0.32	1.91	18.5	3.49	8.06	8.92	28.5
After	6.64	0.13	1.11	0.41	0.00	0.41	62.9	10.7	3.46	0.55	13.7

To better understand the time use of tourists in the three periods, we select one observation day from each period and visualize their time use (Figure 11) and time-activity diagrams (Figure 12):

- February 5, 2018
- February 25, 2018
- February 26, 2018

On February 5, 2018, the visitors to Gangneung had diverse time use characteristics (Figure 11A). The city was observed with a mixing of overnight stayers, incoming visitors, outgoing visitors, and temporary stayers. Their time allocation also varies notably. From the perspective of time-activity diagram (Figure 12A), the intra-urban movements and activities of tourists started to increase after 07:00 and the city remained active during the day time. The proportion of visitors that were out of the city (OUT) was roughly the same at the beginning and end of the day (around 0.4). On the one hand, it suggests that the city tended to maintain its attractiveness by balancing the incoming and outgoing visitors. On the other hand, it shows that the city has a lower "metabolic rate" than the level observed during representative days (Figure 8A). This indicates that the WO has a positive impact on the city's attractiveness.

Feb 25, 2018 is the last day of the 2018 WO, and we observe interesting patterns of tourist time use. As shown in Figure 11B, there are clear synchronizations of tourist movements (TRAVEL) and activities (OTHERS) during particular periods of the day. Three blue stripes in the diagram (Figure 11B) suggest that massive amounts of movements were observed in the morning, around noon, and in the late afternoon. Meanwhile, most of the visitors were conducting "out-of-home" activities (OTHERS) in between. These findings are reaffirmed by looking at the time-activity diagram, where local maxima of tourist movements (TRAVEL) were observed on 8:40, 12:00 and 16:00, respectively (circles in Figure 12B). From the diagram we can see that the proportion of visitors out of Gangneung

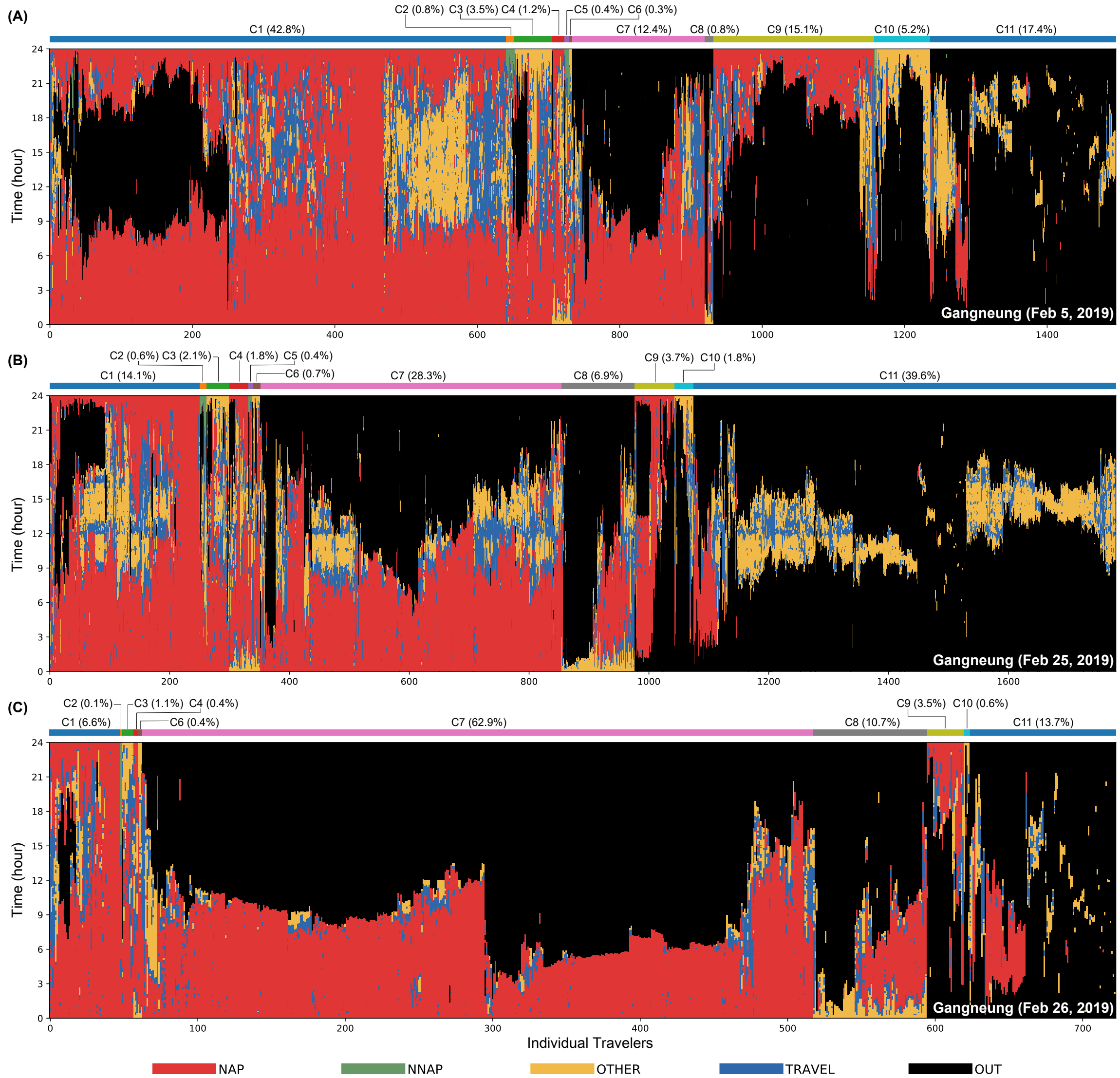

Figure 11: Time use diagrams of Gangneung on: (A) February 5, 2018, (B) February 25, 2018, and (C) February 26, 2018.

(OUT) reaches the maximum value at the end of the day, which indicates that more tourists left than came to the city on Feb 25, 2018.

Feb 26, 2018 is the first day after the closing ceremony of WO. As illustrated in Figure 11C, a significant proportion of visitors left the city during the day (C7: 62.9%; C8: 10.7%), and most of them occurred in the morning. The activity-diagram clearly shows that the city was constantly losing travelers in the day especially before noon (Figure 12C). Overnight stayers accounted for only about 10% of the total visitors. The results in this section demonstrate the effectiveness of the framework for revealing tourist time use during special events, as well as the transitions of their activity patterns across the three periods.

To validate our results, we revisit the three local maxima of tourist movements identified on February 25, 2018 (Figure 12B), and examine the spatial patterns of tourist activities in between: (1) 8:40 - 12:00 and (2) 12:00 - 16:30. We hypothesize that the massive amounts of

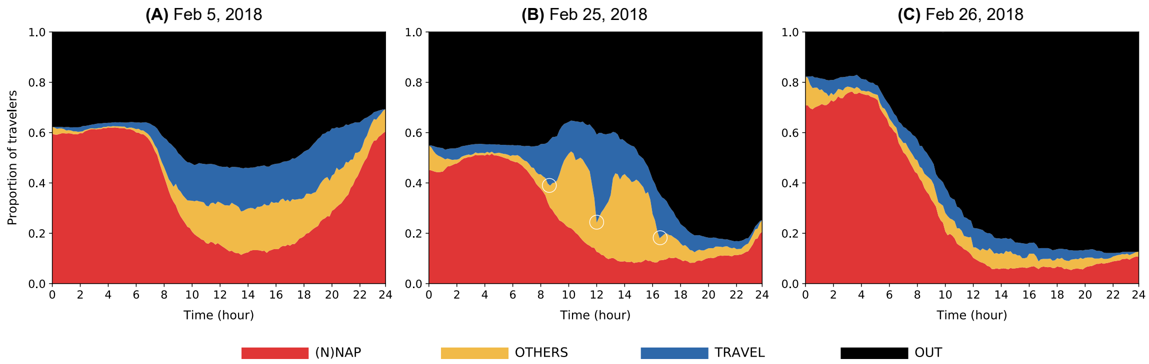

Figure 12: : The time-activity diagrams of Gangneung on: (A) February 5, 2018, (B) February 25, 2018, and (C) February 26, 2018.

synchronized travels around 8:40, 12:00 and 16:00 are affected by special events of WO. By visualizing the spatial patterns of tourist activities (OTHERS) between 8:40 - 12:00, we find that they were mainly concentrated in the city center, especially around the Gangneung Olympic Park (Figure 13A). A close look at the satellite image suggests that the most popular place is around the Gangneung Ice Arena (Figure 13B). While in the afternoon between 12:00 - 16:30, tourist visits were concentrated in the same area (Figure 13C), but the place with the highest visitation frequency has switched to the Gangneung Hockey Center (Figure 13D).

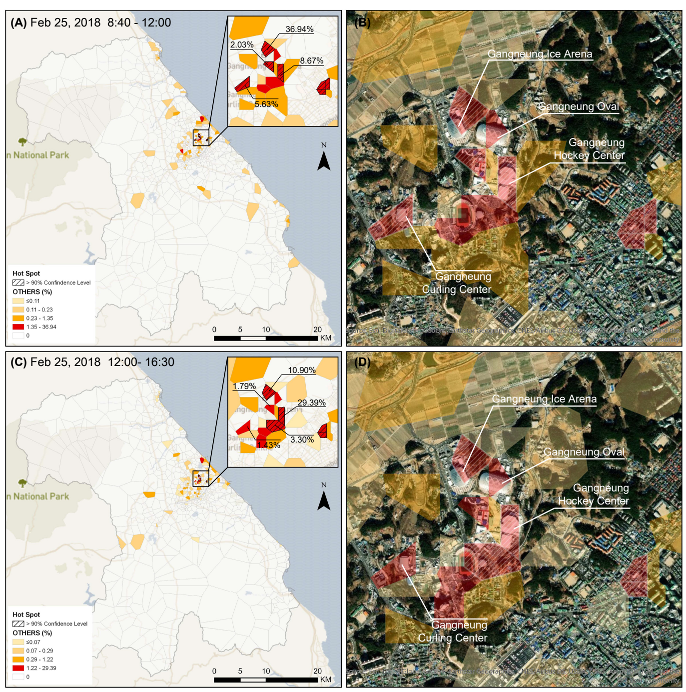

Figure 13: Intensity of tourist activities (OTHERS) in Gangneung and zoom-in view of the popular area through satellite images on February 25, 2018: (A-B) time period of 8:40 - 12:00; (C-D) time period of 12:00 - 16:30.

We compare our findings with the calendar events of the 2018 Winter Olympics. Three events were arranged in Gangneung on Feb 25, 2018 (Wikipedia 2020). These events refer to the exhibition gala of figure skating at the Gangneung Ice Arena (started at 9:30), the women's tournament gold medal game of curling at Gangneung Curling Center (started at 9:05), and the men's tournament gold medal game of ice hockey at Gangneung Hockey Center (started at 13:10). The schedules of these events and the associated venues generally match with the spatio-temporal patterns in Figure 11B, Figure 12B, and Figure 13. The results demonstrate the feasibility of the proposed computational framework and the robustness of the findings. It also suggests the reliability of the mobile phone data in capturing fine-grained aspects of tourist behaviors in cities.

6 Summary

In this research, we develop a computational framework to derive, analyze and visualize tourist time use in cities. It provides a new perspective that connects time geography and tourism through the usage of spatio-temporal big data. We argue that the tourism geography of cities can be better understood as mobility observations of tourists become both massive and granular. By applying the proposed approaches over a fine-grained mobile phone dataset in South Korea, we perform a comparative analysis across three different cities (Gangneung, Jeonju and Chuncheon). The results demonstrate that the framework is effective in revealing the heterogeneity of tourist time use within cities, the interurban variations, as well as how tourists behave during routine days versus special events.

We first develop an anchor point extraction approach, which allows us to convert each individual's cellphone trajectory into a sequence of stays at meaningful activity locations. A semantic labelling process is followed to annotate each individual's stays or movements into one of the five categories (NAP, NNAP, OTHERS, TRAVEL and OUT). The semantics derived in this step can indicate travelers' dwelling places in a city, when they tended to travel, where their activities occurred, and when they were outside of a city. Such information is then used to construct the daily time use diagrams of individuals. The diagrams carry the semantics of tourist activities while preserving the order they occurred in time.

After the daily time use diagrams of individuals are derived, we aggregate them by each observation day to form the daily time use diagrams of a city. Since individuals are organized randomly in the aggregated diagrams, to convey meaningful information about a city, we first perform a coarse segmentation of individuals based on the eleven predefined classes that distinguish different travel styles. A hierarchical clustering algorithm based on a time-use similarity metric is then performed within each class, such that individuals with similar time use characteristics are grouped together. The resulting diagrams turn out to be a powerful tool for visualizing and analyzing the time use of individual travelers, the interpersonal variations, as well as collective patterns that form the signature of a city.

We then process the full dataset to generate the daily time use diagrams of three cities through the data collection period (365 days). By introducing a day-to-day similarity measure, we are able to identify the so-called *representative days* in each city when tourists tended to show more regular patterns, compared to days when they exhibited unique characteristics. By further investigating tourist time use in the three cities during representative days, we find that Gangneung and Jeonju are more similar to each other while Chuncheon shows more distinctive patterns. The results suggest that the computational framework can be useful in capturing the similarities and differences of tourist time use across cities. Following this analysis, we introduce another visualization layout — the time-activity diagram — to illustrate the proportion of individual travelers by activity category in a city and its evolution over time. The two types of diagrams (time-use diagram; time-activity diagram) complement each other in portraying the pulses and rhythm of tourist activities in cities.

Next, we select individuals with dominant travel styles and examine the time allocation and spatial patterns of their activities. We find that although tourists organized time differently across cities, their time allocation patterns, when controlling the traveler group, are highly similar. For example, full-day visitors (C1-full) on average spent 16∼18 hours at night-time anchor points and 4∼5 hours travelling or moving around, a finding that is universal across the three cities. By mapping the location footprints of selected traveler groups, we find that cities show different spatial structures of tourist activities. However, within the same city, the location choice and preference can vary among different traveler groups (e.g., full-day vs. temporary visitors in Chuncheon).

Finally, we employ the framework to explore tourist time use under special events, using Gangneung during the Winter Olympics (WO) as an example. Three periods are selected to analyze the time use of tourists before, during, and after the WO. The results suggest notable transitions of tourist activity patterns across the three stages of this mega event. A close examination of tourist activities on Feb 25, 2018 is performed and the results are validated by comparing the spatio-temporal patterns with the calendar events of WO. The validation demonstrates the robustness of our findings and the efficacy of the proposed computational framework. It also suggests fine-grained mobile phone data as a promising resource to advancing geography and tourism research.

7 Discussions

We want to discuss the broad implications of this study. Traditional studies on human mobility mainly replied on surveys and questionnaires. Despite considerable costs in data acquisition, they usually contain detailed information of travelers, which is often regarded as "ground truth" for producing rigorous research outcomes. This ideology is however disrupted by the emergence of big data (e.g., mobile phone data), which were not intended for (geographical) research when they were collected. As a result, the findings derived from big data are more difficult to validate due to the absence of ground truth. In this study, we demonstrate that fine-grained mobile phone data can be leveraged to produce novel insights of tourist activities at an unprecedented spatio-temporal resolution. There are still issues in the data such as location uncertainty caused by positional inaccuracy (Isaacman et al. 2012; Csáji et al. 2013; Goodchild 1998), and lack of contextual information on travelers (e.g., activity purpose, travel mode). Nevertheless, we think the framework and results are already helpful to decision making and spatial planning, especially when they are used as tools for hypothesis formulation and follow-up studies. For instance, the methods developed and used in this study (time use diagram, time-activity diagram, spatial mapping) can be combined to form smart dashboards for interactive data exploration. Many meaningful questions can be asked. For example, who are the temporary visitors in a city? Why do they come to visit but choose not to stay overnight? What are the places where tourist activities proliferate at night, and what do they tell about night-time economy (Cohen and Hopkins 2019)? Why are activities of some visitors fragmented, and what are the implications for location-based services and recommendation? Some of these questions can be better addressed by conducting surveys with tourists. The design of surveys, and when and where they should be conducted, can be guided and informed by the insights derived from the smart dashboard. In other words, we think big and small data are not two sides of a coin, but can be integrated in organic ways to facilitate geography and tourism research.

The computational framework can also be extended to study time use of large urban populations (e.g., residents). For instance, mobile signaling data could capture the whereabouts of cellphone users at a fine spatio-temporal resolution (Janecek et al. 2015). Such information can be used to infer individuals' meaningful activity locations (e.g., home, workplace, favourite restaurants) and time allocation at these places. Feeding these behavioral patterns into the time use diagram would shed light on the rhythm and paces of urban life (Reades et al. 2009; Yuan and Raubal 2012; Ahas et al. 2015; Chen et al. 2018; Thulin et al. 2020). Many profound issues of cities can be better investigated. For example, the framework can be used to understand the time use of different traveler or sociodemographic groups (Yuan et al. 2012; Xu et al. 2018), and the implications for inequality, population aging, and urban planning. How residents behave under natural disasters (Lu et al. 2012) or public health emergency (e.g., COVID-19 (Kraemer et al. 2020)) can also be visualized and analyzed through the framework to inform policy and decision making. We think this research is an active response to the emerging discussion on "Smart Spaces and Places" in the geography community (Bian 2020), and the computational framework enables a big data and time-oriented perspective to "perceive, analyze, and visualize spaces and places" (pp.335). Of course, there are many challenges to overcome if the framework is applied on massive urban populations. For instance, how to address the increasing computational cost on segmenting and clustering individuals in the time use diagram, and how to design appropriate time-use similarity metrics for given application purposes are important research questions. These are the future directions we aim to pursue. At the meantime, we should bear in mind that mobile phone data, especially individual-level observations, could raise public concerns about people's privacy. Ensuring their proper use with appropriate data protection mechanisms is important for future research (Oliver et al. 2020).

We want to point out a few limitations of this research. Although the resolution of the mobile phone dataset is adequate for quantifying tourist time use, uncertainties still exist in the data. From the spatial perspective, the mobile base stations (MBS) are unevenly distributed in cities, and their densities are usually higher at densely populated areas. Thus, the mobile phone observations would have lower spatial resolutions where MBS are sparsely distributed. This may affect the estimation and interpretation of activity semantics. For instance, it is possible that some short-range movements of phone users in certain areas (e.g., where MBS densities are low) were not captured, and thus represented as stay activities in the time use diagrams. For such cases, the meaning of "stay" needs to be interpreted with caution. From the temporal perspective, the dataset is able to document all the stay activities of individuals. However, their movement traces between consecutive stays were not recorded. This would have an impact on the estimation of individual time spent within/out of a city. Given an individual, the time associated with intra-urban activities in a city (e.g., NAP, NNAP, OTHERS, TRAVEL), according to our definition, started when the first stay activity was observed (e.g., 9 AM). The proportion of time before that was identified as "Out of the City". However, it is likely that the individual was already in the city before 9 AM (e.g. took a train from a neighbouring city at 8 AM and headed to the first destination of the city). In other words, the dataset cannot precisely capture when an individual physically entered a city from another. Thus, our framework would overestimate the time an individual was observed “Out of the City”, and an individual’s time within a city is interpreted in this study as the period between his/her first and last stay activities. Nevertheless, we think the insights derived from the dataset is a notable contribution to the literature of geography and tourism. The computational framework can be applied to compatible datasets to advance time geography, tourism, and urban mobility research.

Notes

[1]The cellular operator has identified international travelers from local residents. The fundamental idea is that the mobile data used in this research refers to inbound tourists who have used mobile roaming services in South Korea.

[2]Defined as the number of phone users with at least one record during the specified day.

[3]Note that individual activities (e.g., NAP, OTHERS) are identified at the level of activity anchor points, which could consist of one or several MBS locations. This introduces an issue when we want to map the density of activities at the level of MBS. To simplify the mapping process, for each activity anchor (of each individual), we identify its representative MBS, defined as the MBS in the anchor point with the highest stay duration. The activity is always allocated to the representative MBS when counting the frequency of activities.

Acknowledgement

We sincerely thank Korea Tourism Organization (particularly to Tourism Big Data Center) for their full support in initiating the project and conducting the data collection. The work described in this article was jointly supported by a PTeC funding (Project Code: P18-0015) from The Hong Kong Polytechnic University, and the Hong Kong Polytechnic University Start-Up Grant (Project Code: 1-BE0J).

References

R. Ahas, A. Aasa, Ü. Mark, T. Pae, and A. Kull. Seasonal tourism spaces in estonia: Case study with mobile positioning data. *Tourism management*, 28(3):898–910, 2007.

R. Ahas, A. Aasa, Y. Yuan, M. Raubal, Z. Smoreda, Y. Liu, C. Ziemlicki, M. Tiru, and M. Zook. Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in harbin, paris, and tallinn. *International Journal of Geographical Information Science*, 29(11):2017–2039, 2015.

M. Bauder and T. Freytag. Visitor mobility in the city and the effects of travel preparation. *Tourism Geographies*, 17(5):682–700, 2015.

R. Becker, R. Cáceres, K. Hanson, S. Isaacman, J. M. Loh, M. Martonosi, J. Rowland, S. Urbanek, A. Varshavsky, and C. Volinsky. Human mobility characterization from cellular network data. *Communications of the ACM*, 56(1):74–82, 2013.

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

L. Bian. Introduction: Smart spaces and places. *Annals of the American Association of Geographers*, 110(2):335–338, 2020.

A. Birenboim and N. Shoval. Mobility research in the age of the smartphone. *Annals of the American Association of Geographers*, 106(2):283–291, 2016.

A. Birenboim, S. Anton-Clavé, A. P. Russo, and N. Shoval. Temporal activity patterns of theme park visitors. *Tourism Geographies*, 15(4):601–619, 2013.

V. D. Blondel, A. Decuyper, and G. Krings. A survey of results on mobile phone datasets analysis. *EPJ data science*, 4(1):10, 2015.

L. Calderwood and M. Soshkin. The travel & tourism competitiveness report 2019. In *World Economic Forum. http://www3. weforum. org/docs/WEF_TTCR_2019. pdf*, 2019.

B. Y. Chen, Y. Wang, D. Wang, Q. Li, W. H. Lam, and S.-L. Shaw. Understanding the impacts of human mobility on accessibility using massive mobile phone tracking data. *Annals of the American Association of Geographers*, 108(4):1115–1133, 2018.

C. Chen, J. Ma, Y. Susilo, Y. Liu, and M. Wang. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation research part C: emerging technologies*, 68:285–299, 2016.

S. A. Cohen and D. Hopkins. Autonomous vehicles and the future of urban tourism. *Annals of tourism research*, 74:33–42, 2019.

B. C. Csáji, A. Browet, V. A. Traag, J.-C. Delvenne, E. Huens, P. Van Dooren, Z. Smoreda, and V. D. Blondel. Exploring the mobility of mobile phone users. *Physica A: statistical mechanics and its applications*, 392(6):1459–1473, 2013.

S. De Cantis, M. Ferrante, A. Kahani, and N. Shoval. Cruise passengers' behavior at the destination: Investigation using gps technology. *Tourism Management*, 52:133–150, 2016.

M. Dijst. Two-earner families and their action spaces: A case study of two dutch communities. *GeoJournal*, 48(3):195–206, 1999.

D. East, P. Osborne, S. Kemp, and T. Woodfine. Combining gps & survey data improves understanding of visitor behaviour. *Tourism Management*, 61:307–320, 2017.

D. Edwards and T. Griffin. Understanding tourists spatial behaviour: Gps tracking as an aid to sustainable destination management. *Journal of Sustainable Tourism*, 21(4): 580–595, 2013.

K. Ellegård. A time-geographical approach to the study of everyday life of individuals-a challenge of complexity. *GeoJournal*, 48(3):167–175, 1999.

A. Franklin and M. Crang. The trouble with tourism and travel theory?, 2001.

C. Gibson. Locating geographies of tourism. *Progress in Human Geography*, 32(3):407–422, 2008.

M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *nature*, 453(7196):779, 2008.

M. F. Goodchild. Uncertainty: the achilles heel of gis. *Geo Info Systems*, 8(11):50–52, 1998.

T. Hägerstrand. What about people in regional science? In *Papers of the Regional Science Association*, volume 24, 1970.

C. M. Hall. Reconsidering the geography of tourism and contemporary mobility. *Geographical Research*, 43(2):125–139, 2005a.

C. M. Hall. *Tourism: Rethinking the social science of mobility.* Pearson Education, 2005b.

C. M. Hall. Framing tourism geography: Notes from the underground. *Annals of Tourism Research*, 43:601–623, 2013.

C. M. Hall and S. J. Page. *The geography of tourism and recreation: Environment, place and space*. Routledge, 2014.

S. Isaacman, R. Becker, R. Cáceres, M. Martonosi, J. Rowland, A. Varshavsky, and W. Willinger. Human mobility modeling at metropolitan scales. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 239–252. Acm, 2012.

A. Janecek, D. Valerio, K. A. Hummel, F. Ricciato, and H. Hlavacs. The cellular network as a sensor: From mobile phone data to real-time road traffic monitoring. *IEEE transactions on intelligent transportation systems*, 16(5):2551–2572, 2015.

B. L. Keeler, S. A. Wood, S. Polasky, C. Kling, C. T. Filstrup, and J. A. Downing. Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. *Frontiers in Ecology and the Environment*, 13(2):76–81, 2015.

Y. Kim, C.-k. Kim, D. K. Lee, H.-w. Lee, and R. I. T. Andrada. Quantifying nature-based tourism in protected areas in developing countries by using social big data. *Tourism Management*, 72:249–256, 2019.

M. U. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage, et al. The effect of human mobility and control measures on the covid-19 epidemic in china. *Science*, 2020.

A. Kuusik, M. Tiru, R. Ahas, and U. Varblane. Innovation in destination marketing: The use of passive mobile positioning for the segmentation of repeat visitors in estonia. *Baltic Journal of Management*, 6(3):378–399, 2011.

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

M. Li, S. Gao, F. Lu, and H. Zhang. Reconstruction of human movement trajectories from large-scale low-frequency mobile phone data. *Computers, Environment and Urban Systems*, 77:101346, 2019.

I. S. Lo, B. McKercher, A. Lo, C. Cheung, and R. Law. Tourism and online photography. *Tourism management*, 32(4):725–731, 2011.

X. Lu, C. Wang, J.-M. Yang, Y. Pang, and L. Zhang. Photo2trip: generating travel routes from geo-tagged photos for trip planning. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 143–152, 2010.

X. Lu, L. Bengtsson, and P. Holme. Predictability of population displacement after the 2010 haiti earthquake. *Proceedings of the National Academy of Sciences*, 109(29):11576–11581, 2012.

B. McKercher, N. Shoval, E. Ng, and A. Birenboim. First and repeat visitor behaviour: Gps tracking and gis analysis in hong kong. *Tourism Geographies*, 14(1):147–161, 2012.

N. Oliver, B. Lepri, H. Sterly, R. Lambiotte, S. Delataille, M. De Nadai, E. Letouzé, A. A. Salah, R. Benjamins, C. Cattuto, et al. Mobile phone data for informing public health actions across the covid-19 pandemic life cycle, 2020.

R. Pettersson and M. Zillinger. Time and space in event behaviour: Tracking visitors by gps. *Tourism Geographies*, 13(1):1–20, 2011.

E. Pultar and M. Raubal. Progressive tourism: Integrating social, transportation, and data networks. In *Tourism informatics: Visual travel recommender systems, social communities, and user interface design*, pages 145–159. IGI Global, 2010.

J. Reades, F. Calabrese, and C. Ratti. Eigenplaces: analysing cities using the space–time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, 36(5):824–836, 2009.

M. H. Salas-Olmedo, B. Moya-Gómez, J. C. García-Palomares, and J. Gutiérrez. Tourists' digital footprint in cities: Comparing big data sources. *Tourism Management*, 66:13–25, 2018.

S. Schönfelder and K. W. Axhausen. Activity spaces: measures of social exclusion? *Transport policy*, 10(4):273–286, 2003.

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

N. Shoval, B. McKercher, E. Ng, and A. Birenboim. Hotel location and tourist activity in cities. *Annals of Tourism Research*, 38(4):1594–1612, 2011.

S. Silm and R. Ahas. Ethnic differences in activity spaces: A study of out-of-home nonemployment activities with mobile phone data. *Annals of the Association of American Geographers*, 104(3):542–559, 2014.

C. Song, T. Koren, P. Wang, and A.-L. Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.

A. Tchetchik, A. Fleischer, and N. Shoval. Segmentation of visitors to a heritage site using high-resolution time-space data. *Journal of Travel Research*, 48(2):216–229, 2009.

E. Thulin, B. Vilhelmson, and T. Schwanen. Absent friends? smartphones, mediated presence, and the recoupling of online social contact in everyday life. *Annals of the American Association of Geographers*, 110(1):166–183, 2020.

M. Tiru, A. Kuusik, M.-L. Lamp, and R. Ahas. Lbs in marketing and tourism management: measuring destination loyalty with mobile positioning data. *Journal of Location Based Services*, 4(2):120–140, 2010.

H. Q. Vu, G. Li, R. Law, and B. H. Ye. Exploring the travel behaviors of inbound tourists to hong kong using geotagged photos. *Tourism Management*, 46:222–232, 2015.

Wikipedia. Chronological summary of the 2018 Winter Olympics, Feb. 2020. URL `https://en.wikipedia.org/w/index.php?title=Chronological_summary_of_the_2018_Winter_Olympics&oldid=940089111`. Page Version ID: 940089111.

S. Williams. *Tourism geography*. Psychology Press, 1998.

S. A. Wood, A. D. Guerry, J. M. Silver, and M. Lacayo. Using social media to quantify nature-based tourism and recreation. *Scientific reports*, 3:2976, 2013.

H. Xiao-Ting and W. Bi-Hu. Intra-attraction tourist spatial-temporal behaviour patterns. *Tourism Geographies*, 14(4):625–645, 2012.

Y. Xu, S.-L. Shaw, Z. Zhao, L. Yin, Z. Fang, and Q. Li. Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. *Transportation*, 42(4):625–646, 2015.

Y. Xu, S.-L. Shaw, Z. Zhao, L. Yin, F. Lu, J. Chen, Z. Fang, and Q. Li. Another tale of two cities: Understanding human activity space using actively tracked cellphone location data. *Annals of the American Association of Geographers*, 106(2):489–502, 2016.

Y. Xu, A. Belyi, I. Bojic, and C. Ratti. How friends share urban space: An exploratory spatiotemporal analysis using mobile phone data. *Transactions in GIS*, 21(3):468–487, 2017.

Y. Xu, A. Belyi, I. Bojic, and C. Ratti. Human mobility and socioeconomic status: Analysis of singapore and boston. *Computers, Environment and Urban Systems*, 72:51–67, 2018.

Y. Xu, A. Belyi, P. Santi, and C. Ratti. Quantifying segregation in an integrated urban physical-social space. *Journal of the Royal Society Interface*, 16(160):20190536, 2019.

L. Yan, D. Wang, S. Zhang, and D. Xie. Evaluating the multi-scale patterns of jobs-residence balance and commuting time–cost using cellular signaling data: a case study in shanghai. *Transportation*, 46(3):777–792, 2019.

Y. Yuan and M. Raubal. Extracting dynamic urban mobility patterns from mobile phone data. In *International Conference on Geographic Information Science*, pages 354–367. Springer, 2012.

Y. Yuan and M. Raubal. Analyzing the distribution of human activity space from mobile phone usage: an individual and urban-oriented study. *International Journal of Geographical Information Science*, 30(8):1594–1621, 2016.

Y. Yuan, M. Raubal, and Y. Liu. Correlating mobile phone usage and travel behavior–a case study of harbin, china. *Computers, Environment and Urban Systems*, 36(2):118–130, 2012.

Y.-T. Zheng, Z.-J. Zha, and T.-S. Chua. Mining travel patterns from geotagged photos. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(3):1–18, 2012.