

Article

# Exploring Travel Patterns during the Holiday Season—A Case Study of Shenzhen Metro System During the Chinese Spring Festival

Jianxiao Liu <sup>1</sup> , Wenzhong Shi <sup>1,\*</sup> and Pengfei Chen <sup>1,2</sup> 

<sup>1</sup> Department of Land Surveying and Geo-Informatics, Hong Kong Polytechnic University, Hong Kong 999077, China; jianxiao.liu@connect.polyu.hk (J.L.); chenpf9@mail.sysu.edu.cn (P.C.)

<sup>2</sup> School of Geospatial Engineering and Science, Sun Yat-Sen University, Guangzhou 510085, China

\* Correspondence: john.wz.shi@polyu.edu.hk; Tel.: +852-9799-5902

Received: 21 September 2020; Accepted: 21 October 2020; Published: 30 October 2020



**Abstract:** Research has shown that the growing holiday travel demand in modern society has a significant influence on daily travel patterns. However, few studies have focused on the distinctness of travel patterns during a holiday season and as a specified case, travel behavior studies of the Chinese Spring Festival (CSF) at the city level are even rarer. This paper adopts a text-mining model (latent Dirichlet allocation (LDA)) to explore the travel patterns and travel purposes during the CSF season in Shenzhen based on the metro smart card data (MSC) and the points of interest (POIs) data. The study aims to answer two questions—(1) how to use MSC and POIs inferring travel purpose at the metro station level without the socioeconomic backgrounds of the cardholders? (2) What are the overall inner-city mobility patterns and travel activities during the Spring Festival holiday-week? The results show that six features of the CSF travel behavior are found and nine (three broad categories) travel patterns and trip activities are inferred. The activities in which travelers engaged during the CSF season are mainly consumption-oriented events, visiting relatives and friends and traffic-oriented events. This study is beneficial to metro corporations (timetable management), business owners (promotion strategy), researchers (travelers' social attribute inference) and decision-makers (examine public service).

**Keywords:** vacation for holidays; travel patterns; urban mobility; travel activities; LDA; smart card

## 1. Introduction

Previously, traditional traffic surveys mainly collected data on people's workdays [1,2] and there were few surveys specifically for holidays (e.g., the 1995 American Travel Survey). As a consequence, research on travel behavior had focused on relatively habitual travel behavior, while some types of travel are more flexible and freer both in time and space, such as occasional weekend trips or holiday trips [3] and their characteristics and motivations were not fully understood [4]. However, with sightseeing, shopping and family gatherings have become the mainstream lifestyle of modern society, holiday travel demand has dramatically increased [5] and hence come into sight of policy-making organs and researchers. For example, in Germany, the National Household Travel Survey (MiD (Mobility in Germany)) includes long-distance travel information about public holidays such as Christmas. Likewise, the Reiseanalyse (RA) collects holiday behavior as well as the holiday interests and motivations of the German-speaking people in Germany [6]. Since behaviors like long-distance travel and leisure consumption are largely coupled with holidays, in many European countries, more and more large-scale household travel surveys have covered information on long-distance travel, such as the INVERMO (Germany), Micro Census (Switzerland) and MEST/TEST (France, Portugal, Sweden, UK) [7]. To be

more specific, LaMondia et al. [8] found that demographic factors, education level, employment factors and household income play important roles for leisure long-distance travel, which were helpful for long-distance survey makers to include better inquiries in their surveys. Reichert and Holz-Rau [9–11] found that, in Germany, people living in large cities are more likely to make long-distance travel for leisure purposes, while in Swiss, airport accessibility positively influences the probability to fly for leisure purposes [12]. A systematic review of urban form and long-distance leisure travel can be found in Reference [13]. Overall, prior studies and surveys enrich our understanding of the tourism-based holiday travel behavior (interurban, interprovincial or international).

However, there exists a distinction between daily travel behavior and tourism-related travel behavior. The abovementioned tourism-related long-distance travel, according to the World Tourism Organization, usually take place outside the usual environment of people [14], paying more attention to ‘tourists’ rather than ‘passengers’ on holidays, while the study of public transport in a usual environment (within the city) during vacations is poorly studied. Few efforts have been made to explore the differences in travel behavior between holidays and weekdays at the intra-urban level [15] and the travel patterns and travel activities of travelers on holidays remain unknown, especially for some important and long-lasting festivals, such as the Chinese Spring Festival (CSF) or Christmas Day. But considering the pervasive influence of public holidays on daily travel patterns [16], it is necessary to include the holiday effect in travel behavior studies [5,17]. To this end, this study benefiting from the metro transaction dataset collected by the automated fare collection (AFC) system in the urban metro system explores the travel patterns and travel activities in Shenzhen City during the CSF holiday season.

Nevertheless, there are still some challenges that have to be faced when applying metro transaction data to intra-urban holiday travel behavior studies. One obvious example is that, due to the lack of cardholders’ information and limited coverage of the metro stations in the urban area, it is hard to infer passenger’s travel activities at the metro station level [18]. Although a few attempts had been made to infer travel activities at the metro station level [19–21], relevant work almost had to rely on personal travel survey data, whose collection process is extremely time- and labor-consuming. A relatively quick and reproducible approach for this task is expected. Given this, we allocate the point of interests (POIs) data to each metro station to infer the probability of station-level travel activities, although this approach is by no means above perfect. The final destination or exact travel activities cannot be precisely determined due to people’s potential extra movements by other means of transportation (shared bike or bus) when they leave metro stations. However, via the POIs, the station-level travel activities can be further inferred from uncertain events to multi-events with a probability distribution (e.g., working: 60%, shopping: 30% and business: 10%).

This study seeks to answer two questions—(1) how can metro smart card (MSC) data and POIs be used to infer travel activities at the metro station level without the additional information of cardholders? (2) What are the overall inner-city mobility pattern and travel activities during the Spring Festival holiday-week? Answering these questions can help us better understand travel activities during the CSF season and provide timely information for adjusting metro service within the city in the holiday period from the perspective of metro passengers. Using a text mining technique to explore the travel patterns and trip activities of an important Chinese holiday season, the research aims to achieve the following contributions—(1) based on the various passenger groups, the overall inner-city mobility characteristics of the Spring Festival are described from three levels; (2) With the POIs-appended metro stations, travelers’ trip activities at the station level are inferred and passengers’ travel pattern difference between the holiday-week and the other two normal-weeks is revealed; (3) Latent Dirichlet allocation, a text mining technique is applied to explore the travel patterns and several policy implications are discussed.

The rest of this paper is organized into six parts. Section 2: a literature review of travel behavior studies about the holiday, the CSF and travel purpose are presented. Section 3: the study area and dataset are briefly introduced. Section 4: methodology, including data preprocessing, descriptive

analysis of holiday mobility features and trip activities exploring are discussed. Section 5: the results are presented and interpreted. Section 6: the essential findings and potential policy implications are summarized. Section 7: research limitations and lines of future research are identified.

## 2. Related Works

### 2.1. Holiday Travel Behavior

Previous holiday travel behavior studies are mainly carried out from a tourism research perspective. Consequently, space was not previous studies' concern because the space involved was usually not limited to a certain spatial scope due to tourists' various travel destinations (home or abroad). Instead, through online or offline survey data, relevant studies cared more about analyzing holiday traffic characteristics or holiday travel choice behaviors, such as the influence factors of tourists' travel destination/visiting choice [22–24], travel mode choice [23], travel planning [25,26], travel periodicity [22], travel purposes [27] and travel consumer behavior [28]. These tourism-based research were aimed to measure tourists' trip satisfaction, subjective consciousness or behavior patterns during holiday seasons rather than the spatiotemporal regularity of passengers, which is the concern of public transit.

Recently, holiday travel pattern studies from a perspective of public transit have received increasing attention as a matter of fact that daily travel pattern is inevitably influenced by holidays [16]. For example, during holiday seasons the traffic counts are extremely lower than workdays [17]. Also, travel decision-making, travel sensitive factors and travel dependence are different between daily commute travel and holiday non-commuter travel [15,29,30]. However, the 'travel pattern' in the context of relevant studies is mostly referred to as the travel mode choice [4,15,31], travel flow/counts [17,32] or travel time expenditure [16], whereas the spatiotemporal differences in the use of urban space and passengers' travel activities between holiday and common-day are seldom investigated. In general, current travel behavior studies about holidays, especially the city-level studies, are lack of the activity- and space-based perspective. That is, it is hard to answer what activities are passengers involved in during the holiday and common-days, respectively? And when and where these activities happen in the city? Filling this gap will help urban and transport planners and decision-makers further understand the use of urban space and passengers' corresponding activities in holidays and normal days because holiday pattern is a real running status that may describe some unconventional but important and easily neglected elements. In this sense, the holiday pattern offers an alternative view and an opportunity for transport planners and policymakers to examine and rethink the current public transit service they offered. This study takes the CSF as an example for further research, so previous CSF travel behavior related studies are reviewed below.

### 2.2. Chinese Spring Festival and Its Travel Behavior Studies

The reason we chose the CSF as the holiday season in this study is not merely due to the availability of research data; another significant reason is that the CSF for all the Chinese is the most representative and crucial holiday. In addition, many internationals could easily and correctly associate the CSF with the color red, lion dances, fireworks and red packets, while few could name other major Chinese holidays such as the Mid-Autumn Festival or the Dragon Boat Festival. Although the CSF celebration custom is slightly different depending on places in China, its uniqueness can be generally reflected in the following six aspects, which may affect mobility patterns. Firstly, the CSF has the longest period off usually at least for seven days officially, which produces abundant activities and travel during this period. Secondly, the CSF might be one of the few but important opportunities for a family reunion over a whole year [31] and this consciousness is rooted in the Chinese and prompts them to return to their hometowns from afar. Thirdly, the CSF has an impact beyond the official holiday because people usually begin preparing for it one or two weeks before it starts [33]. Fourth, everybody goes out to purchase necessities such as food materials, couplets and new clothes before or during the CSF.

Fifth, social interaction between people becomes physically frequent and intimate during the CSF. This is caused by the custom that people should visit their parents-in-law, relatives and close friends on the first or second day of the New Year. Lastly, there are varied activities that people would participate in such as going to the temple to burn joss sticks, strolling around the flower market and watching a fireworks show. The CSF periods are non-working and family-gathering days, so people may make some fully independent and unusual travel activities.

From the above, there is supposed to be a mobility variation during the CSF, while as a specific holiday mobility study, the CSF mobility studies are even rarer previously. Until 2014, with the advent of location-based service, the first study of the CSF travel rush was preliminarily carried out based on the Baidu migration data [33]. Through visualization and statistics of the travel flows between cities, Wang et al. (2014) [33] found the overall migration trend has a big fluctuation between one week before and one week after New Year's Day (which serves as the cut-off point). Besides, taking Guangdong Province, Beijing and Shanghai as instances, they found that the migration source and destination regions have characteristics of geographical proximity. Subsequently, based on location-based service data of the Baidu, Tencent and Qihoo platforms, Li et al. (2016) [31] applied the complex network and time-sequence analysis method to study the spatiotemporal characteristics of the travel peak during the CSF. They found that the CSF travel network at the provincial scale showed a multicenter and geographic clustering characteristic instead of the small-world and scale-free characteristics. Moreover, they noted the CSF travel network was more influenced by the socioeconomic factors rather than geographical location factors. Using complex network analysis and data mining techniques, Hu et al. (2017) [34] built an urban network of the population based on the Weibo social media data. They visualized the spatial and temporal network structure characteristics of human mobility from the perspective of society as a whole and explored the relationship between human mobility patterns and urban economic development. They found the CSF customs and traditions indeed have an influence on people's travel behavior and the key attraction to the floating population is from the eastern region of China, which showed that people tend to move from/to areas with a higher level of economic development. Similarly, Wei et al. (2018) [35] used the weighted network's rich club coefficient and normalized imbalance coefficient method to analyze the phenomenon and imbalance of the rich clubs in the population movement network during the Spring Festival of China in 2015.

Apparently, as a kind of specific example of holiday mobility study, the previous CSF mobility studies are mainly concentrated on a relatively large-scale study area (interprovincial or intercity) to reveal the phenomenon of regional economic unbalanced development during the rapid urbanization process of China, while few CSF studies have performed analyses at the inner-city level. Accordingly, policy implications derived from the large-scale area were usually limited at the large-scale area (national level) such as the household registration policy, industrial structure adjustment policy and so on. In turn, why the inner-city CSF study is important is that it may offer some insights to policymakers at the city level for inspecting the public services provided within the city.

### 2.3. Metro Travel Purpose Inference

The MSC dataset collected by the AFC system can be regarded as appropriate data with which to study the inner-city CSF mobility because the metro network is extensive and its demand is high. For example, in Shenzhen, the traffic volume of the metro system accounts for 14% of the total share. However, an intrinsic limitation of the MSC data is that it is hard to estimate metro passengers' final destinations, trip purpose and activity information [18], whereas they are important information to predict travel demand, to model travel behavior, to adjust transportation planning decision and so on. Recently, various new datasets were used to infer travelers' trip purposes, for example, the social media and online service data [36,37], mobile phone data [38–40], taxi trajectory data [41] and bike-sharing data [42,43], which have a relative more information (in space, time and flexibility) for inferring trip purpose compared to the MSC data. Meanwhile, due to the limited data information, many research methods applied to the above datasets are hard to be used for the MSC data.

Overall, inferring the trip purpose at the metro station level is a difficult task since the MSC data itself has limited information, so the MSC data is usually combined with personal travel survey (PTS) data to accomplish this task. In this case, combined with MSC, PTS and land-use datasets, Chakirov and Erath (2012) [19] applied the rule-based model and discrete choice model to detect activities of public transport passengers in the city-state Singapore. Taking the activity duration, activity start-time and land-use into account, they identified home and work activities and their locations. With the same data (MSC, TPS and land-use data), Alsger et al. (2018) [20] used the rule-based model to predict five trip purposes (work, education, shopping, home and recreational) in Brisbane, Queensland with an overall 78% accuracy and among them, the inference accuracy of work and home trips are up to 92% and 96%. Kusakabe and Asakura (2014) [21] did the task a little bit differently, they fused the PTS data with MSC data to estimate the trip purpose based on the naïve Bayes classifier and five major travel activities (go to work, go to school, leisure, business and returning home) were identified with a 76.8% accuracy.

Regarding the metro travel purpose inference, although a few attempts had been made previously, relevant work almost had to rely on the PTS and land-use data. But the collection process of the PTS data is time- and labor-consuming, while the land-use data has a relatively low spatial-resolution for trip purpose inference. On the other hand, previous works were mainly concentrated on inferring passengers' trip purposes on workdays, while holiday ones that might be the dead zone of urban public service were seldom investigated. Therefore, a method that relies on fewer datasets to infer travel purposes at the metro station level on holiday seasons is expected.

### 3. Study Area and Datasets

#### 3.1. Shenzhen and Shenzhen Metro System

Our study area, Shenzhen, is a highly developed city in China, with a total area of 1997 square kilometers. Shenzhen is located in the southern part of Guangdong province (Figure 1) and it is a link and a bridge connecting Hong Kong and the Chinese mainland. According to the Shenzhen Statistical Yearbook of 2017, there are approximately 12 million people living in this city. The first metro line in Shenzhen officially opened on 28 December, 2004. Presently, there are 8 metro lines in Shenzhen with 166 stations and a total length of 285 km (Figure 1). According to the Shenzhen Transport Annual Report 2016, the annual metro passenger traffic volume is 1297.13 million person-times and the daily average transport volume is 3.55 million person-times, accounting for near 14% of the total travel volume in Shenzhen. The share of the metro ridership (accounts for the total travel volume) may help readers to understand the impact of relevant conclusions in the following parts.



Figure 1. Location of Shenzhen and its metro lines.

### 3.2. Metro Smart Card and Pois Datasets

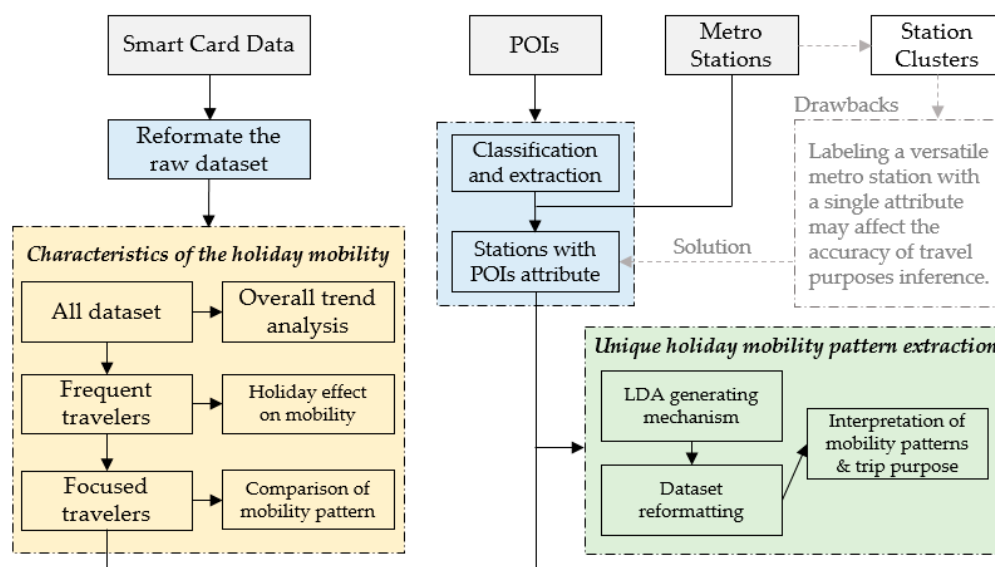
The metro smart card (MSC) dataset used in this study is the 3-week metro transaction records of 4,901,073 cardholders in Shenzhen (the second week is the Spring Festival holiday week). There are more than 50 million transaction records, covering 21 consecutive weekdays from 20 January to 9 February 2017 (27 January to 2 February is the holiday-week). Every time a traveler passes through the metro gantry, a transaction record is automatically collected. The 6 attributes contained in the records are shown in Table 1. The points of interest (POIs) data of Shenzhen City used in our research were collected from the Amap (<https://www.amap.com/>) in March 2019, with a total number of more than 700,000 data points, which is similar to a relevant study with 611,122 records [44]. The collected POIs data are separated into 17 categories and 140 subcategories.

**Table 1.** Record format of the smart card dataset.

Field	Value
Card_ID	Identifier of a unique cardholder
Trmnl_ID	Represents a unique subway station
Transaction_Time	Transaction time
Transaction_Type	Enter or exit station
Line	The corresponding line to the station

## 4. Methodology

Figure 2 shows the overview methodology framework of this research, which includes three parts (highlighted in blue, orange and green): data preprocessing; descriptive analysis of the holiday mobility characteristics; and holiday travel patterns and trip activities exploring.

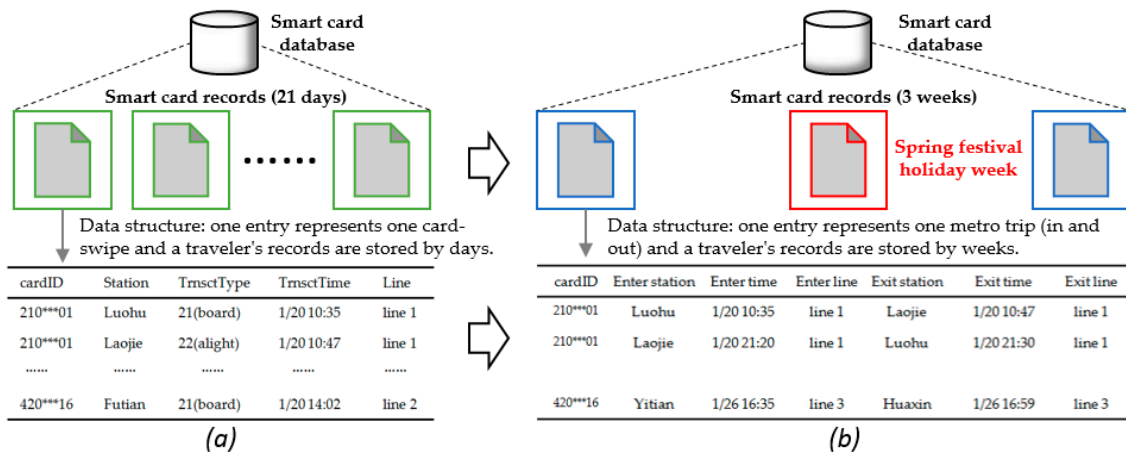


**Figure 2.** Research methodology framework of this paper.

### 4.1. Data Preprocessing

The basic step to derive the mobility pattern from the smart card dataset is to reconstruct the original data format. As described in Section 3.2, the original smart card data are stored by day and each entry represents one card-swiping act. However, this kind of data storage format is not efficient and convenient for knowledge mining. For instance, it is hard to identify a passenger's complete transaction record of multiday, which is listed in the time sequence. The objective of data reconstruction

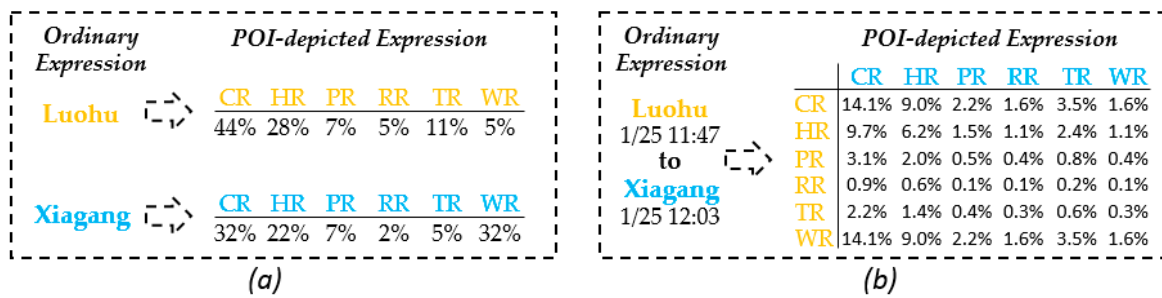
is to combine the weekly card-usage information of passengers. Thus, datasets are separated into 3 weeks (Figure 3b) from 21 separate days (Figure 3a).



**Figure 3.** Examples of the smart card data reconstruction. (a) original data storage form; (b) reconstruction of the data storage format.

Furthermore, considering the appropriate walking time (6–10 min) and the coverage of metro stations in Shenzhen (Figure 1), six types of POIs with a radius of 500 m centered on metro stations are counted. Thus, we can differentiate the characteristics of the 166 stations and inferring the travel purpose of passengers at the metro station level becomes possible, which makes the inferred travel purposes change from a single event (e.g., working) to multi-events with a probability distribution (e.g., working: 60%, shopping: 30% and business: 10%). Some key issues related to the processing of the POIs dataset are explained as follows:

- **Why POIs data instead of land use data?** The POIs categories are closely related to the current land-use types in China [45], although they are not exactly the same, the POIs can reflect the type of land use [46]. Meanwhile, compared to land use data, POIs data have several advantages: (1) the POIs data are much finer than land use data, which contain more useful information to further personal preference studies. (2) POIs data could represent the mixed-use situation instead of a single land-use type; therefore, potential activities for a certain land use are expanded and specified. (3) For researchers, POIs data are easier to obtain to carry out academic research.
- **What are the POIs-appended stations like?** As shown in Figure 4a, every metro station is changed from the text to a POIs feature vector. For example,  $S_1 = (CR_1, HR_1, PR_1, RR_1, TR_1, WR_1)$ , where  $CR_1, HR_1, PR_1, RR_1, TR_1$  and  $WR_1$  are the proportion of each POIs category around station  $S_1$  with a radius of 500 m. Therefore, the travel between two stations forms a matrix with 36 variables, which correspond to 36 potential types of activities (Figure 4b).
- **How many POIs categories are used in this study?** According to the Urban Land Classification and Construction Land Planning Standards (GB50137-2011), we re-categorize the 17 primary categories (140 subcategories) POIs into 6 types, namely, housing-related POIs (HR), work-related (WR), consumption-related (CR), recreation-related (RR), public-service related (PR) and traffic-related (TR).

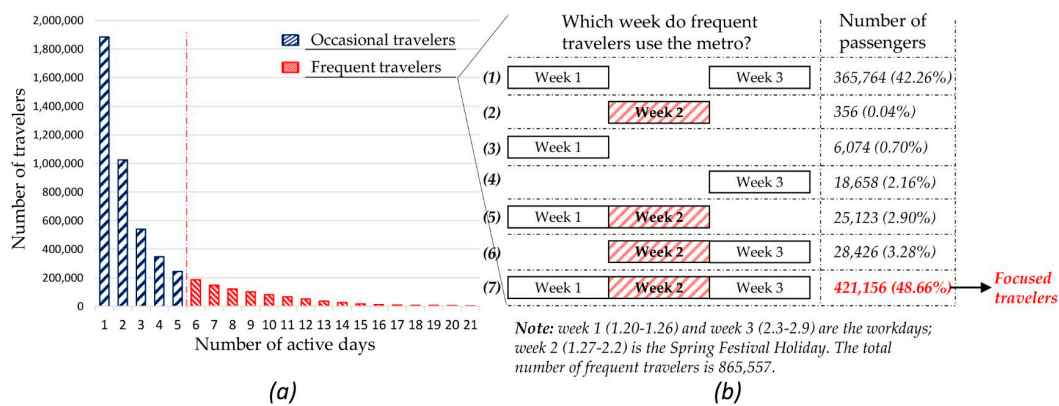


**Figure 4.** Examples of the points of interest (POI)-depicted metro stations. (a) metro station expressed by the POIs feature vector; (b) travel between two stations expressed by a 36-variable matrix.

#### 4.2. Descriptive Analysis of Holiday Mobility Characteristics

##### 4.2.1. Definition of Frequent and Focused Travelers

Occasional and infrequent metro passengers with a relatively infrequent travel frequency and lower periodicity are not informative for exploring their trip patterns. There is not enough information to extract or reveal any meaningful spatiotemporal patterns from their rare metro trips. To address this issue, we filter passengers from the whole dataset according to their metro travel frequency (active travel days) during the 21-day research period. For example, if a passenger uses the metro every day during these 21 days, his/her number of active days will be 21. Figure 5a shows the number of distributions of active days for all cardholders who used the metro once during the data collection period, which indicates that nearly 59.2% of passengers travel by metro only 1 or 2 days over these 21 days.



**Figure 5.** Distribution of active days and 7 types of frequent travelers. (a) Filter frequent travelers from all metro passengers; (b) among the selected frequent travelers, further filter the focused travelers who use the subway at least one time of each week.

However, due to the different dataset structure, study period and study area, the active day threshold values to filter frequent travelers from the whole dataset are varied in previous studies. For example, 6/21 (6 days out of 21 days), 7/30, 10/30, 2/30 and 10/28 are used in relevant studies [47–51]. By contrast, k-means clustering is a more objective method regarding its application in handling passengers with different travel frequencies [51,52]. Referring to the operation in previous studies, we apply the k-means algorithm to cluster passengers into two groups—one group is frequent travelers and the other is infrequent travelers. Consequently, frequent/infrequent travelers are divided by 6 (days) out of 21 days, indicating that a frequent traveler should use the metro at least 28.6% of the days during the 21-day research period. The red line in Figure 5a differentiates frequent/infrequent passengers; the figure shows that among the 4,901,073 passengers, 865,557 of them (17.7%) are frequent travelers and 4,035,516 of them (82.3%) are infrequent travelers.

According to whether people use the subway each week (which week do they take the subway), among the 865,557 frequent travelers, the metro-usage pattern can be divided into 7 types (Figure 5b). We further filter focused travelers who use the metro at least one day every week (Figures 5b, 6 and 7) from frequent travelers to further explore the variation of the mobility patterns of various weeks. In summary, from the whole database of 21 days (4,901,073 travelers), we define frequent travelers (865,557 travelers) and focused travelers (421,156 travelers) as follows:

- **Frequent travelers:** traveling by subway at least 6 days out of 21 days. ( $N > 5$ )
- **Focused travelers:** traveling by subway at least one day of every week. ( $N > 5$  and  $A > 0$  and  $B > 0$  and  $C > 0$ ), where  $N$  represents the total active days and  $A$ ,  $B$  and  $C$  are the active days in week 1, week 2, week 3, respectively.

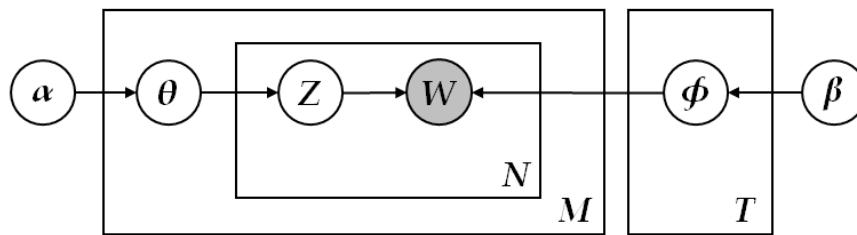


Figure 6. Graphical model representation of the latent Dirichlet allocation (LDA) generating mechanism.

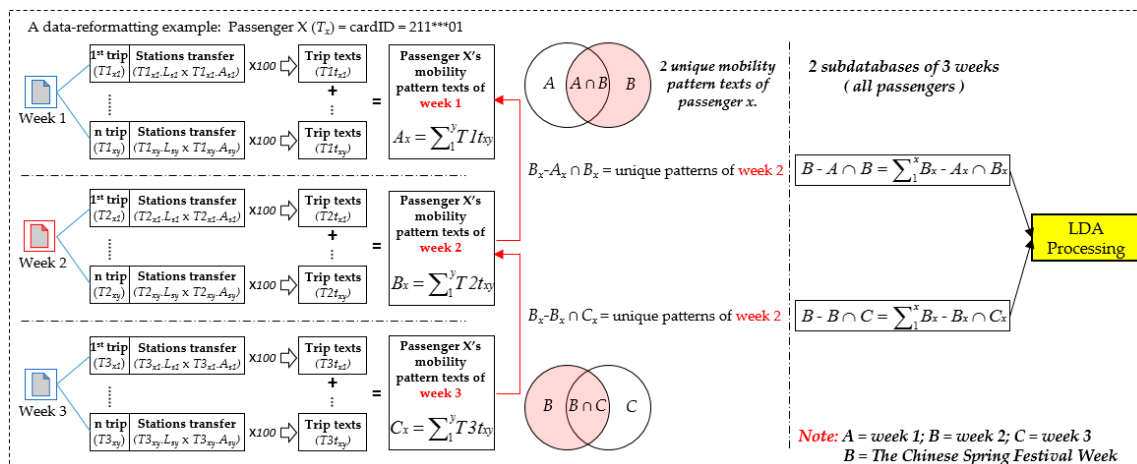


Figure 7. Process of reformatting the mobility pattern datasets.

#### 4.2.2. Evaluation of the Holiday Mobility Features

The holiday mobility characteristic is measured from three levels for three different groups of people: the holiday overall mobility trend of all travelers (4,901,073); the overall holiday effect on the mobility of frequent travelers (865,557); and the overall holiday mobility pattern of focused travelers (421,156).

Firstly, three indicators illustrate the overall mobility trend features of the holiday season from a macro perspective. The first indicator is the travel flow volume of each day, which is represented by the daily number of card-swiping records of all passengers. The second one is all passengers' daily average time cost on the metro for each trip. The last one is the influence scope of the holiday, which is reflected by the flow patterns consisting of the hourly passenger flow volume per day. Secondly, the travel frequency of frequent travelers is used to describe the overall holiday effect, which can help to understand how many frequent travelers are influenced by the holiday. Lastly, focused travelers' overall mobility patterns of three different weeks are compared. The overall mobility patterns are

represented by the POIs-depicted travel purpose, which is estimated by the shift between two metro stations with the appended POIs attribute.

### 4.3. Holiday Trip Activities Exploring through LDA Clustering

#### 4.3.1. Latent Dirichlet Allocation

The latent Dirichlet allocation (LDA) is a three-level hierarchical Bayesian model that refines probabilistic latent semantic analysis [53]. Unlike the decision tree, Bayes classifier and support vector machine (SVM), which have to predefine the classes, LDA generates possible patterns from the data itself and is a data-driven, unsupervised learning method. Furthermore, unlike other approaches, which are sensitive to outliers, LDA is insensitive to data noise and has efficient computing power for big data. Therefore, in our research context, LDA is regarded as an efficient and robust method for mining travel patterns and trip purposes. More information about the advantages of LDA against other unsupervised learning approaches is described by Bao et al. (2017) and Blei et al. (2003) [42,54]. LDA was initially used for mining text topics, while recently, researchers have applied it to transportation, mobility and urban studies. For example, Pereira et al. (2013) [55] used LDA to predict the traffic incident duration, Côme et al. (2014) [56] applied LDA to analyze the dynamic origin-destination and Hasan et al. (2014) [57] classified the urban activities with this method. In addition, Steiger et al. (2015) [58] applied LDA and spatial autocorrelation analysis to geo-referenced tweets to explore human social activities. Similarly, Fu et al. (2018) [59] used LDA to analyze the tweet data to differentiate different activity types.

The theory behind the LDA technique is that each document is treated as a random mixture of words and the latent topic is extracted by the word distribution probability in each document [54]. In the LDA model, a *document* consists of several *words* and the *words* of all documents form the *corpus*. Each *document* of the *corpus* has multiple *topics*, while each *word* of a *document* supports certain *topics* [60]. In the LDA model, there are several parameters used in the generating process.  $T$  is the number of topics in each document.  $V$  is the total number of words in the corpus  $D$ .  $M$  is the number of all documents.  $\theta$  is the  $T \times M$  matrix, which is the topic proportions for the  $T$  topics in each document.  $\phi$  is the  $V \times T$  matrix, which is the distribution of  $V$  in each topic.  $Z$  is the topic assignment for each document and  $W$  is the observed words for each document.  $\alpha$  and  $\beta$  are the prior-parameters for  $\theta$  and  $\phi$ , respectively, which both follow the Dirichlet distribution and  $\alpha = 50/T$  and  $\beta = 0.1$  are the most commonly used values in related studies (e.g., References [42,57,61]). According to the notations above, the generating mechanism of LDA can be explained as a graphical model representation (Figure 6):

- For each topic  $t \in \{1, 2, \dots, T\}$ , draw  $\theta_t \sim \text{Dirichlet}(\beta)$
- For each document  $m \in \{1, 2, \dots, M\}$  in the corpus, draw  $\theta_m \sim \text{Dirichlet}(\alpha)$
- For each word  $n \in \{1, 2, \dots, N\}$  in a certain document:
  1. Draw  $Z_n \sim \text{Multinomial}(\theta_m)$
  2. Draw  $W_n \sim \text{Multinomial}(\phi_{Z_n})$

Except for  $\alpha$  and  $\beta$ , the topic number ( $T$ ) also needs to be predetermined. According to previous studies, perplexity can be applied to infer a reasonable value of  $T$  and it is one of the most widely used evaluations of LDA [54]. In general, a better LDA clustering result matches a lower value of perplexity, indicating a better predictive performance and fewer uncertainties in the model. Perplexity can be calculated by the following formula:

$$\text{Perplexity} = \exp \left\{ - \frac{\sum_{m=1}^M \log p(w_m)}{\sum_{m=1}^M N_m} \right\}, \quad (1)$$

where  $M$  is the number of documents,  $N_m$  is the document length (document-specific) and  $p(w_m)$  is the likelihood of a text document of the corpus.

#### 4.3.2. Data Reformatting and Mobility Pattern Clustering

Most previous research merely applies LDA for one dataset to discover pattern (topics) characteristics. Few researchers have tried to explore variations for multiple datasets through LDA. Based on the POIs-appended metro stations and the reconstructed data over three weeks presented in Section 4.1, we apply LDA twice to detect the mobility patterns of three separate weeks (week 1–3) and to explore the travel purpose during the holiday-week (week 2).

Before applying LDA, the previously processed datasets have to be reformatted. Slightly different from previous studies, we apply the LDA twice (in total) to two sub-databases. The first application is to detect the holiday mobility pattern based on the comparison of week 1 & week 2 and the second application is based on the comparison of week 2 & week 3. Two sub-databases and how they are reformatted from the processed data are introduced. The preliminaries described below and Figure 7 aid in understanding this process.

**Definition 1 (Trip).** A trip  $T$  contains four items: leaving station  $L_s$ , arriving station  $A_s$ , leaving time  $L_t$  and arriving time  $A_t$ .  $T_1$ ,  $T_2$  and  $T_3$  represent the trips that happened in the first, second and third weeks of the research period, respectively. For example, the second trip in the first week of passenger  $X$  is  $T1_{x2} = (L1_{s2}, A1_{s2}, L1_{t2}, A1_{t2})$ .

**Definition 2 (POIs-depicted shift between two stations).** A POIs-depicted shift between two stations is  $ST = (L_s \times A_s)$ .  $L_s = (C_L: \_, H_L: \_, P_L: \_, R_L: \_, T_L: \_, W_L: \_)$ , where  $C_L$ ,  $H_L$ ,  $P_L$ ,  $R_L$ ,  $T_L$  and  $W_L$  represent the percentage of consumption, housing, public service, recreation, traffic and working related POIs within the radius of 500 m with the leaving station as the center, respectively. Similarly,  $A_s = (C_A: \_, H_A: \_, P_A: \_, R_A: \_, T_A: \_, W_A: \_)$ . Thus, a POIs-depicted shift between two stations is a matrix with 36 variables.

**Definition 3 (Trip texts).** The trip texts  $T_e$  are generated from the  $ST$  (see definition 2) and the arriving time  $A_t$ .  $T_e = \text{Textualize} \{ [\text{time slot}(A_t) + ST] \times 100 \}$ , where **Textualize** is the process of translating the numeric variables into text characters and **time slot** is the process of making the timestamp into temporal intervals (e.g., 08:30 change to 08:00–09:00 and the 08:00–09:00 is expressed by 08 for later processing).

**Definition 4 (Travel pattern texts).** The travel pattern texts are the sum of the trip texts of a certain passenger in a week. The travel pattern texts describe the travel pattern by the texts that contain the POIs-appended shift between two stations and the corresponding arriving time. The pattern texts can help to explore the possibility of travelers' potential activities. Generally, the more patterns there are for a certain type of text, the more likely that kind of activity is engaged. For example, '17WH' means arriving at a housing-related station from a working-related station at 17:00–18:00.

Based on the above definitions, any specific trip can be represented by several texts and the number of texts represents the possibility of potential activity. The larger the number of a certain type of text, the higher possibility of a corresponding activity. An example reveals this process as follows and the value on the upper right of the text represents the occurrence number of the text and this number is a reflection of the probabilities.

$$\text{Textualize} \{[\text{time slow} (14 : 23) + \begin{bmatrix} - & C_L & H_L & P_L & R_L & T_L & W_L \\ C_A & 14\% & 9\% & 2\% & 2\% & 4\% & 2\% \\ H_A & 10\% & 6\% & 1\% & 1\% & 2\% & 1\% \\ P_A & 3\% & 2\% & 1\% & 0 & 1\% & 0 \\ R_A & 1\% & 1\% & 0 & 0 & 0 & 0 \\ T_A & 2\% & 1\% & 0 & 0 & 1\% & 2\% \\ W_A & 14\% & 9\% & 2\% & 2\% & 4\% & 2\% \end{bmatrix} \times 100]\} \quad (2)$$

$$= \begin{bmatrix} 14CC^{14}; 14CH^9; 14CP^2; 14CR^2; 14CT^4; 14CW^2 \\ 14HC^{10}; 14HH^6; 14HP^1; 14HR^1; 14HT^2; 14HC^1 \\ 14PC^3; 14PH^2; 14PP^1; 14PT^1 \\ 14RC^1; 14RH^1 \\ 14TC^2; 14TH^1; 14TT^1; 14TC^2 \\ 14WC^{14}; 14WH^9; 14WP^2; 14WR^2; 14WT^4; 14WC^2 \end{bmatrix}$$

**Definition 5 (Mobility pattern variation texts).** The mobility pattern variation texts are the difference of a certain traveler's mobility pattern texts between two weeks. These texts can describe the difference in travel behavior between two weeks. For example, for a passenger, if the mobility text '23CH' does not exist in week 1 but appears in week 2, then '23CH' is his/her mobility pattern variation text between week 1 and week 2, which can also be regarded as his/her mobility pattern in week 2. Following this definition, we generate 2 kinds of variation texts to reflect the mobility pattern of the holiday-week: (1) mobility pattern texts in week 2 based on the comparison of week 1 and week 2; (2) mobility pattern texts in week 2 based on the comparison of week 2 and week 3. Accordingly, we can perform a pairwise comparison (week 1 & week2 and week 2 & week 3) to obtain the holiday mobility patterns and then explore their corresponding travel purpose.

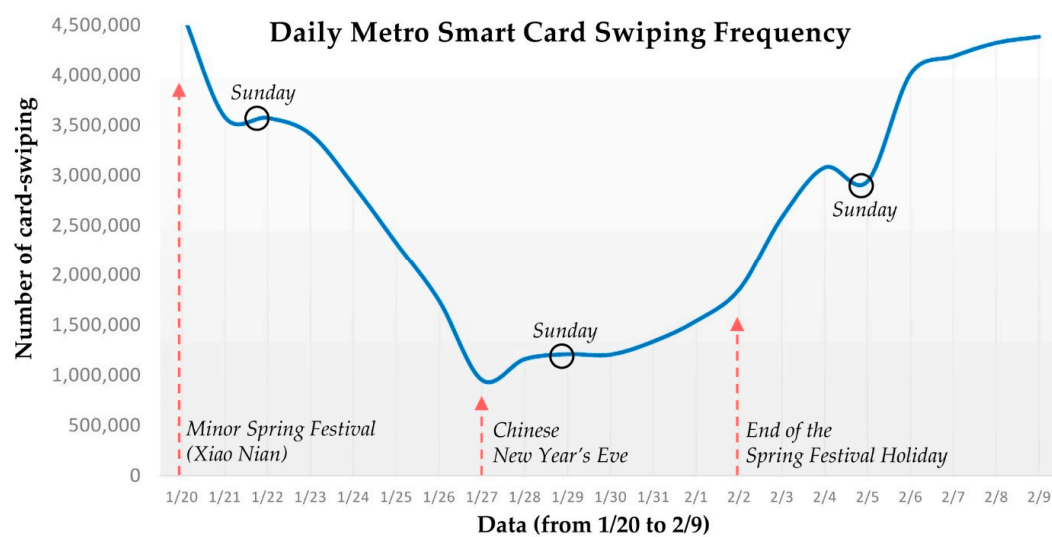
Based on the above definitions, in our research context, the variation texts derived from all passengers' mobility patterns make up 2 sub-databases for LDA processing: (1) *U21*: all focused passengers' pattern variation texts of week 2 compared with week 1; (2) *U23*: all focused passengers' pattern variation texts of week 2 compared with week 3. Every passenger (*cardID*) is a **document**, the variation trip text is a **word** of a **document** and all trip texts form the **corpus**. After clustering the mobility patterns through LDA, 2 groups of LDA clustering results are generated. The patterns of these two groups are mobility patterns and only appear during the holiday season. Through the interpretation of these mobility patterns, we can determine what travel purpose and activities passengers are most likely engaged in.

## 5. Result

### 5.1. Overall Holiday Mobility Characteristics

#### 5.1.1. Lower Travel Frequency

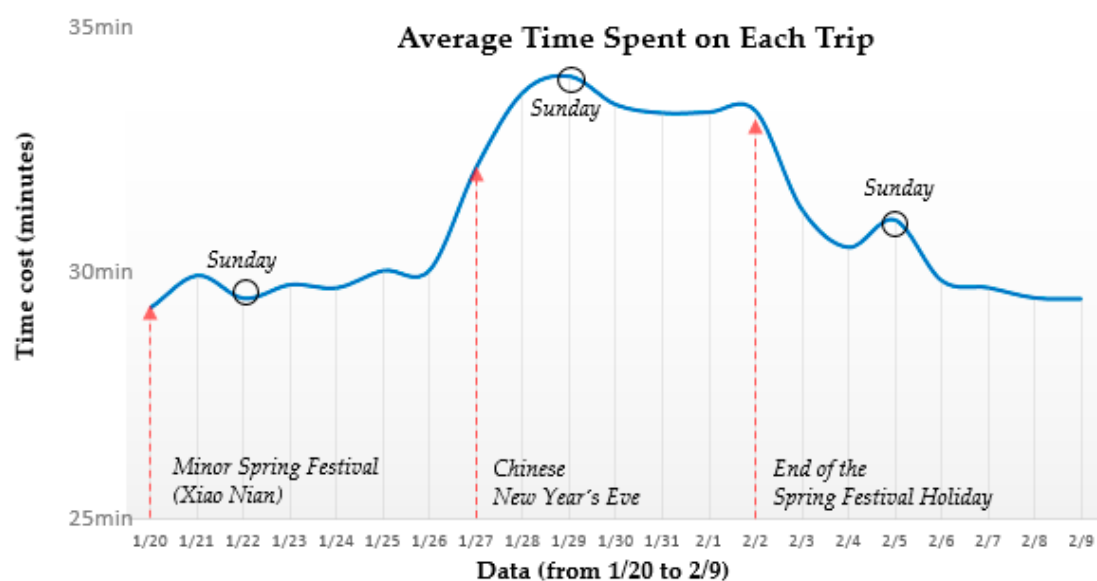
The mobility volume of each day according to the card-swiping frequency is counted (Figure 8), which shows that the travel frequency is significantly reduced since the Minor Spring Festival (1/20, also called Xiao Nian), reaching the lowest level on Chinese New Year's Eve (1/27) and gradually recovering after the end of the holiday. In general, the travel frequency during this holiday is much lower than that of normal weeks, indicating a large influence of this holiday on travel decisions and the findings are consistent with previous studies [33].



**Figure 8.** Daily metro smart card swiping frequency of 21 days.

### 5.1.2. Longer Time for Each Trip

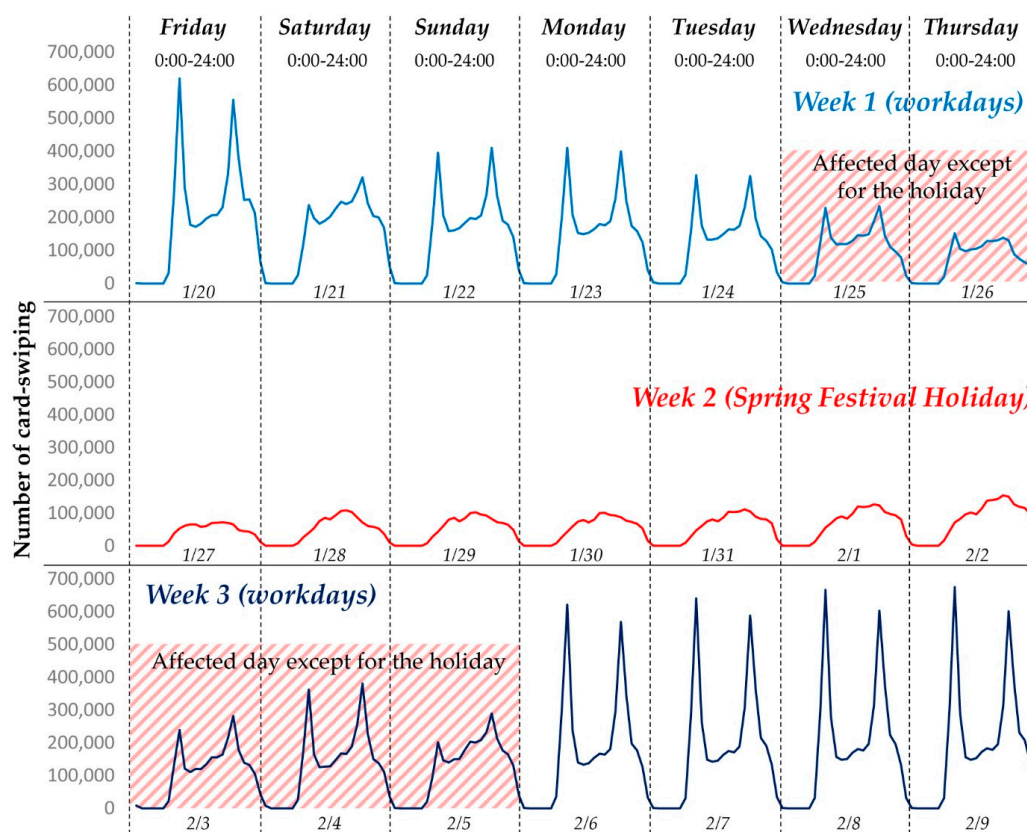
Another feature of Spring Festival travel is the average time travelers spent on each trip. Figure 9 shows that the travel time reaches a peak during the Spring Festival holiday season, which may indicate passengers travel longer distances in the city. Passengers spend a significantly longer time per trip. Besides, small samples can examine our above findings. We randomly select 1000 focused travelers from the datasets and we find their average travel frequency and travel time in the normal week (week 1) is 6.5 times/week and 27.2 min/week, respectively. However, during the holiday the travel frequency and travel time is 4.0 times/week and 32.0 min/week. Although according to holiday regulations, employees cannot take leave on the weekend before and after the holiday, in Figure 9 there still a little bit fluctuation in the travel time. This may due to some private enterprises and self-employed individuals who do not have to follow the deferred-holiday rules.



**Figure 9.** Daily average time spent on each trip of 21 days.

### 5.1.3. Influence Beyond the Holiday Season

In terms of the hourly number of passengers each day, according to prior studies [62–64], the travel volume patterns from Monday to Friday should be roughly the same, indicating a normal workday. However, in the 3-week study period, only 1/20 and 2/6 to 2/9 maintain a relatively regular state, the volume patterns of other dates are obviously interfered with by the Spring Festival holiday (Figure 10). It shows that before the official start or end of the holiday, even if it is still a working day, the volume pattern has been affected, such as 1/25 and 1/26 in week 2 and 2/3 to 2/5 in week 3. This indicates that the impact of the Spring Festival is continuous and that the scope of the impact is not limited to the holiday period. One possible reason for this finding is that passengers stop working to prepare for the festival before its official start and gradually return to work after its official end. This phenomenon, in which the travel volume pattern is disturbed for nearly half a month, has not been studied in related research.

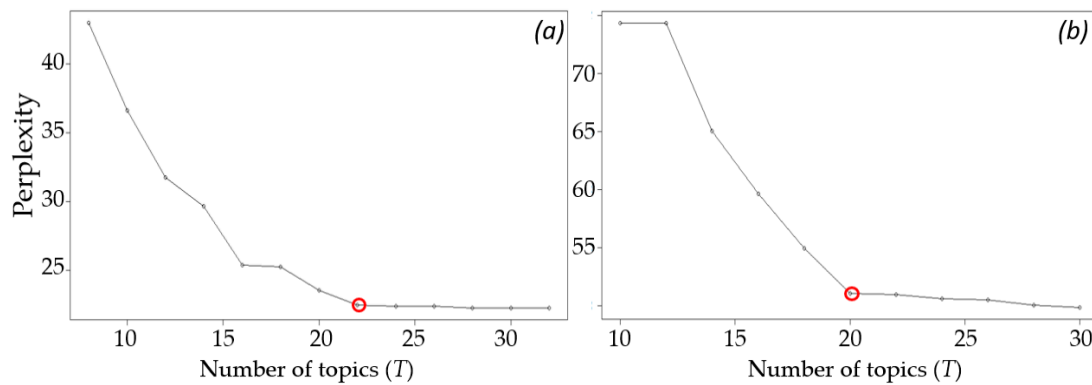


**Figure 10.** The hourly number of passengers on each day of the 3 weeks.

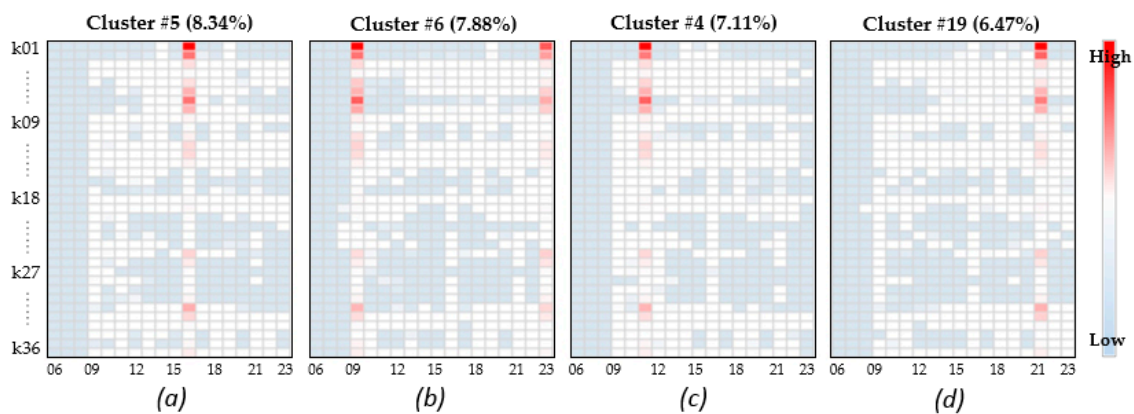
## 5.2. Travel Purpose Inference by the Mobility Patterns of the Holiday-Week

### 5.2.1. Holiday Mobility Patterns Compared to Week 1

In the mobility pattern comparison between week 1 and week 2 (U21), among the 421,156 focused travelers, 409,410 (97.2%) have mobility patterns during the holiday-week (week 2). The perplexity reaches the minimum when  $T$  is set to  $T = 22$  for the U21 (Figure 11a), which means 22 mobility patterns in the holiday-week do not exist in the week before the holiday (week 1). Although the LDA generated 22 mobility patterns, we only interpret the details of the most important 4 mobility patterns in Figure 12. The 18 remaining patterns have a similar structure with the four above and their interpretation can be easily understood from the figures.



**Figure 11.** Perplexity value for U21 and U23. (a) Perplexity value for U21; (b) Perplexity value for U23.



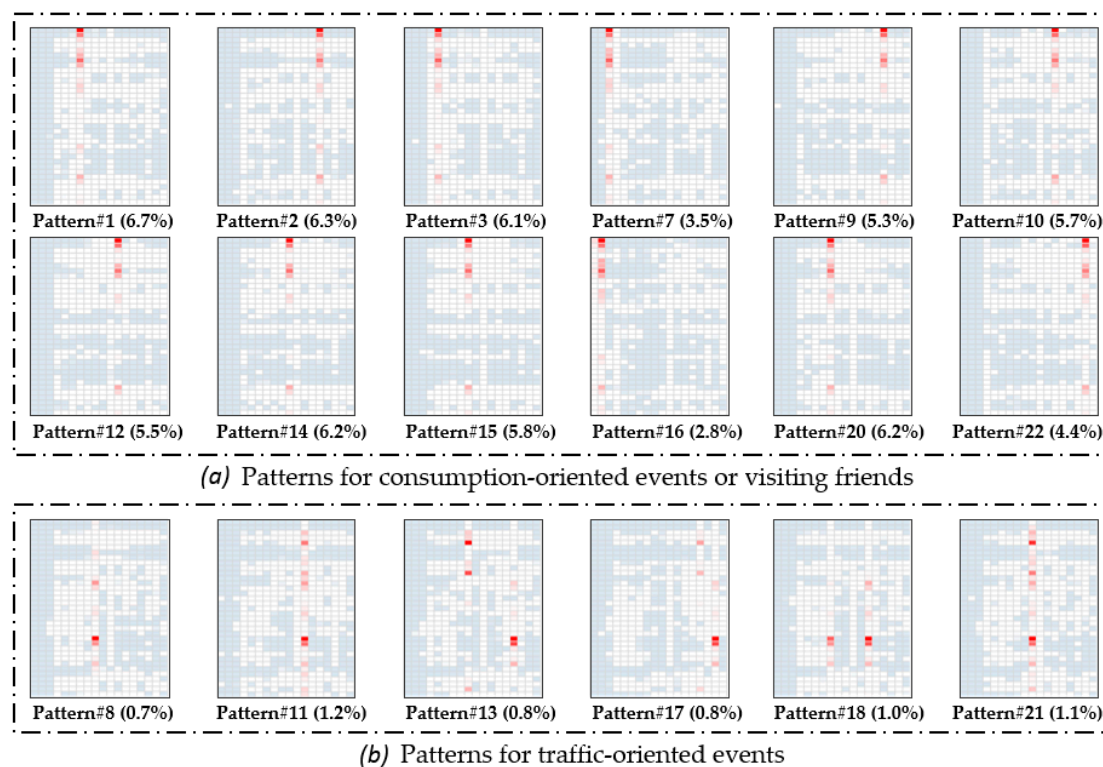
**Figure 12.** Examples of the mobility patterns on the holiday. (a) mobility pattern of cluster 5; (b) mobility pattern of cluster 6; (c) mobility pattern of cluster 4; and (d) mobility pattern of cluster 19.  
**Note:** (1) k1 = C-C, k2 = C-H, k3 = C-P, k4 = C-R, k5 = C-T, k6 = C-W, k7 = H-C, k8 = H-H, k9 = H-P, k10 = H-R, k11 = H-T, k12 = H-W, k13 = P-C, k14 = P-H, k15 = P-P, k16 = P-R, k17 = P-T, k18 = P-W, k19 = R-C, k20 = R-H, k21 = R-P, k22 = R-R, k23 = R-T, k24 = R-W, k25 = T-C, k26 = T-H, k27 = T-P, k28 = T-R, k29 = T-T, k30 = T-R, k31 = W-C, k32 = W-H, k33 = W-P, k34 = W-R, k35 = W-T, k36 = W-W;  
 (2) every grid in the time axis represents a time-slot, e.g., 06 = 06:00–07:00.

As shown in Figure 12a, mobility pattern #5 reveals that on the holiday, passengers have the greatest possibility of arriving at consumption-related metro stations from housing-related stations (H-C) or other consumption-related stations (C-C) at 16:00–17:00, which implies that the travelers' travel purpose at this time is mainly consumption-oriented events. Moreover, other high possibility shifts are also included. (1) Arriving at housing-related stations from consumption-related stations (C-H) at 16:00–17:00. (2) Arriving at housing-related stations from other housing-related stations (H-H) at 16:00–17:00. (3) Arriving at consumption-related stations from traffic-related stations (C-T) at 16:00–17:00. Mobility pattern #5 indicates that the activities in which people commonly participate in at 16:00–17:00 during the holiday-week are consumption-related events (shopping, eating and so on) or visiting friends and/or family (H-H). On the other hand, the time from 16:00–17:00 is when travelers are often working in a normal week, which leads to a particular mobility pattern during the holiday-week.

Compared with the mobility pattern #5, patterns #6, #4 and #19 (Figure 12b–d) show similar travel purposes, namely, consumption-related activities (reflected by C-C, H-C, etc.), going home after consuming (C-H) and visiting friends (H-H). The difference among them is the choice of arriving time for these activities. For example, pattern #6 makes it clear that passengers arriving at one place at 9:00–10:00 and then return home at 23:00–24:00. Cluster #4 implies people would go out slightly late (11:00–12:00) on the holiday and Cluster #19 indicates people would come home late at night

(21:00–22:00). These patterns are not only mobility patterns that the week before the holiday does not have but also reveal the reasons why travelers might travel on a holiday.

Furthermore, the remaining 18 holiday mobility patterns are listed in Figure 13. Compared to week 1, the time passengers choose to travel on holiday is relatively balanced and few travelers on the holiday would take the metro at 07:00–09:00. Instead, 9:00–10:00, 16:00–17:00 and 23:00–24:00 are the three peak periods when most of the travelers choose to travel, revealing that few people took the metro during these periods during week 1. Overall, these 22 mobility pattern clusters can be divided into 3 types. These 3 types can also explain why holiday-travel is distinct as well as the trip purposes of travelers.



**Figure 13.** Remaining 18 mobility patterns on the holiday (U21).

In the first and second types, travelers choose other time to travel on holidays and their trip purposes are mainly consumption-oriented events (C-C, H-C, T-C, etc.) or visiting friends (H-H). These two types are represented by patterns #1, #2, #3, #4, #5, #6, #7, #9, #10, #12, #14, #15, #16, #19, #20 and #22 (Figures 12 and 13a) and passengers associated with these patterns account for 94.5% (386,748/409,410). Two factors cause the uniqueness of these mobility patterns. One is that the time corresponding to travelers taking the metro on the holiday is different from one week before the holiday and the other is the different travel purposes. For example, people in pattern #5 did not travel at 16:00–17:00 during week 1 and most of their trip purposes were not consumption-oriented events or visiting friends. They may travel at the same time (16:00–17:00) for other purposes (e.g., working: H-W) or travel at another time for the same purposes.

In the third type, travelers change their trip purposes on the holiday and most of the trip purposes on the holiday are traffic-oriented events (T-C, T-H, C-T and H-T). This type is represented by patterns #8, #11, #13, #17, #18 and #21 (Figure 13b) and travelers involved in these patterns account for 5.5% (22,662/409,410). The reason why these patterns are distinct on the holiday is mainly due to travelers' trip purposes being different from week 1. These travelers' trip purposes are mainly traffic-oriented events, indicating leaving or arriving in Shenzhen during the CSF holiday week.

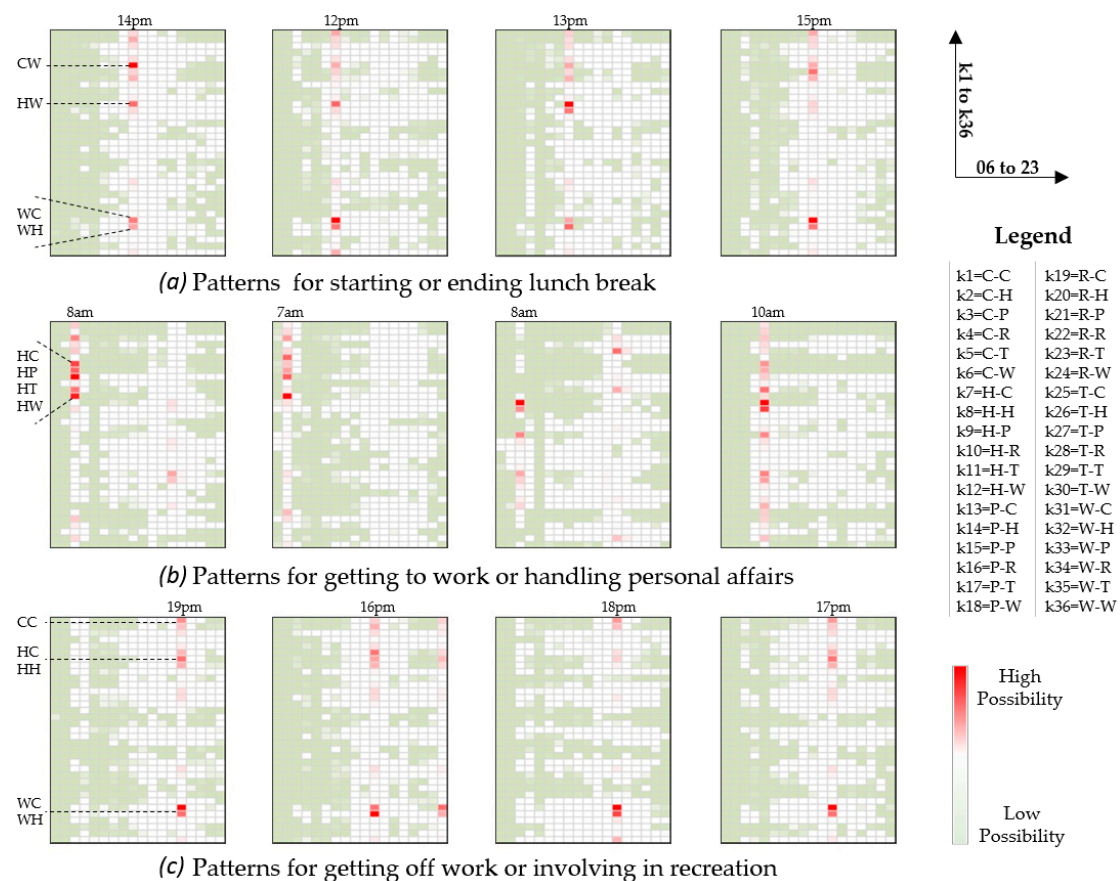
### 5.2.2. Holiday Mobility Patterns Compared to Week 3

By comparing the mobility patterns between week 1 and week 2, we obtained some of the primary holiday travel purposes from 22 holiday mobility patterns (U21). However, a concern is raised here: do these mobility patterns and travel purposes remain the same between week 2 and week 3 (U23)? To address this concern, we also compare the mobility patterns between week 2 and week 3 to check the holiday mobility patterns and travel purposes. In the LDA clustering processing for U23, the perplexity reaches the minimum when  $T$  is set as  $T = 20$ . Thus, 20 distinct holiday mobility patterns are generated between week 2 and week 3 and most are consistent with the previous 22 holiday mobility patterns (U21). This finding means that, similarly, the 20 mobility patterns can also be divided into 3 types and 96.9% (395,618/408,258) are the consumption-oriented and visiting-related types and 3.1% (12,667/408,258) are the traffic-oriented type. These three approximate mobility types of U21 and U23 indicate that the trip purposes during the CSF in Shenzhen between week 1 and week 2, week 2 and week 3 are almost the same.

Therefore, we can state that the holiday travel purposes of focused travelers in order of possibility mainly include the following: (1) C-C: consuming or relaxing at various commercially oriented places, such as different shopping malls. (2) H-C: going to consumption-oriented places from home. (3) C-H: coming back home or visiting friends after consuming. (4) C-W: going to work after consuming. (5) W-C: going to consumption-oriented places after working. (6) H-H: going to visit friends. (7) T-C: going to consumption-oriented places from transportation hubs. (8) T-H: going back home from transportation hubs. (9) C-T: going to transportation hubs after consuming. The (1)–(5) are the consumption-related type, (6) is the visiting-related type and the (7)–(9) are the traffic-related type.

### 5.3. Mobility Patterns of the Normal-Day (Week 1)

To illustrate the different patterns of the Holiday and Normal-day, we extract the mobility pattern of weekday (week 1) with the same clustering method as well. Except for some similar patterns of consumption events, Figure 14 shows some weekday patterns have obvious working- and housing-related attributes, which are not prominent in holiday patterns. These weekday patterns could be roughly divided into three groups. The first group is the patterns for passengers starting or ending their lunch break (C-W, H-W, W-C, W-H) and the time is usually distributed between 12:00 to 15:00 (Figure 14a). The second group is the patterns for passengers getting to work (H-W) or going out to handle their personal affairs (H-C, H-P, H-T) and these activities usually happen from 7:00 to 10:00 (Figure 14b). The last group is the patterns for passengers getting off work (W-C, W-H) or involving in recreation at night (H-C, C-C) and they usually happen from 16:00 to 19:00 (Figure 14c). This comparison shows that combining the multiply result of the POIs ratio of OD station pairs, the arriving time of the station pairs and the day type (holiday or weekday), our proposed method could help to estimate the different travel patterns and activities at the metro station level.



**Figure 14.** Working- and housing-related mobility patterns of the weekday (week 1).

## 6. Summary of Results and Policy Implications

To date, public transport during vacations is poorly studied. Accordingly, this study performs experiments on two datasets, namely, the metro smart card (MSC) dataset and the POIs dataset, to explore travel patterns and travel activities during the CSF holiday season in Shenzhen. The main findings of this study in terms of our research question can be summarized as follows:

Firstly, with MSC and POIs datasets only, the proposed strategy appending the POIs attribute to metro stations with a radius of 500 m can make the trip purpose inference change from a single event to multi-events with a possibility distribution at the station level. The holiday travel purposes of focused travelers are inferred through clustering analysis and the results highlight the uniqueness of the Spring Festival travel compared to one week before and one week after the holiday. Three general types of travel patterns are revealed: consumption-oriented or visiting-friends events and traffic-oriented events. Among the three types of patterns, nine primary travel purposes are discovered. Secondly, we define frequent travelers and focused travelers to separate them from all passengers according to their travel active days during the Spring Festival. Then, six characteristics of holiday mobility are measured in these three groups of passengers: (1) Extremely lower travel frequency; (2) Longer trip time. (3) Relatively long-lasting influence beyond the holiday; (4) About 50% of frequent travelers stop their metro travel on holidays; (5) The holiday mobility pattern of focused travelers is distinct from those of the other two normal-weeks in three aspects, namely, travel purposes, travel stability and travel peak time; (6) The time of going out and coming back home are both late on holidays.

Our analysis is beneficial to metro corporations (timetable management), business owners (promotion strategy), researchers (travelers' social attribute inference) and decision-makers (to examine public services). Firstly, the analysis could provide information to adjust the MTR holiday services. Although traffic-oriented activities have a great occurrence during the CSF is nearly a common-sense

phenomenon, while beyond that the MTR corporations could identify where and when these activities are generating within the city from the passenger clusters of the analysis, then making corresponding countermeasures. Secondly, from an urban planning perspective, this study could offer help to optimize the pattern of urban retail business centers. Because for most urban planners (at least in China), holiday demand is what barely considered factors during the planning process. So if planners consider this factor, they may find some potential gaps lies in the current planning and then to better optimize the planning. After all, except for CSF, there are many other public holidays in China. Thirdly, the study helps researchers and policymakers to better detect different social groups and then deliver targeted social welfare, because travelers' trip purposes during a special period (holiday season) can reflect their socioeconomic classes in some ways. For instance, since the CSF is very important to all the Chinese and it is a rare opportunity for people to relax and reunion, so unusual travel behavior during this period would capture people's attention, which may also imply some essential underlying mechanism. Therefore, the analysis could be considered in the process of allocating social welfare such as the application of affordable housing and concessionary metro fare.

## 7. Limitations and Further Steps

Nevertheless, our study is an initial step of exploring different travel patterns between weekdays and a long holiday season. Based on our work presented herein, several improvements could be made. First, since the traffic volume of the metro system in Shenzhen only accounts for 14% of the total share, our results and conclusion may be confined to the subway system and have the limitation of generalization. Second, the usage of the POIs in the study has its innate limitations. Although the POIs might be more appropriate data than land-use data to infer travel activities, while POIs are applied only by the number is not accurate enough. Because information such as the scale and the size of POIs are not included. For example, a train station may be represented by only a few POIs but they contain more meaning than lots of restaurants' POIs. Third, due to the special development background of Shenzhen City (a large number of migrants), the CSF mobility pattern of other Chinese cities is worth exploring because they might have some common features or differences that need to be further discussed. Lastly, since the proposed approach to infer travel activities Future studies can proceed from the following aspects: (1) Adding some other mobility data sources such as the taxi trajectory, bus smart card and bike-sharing or combine some personal travel survey data to better understand the overall CSF inner-city mobility; and (2) Extending the holiday travel studies into a large-scale area such as the national level and also includes other public holiday seasons.

**Author Contributions:** Jianxiao Liu and Pengfei Chen conceived the presented idea; Jianxiao Liu wrote the manuscript with support from Wenzhong Shi. All the authors have approved the final version of this manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the Ministry of Science and Technology of the People's Republic of China (2019YFB2103102), The Hong Kong Polytechnic University (4-BCF7, 1-ZVN6; Smart Cities Research Institute, The Hong Kong Polytechnic University) and The State Bureau of Surveying and Mapping (1-ZVE8).

**Acknowledgments:** The authors are grateful to the four reviewers for their helpful comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Szeto, W.; Yang, L.; Wong, R.; Li, Y.; Wong, S. Spatio-temporal travel characteristics of the elderly in an ageing society. *Travel Behav. Soc.* **2017**, *9*, 10–20. [[CrossRef](#)]
2. Yang, L. Modeling the mobility choices of older people in a transit-oriented city: Policy insights. *Habitat Int.* **2018**, *76*, 10–18. [[CrossRef](#)]
3. Vilhelmson, B. The Use of the Car-mobility Dependencies of Urban Everyday Life. In *Threats from Car Traffic to the Quality of Urban Life*; Emerald Group Publishing Limited: Bingley, West Yorkshire, UK, 2007; pp. 143–164.

4. Große, J.; Olafsson, A.S.; Carstensen, T.A.; Fertner, C. Exploring the role of daily “modality styles” and urban structure in holidays and longer weekend trips: Travel behaviour of urban and peri-urban residents in Greater Copenhagen. *J. Transp. Geogr.* **2018**, *69*, 138–149. [\[CrossRef\]](#)
5. Liu, Z.; Sharma, S. Statistical investigations of statutory holiday effects on traffic volumes. *Transp. Res. Rec.* **2006**, *1945*, 40–48. [\[CrossRef\]](#)
6. Pfefferkorn, U. Approaches to create a data basis for modelling of long-distance travel behaviour. In *AET Papers Repository. ETC—European Transport Conference; Association for European Transport (AET)*: London, UK, 2016.
7. Frei, A.; Kuhnimhof, T.G.; Axhausen, K.W. *Long Distance Travel in Europe Today: Experiences with a New Survey*; Arbeitsberichte Verkehrs-und Raumplanung: Zürich, Switzerland, 2010; Volume 611.
8. LaMondia, J.J.; Aultman-Hall, L.; Greene, E. Long-distance work and leisure travel frequencies: Ordered probit analysis across non-distance-based definitions. *Transp. Res. Rec.* **2014**, *2413*, 1–12. [\[CrossRef\]](#)
9. Holz-Rau, C.; Scheiner, J.; Sicks, K. Travel distances in daily travel and long-distance travel: What role is played by urban form? *Environ. Plan. A* **2014**, *46*, 488–507. [\[CrossRef\]](#)
10. Reichert, A.; Holz-Rau, C. Mode use in long-distance travel. *J. Transp. Land Use* **2015**, *8*, 87–105. [\[CrossRef\]](#)
11. Reichert, A.; Holz-Rau, C.; Scheiner, J. GHG emissions in daily travel and long-distance travel in Germany—Social and spatial correlates. *Transp. Res. Part D Transp. Environ.* **2016**, *49*, 25–43. [\[CrossRef\]](#)
12. Enzler, H.B. Air travel for private purposes. An analysis of airport access, income and environmental concern in Switzerland. *J. Transp. Geogr.* **2017**, *61*, 1–8. [\[CrossRef\]](#)
13. Czepkiewicz, M.; Heinonen, J.; Ottelin, J. Why do urbanites travel more than do others? A review of associations between urban form and long-distance leisure travel. *Environ. Res. Lett.* **2018**, *13*, 073001. [\[CrossRef\]](#)
14. Magdolen, M.; Ecke, L.; Hilgert, T.; Chlond, B.; Vortisch, P. Identification of Non-Routine Tours in Everyday Travel Behavior. In Proceedings of the 99th Transportation Research Board Annual Meeting, Washington, DC, USA, 12–16 January 2020. in press.
15. Yang, L.; Shen, Q.; Li, Z. Comparing travel mode and trip chain choices between holidays and weekdays. *Transp. Res. Part A Policy Pract.* **2016**, *91*, 273–285. [\[CrossRef\]](#)
16. Cools, M.; Moons, E.; Wets, G. Assessing the impact of public holidays on travel time expenditure: Differentiation by trip motive. *Transp. Res. Rec.* **2010**, *2157*, 29–37. [\[CrossRef\]](#)
17. Cools, M.; Moons, E.; Wets, G. Investigating effect of holidays on daily traffic counts: Time series approach. *Transp. Res. Rec.* **2007**, *2019*, 22–31. [\[CrossRef\]](#)
18. Bagchi, M.; White, P.R. The potential of public transport smart card data. *Transp. Policy* **2005**, *12*, 464–474. [\[CrossRef\]](#)
19. Chakirov, A.; Erath, A. Activity Identification and Primary Location Modelling Based on Smart Card Payment Data for Public Transport. In Proceedings of the 13th International Conference on Travel Behaviour Research (IATBR 2012), Toronto, ON, Canada, 15–20 July 2012.
20. Alsger, A.; Tavassoli, A.; Mesbah, M.; Ferreira, L.; Hickman, M. Public transport trip purpose inference using smart card fare data. *Transp. Res. Part C Emerg. Technol.* **2018**, *87*, 123–137. [\[CrossRef\]](#)
21. Kusakabe, T.; Asakura, Y. Behavioural data mining of transit smart card data: A data fusion approach. *Transp. Res. Part C Emerg. Technol.* **2014**, *46*, 179–191. [\[CrossRef\]](#)
22. Lyons, S.; Mayor, K.; Tol, R.S. Holiday destinations: Understanding the travel choices of Irish tourists. *Tour. Manag.* **2009**, *30*, 683–692. [\[CrossRef\]](#)
23. LaMondia, J.; Snell, T.; Bhat, C.R. Traveler behavior and values analysis in the context of vacation destination and travel mode choices: European Union case study. *Transp. Res. Rec.* **2010**, *2156*, 140–149. [\[CrossRef\]](#)
24. Kavoura, A.; Stavrianeas, A. The importance of social media on holiday visitors’ choices—the case of Athens, Greece. *EuroMed J. Bus.* **2015**, *10*, 360. [\[CrossRef\]](#)
25. Fotis, J.N.; Buhalis, D.; Rossides, N. *Social Media Use and Impact during the Holiday Travel Planning Process*; Springer: Vienna, Austria, 2012.
26. Ráthonyi, G.; Ráthonyi-Odor, K.; Várallyai, L.; Botos, S. Influence of social media on holiday travel planning. *J. Ecoagritour.* **2016**, *12*, 57–62.
27. Amaro, S.; Duarte, P. Social media use for travel purposes: A cross cultural comparison between Portugal and the UK. *Inf. Technol. Tour.* **2017**, *17*, 161–181. [\[CrossRef\]](#)

28. Fotis, J.N. *The Use of Social Media and Its Impacts on Consumer Behaviour: The Context of Holiday Travel*; Bournemouth University: Bournemouth, UK, 2015.
29. Hensher, D.A.; Reyes, A.J. Trip chaining as a barrier to the propensity to use public transport. *Transportation* **2000**, *27*, 341–361. [\[CrossRef\]](#)
30. Ye, X.; Pendyala, R.M.; Gottardi, G. An exploration of the relationship between mode choice and complexity of trip chaining patterns. *Transp. Res. Part B Methodol.* **2007**, *41*, 96–113. [\[CrossRef\]](#)
31. Li, J.; Ye, Q.; Deng, X.; Liu, Y.; Liu, Y. Spatial-temporal analysis on Spring Festival travel rush in China based on multisource big data. *Sustainability* **2016**, *8*, 1184. [\[CrossRef\]](#)
32. Liu, Z.; Sharma, S. Nonparametric method to examine changes in traffic volume pattern during holiday periods. *Transp. Res. Rec.* **2008**, *2049*, 45–53. [\[CrossRef\]](#)
33. Wang, X.; Liu, C.; Mao, W.; Hu, Z.; Gu, L. Tracing the largest seasonal migration on earth. *arXiv* **2014**, arXiv:1411.0983.
34. Hu, X.; Li, H.; Bao, X. Urban population mobility patterns in Spring Festival Transportation: Insights from Weibo data. In Proceedings of the 2017 International Conference on Service Systems and Service Management, Dalian, China, 16–18 June 2017; pp. 1–6.
35. Wei, Y.; Song, W.; Xiu, C.; Zhao, Z. The rich-club phenomenon of China's population flow network during the country's spring festival. *Appl. Geogr.* **2018**, *96*, 77–85. [\[CrossRef\]](#)
36. Ermagun, A.; Fan, Y.; Wolfson, J.; Adomavicius, G.; Das, K. Real-time trip purpose prediction using online location-based search and discovery services. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 96–112. [\[CrossRef\]](#)
37. Cui, Y.; Meng, C.; He, Q.; Gao, J. Forecasting current and next trip purpose with social media data and Google Places. *Transp. Res. Part C Emerg. Technol.* **2018**, *97*, 159–174. [\[CrossRef\]](#)
38. Widhalm, P.; Yang, Y.; Ulm, M.; Athavale, S.; González, M.C. Discovering urban activity patterns in cell phone data. *Transportation* **2015**, *42*, 597–623. [\[CrossRef\]](#)
39. Alexander, L.; Jiang, S.; Murga, M.; González, M.C. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 240–250. [\[CrossRef\]](#)
40. Jiang, S.; Ferreira, J.; Gonzalez, M.C. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Trans. Big Data* **2017**, *3*, 208–219. [\[CrossRef\]](#)
41. Gong, L.; Liu, X.; Wu, L.; Liu, Y. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartogr. Geogr. Inf. Sci.* **2016**, *43*, 103–114. [\[CrossRef\]](#)
42. Bao, J.; Xu, C.; Liu, P.; Wang, W. Exploring bikesharing travel patterns and trip purposes using smart card data and online point of interests. *Netw. Spat. Econ.* **2017**, *17*, 1231–1253. [\[CrossRef\]](#)
43. Zhang, Y.; Brussel, M.J.; Thomas, T.; van Maarseveen, M.F. Mining bike-sharing travel behavior data: An investigation into trip chains and transition activities. *Comput. Environ. Urban Syst.* **2018**, *69*, 39–50. [\[CrossRef\]](#)
44. Wang, S.; Xu, G.; Guo, Q. Street centralities and land use intensities based on points of interest (POI) in Shenzhen, China. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 425. [\[CrossRef\]](#)
45. Wu, C.; Ye, X.; Ren, F.; Du, Q. Check-in behaviour and spatio-temporal vibrancy: An exploratory analysis in Shenzhen, China. *Cities* **2018**, *77*, 104–116. [\[CrossRef\]](#)
46. Yue, Y.; Zhuang, Y.; Yeh, A.G.; Xie, J.-Y.; Ma, C.-L.; Li, Q.-Q. Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 658–675. [\[CrossRef\]](#)
47. Zhao, J.; Tian, C.; Zhang, F.; Xu, C.; Feng, S. Understanding Temporal and Spatial Travel Patterns of Individual Passengers by Mining Smart Card Data. In Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, China, 8–11 October 2014; pp. 2991–2997.
48. Zhao, J.; Qu, Q.; Zhang, F.; Xu, C.; Liu, S. Spatio-temporal analysis of passenger travel patterns in massive smart card data. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 3135–3146. [\[CrossRef\]](#)
49. El Mahrssi, M.K.; Côme, E.; Baro, J.; Oukhellou, L. Understanding Passenger Patterns in Public Transit through Smart Card and Socioeconomic Data: A Case Study in Rennes, France. In Proceedings of the ACM SIGKDD Workshop on Urban Computing, New York, NY, USA, 24 August 2014.
50. Lathia, N.; Quercia, D.; Crowcroft, J. The Hidden Image of the City: Sensing Community Well-being from Urban Mobility. In Proceedings of the International Conference on Pervasive Computing, Seoul, Korea, 21–24 September 2008; pp. 91–98.
51. Liu, Y.; Cheng, T. Understanding public transit patterns with open geodemographics to facilitate public transport planning. *Transp. A: Transp. Sci.* **2020**, *16*, 76–103. [\[CrossRef\]](#)

52. Kieu, L.M.; Bhaskar, A.; Chung, E. Transit Passenger Segmentation Using Travel Regularity Mined from Smart Card Transactions Data. In Proceedings of the Transportation Research Board (TRB) 93rd Annual Meeting Compendium of Papers, Washington, DC, USA, 12–16 January 2014; Volume 7013, pp. 1–17.
53. Hofmann, T. Probabilistic latent semantic analysis. *arXiv* **2013**, arXiv:1301.6705.
54. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
55. Pereira, F.C.; Rodrigues, F.; Ben-Akiva, M. Text analysis in incident duration prediction. *Transp. Res. Part C Emerg. Technol.* **2013**, *37*, 177–192. [[CrossRef](#)]
56. Come, E.; Randriamanamihaga, N.A.; Oukhellou, L.; Aknin, P. Spatio-temporal analysis of dynamic origin-destination data using latent dirichlet allocation: Application to vélib’bike sharing system of paris. In Proceedings of the TRB 93rd Annual Meeting, Washington, DC, USA, 12–16 January 2014.
57. Hasan, S.; Ukkusuri, S.V. Urban activity pattern classification using topic models from online geo-location data. *Transp. Res. Part C Emerg. Technol.* **2014**, *44*, 363–381. [[CrossRef](#)]
58. Steiger, E.; Westerholt, R.; Resch, B.; Zipf, A. Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Comput. Environ. Urban Syst.* **2015**, *54*, 255–265. [[CrossRef](#)]
59. Fu, C.; McKenzie, G.; Frias-Martinez, V.; Stewart, K. Identifying spatiotemporal urban activities through linguistic signatures. *Comput. Environ. Urban Syst.* **2018**, *72*, 25–37. [[CrossRef](#)]
60. Yuan, J.; Zheng, Y.; Xie, X. Discovering Regions of Different Functions in a City Using Human Mobility and POIs. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 186–194.
61. Farrahi, K.; Gatica-Perez, D. Discovering routines from large-scale human locations using probabilistic topic models. *ACM Trans. Intell. Syst. Technol. (TIST)* **2011**, *2*, 1–27. [[CrossRef](#)]
62. Zhong, C.; Manley, E.; Arisana, S.M.; Batty, M.; Schmitt, G. Measuring variability of mobility patterns from multiday smart-card data. *J. Comput. Sci.* **2015**, *9*, 125–130. [[CrossRef](#)]
63. Nishiuchi, H.; King, J.; Todoroki, T. Spatial-temporal daily frequent trip pattern of public transport passengers using smart card data. *Int. J. Intell. Transp. Syst. Res.* **2013**, *11*, 1–10. [[CrossRef](#)]
64. Kim, M.-K.; Kim, S.; Sohn, H.-G. Relationship between spatio-temporal travel patterns derived from smart-card data and local environmental characteristics of Seoul, Korea. *Sustainability* **2018**, *10*, 787.

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).