Sustainable Cities and Society 69 (2021) 102862

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Does the returners and explorers dichotomy in urban human mobility depend on the observation duration? An empirical study in Guangzhou, China

Ruoxi Wang [a], Nan Li [a,b,*], Yan Wang [c]

[a] Department of Construction Management, Tsinghua University, China
[b] Hang Lung Center for Real Estate, Tsinghua University, Beijing, 100084, China
[c] Department of Urban and Regional Planning and Florida Institute for Built Environment Resilience, University of Florida, USA

ARTICLE INFO

Keywords:
Human mobility
Returner and explorer
Observation duration
GPS trajectory
Travel behavior
Urban analytics

ABSTRACT

The increasing accessibility to digital traces of human whereabouts in cities has offered numerous new opportunities for exploring patterns of human mobility in urban spaces. Prior research pointed out that there exist two distinct subpopulations in cities, namely returners and explorers, whose mobility patterns differ in the extent to which their characteristic traveled distance is impacted by their recurrent mobility. However, the potential dependence of the returners and explorers dichotomy on the observation duration has been largely ignored in prior research, which may cause biased understanding of the returners and explorers dichotomy in urban mobility patterns. By analyzing the daily trajectory data of 21,240 individuals in Guangzhou over 111 weekdays, this study evidenced that the returners and explorers dichotomy is significantly dependent on the duration within which individuals' trajectories are observed. This study further revealed that such dependence could be interpreted by three underlying explanations, which are respectively related to information accumulation, individuals' spatial exploration behaviors and changes in individuals' important locations. The findings provide fundamental knowledge for studying urban human mobility patterns for disease prediction, population behavioral modeling, and understanding dynamic human-environment interactions at urban scales.

1. Introduction

The increasing accessibility to digital traces of human whereabouts in cities, made available by recent advancements of mobile and ubiquitous computing technologies, has offered numerous new opportunities for exploring patterns and applications of urban human mobility. The amount of literature on human mobility has increased exponentially over recent years. One remarkable phenomenon that was repeatedly reported in prior research is that there is a surprising coexistence of variability and regularity in individuals' mobility characteristics. For instance, while individuals' daily mobility ranges are highly diverse, most of them repeat certain daily activities, such as commuting between home and workplace and socializing with friends, which are dominated by routines. Motivated to explain this coexistence of variability and regularity, Pappalardo et al. (2015) analyzed two trajectory datasets, and identified two distinct classes of individuals, whom they referred to as returners and explorers. The mobility patterns of returners and explorers differ in the extent to which their characteristic traveled distance, which is usually measured by the radius of gyration (Pappalardo et al., 2015), is impacted by their recurrent mobility, i.e. mobility between a few important locations such as home and working places (Song, Qu, Blumm, & Barabási, 2010). This dichotomous classification of the population sheds light on how individuals' mobility patterns are affected by their personal preferences to return to previously visited locations or explore new locations (De Nadai, Cardoso, Lima, Lepri, & Oliver, 2019). Moreover, the returners and explorers dichotomy has inspired a bulk of research that examines a spectrum of relevant topics ranging from urban population's job and housing dynamics (Huang, Levinson, Wang, Zhou, & Wang, 2018; Zhou & Long, 2014) to correlation between social and spatial behaviors (Alessandretti, Lehmann, & Baronchelli, 2018; Fan, Liu, Huang, Rong, & Zhou, 2017), and human mobility in cyber-physical spaces (De Nadai et al., 2019).

One factor, however, that may have a significant impact on the reliability of the returns and explorers dichotomy, but was originally

* Corresponding author.
E-mail address: nanli@tsinghua.edu.cn (N. Li).

https://doi.org/10.1016/j.scs.2021.102862
Received 20 October 2020; Received in revised form 11 February 2021; Accepted 16 March 2021
Available online 22 March 2021

unaddressed in Pappalardo et al.'s study (2015), is the potential dependence of the returners and explorers dichotomy on the observation duration namely the length of the time window during which the individuals' mobility behaviors are observed and recorded. The two datasets used by Pappalardo et al. (2015) contained call records collected over three months and vehicle GPS data collected over one month, respectively. Both datasets were processed, analyzed and interpreted identically, with an implicit assumption that the returns and explorers dichotomy is not dependent on the observation duration.

However, the above assumption has yet to be tested. Knowing the validity of this assumption is critical, the reason for which is twofold. First, it determines whether the observation duration is an important variable that should be considered in various applications of the returner-explorer dichotomy. One recent example of such applications being that, considering the distinct trajectory diffusion properties of returners and explorers, populations with higher proportions of explorers tend to have higher chances to be globally invaded by epidemics (Shoghri, Liebig, Gardner, Jurdak, & Kanhere, 2019). Therefore, depicting the returners and explorers' pattern is crucial for forecasting the transmission of diseases and developing epidemic control measures (Balcan & Vespignani, 2011), for which the health authorities may need to know whether and how they should factor in the impact of observation duration. Second, the above implicit assumption has been carried on by many follow-up studies (Barbosa, de Lima-Neto, Evsukoff, & Menezes, 2016; De Nadai et al., 2019; Fan et al., 2017). This highlights the importance of testing this assumption, so as to strengthen the theoretical basis of human mobility knowledge derived from these follow-up studies as well as future studies that aim to look into relevant subjects.

As a matter of fact, there exist at least the following three reasons to hypothesize that the above assumption may not always be true, and that the returners and explorers dichotomy may be dependent on the observation duration. First, it is evident that individuals' mobility patterns usually differ between short-term and long-term (Schneider, Belik, Couronné, Smoreda, & González, 2013). For instance, the phenomenon that most individuals' daily mobility networks can be described with 17 different motifs (Schneider et al., 2013) is only observable at a daily scale, whereas the distributions of the number of visited location (Song, Qu et al., 2010) and the radius of gyration (Gonzalez, Hidalgo, & Barabasi, 2008) show significant statistical characteristics only in long-term observations that usually span over a few months. It is hence possible that datasets collected over short and long periods may not necessarily lead to consistent conclusions regarding the returners and explorers dichotomy. Second, a prior study (Stanley, Yoo, Paul, & Bell, 2018) revealed that increasing the amount of mobility data can provide incremental information gains for capturing individuals' spatial behavior, until the observation duration reaches a cut-off value. This finding suggests that the precision of human mobility characterization may not be assured when the amount of mobility data is insufficient, while the data amount is closely tied to the observation duration. Third, Pappalardo et al. (2015) observed that there was no balance point of the returners and explorers dichotomy using the GPS dataset in the same study in which they proposed the returners and explorers dichotomy. They could not find any k value where the population would reach a balance between k -returns and k -explorers, as there were always more k -explorers than k -returners. This contradicted the outcomes from their dataset of mobile phone geolocations, suggesting an inherent heterogeneity between the two datasets that could impact their classification results. The observation duration, among other possible factors, is highly susceptible.

The possibility that the validity of existing knowledge about urban human mobility that prior studies have derived based on the returners and explorers dichotomy may be brought into question, due to the potential dependence of the dichotomy on the observation duration, is the primary motivation of this study. What further motivates this study is the prospect to advance the understanding of distinctions between individuals' mobility in the short term and the long term, which may have significant implications for transferring the knowledge about urban human mobility into smart and resilient city applications. This study aims to answer two specific research questions: (1) Are returner-explorer classifications dependent on the study duration within which individuals' trajectories are observed? If so, does the observation duration-dependence of classification results change over time, and how? (2) What are the underlying explanations behind the observation duration-dependence of the returners and explorers dichotomy?

A mobility dataset collected in Guangzhou, China was employed in this study to explore these questions. The dataset contained daily trajectories of 21,240 individuals in the city over five and a half months. The returner-explorer classification was conducted on different observation durations in the above dataset. The resulting classification results were compared to assess the potential observation duration-dependence of the returners and explorers dichotomy at both the population and the individual levels. Moreover, three possible explanations were analyzed to explain the observation duration-dependence of the returners and explorers dichotomy. The findings are expected to offer new insights into human urban mobility behavior modeling and prediction by providing a dynamic and temporal perspective, which would in turn enhance the intelligence and resilience in cities and support more informed urban development practices.

2. Background

2.1. *Urban human mobility patterns and implications for smart and resilient cities*

The emerging knowledge about urban human travel behaviors and mobility patterns, made available by mining human trajectory data, provides important insights into the relationship between urban environments and human activities, which have profound implications for improving the intelligence and resilience of modern cities.

Knowledge about urban human mobility patterns is valuable for a wide range of smart city applications (Bibri & Krogstie, 2017; Lopez-Carreiro & Monzon, 2018; Peprah, Amponsah, & Oduro, 2019). For instance, an array of studies have used individuals' trajectories to determine people's travel modes (Shin et al., 2015), model travel routes (Jiang et al., 2016), and forecast travel demands (Wang, Yang, Sun, & Gao, 2017). The outcomes have important implications for urban road network planning (Tsiotas & Polyzos, 2017) and traffic congestion management (Qian, Lei, Xue, Lei, & Ukkusuri, 2020). These studies have also inspired innovative solutions to various urban challenges such as transport policy making, public resources planning, and urban land use planning (Forghani & Karimipour, 2018; Zhang, Liu, Tang, Cheng, & Wang, 2019).

Understanding urban human mobility patterns also informs building resilience of cities and their infrastructures (Nogal & Honfi, 2019). There is a growing body of literature aimed at reducing disaster risk and enhancing resilience of cities and communities by understanding and predicting human mobility. For instance, disaster-induced perturbations to human mobility are investigated widely (Wang & Taylor, 2014; Wang, Wang, & Taylor, 2017), in order to inform governments and policymakers to improve disaster risk assessments, responses and relief plans (Guadagno, 2016). Human movements under disasters are modeled and predicted (Solmaz & Turgut, 2017; Song et al., 2016), which can be used to support disaster evacuation planning (Zamichos, Theodorou, Drosou, & Tzovaras, 2018) and emergency shelter site selection (Tai, Lee, & Lin, 2010). Tracking human mobility perturbations has also been widely considered as a novel and effective approach for assessing the resilience of urban infrastructures (Nogal & Honfi, 2019; Saja, Teo, Goonetilleke, Ziyath, & Nianthi, 2020; Zhang, Li et al., 2019), leading to better understanding of the strengths and vulnerabilities of the infrastructures against disaster impacts.

In sum, prior studies have demonstrated the broad applications and significant values of incorporating the urban human mobility knowledge to support the development of smart and resilient cities. However, one issue that is potentially impactful on the use of human mobility knowledge in urban studies but has drawn little attention thus far is the temporal characteristics of human mobility patterns. There is abundant evidence indicating that individuals' mobility patterns usually differ between short term and long term (Schneider et al., 2013), and may be dependent on the temporal duration within which the mobility is observed (Stanley et al., 2018). In reality, different applications of human mobility knowledge are associated with different timespans. For instance, travel demand forecast (Wang, Yang et al., 2017) and urban hotspots detection (Yang, Zhao, & Lu, 2016) usually require day-to-day mobility patterns, whereas urban social dynamics analysis and prediction (Liu, Qiao, Tao, Lin, & Yang, 2017) and epidemics modeling and prediction (Xu et al., 2017) are based on mobility patterns that are only significant in long-term observations that usually span over a few months. It is not uncommon to see mismatches between mobility datasets or human mobility findings over different temporal scales and their real-world applications. The impact of such mismatches on the usability of human mobility knowledge in addressing challenges in smart and resilient cities has largely remained to be examined.

2.2. Returners and explorers dichotomy in urban human mobility

Based on the demonstration of the truncated power law of trip distance distribution, Gonzalez et al. (2008) posited that people have the tendency to return to primary locations. Song, Qu et al. (2010) used mobility networks to represent trajectories of mobile phone users, and found that people tended to spend most of their time at a few important locations, which suggested a recurrent nature of human mobility. Furthermore, considering the high probability for individuals to return to a few important locations, Song, Koren, Wang, and Barabási (2010) proposed two generative mechanisms, namely exploration and preferential return (EPR), to capture two different tendencies of individuals, either to return to previously visited locations or to explore new locations. This formed the basis of the widely referenced EPR human mobility model. In addition, by integrating additional movement choice-making mechanism, several variants of the EPR model were introduced in later studies and achieved improved approximations of empirical human mobility data. For instance, Jiang et al. (2016) proposed the r-EPR model, which is a rank-based exploration and preferential return model that incorporates a ranking-based selection method for spatial choices when exploring new locations. Similarly, based on the exploration-exploitation dichotomy of individual' movement choices, Barbosa, de Lima-Neto, Evsukoff, and Menezes (2015) proposed the recency model, which considers both recently visited locations and frequently visited locations in movement decisions.

The research on the exploration and preferential return mechanisms also inspired Pappalardo et al. (2015) to propose the returners and explorers dichotomy, which posits that there exist two classes of individuals in any given population, namely returners and explorers. These two classes of individuals have distinct mobility patterns: returners are people who spend most of their time at a few frequently visited locations, therefore, the overall mobility of returners is largely determined by their recurrent mobility; whilst explorers are those who have a strong tendency to explore new locations, so that frequently visited locations extracted from explorers' mobility histories have little contribution to their overall mobility. The radius of gyration, which characterizes an individual's tendency to deviate from the center of his or her own movements, has been widely used to quantify the spatial range of the individual's trajectories. Adapted from the definition of radius of gyration, the k -radii of gyration was introduced by Pappalardo et al. (2015) to represent the individual's mobility range restricted to the top k frequently visited locations. Then, an individual can be determined as either returner or explorer, by calculating the contribution of his or her recurrent mobility between top k frequently visited locations to his or her overall mobility.

The classification of returners and explorers has been adopted and found useful in various applications. For instance, Pappalardo et al. (2015) revealed that these two classes of individuals play distinct roles in epidemic spreading. The chances that an epidemic would invade globally increases with the fraction of explorers in the population. By considering geographical characteristics, Liao, Yeh, and Jeuken (2019) expanded the dichotomy and classified a population into four distinct groups, including local explorers, local returners, global explorers and global returners, of travelers, which led to better understanding of the population heterogeneity in travel behaviors. Individuals in the local groups usually visit nearby locations while individuals in the global groups have high proportions of international trips. Moreover, a number of studies found that, for individuals belonging to the same dichotomous class, their spatiotemporal dynamics exhibit highly similar scaling properties in various aspects. For instance, Fan et al. (2017) found that the social proximity and mobility similarity between individuals in the same class are significantly higher than those between individuals in different subpopulations. This echoed Pappalardo et al.'s study (2015), which reported that individuals are more likely to construct social ties with those in the same class. In addition, Liao et al. (2019) found that the behaviors of exploration and return are correlated in both social and spatial domains. The apparent correlation between human behaviors in physical space and cyber space indicates that there also exist two distinct classes, namely returners and explorers, with respect to mobile application usage (De Nadai et al., 2019) and web browsing (Barbosa et al., 2016).

In sum, the returner-explorer classification has been widely used to reflect individuals' behavioral patterns. In prior studies, the classification of a population into returners and explorers was conducted on datasets with different durations. The potential observation duration-dependence of the returners and explorers dichotomy was largely ignored, which could lead to biased interpretation of observed mobility patterns and inaccurate understanding of the distinction between individuals' mobility in the short term and the long term. Therefore, there is an urgent need to investigate the significance of the potential observation duration-dependence of the returners and explorers dichotomy, and reveal the underlying explanations of such dependence, in order to better explain the mobility patterns extracted from human traces and enhance the validity of related findings.

3. Data and methods

3.1. Data collection and preprocessing

This study uses anonymized geolocation data collected from mobile devices used by individuals within the City of Guangzhou in Guangdong Province, China. The dataset includes 98,652 distinct individuals and covers 167 days from May 1 to October 15 in 2018. Trajectories in the dataset include a total of 453 million geolocations. These geolocations were preprocessed before analyses as follows: First, only individuals whose observed trajectories covered at least 20 days in each month and five hours each day were included, and trajectories of individuals that did not meet the requirement were removed. This preprocessing step was conducted to exclude temporary visitors and ensured that the data for all individuals included in the study were adequate and properly distributed. As such, trajectories of 78 % of individuals were removed from the original dataset; Second, trajectories collected over weekends and holidays, accounting for 32.64 % of the remaining geolocations in the dataset, were also excluded, because this study focuses on individuals' mobility on weekdays, which is more regular than mobility on weekends and is usually studied separately from mobility on weekends (Feng, Wang, Kong, Wang, & Liu, 2018; Zhang, Liu et al., 2019); Third, trajectories on Monday, September 17, 2018, which accounted for 0.84 % of the remaining geolocations in the dataset, were also excluded, as the city was under the impact of Typhoon Mangkhut on that day, and mobility in the city may have been perturbed by the extreme weather (Brum-Bastos, Long, & Demšar, 2018; Wang, Wang et al., 2017).

The preprocessed dataset includes the trajectories of 21,240 distinct individuals, representing approximately 0.15 % of the city's 14.5-million population, over a total of 111 weekdays. The dataset contains a total of 303 million geolocations. On average, each individual in the dataset has about 128 recorded geolocations per weekday. The spatial resolution of the geolocations is approximately 20 m. The spatial distribution of these geolocations is illustrated in Fig. 1, which is generally consistent with the distribution of the population in the city. The median temporal resolution of the dataset is approximately 7 min per geolocation. The temporal distribution of the geolocations, largely determined by the usage patterns of the mobile devices which record geolocations less frequently when they are on standby, varies over time and differs between different individuals. Notably, the studied dataset does not contain any personally identifiable information of the individuals, and the study only uncovers empirical findings in an aggregated manner.

3.2. Identifying locations

3.2.1. Detecting mobility phases

Rhee et al. (2011) observed that GPS points tend to gather in a few limited areas, which aligns with the common sense that people tend to gather in places that attract them during the day. These places could be homes, work places, restaurants, coffee shops and so on, and in general, any location people tend to spend time at for a while. For a GPS trajectory dataset, a place where an individual is standing still or moving around very slowly can be identified as a location (Papandrea et al., 2016).

According to this view, human mobility can be separated into two phases: a static phase where an individual spends some time in a place, and a movement phase where an individual moves towards a place. This study focused on locations during the static phase because the classification of the returners and explorers is based on visited locations where individuals spend time during the static phase.

A filter proposed in previous studies (Papandrea et al., 2016; Zignani, Gaito, & Rossi, 2013) was applied to extract the static phase from an individual's GPS trace. If two consecutive points p_i and p_{i+1} from an individual's GPS trace, with timestamps $t(p_i)$ and $t(p_{i+1})$ respectively, do not satisfy

$$\frac{\|p_{i+1} - p_i\|}{t(p_{i+1}) - t(p_i)} \leq \Delta \quad (1)$$

then the point p_{i+1} belongs to the movement phase and was removed from the trajectories. The threshold was set as Δ =1.4 m/s, based on the fact that observed human walking speed is generally at 1.1–1.4 m/s (Papandrea et al., 2016). In this way, points belonging to the movement phase, which accounted for 24.6 % of the dataset, were removed (Papandrea et al., 2016; Zignani et al., 2013).

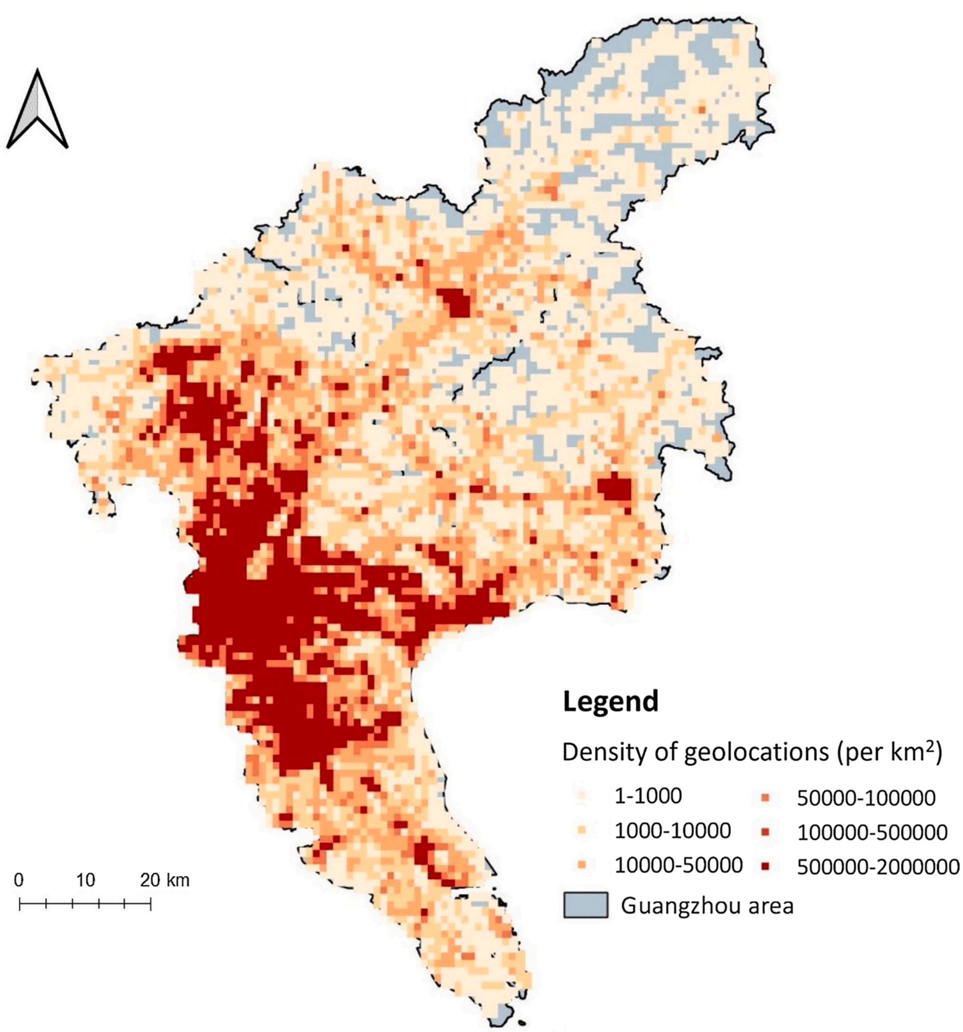

Fig. 1. Spatial distribution of geolocations in the preprocessed dataset.

3.2.2. Extracting locations from GPS trajectories

Based on trajectories belonging to the static phase, individuals' location clusters were captured using DBSCAN algorithm. The algorithm removed abnormal data points from the dataset, and extracted the locations that each individual visited based on their trajectories. DBSCAN is a density-based algorithm widely used to find the high-density areas in space (Hu et al., 2015; Luo, Zheng, Xu, Fu, & Ren, 2017). Compared with other clustering methods, such as K-Means and hierarchical clustering, DBSCAN does not require a pre-defined number of output clusters and can detect clusters with different shapes (Dixon, 2017). Two key input parameters in DBSCAN were set as follows: the maximum search radius was set as 50 m and the minimum number of points to form a cluster was set as two, based on recommendations in prior research (Cuttone, Lehmann, & González, 2018). Each point cluster extracted by DBSCAN may include GPS points from different visits paid by the same individual. In addition, the location extraction should also consider the stay time feature (Zignani et al., 2013). It was observed that in individuals' trajectories there were many point clusters where individuals only stayed for a short period of time. These point clusters probably represented small pauses in individuals' movement towards real destinations, rather than meaningful locations (Papandrea et al., 2016; Zignani et al., 2013). Therefore, point clusters where an individual never stayed for more than 5 min during a single visit were removed. The centroids of the remaining point clusters were then extracted, and used to represent the locations that were visited by the individual. These locations were used for following analyses.

3.3. Returner-explorer classification

Returners are individuals whose mobility range is dominated by a few important locations such as home and workplace, while explorers have a strong tendency to explore a larger number of different locations. Following Pappalardo et al.'s definition (2015), individuals can be classified into returners or explorers based on the relationship of their overall mobility, measured by radius of gyration (r_g), and their recurrent mobility, measured by the k -radius of gyration ($r_g^{(k)}$).

Specifically, radius of gyration has been widely used as the characteristic distance covered by an individual's trajectories. It can be calculated based on the following equation (Liao et al., 2019; Pappalardo et al., 2015):

$$r_g = \sqrt{\frac{1}{N}\sum_{i\in L} n_i\left[2r\times \sin^{-1}\left(\sqrt{\sin^2\left(\frac{\phi_i-\phi_c}{2}\right)+\cos\phi_i\cos\phi_c\sin^2\left(\frac{\varphi_i-\varphi_c}{2}\right)}\right)\right]^2} \quad (2)$$

where L is the set of locations visited by the individual, r is the radius of the earth, ϕ_i and φ_i is the latitude and longitude of location i, ϕ_c and φ_c are the latitude and longitude of the center of mass computed on the individual's visited locations, n_i is the visitation frequency by the individual in location i, $N = \sum_{i\in L} n_i$ is the total number of visits.

The concept of the k -radius of gyration was introduced by Pappalardo et al. (2015) to measure the recurrent mobility range of an individual dominated by the individual's top k frequently visited locations $L_1, \ldots, L_k$. It can be calculated based on the following equation (Pappalardo et al., 2015):

$$r_g^{(k)} = \sqrt{\frac{1}{N_k}\sum_{i=1}^{k} n_i\left[2r\times \sin^{-1}\left(\sqrt{\sin^2\left(\frac{\phi_i-\phi_c^{(k)}}{2}\right)+\cos\phi_i\cos\phi_c^{(k)}\sin^2\left(\frac{\varphi_i-\varphi_c^{(k)}}{2}\right)}\right)\right]^2} \quad (3)$$

where $\phi_c^{(k)}$ and $\varphi_c^{(k)}$ are the latitude and longitude of the center of mass computed on individual's top k most frequently visited locations, N_k is the sum of the weights assigned to the top k most frequently visited locations.

Based on the radius of gyration and the k -radius of gyration, the classification of returners and explorers can be conducted using a bisector method (Pappalardo et al., 2015), where individuals whose mobility pattern satisfies the following criterion are coded as k -returners, or otherwise as k -explorers:

$$r_g^{(k)} > \frac{r_g}{2} \quad (4)$$

The trajectories of two anonymous individuals, classified as 2-returner and 2-explorer respectively, over one week are shown in Fig. 2 to illustrate the typical difference between these two classes of individuals. The trajectories shown in the figure were aggregated with the location-extraction algorithm.

Moreover, to test the impact of the observation duration on the returner-explorer classification results, the above classification method is applied to a total of 111 different durations extracted from the dataset. Specifically, as the value of n traverses from 1 to 111, the population is classified into returners and explorers based on the n -day duration, which includes the first n days covered by the dataset. A fixed value $k = 2$ is used in the classification. As such, a total of 111 sets of classification results are obtained, each associated with a distinct dataset duration.

4. Analyses and findings

The returner-explorer classification results on different observation durations are plotted in Fig. 3. The results showed that the proportion of returners was higher than that of explores for relatively short durations, with the former having a steady trend to shrink and the latter having a steady trend to increase as the observation duration expanded. The proportions of the two classes reached a balancing point at the duration of 21 days, beyond which the proportion of returners surpassed that of explorers. As the observation duration continued to expand, the changing rate of the proportion of both returners and explorers kept decreasing. To arrive at a detailed understanding of how the dichotomy of explorers or returners varies with different observation durations, the proportions of explorers or returners in the population, with the expansion of the observation duration from 1 day to 111 days, were further fitted into general function models, including Gaussian, polynomial, exponential and power, using MATLAB. The results, details of which are provided in the supplementary materials, showed that the power function model could best characterize this dependence, with a high R-squared measure of goodness of fit (0.998). Moreover, the power function model suggested that the proportions of returners and explorers had a clear tendency to converge. In addition, the bootstrapping method was used to estimate the confidence intervals of the parameters of the above power function model. The results, details of which are provided in the supplementary materials, showed that the proportions of returners and explorers among the entire population could also be well fitted by the power function and showed a clear tendency to converge, suggesting that the above conclusion about the observation duration-dependency of the dichotomy would hold regardless of the limited sampling rate of the population.

The above result suggested two important findings: First, there is

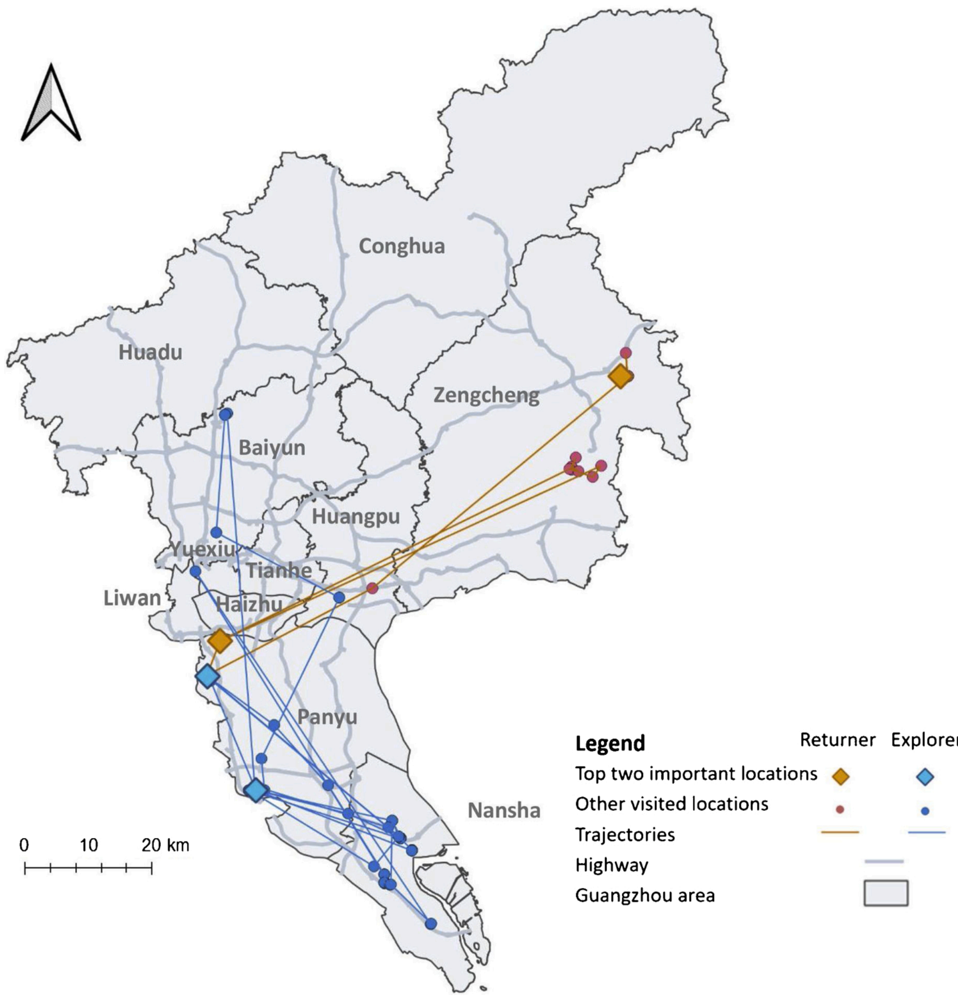

Fig. 2. The trajectories of a 2-returner and a 2-explorer.

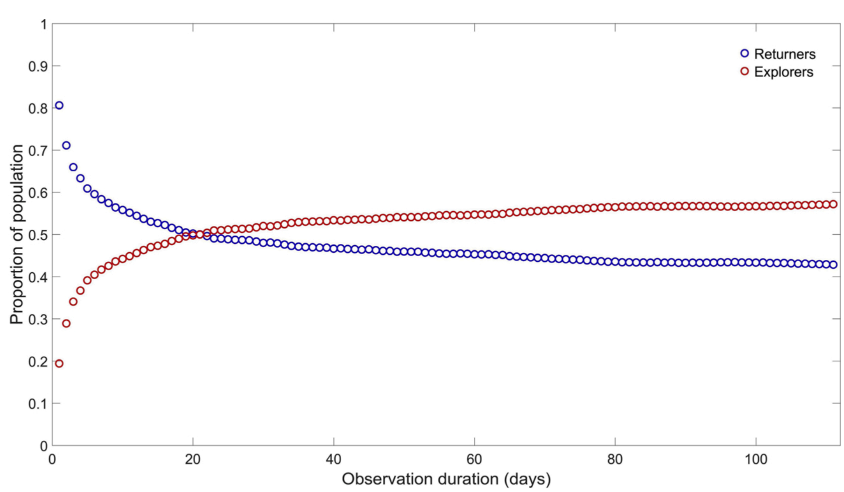

Fig. 3. Returner-explorer classification results under different observation durations.

notable observation duration-dependence of the returners and explorers dichotomy, since the dichotomous classification results are heavily impacted by the observation duration. This finding directly addresses the first research question of this study. Second, the classification results become generally stable when the observation duration is relatively long. Additional sensitivity analyses, the details of which can be found in the supplementary materials, suggested that these findings are robust to reasonable variations of a few key parameters of the methods explained in Section 3.2.

Moreover, as Fig. 3 shows, changes in the classification results caused by the expansion of the observation duration converged to a low level after several weeks. To further examine this phenomenon, for any given n-day duration, the median value of radius of gyration (r_g) and 2-

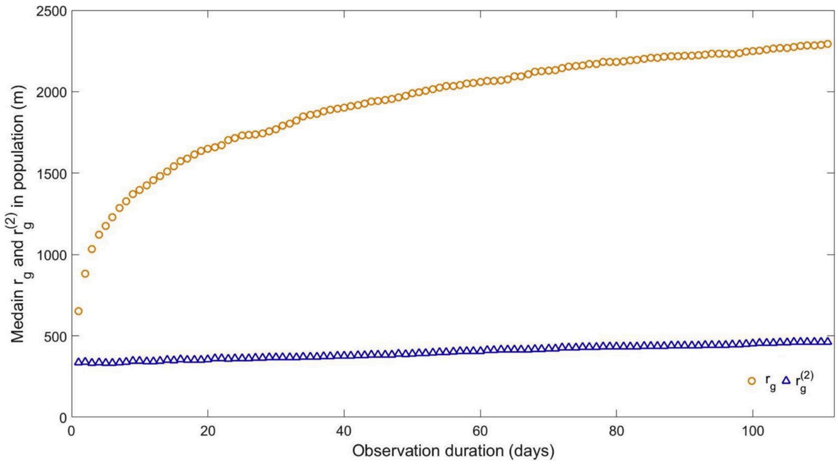

Fig. 4. Median radius of gyration and 2-radii of gyration of the population under different durations.

radii of gyration ($r_g^{(2)}$) of all individuals were calculated and plotted in Fig. 4. As shown in Fig. 4, at the population level, as the observation duration increased, the average distance between individuals' top two most frequently visited locations slightly decreased, whereas their average overall mobility range increased at a relatively higher rate. Based on Eq. (4), the above result suggested that, as the observation duration increases, the continuous increase of the proportion of explorers in the population is mainly driven by the expansion of individuals' overall mobility range over time. That being said, intuitively, an individual's radius of gyration cannot expand infinitely. Rather, it has a limit and may ultimately converge, which can be considered as a saturation process (Gonzalez et al., 2008). This explains why the classification results, as illustrated in Fig. 3, tend to become stable in the long term.

In addition, the classification results were further examined to investigate whether their changes also decayed at the individual level. In Fig. 5, each bar shows the number of individuals whose classification flipped from returner to explorer or vice versa, when the observation duration expanded from $n - 1$ days to n days. The size of the 'swinging' group generally shrank as n increased, and it remained below 160 individuals (or 0.8 % of the population) after n exceeded 70. Among all individuals, 92 % had consistent classifications as n increased from 70 to 111. The above result showed that most individuals' classifications would not change in relatively long observation durations, suggesting that the changes of classification results also decayed at the individual level.

5. Discussions

5.1. Explanation one: information accumulation

Given the observation duration-dependence of the returners and explorers dichotomy discovered from the above results, the next question to answer is: what are the underlying explanations that have led to such observation duration-dependence. Due to the sampling nature of the trajectory data, individuals' mobility patterns may not be completely captured based on the recorded trajectories (Song et al., 2016), especially when the length of observation duration is short (e.g. several days). Stanley et al. (2018) examined GPS trajectories collected from over a hundred college students for 21 days, and revealed that knowledge gained from short-term datasets is often unstable and insufficient to form a complete picture of people's routine spatial behaviors, considering the sampling nature of used data sources. They argued that information about people's mobility behaviors accumulates, and that the mobility patterns derived from the information become more stable and accurate as the observation duration expands. They defined the 'completeness' of captured routine human spatial behaviors as the point at which marginal information gained from extra trajectory data becomes negligible according to the temporal analysis of the KL divergence. Based on this concept, they concluded that information divergence converged to a low level within two weeks for their dataset, and that information gains after two weeks could be negligible. In other words, two weeks of trajectory data of a homogeneous group (college

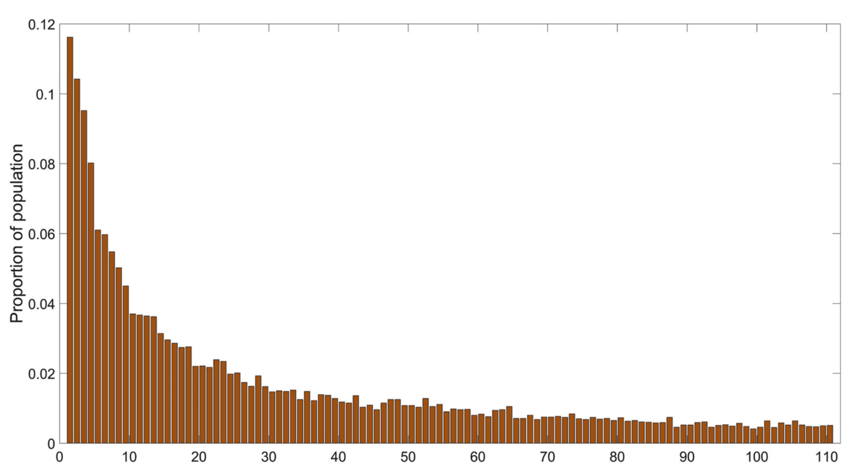

Fig. 5. Proportion of swinging group in population under different observation durations.

students) was found to be the minimum data required to completely capture routine human mobility.

This study aimed to test whether the above information accumulation explanation also exists in the returner-explorer classification, as it would provide an explanation of the observation duration-dependence of the returners and explorers dichotomy. More specifically, this explanation would suggest that the observation duration-dependence of the returners and explorers dichotomy is the consequence of the accumulation of information used to infer people's mobility patterns, and the accumulation process is completed when marginal information gains from extra data become negligible.

Should the above explanation apply, it could be inferred that down sampling the trajectory data would impede the information accumulation process and hence decrease the change speeds of the proportions of the two classes. Therefore, to test this explanation, each individual's original trajectories in the dataset were uniformly resampled at a two-thirds sampling rate to generate a resampled dataset, and then at a one-third sampling rate to generate a second resampled dataset. The two resultant datasets had the same duration as the original dataset, i.e. 111 days, but differed in terms of the amount of data. All individuals' classification results on different durations of the two resampled datasets were computed using the method explained in Section 3.3, and the results were compared with those from the original dataset. The comparison is illustrated in Fig. 6. As evidenced in the figure, smaller datasets were associated with slower convergence processes. While the proportion of returners and that of explorers reached a balance on Day 21 in the original dataset, this point was pushed back to Day 28 in the first resampled dataset, and further to Day 58 in the second resampled dataset. This result suggested that the returner-explorer classification is subject to a similar impact of the information accumulation process reported in Stanley et al.'s study (2018).

In addition, the KL divergence series, whose computational method is detailed in the supplementary materials, were calculated for each individual in the dataset to determine the minimum days required to completely capture the individuals' routine mobility (Stanley et al., 2018). The result showed that, at the individual level, a median minimum of 37 days was required to accumulate sufficient information to completely capture the individuals' routine spatial behaviors.

However, as shown in Figs. 3 and 4, changes in the classification results were observed among a good portion of individuals after the observation duration exceeded 37 days. This motivated us to examine whether the observation duration-dependence of the returners and explorers dichotomy is solely caused by the information accumulation process. Specifically, should the information accumulation process be the only cause, then randomizing the data would cause little effect on the minimum observation duration required to completely capture each individual's routine mobility. To investigate whether the above inference is true, the time sequence of the original dataset was randomized, by preserving individual routines only up to the daily level. Using the randomized dataset, the population was classified into returners and explorers on the n-day duration, where n traversed from 1 to 111. Then, for each individual the minimum observation duration required to reach stable classification results was calculated, and compared with that obtained from the original dataset. The difference between these two durations was calculated for each individual, and the distribution of this difference within the entire population is illustrated in Fig. 7. As can be seen in the figure, for a sizable portion of the population, there was significant difference before and after the randomization. Specifically, over 50 % individuals had a difference over 10 days. Therefore, the above inference was proven false, which indicated that the trajectories should not be purely interpreted as time-series data, and that the observed observation duration-dependence of the returners and explorers dichotomy is a consequence resulting from human behaviors instead of a simple consequence of time constraints. Moreover, the above finding suggests that the information accumulation process is not the only explanation that dominates the observation duration-dependence the returners and explorers dichotomy.

5.2. Explanation two: exploration of new locations

A second possible explanation that may be responsible for the observation duration-dependence of the returners and explorers dichotomy is related to individuals' visits to new locations. 'Exploration' refers to a visit to a place that has not been visited by an individual before. When an exploration occurs, it could impact the individual's radius of gyration, especially when the distance of the new location largely differs from the individual's previous characteristic traveled distance. This may in turn change the relationship between the individual's recurrent mobility and overall mobility, hence flipping the individual's returner-explorer classification.

Individuals tend to explore new locations out of their regular paths, which may cause changes in their returner-explorer classification results (Quadri, Zignani, Gaito, & Rossi, 2018). Contrary to a common notion that individuals usually return to previously visited locations and explore new locations once in a while (Quadri et al., 2018), empirical evidence has shown that individuals' probability for exploration every time they select the location of their next visit is between 0.2 and 0.25, in other words, on average an individual discovers a new place in every 4 or 5 visits (Cuttone et al., 2018). For the dataset used in this study, Fig. 8 shows the average number of locations that each individual has visited since the beginning of the study period. This number grew steadily over time, and reached approximately 65 at the end of the study

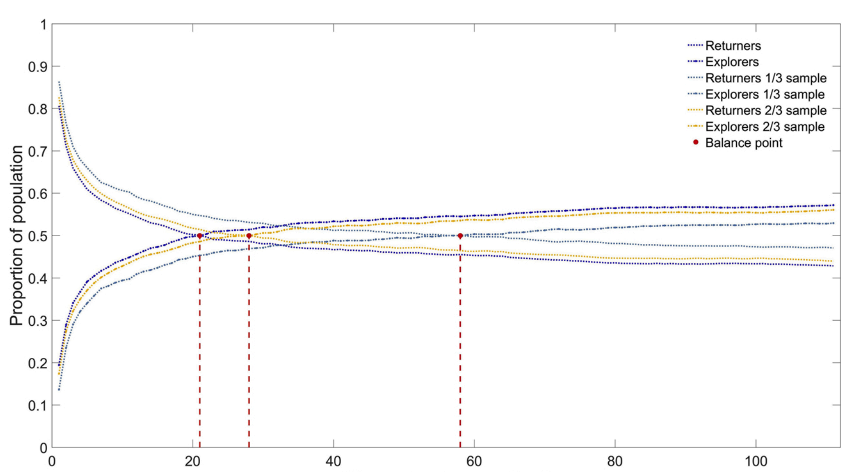

Fig. 6. Comparison of classification results from the original dataset and two resampled datasets.

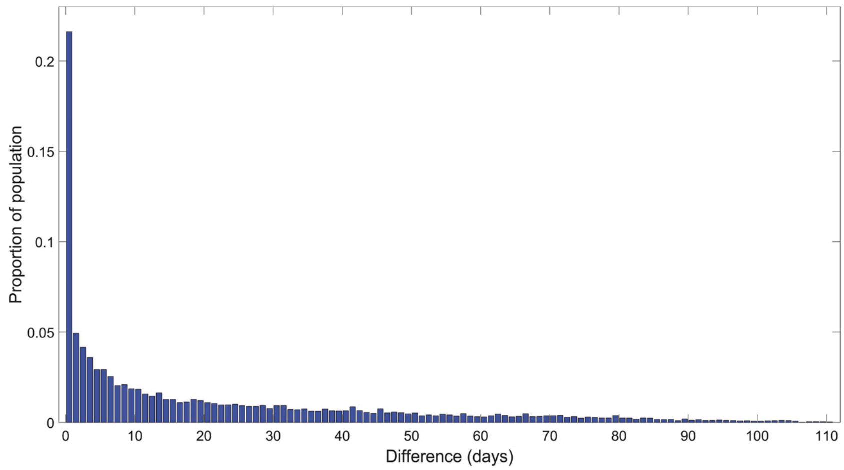

Fig. 7. The distribution of individuals' differences in the minimum duration to complete information accumulation between the original and randomized datasets.

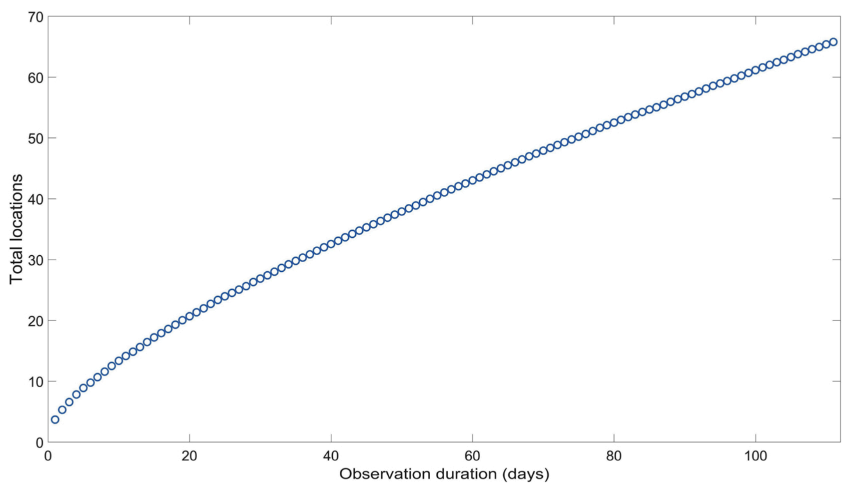

Fig. 8. Average number of total visited locations of individuals in the dataset.

period, which is consistent with results reported by Cuttone et al. (2018).

For each individual, this study further identified the days when explorations occurred, and examined whether the individual's returner-explorer classification changed on these days. If a change in the individual's classification was observed on the same day when the individual explored at least one new location, the change in the classification could be considered to be possibly caused by the exploration. Using this criterion, it was found that 64 % changes in the classifications of all 21,240 individuals during the entire study period were possibly associated with the exploration of new locations. This finding indicated that individuals' exploration behaviors, which were found to be common and consistent, may be a main cause for the observation duration-dependence of the returner-explorer dichotomy, particularly after the information process completes.

Moreover, although individuals have a steady tendency to explore new locations, their mobility ranges do not diffuse with the same trend. Findings in prior research showed a bound nature of human trajectories, suggesting that individuals' radius of gyration would become 'saturated' after several months' observation and can be approximated by a manifestly slower logarithmic growth (Gonzalez et al., 2008; Song, Koren et al., 2010). This saturation process was also observed in this study, as illustrated in Fig. 4. This finding suggested that the influence of individuals' exploration behaviors on their returner-explorer classifications would gradually decrease as the duration of observation expands. This inference helps explain the decreasing changing rate of the proportions of the two subpopulations observed beyond 37 days in the dataset.

5.3. Explanation three: changes in important locations

A third possible explanation that may be responsible for the observation duration-dependence of the returners and explorers dichotomy is related to changes in individuals' routine activities. Individuals' routine activities, such as commuting, working and social activities, are usually associated with important locations such as home and workplace (Song et al., 2016). Moreover, they are inclined to spend the majority of their time in these few important locations. This means when individuals change their routine activities, their mobility behaviors may be affected, which in turn may cause changes in their returner-explorer classifications.

Taking individuals' top two most frequently visited locations as an example, this study aimed to investigate the influence of changes in important locations on individuals' mobility patterns. The top two most frequently visited locations are usually individuals' home and workplace, which can be identified from their trajectories with relatively high accuracy by considering the time and frequency of the visits (Jiang et al., 2013). Specifically, based on the notion that individuals are normally at their workplace during the daytime on weekdays, whilst they are most likely to be found at home during the nighttime, the periods of 9 a.m. to

5 p.m. and 0 a.m. to 6 a.m. were used to define daytime and nighttime (Jiang et al., 2016), respectively. For each individual, the DBSCAN method was first used to identify the frequently visited locations. Then, for each week (five consecutive weekdays), these locations were ranked according to the visitation frequencies. The most frequently visited locations during the daytime and nighttime were identified as workplace and home for this individual. As the dataset included 22 weeks, for each individual there were up to 22 consecutive estimates of his/her home and workplace locations, which allowed for the detection of changes in these two locations.

Based on the dataset, the home and workplace locations were identified for 20,626 individuals, among whom a total of 15,138 individuals had their recurrent mobility ($k = 2$) dominated by the home and workplace locations. For these 15,138 individuals, changes in their home and workplace location with distance over 500 m (Long & Thill, 2015) were captured and recorded, by comparing the consecutive weekly estimates, in order to investigate whether and to what extent these changes possibly affected the individuals' classifications. Among the 15,138 individuals, 6,197 of them changed their home or workplace locations, and 49 % of the changes in their returner-explorer classifications were preceded by at least one home or workplace change within the previous two weeks. Conversely, 33 % of home and workplace changes associated with these 15,138 individuals were followed by at least one subsequent change in the corresponding individuals' returner-explorer classifications within the following two weeks. The above findings, although not amounting to definitive conclusions, showed substantial correlations between changes in individuals' important locations and changes in their returner-explorer classifications, suggesting the impact of the third explanation. This was further corroborated by the observation that the classification changes that followed the important location changes were generally lasting (they lasted 42 days on average), which was probably because changes in important locations tend to have a long-term and decisive impact on an individual's mobility pattern. In sum, the above findings indicated that changes in peoples' most visited locations, such as home and workplace, which are infrequent for each individual but occur consistently at the population level, may be largely responsible for the observed dynamics in returner-explorer classification results, particularly in the long term when the applicability of the other two explanations becomes limited.

6. Conclusions, implications and future research

The returners and explorers dichotomy has been widely applied to describe individuals' mobility patterns in previous studies, in which the classification was conducted based on mobility datasets of different temporal durations, such as one month (Pappalardo et al., 2015), six months (De Nadai et al., 2019; Liao et al., 2019), or two years (Alessandretti et al., 2018). The differing study durations may cause anomalies when the observation duration-dependence of the returners and explorers dichotomy is ignored. For example, the GPS dataset used in Pappalardo et al.'s study (2015) contained around one-month vehicle GPS trajectories, which was insufficient to reach reliable classification results. In fact, the researchers realized that there were always more k -returners than k -explorers regardless of the value of k, which contradicted the outcomes from another dataset with much longer duration used in the same study. The researchers argued that the unexpected result may have been caused by the unbalanced representation of transportation modes in the collected GPS data. Yet, the above hypothesis was not tested in the paper. The existence of the observation duration-dependence of the returner-explorer dichotomy, however, provides an alternative and more compelling explanation for the inconsistency in the above study: the proportions of returners and explorers are dependent on the observation duration, and the one-month GPS dataset contained insufficient information to reliably characterize individuals' mobility patterns. Moreover, based on findings of this study related to the first and second explanations, when the dataset size is small and information is insufficient, as was the case with the above GPS dataset, the proportions of returners and explorers in a population would be overestimated and underestimated, respectively.

The existence of the observation duration-dependence in the returners and explorers dichotomy suggests that prior studies' outcomes derived from the returner-explorer classifications may need to be reexamined. Particularly, prior findings based on short-term mobility datasets may yield inaccurate and unreliable characterization of individuals' and population's mobility patterns. Meanwhile, it was found in this study that the classification results would become stable, at both the population- and individual-level, suggesting that prior findings based on long-term mobility datasets may be more generalizable and reliable. The specific boundary between the short term and the long term may vary, depending on several factors such as the precision of trajectory data and the rate of data sampling. In addition, knowing that the classification results would become stable in the long term has another important implication. While reliable urban mobility studies prefer more and long-duration mobility data, data availability is always a challenge. Findings in this study suggest that a cut-off value of the amount of data can be identified for data collection to ensure reliable classification results while avoiding excessive data collection efforts.

Three underlying explanations of the returner and explorer dichotomy were revealed and tested in this study, which altogether uncover the observation duration-dependence of the dichotomy. First, for the information accumulation explanation, the result of the KL series analysis showed that the average minimum observation duration required to capture the complete mobility space of the population was 37 days, which is almost the twice of the average minimum days in the study of Stanley et al. (2018). This suggests that the minimum observation duration required to reliably capture human mobility patterns should be determined on a case-by-case basis, subject to impacts of the temporal and spatial resolutions of the dataset (Stanley et al., 2018). Also, the dataset used in Stanley et al. (2018)'s study involved students only, while the dataset used in this study was sampled from the general population of a megacity. The heterogeneity in urban demographics is another important factor to consider when determining the minimum observation duration. Second, regarding the exploration explanation, the findings in this study suggest that, although individuals have a steady tendency to explore new locations, their radii of gyration will not grow infinitely because of the recurrent nature of human mobility, as many of the newly explored locations are not far from the few import locations, such as home and workplace (Quadri et al., 2018). In addition, the findings highlight the significant impact of the exploration human mobility patterns, and indicated that it should be properly modeled to improve existing mobility predictors (Cuttone et al., 2018). Lastly, the third explanation suggests that changes in individuals' daily activities such as commuting and social activities may cause changes in their most visited locations, hence impacting their returner-explorer classifications. For instance, urban residents may change home locations or workplaces for more balanced jobs-housing relationships. Using Beijing as an example, Huang, Levinson et al. (2018) found that over 60 % of the residents have more than two home moves and/or job changes over seven years. Such permanent changes in individuals' mobility patterns are almost certain to be reflected in their return-explorer classifications. This also suggests that the observation of changes in individuals' important locations could be used as a predictor of possible subsequent changes in their returner-explorer classifications.

To sum up, based on the analysis of high-granularity trajectory data collected from a large and heterogeneous population, this study examined the observation duration-dependence of the returners and explorers dichotomy, and investigated three possible explanations that could explain the observation duration-dependence. By achieving these objectives, this study advances the understanding of the differences of individuals' mobility patterns between the short term and the long term, and contributes to the existing literature on urban human mobility patterns and their dynamics.

The findings of this study could also shed light on a few practical challenges that cities are faced with. First, the revealed observation duration-dependence of human mobility patterns could underline and help avoid the often-seen mismatches between human mobility tracking technologies or mobility datasets over different temporal scales and their real-world applications in cities. Second, the explanations found to be dominating the observation duration-dependence of individuals' mobility patterns could enable urban planners and policymakers to better understand and predict the mobility behaviors of the urban population, so that they can make more informed decisions regarding the provision and management of public goods (Zhang, Liu et al., 2019). Third, the temporal dynamics of individuals' mobility patterns identified in this study could be used to improve upon existing contagion dynamics modeling and prediction approaches (Balcan & Vespignani, 2011; Xu et al., 2017), which are critical to the control of infectious diseases, a major threat to public health faced by most cities as remarkably evidenced by the recent coronavirus pandemic. Fourth, based on the correlation between individuals' spatial mobility and their social ties found in prior research (Fan et al., 2017; Pelechrinis & Krishnamurthy, 2016), the observation duration-dependence of individuals' mobility patterns revealed in this study could help identify, explain and predict similar dynamics of individuals' activities in the cyberspace.

Lastly, this study bears three noteworthy limitations. First, this study only examined individuals' trajectories in the city of Guangzhou. Future research could extend the investigation to other cities with differing demographic structures and urban forms to provide more generalizable findings. Second, this study only considered individuals' trajectories on weekdays, thus the findings may not apply to urban mobility over weekends, holidays, or during extreme weather events. Additional volunteered geographic information (VGI)-based empirical studies could be conducted to investigate and compare the observation duration-dependence of the dichotomy within different temporal contexts. Third, this study mainly focused on the effects of the three explanations on the observation duration-dependence of the dichotomy at the population level, and did not examine whether such effects would vary across different groups of individuals. This could be addressed in future research by considering certain demographic characteristics such as gender, age, occupation, and income, should relevant data become available.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This material is based upon work supported by the National Natural Science Foundation of China (NSFC) under Grant No. 71974105. The authors from Tsinghua University are grateful for the support of the NSFC. Any opinions, findings, conclusions or recommendations expressed in the paper are those of the authors and do not necessarily reflect the views of the funding agency.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.scs.2021.102862.

References

Alessandretti, L., Lehmann, S., & Baronchelli, A. (2018). Understanding the interplay between social and spatial behaviour. *EPJ Data Science, 7*(1), 36. https://doi.org/10.1140/epjds/s13688-018-0164-6

Balcan, D., & Vespignani, A. (2011). Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nature Physics, 7*(7), 581–586. https://doi.org/10.1038/nphys1944

Barbosa, H., de Lima-Neto, F. B., Evsukoff, A., & Menezes, R. (2015). The effect of recency to human mobility. *EPJ Data Science, 4*(1), 21. https://doi.org/10.1140/epjds/s13688-015-0059-8

Barbosa, H., de Lima-Neto, F. B., Evsukoff, A., & Menezes, R. (2016). *Returners and explorers dichotomy in web browsing behavior—A human mobility approach.* Springer International Publishing.

Bibri, S. E., & Krogstie, J. (2017). Smart sustainable cities of the future: An extensive interdisciplinary literature review. *Sustainable Cities and Society, 31*, 183–212. https://doi.org/10.1016/j.scs.2017.02.016

Brum-Bastos, V. S., Long, J. A., & Demšar, U. (2018). Weather effects on human mobility: A study using multi-channel sequence analysis. *Computers, Environment and Urban Systems, 71*, 131–152. https://doi.org/10.1016/j.compenvurbsys.2018.05.004

Cuttone, A., Lehmann, S., & González, M. C. (2018). Understanding predictability and exploration in human mobility. *EPJ Data Science, 7*(1), 2. https://doi.org/10.1140/epjds/s13688-017-0129-1

De Nadai, M., Cardoso, A., Lima, A., Lepri, B., & Oliver, N. (2019). Strategies and limitations in app usage and human mobility. *Scientific Reports, 9*(1), 10935. https://doi.org/10.1038/s41598-019-47493-x

Dixon, B. (2017). Investigating clustering algorithm DBSCAN to self select locations for power based malicious code detection on smartphones, 11-12 Feb. 2017 *Paper Presented at the 2017 Third International Conference on Mobile and Secure Services (MobiSecServ)*.

Fan, C., Liu, Y., Huang, J., Rong, Z., & Zhou, T. (2017). Correlation between social proximity and mobility similarity. *Scientific Reports, 7*(1), 11975. https://doi.org/10.1038/s41598-017-12274-x

Feng, X., Wang, J., Kong, X., Wang, Z., & Liu, C. (2018). Exploring human mobility patterns in urban scenarios: A trajectory data perspective. *IEEE Communications Magazine, 56*(3), 142–149.

Forghani, M., & Karimipour, F. (2018). Interplay between urban communities and human-crowd mobility: A study using contributed geospatial data sources. *Transactions in GIS, 22*(4), 1008–1028. https://doi.org/10.1111/tgis.12465

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779–782. https://doi.org/10.1038/nature06958

Guadagno, L. (2016). Human mobility in the Sendai framework for disaster risk reduction. *International Journal of Disaster Risk Science, 7*(1), 30–40. https://doi.org/10.1007/s13753-016-0077-6

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems, 54*, 240–254. https://doi.org/10.1016/j.compenvurbsys.2015.09.001

Huang, J., Levinson, D., Wang, J., Zhou, J., & Wang, Z.-j. (2018). Tracking job and housing dynamics with smartcard data. *Proceedings of the National Academy of Sciences, 115*(50), 12710–12715. https://doi.org/10.1073/pnas.1815928115

Jiang, S., Fiore, G. A., Yang, Y., Ferreira, J., Frazzoli, E., & González, M. C. (2013). A review of urban computing for mobile phone traces: Current methods, challenges and opportunities. *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, 1–9.

Jiang, S., Yang, Y., Gupta, S., Veneziano, D., Athavale, S., & González, M. C. (2016). The TimeGeo modeling framework for urban motility without travel surveys. *Proceedings of the National Academy of Sciences of the United States of America*, Article 201524261.

Liao, Y., Yeh, S., & Jeuken, G. S. (2019). From individual to collective behaviours: Exploring population heterogeneity of human mobility based on social media data. *EPJ Data Science, 8*(1), 1–22. https://doi.org/10.1140/epjds/s13688-019-0212-x

Liu, Z., Qiao, Y., Tao, S., Lin, W., & Yang, J. (2017). Analyzing human mobility and social relationships from cellular network data. In *13th International Conference on Network and Service Management (CNSM)*.

Long, Y., & Thill, J.-C. (2015). Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing. *Computers, Environment and Urban Systems, 53*, 19–35.

Lopez-Carreiro, I., & Monzon, A. (2018). Evaluating sustainability and innovation of mobility patterns in Spanish cities. Analysis by size and urban typology. *Sustainable Cities and Society, 38*, 684–696. https://doi.org/10.1016/j.scs.2018.01.029

Luo, T., Zheng, X., Xu, G., Fu, K., & Ren, W. (2017). An improved DBSCAN algorithm to detect stops in individual trajectories. *ISPRS International Journal of Geo-Information, 6*, 63. https://doi.org/10.3390/ijgi6030063

Nogal, M., & Honfi, D. (2019). Assessment of road traffic resilience assuming stochastic user behaviour. *Reliability Engineering & System Safety, 185*, 72–83. https://doi.org/10.1016/j.ress.2018.12.013

Papandrea, M., Jahromi, K. K., Zignani, M., Gaito, S., Giordano, S., & Rossi, G. P. (2016). On the properties of human mobility. *Computer Communications, 87*, 19–36.

Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A. (2015). Returners and explorers dichotomy in human mobility. *Nature Communications, 6*(8166), 8166.

Pelechrinis, K., & Krishnamurthy, P. (2016). Socio-spatial affiliation networks. *Computer Communications, 73*, 251–262. https://doi.org/10.1016/j.comcom.2015.06.002

Peprah, C., Amponsah, O., & Oduro, C. (2019). A system view of smart mobility and its implications for Ghanaian cities. *Sustainable Cities and Society, 44*, 739–747. https://doi.org/10.1016/j.scs.2018.10.025

Qian, X., Lei, T., Xue, J., Lei, Z., & Ukkusuri, S. V. (2020). Impact of transportation network companies on urban congestion: Evidence from large-scale trajectory data. *Sustainable Cities and Society, 55*, Article 102053. https://doi.org/10.1016/j.scs.2020.102053

Quadri, C., Zignani, M., Gaito, S., & Rossi, G. P. (2018). On non-routine places in urban human mobility, 1-3 Oct. 2018 *Paper Presented at the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*.

Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S. J., & Chong, S. (2011). On the levy-walk nature of human mobility. *IEEE/ACM Transactions on Networking, 19*(3), 630–643.

Saja, A. M. A., Teo, M., Goonetilleke, A., Ziyath, A. M., & Nianthi, K. W. G. R. (2020). Surrogate measures to assess mobility of people as a resilience indicator in disaster management: An exploratory study in Southeastern Sri Lanka. *International Journal of Disaster Risk Science, 11*(1), 13–31. https://doi.org/10.1007/s13753-020-00251-4

Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of the Royal Society, Interface, 10* (84), Article 20130246. https://doi.org/10.1098/rsif.2013.0246

Shin, D., Aliaga, D., Tunçer, B., Arisona, S. M., Kim, S., Zünd, D., & Schmitt, G. (2015). Urban sensing: Using smartphones for transportation mode classification. *Computers, Environment and Urban Systems, 53*, 76–86. https://doi.org/10.1016/j.compenvurbsys.2014.07.011

Shoghri, A. E., Liebig, J., Gardner, L., Jurdak, R., & Kanhere, S. (2019). How mobility patterns drive disease spread: A case study using public transit passenger card travel data. *2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*.

Solmaz, G., & Turgut, D. (2017). Modeling pedestrian mobility in disaster areas. *Pervasive and Mobile Computing, 40*, 104–122. https://doi.org/10.1016/j.pmcj.2017.05.005

Song, X., Zhang, Q., Sekimoto, Y., Shibasaki, R., Yuan, N. J., & Xie, X. (2016). Prediction and simulation of human mobility following natural disasters. *ACM Transactions on Intelligent Systems and Technology (TIST), 8*(2), 1–23.

Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature Physics, 6*(10), 818–823. https://doi.org/10.1038/nphys1760

Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010). Limits of predictability in human mobility. *Science, 327*(5968), 1018–1021. https://doi.org/10.1126/science.1177170

Stanley, K., Yoo, E.-H., Paul, T., & Bell, S. (2018). How many days are enough?: Capturing routine human mobility. *International Journal of Geographical Information Science, 32*(7), 1485–1504. https://doi.org/10.1080/13658816.2018.1434888

Tai, C., Lee, Y., & Lin, C. (2010). Urban disaster prevention shelter location and evacuation behavior analysis. *Journal of Asian Architecture and Building Engineering, 9* (1), 215–220.

Tsiotas, D., & Polyzos, S. (2017). The topology of urban road networks and its role to urban mobility. *Transportation Research Procedia, 24*, 482–490. https://doi.org/10.1016/j.trpro.2017.05.087

Wang, Q., & Taylor, J. E. (2014). Quantifying human mobility perturbation and resilience in Hurricane Sandy. *PloS One, 9*(11), Article e112608. https://doi.org/10.1371/journal.pone.0112608

Wang, Y., Wang, Q., & Taylor, J. E. (2017). Aggregated responses of human mobility to severe winter storms: An empirical study. *PloS One, 12*(12), Article e0188734.

Wang, M., Yang, S., Sun, Y., & Gao, J. (2017). Human mobility prediction from region functions with taxi trajectories. *PloS One, 12*(11). https://doi.org/10.1371/journal.pone.0188735

Xu, Z., Glass, K., Lau, C. L., Geard, N., Graves, P., & Clements, A. (2017). A synthetic population for modelling the dynamics of infectious disease transmission in American Samoa. *Scientific Reports, 7*(1), 16725. https://doi.org/10.1038/s41598-017-17093-8

Yang, X., Zhao, Z., & Lu, S. (2016). Exploring spatial-temporal patterns of urban human mobility hotspots. *Sustainability, 8*(7). https://doi.org/10.3390/su8070674

Zamichos, A., Theodorou, T., Drosou, A., & Tzovaras, D. (2018). An evacuation management framework based on urban data and personalized mobility information. *2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService)*.

Zhang, F., Li, Z., Li, N., & Fang, D. (2019). Assessment of urban human mobility perturbation under extreme weather events: A case study in Nanjing, China. *Sustainable Cities and Society, 50*, Article 101671. https://doi.org/10.1016/j.scs.2019.101671

Zhang, S., Liu, X., Tang, J., Cheng, S., & Wang, Y. (2019). Urban spatial structure and travel patterns: Analysis of workday and holiday travel using inhomogeneous Poisson point process models. *Computers, Environment and Urban Systems, 73*, 68–84. https://doi.org/10.1016/j.compenvurbsys.2018.08.005

Zhou, J., & Long, Y. (2014). Jobs-housing balance of bus commuters in Beijing: Exploration with large-scale synthesized smart card data. *Transportation Research Record, 2418*(1), 1–10. https://doi.org/10.3141/2418-01

Zignani, M., Gaito, S., & Rossi, G. (2013). Extracting human mobility and social behavior from location-aware traces. *Wireless Communications and Mobile Computing, 13*(3), 313–327.