



Modeling the scaling properties of human mobility in virtual space

Mingyan Wang · An Zeng · Xiaohua Cui 

Received: 30 March 2023 / Accepted: 2 June 2023 / Published online: 17 June 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract People are increasingly involved in online activities. Online activities can be regarded as movements in virtual space, such as jumping from webpage to webpage while surfing online, switching channels while watching TV, and browsing commodities while shopping online, which can affect information propagation, innovation spread, social activities, etc. Most previous efforts have been devoted to modeling the scaling properties of human mobility in physical space. Few studies aim to establish a unified and integral model to understand the fundamental dynamics underlying human virtual mobility. In this paper, we study human mobility in virtual space empirically and theoretically based on two datasets involving TV watching and online shopping and attempt to answer three unsolved issues. First, human virtual mobility shares common features, supported by the fact that striking agreements appear in the scaling properties of both datasets. Second, there exists a universal rule governing

an individual's choice in virtual mobility, which is distinct from that in the real world due to travel restrictions. Third, there exists a unified model incorporating the behavior rule unique to virtual space under the framework of Exploration and Preferential Return, which can be used to reproduce the scaling properties of virtual mobility. We reveal the mechanism behind human virtual mobility through consistent scaling properties and develop a corresponding dynamic model based on empirical data.

Keywords Human virtual mobility · Scaling properties · Memory effect · Behavior mechanism

1 Introduction

In the real world, individuals constantly travel from one location to another, forming trajectories. Uncovering the statistical patterns that characterize these trajectories [1–4] is of importance for disease control [5, 6], congestion alleviation [7, 8], information propagation [9, 10], etc. The key to understanding human mobility dynamics is to establish models to reproduce the scaling properties of empirical data such as the waiting time distribution $P(\Delta t)$, the distance distribution $P(\Delta r)$, and the visitation frequency f_r , where Δt denotes the time spent by an individual at a location, Δr is distances covered by an individual between consecutive sightings, and f_r is the frequency f of the r_{th} most visited location [11]. Moreover, the model must be

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11071-023-08642-0>.

M. Wang · A. Zeng · X. Cui (✉)
School of Systems Science, Beijing Normal University, Beijing 100875, People's Republic of China
e-mail: xhcui@bnu.edu.cn

M. Wang
e-mail: 202021250014@mail.bnu.edu.cn

A. Zeng
e-mail: anzeng@bnu.edu.cn

self-consistent, which requires that it can reflect the behavior rules of travel microscopically and can be unified with kinds of scaling properties macroscopically. A representative work is the Exploration and Preferential Return (EPR) model proposed in 2010 [12]. Two generic mechanisms, exploration and return, governing human trajectories are introduced to account for the empirically observed scaling exponents. In many subsequent studies, rules within the EPR mechanism were added or adjusted to make predictions more consistent with empirical data. The representative models include the radiation EPR model [13, 14], memory-preferential random walk model [15], and EPR model conditional on current location [16]. These quantifiable human mobility models have evolved from general random walk models [17] to models based on travel mechanisms or memory effects, etc., which match better with reality [12, 18]. The involved fields have also expanded from human travel to urban mobility [19], income segregation [20] and even mobility in virtual space [21].

With the rapid advances in information and communication technologies (ICT), people are increasingly involved in online activities. Online activities can be regarded as virtual movements [21, 24]. Examples include continuous browsing from commercial websites in a single online session, continuous channel switching on a digital TV, and consecutive posting on social networking sites. Facilitated by massive online datasets, an increasing number of studies are devoted to revealing the intrinsic patterns underlying human virtual mobility. Quantities concerned in the virtual space are similar to those in the physical world, including the waiting time distribution $P(\Delta t)$ (Δt is the time spent by an individual at a site), the visitation frequency f_r (the frequency f of the r_{the} most visited site), and the number of distinct sites $S(t)$ (the number of sites visited). From the real world to virtual space, some studies have explored the characteristics of human virtual mobility. Firstly, the visitation frequency f_r of human virtual mobility was found to follow Zipf's law, and an EPR model regardless of waiting time Δt was constructed to reproduce f_r 's scaling properties [21]. Then, f_r was further confirmed to follow Zip's law, and the number of distinct sites $S(t)$ is challenged not to be power law like in the real world. Moreover, a model integrating the preference attachment mechanism and EPR framework was established to interpret the exponents of these quantities [22]. Further, the scaling prop-

erties of real travel and virtual movements were compared. It was found that the waiting time distribution $P(\Delta t)$ of human virtual mobility is heavy-tailed and a model was proposed to link the critical exponents characterizing the spatial dependencies in human mobility and social networks [23]. Recently, there have been discussions on the scaling properties of virtual movement at different scales. Reddit posting data proves that f_r , $S(t)$, and $P(\Delta t)$ are power-law at the community level [24], while cross-app usage data indicates the power-law properties of f_r and the non-power-law properties of $S(t)$ [25].

However, there is still a lack of an empirical, unified, and integral dynamic model for human virtual mobility, which is manifested in a series of unsolved issues. First, most studies only involved scaling properties of one specific online activity, without comparing exponents across different datasets [24]. Unlike the empirical exponents of real-world travel, scaling properties have been proven highly uniform in dollar-bill tracking [1], mobile-phone data [26–28], taxi data [29], etc. Second, the relationship between the waiting time Δt , the number of distinct sites $S(t)$, and the visitation frequency f_r are not fully discussed. Some models ignore the waiting time Δt and take the number of movements n as the only key parameter to derive the scaling exponents [21], and some other models only explain the characteristics of visitation frequency f_r [22, 25]. Third, mobility dynamics are mostly described as a memory-based random walk process [30–33]. Various assessments and quantifications of the memory effect are based on their assumptions and are not verified in empirical data [22, 24]. Therefore, three questions regarding human mobility in virtual space naturally arise. Does human mobility in virtual space share common features? Does online behavior data contain generic rules? Does there exist a unified model that can reproduce the properties of the waiting time distribution $P(\Delta t)$, the number of distinct sites $S(t)$, and the visitation frequency f_r ? In this paper, we discuss human virtual mobility based on two typical datasets, from phenomena to rule and further to model, to answer the three issues raised.

2 Data description

Individuals' changes in physical locations are known as mobility in the real world [1–4]. Accordingly, the

changes in virtual locations, that is, online activities, are called virtual mobility [21,24]. In the past few years, network technology has made it possible to track humans' online footprints. Phone records, online shopping records, web browsing records, and other datasets have provided a new momentum to study human mobility in virtual space. Although these datasets vary greatly in terms of their fields and sources, the results seem to agree on a number of quantitative characteristics of human mobility [34].

In this paper, we use two datasets from different domains to uncover the scaling properties characterizing individual mobility. Dataset D1 contains TV viewing records of 30,000 anonymous users in a large city in China from July 2015 to September 2015. Dataset D2 is obtained from a large multicategory online store capturing 60,000 anonymous users' click records in October 2019. In D1, a user's switch from one channel to another is considered a movement in virtual space. In D2, a movement between commodities consists of two consecutive browsing activities (see Fig. 1). Unlike travel in the real world, travel in virtual space is instantaneous [35]. The memory effect in real travel can last for months or years, but viewing/clicking effects after one day or even one hour are almost independent of previous ones. If a user's 3-month viewing records are regarded as one trajectory, users' footprints are too random to truly characterize scaling properties. Therefore, we segment the sequence for each user and take continuous records as one trajectory [21]. A user may have multiple trajectories in the observation window (see Supplementary Section S2).

3 Results

3.1 Empirical scaling properties

Previous studies have discovered a series of scaling properties of virtual mobility in online data. For example, mobile phone data show that the waiting time distribution $P(\Delta t)$ is heavy-tailed, that is, $P(\Delta t) \propto \Delta t^{-1-\beta}$ [23]. Taobao shopping records and Reddit records both indicate that the frequency f of a user visiting a given location and its rank r follow Zipf's law: $f_r \propto r^{-\zeta}$ [22, 24]. These characteristics seem to imply that mobility in virtual space can be explained by models for the real world, especially the EPR model [12]. However, some studies have questioned that the number of distinct sites

S in virtual space is different from $S(t) \propto t^\mu$ in the real world. Some researchers proposed $S(t) \propto \frac{A}{B+t}^\mu$, while others believed that $S(n) \propto n^\mu$, where n is the number of steps. Another significant statistical characteristic, P_{new} , the probability of exploring new locations, is also controversial and is assumed to follow: $P_{new} \propto S^{-\gamma}$, $P_{new} \propto \frac{A}{B+t}$ or $P_{new} \propto n^{-\gamma}$ [21,22], etc.

We comprehensively explore the scaling properties using two datasets involving TV watching and online shopping. First, we measure the waiting time distribution $P(\Delta t)$. In the TV viewing record dataset (D1), Δt is the viewing time of the audience on a TV channel, calculated as the time difference between the start time and end time of a record. In the online shopping dataset (D2), Δt is the browsing time for a product, approximately the time interval between two consecutive browsing records. We find that $P(\Delta t) \propto \Delta t^{-1-\beta}$ with $\beta_1 = 0.78 \pm 0.11$, $\beta_2 = 0.88 \pm 0.08$ (see Fig. 2(a)), which is consistent with the form in the real world, where $\beta = 0.8 \pm 0.1$ [12]. Then, we discuss the number of distinct sites $S(t)$, finding that $S(t) \propto t^\mu$, where $\mu_1 = 0.52 \pm 0.007$, $\mu_2 = 0.57 \pm 0.011$ (see Fig. 2(b)). In the real world, $\mu = 0.6 \pm 0.02$ [12]. It is worth noting that $\beta \neq \mu$, indicating that travel in virtual space does not follow a continuous-time random walk (CTRW) [31]. Note that there is a slowdown in exploring new locations at large timescales because $\mu < 1$, suggesting a decreasing tendency to visit unvisited locations. Another quantity closely related to S is P_{new} , the probability of exploring a new location. P_{new} deviates from the expectation: $P_{new} \propto S^{-\gamma}$ (see Fig. 2c), suggesting that it may be inappropriate to directly apply the classical EPR model to model movements in virtual space. Finally, we plot the visitation frequency f_r and rank r , which can be written as $f_r \propto r^{-\zeta}$, $\zeta_1 = \zeta_2 = 0.94 \pm 0.02$ (see Fig. 2d). The visitation pattern of virtual mobility is uneven, but this unevenness is weaker than that in the real world, where $\zeta = 1.2 \pm 0.1$.

Naturally, we can draw two conclusions. First, the empirical characteristics or exponents in virtual space are not the same as those in physical space, suggesting that the mechanism behind travel in virtual space differs from that in reality. Second, there are similarities between the scaling properties of the two datasets. We notice that similar P_{new} and $\zeta_1 = \zeta_2$. Although time-related exponents $\beta_1 \neq \beta_2$, $\mu_1 \neq \mu_2$, $\mu = \frac{2}{3}\beta$ hold for both. This suggests that the scaling properties are determined by both the unified behavior mechanism and the system features.

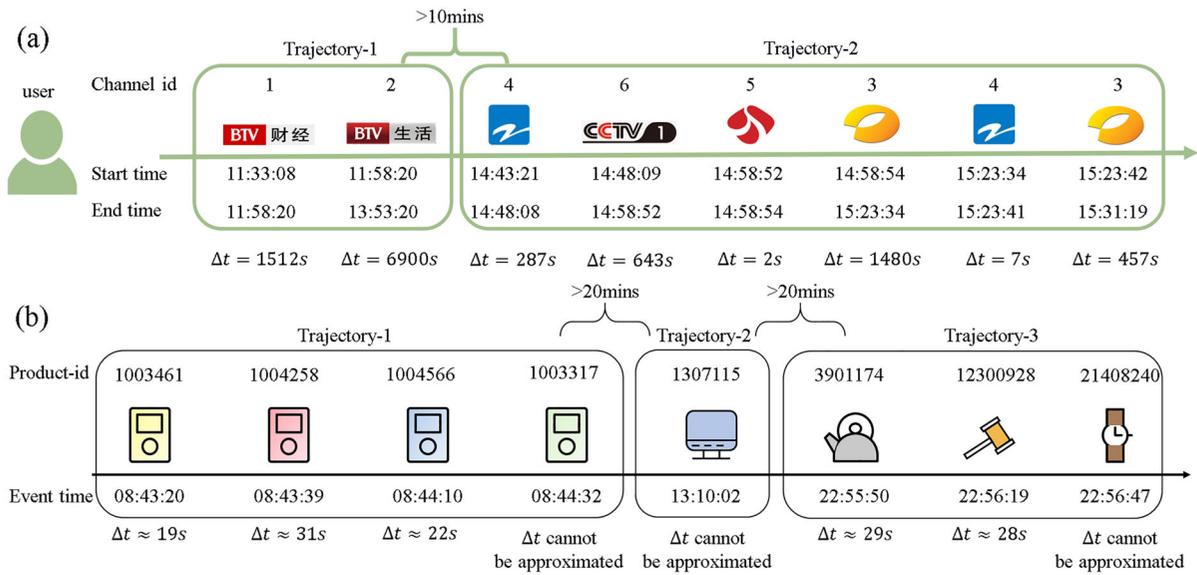


Fig. 1 Human virtual mobility in TV watching and online shopping. **a** In the D1 dataset, switching between channels are considered movements in virtual space. **b** In the D2 dataset, browsing

different products is considered virtual movement. Other figures in the paper are based on the two datasets

3.2 Human mobility model in virtual space

Before introducing our model, we first emphasize a recognized premise that the waiting time distribution $P(\Delta t)$ in virtual space is heavy-tailed, as addressed by previous research [23]. However, there is still a lack of a unified interpretation to derive other scaling properties: $S(t)$, P_{new} , and f_r . With our framework, individual human trajectories are characterized using two generic mechanisms: exploration (visiting a new location) and preferential return (visiting an already visited location). At each time step t , the individual explores a new site or returns to a previously visited site. The probability of exploration is generally assumed to follow a power law as $P_{new} = \rho S^{-\gamma}$ [12]. For the probability P for an individual moving to a previously visited site, there are many configurations such as $P_r = f_r \propto \frac{1}{r}$, $P_r \propto \frac{1+\lambda k_i}{\sum 1+\lambda k_n}$ [12, 22]. However, the above forms are only assumptions. The forms of P_{new} and P_r in virtual mobility have not been verified.

Particularly, in our model, the specific rules governing the return are summarized from the empirical data. We name it the Virtual Exploration and Preferential Return (VEPR) model. We first verify the relationship between P_r and r . It is worth noting that $P_r \propto \log_{10} \frac{1}{r}$,

which is different from $P_r \propto \frac{1}{r}$ in the real world (see Fig. 3a). This discrepancy indicates that the memory effect of return in virtual space is weaker. In the real world, considering time and distance, travel costs are relatively high, so the memory effect is particularly strong. People often return to a few locations, such as home and company, manifested as $P_r \propto \frac{1}{r}$. However, there are no such restrictions in virtual space, where almost every movement is free and instant. Even infrequently visited sites have a certain probability of return traffic, shown as $P_r \propto \log_{10} \frac{1}{r}$ (see Fig. 3b).

As shown in Fig. 3(a), $P_r \propto \log_{10} \frac{1}{r}$. We use the ordinary least square (OLS) method to obtain the coefficient k of the equation $P_r \approx k \log_{10} \frac{1}{r} + c$. However, the coefficient k varies with S , the number of distinct locations visited previously (see Fig. 5a). This means that $P_r \propto \log_{10} \frac{1}{r}$ is still not the most concise form, which is also not conducive to deriving scaling exponents theoretically. Therefore, we need to rescale the parameters in the regression equation. A surprising finding is that k and $\frac{1}{S}$ are almost perfectly linear (see Fig. 4a). Therefore, we change the equation to $P_r \propto \frac{1}{S} \log_{10} \frac{1}{r}$ for regression. There is no significant correlation between the new coefficient k' and S . For the intercept c' , we find that it is proportional to $\log_{10} S$, that is, $c' \approx c'' \log_{10} S$

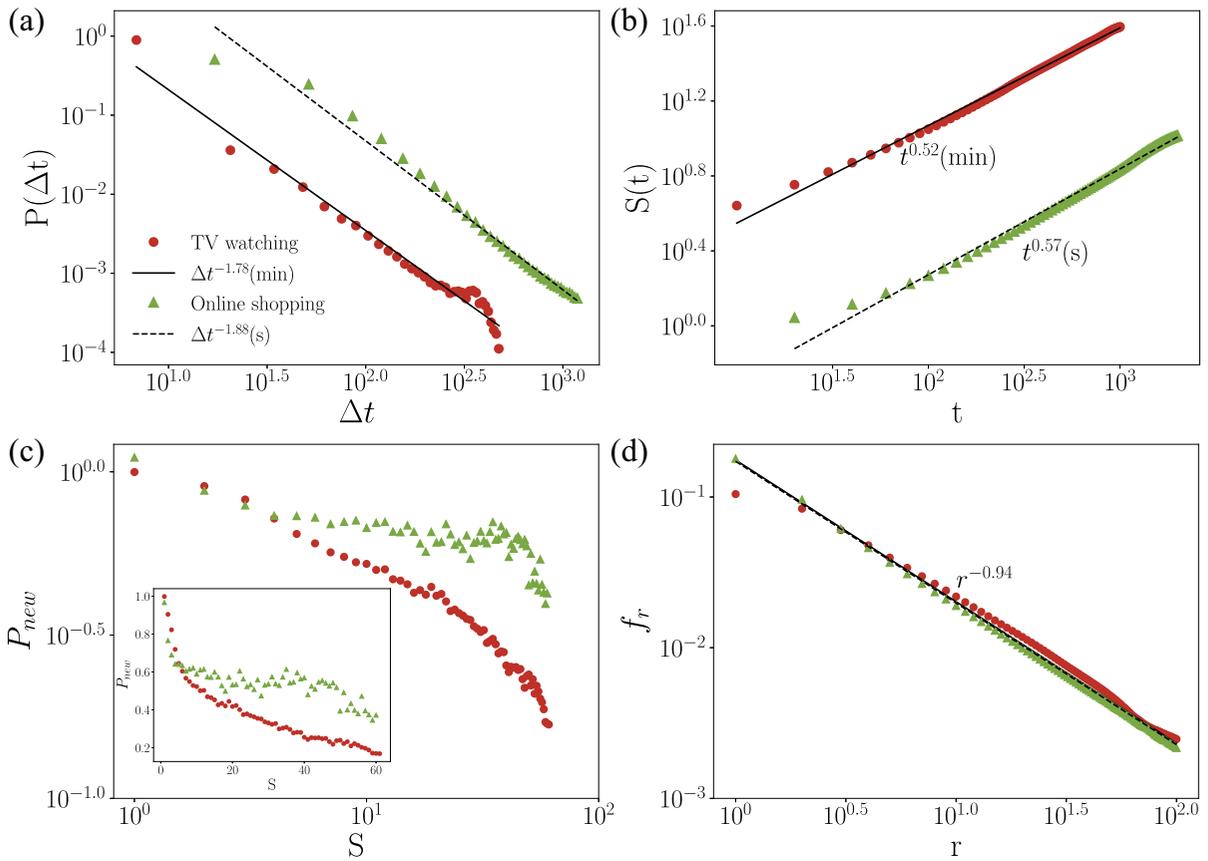


Fig. 2 Scaling properties of empirical data. D1 is the 3-month TV watching records, and D2 is the 1-month browsing records of a shopping site, represented by red circles and green triangles, respectively. **a** The waiting time distribution $P(\Delta t)$. The fitted straight lines represent $\Delta t^{-\beta}$ with $\beta_1 = 0.78 \pm 0.11$, $\beta_2 = 0.88 \pm 0.08$. **b** The number of visited distinct sites S versus time t (TV watching data in minutes, shopping data in

seconds). The fitted straight line indicates that $S(t) \propto t^\mu$, where $\mu_1 = 0.52 \pm 0.007$, $\mu_2 = 0.57 \pm 0.011$. **c** For the probability of exploring a new location P_{new} versus S , the power law is not obvious compared with real travel, indicating that the exploration mechanism is different from that in the real world. **d** The visitation frequency f_r versus rank r , showing $f_r \propto r^{-\zeta}$, with $\zeta_1 = \zeta_2 = 0.94 \pm 0.02$

(see Fig. 4b). Another surprising finding is $k' \approx c''$ (see Fig. 4c). Now, we have

$$P_r \approx k \log_{10} \frac{1}{r} + c \approx \frac{1}{S} (k' \log_{10} \frac{1}{r} + c') \approx \frac{1}{S} (k' \log_{10} \frac{1}{r} + c'' \log_{10} S) \approx \frac{1}{S} (k' \log_{10} \frac{S}{r}), \tag{1}$$

which can be written as

$$P_r \approx \frac{\lambda}{S} \log_{10} \frac{S}{r} + \sigma. \tag{2}$$

The OLS results of Eq. (2) show that there is a linear P_r and $\frac{1}{S} \log_{10} \frac{S}{r}$ at a 95% confidence level (see S4 for detailed OLS results). λ and σ under different S , obtained by Eq. (2), obey normal distribution with $\bar{\lambda}_1 = \bar{\lambda}_2 = 1.44 \pm 0.03$ (see Fig. 4e), $\bar{\sigma}_1 = 0.0026 \pm 0.0003$,

$\bar{\sigma}_2 = 0.0040 \pm 0.0023$ (see Fig. 4f) and are uncorrelated with S (see Fig. 4d). $\bar{\lambda}_1 = \bar{\lambda}_2$ illustrates that the rules governing peoples' return in human virtual mobility are common. $\bar{\sigma}_1$ and $\bar{\sigma}_2$ are close to 0, consistent with the previous approximation that $k' \approx c''$. In conclusion, the probability of returning to the r_{th} location is $P_r \approx \frac{\lambda}{S} \log_{10} \frac{S}{r} + \sigma$ (see S4 for the detailed process of rescaling). After rescaling, all P_r and $\frac{1}{S} \log_{10} \frac{S}{r}$ fall near the straight line with a slope of 1.44 (see Fig. 5b).

We develop our VEPR model according to Eq. (2). A typical step for a person browsing in virtual space is schematically illustrated in Fig. 6. An individual can perform one of the two complementary processes at each step: return or exploration. If the individual

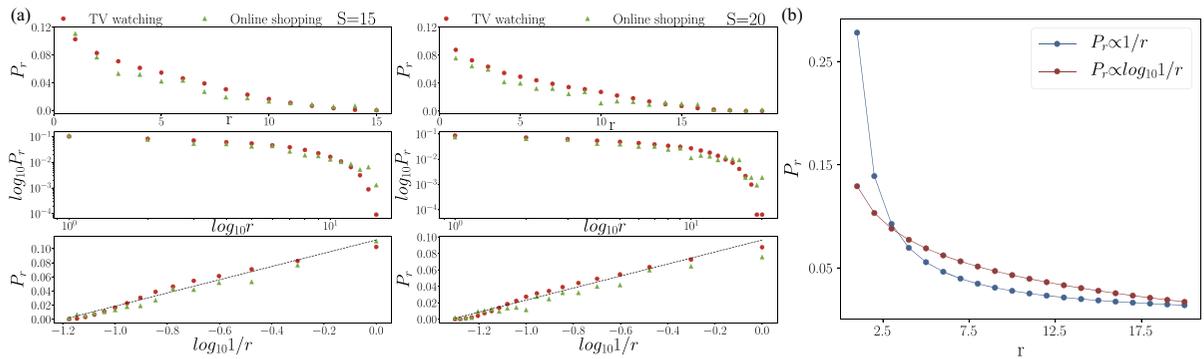


Fig. 3 The rules governing preferential return summarized from empirical data. **a** Verification of the regression forms of P_r and r . We plot P_r versus r , P_r versus $\frac{1}{r}$, and P_r versus $\log_{10}\frac{1}{r}$ when $S = 15$ and $S = 20$, finding that $P_r \propto \log_{10}\frac{1}{r}$. **b** Comparison of two return rules. The return probability distribution is relatively uniform when $P_r \propto \log_{10}\frac{1}{r}$

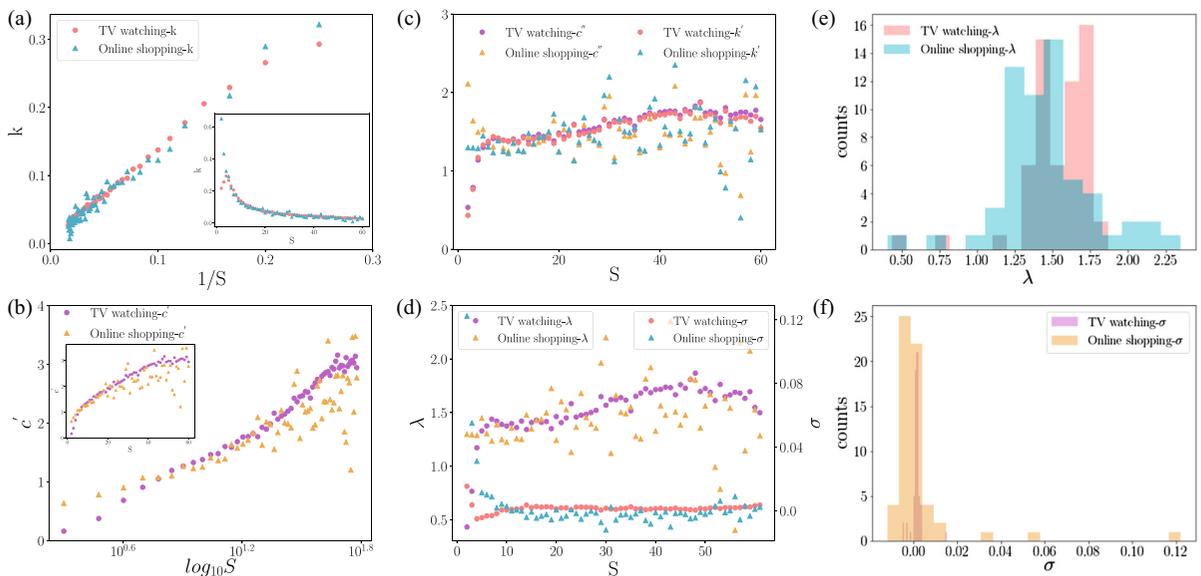


Fig. 4 Rescaling key parameters in the preferential return rules. **a** The slope k of $P_r \approx k \log_{10}\frac{1}{r} + c$ under different S varies greatly and is proportional to $\frac{1}{S}$. **b** The intercept c' of $P_r \approx \frac{1}{S}(k' \log_{10}\frac{1}{r} + c')$ is almost a linear function of $\log_{10}S$, written as $c' \approx c'' \log_{10}S$. **c** $k' \approx c''$. **d** λ and σ of $P_r \approx \frac{\lambda}{S} \log_{10}\frac{S}{r} + \sigma$ are not significantly correlated with S . **e** Distribution of λ with $\bar{\lambda}_1 = \bar{\lambda}_2 = 1.44 \pm 0.03$. **f** Distribution of σ with $\bar{\sigma}_1 = 0.0026 \pm 0.0003$, $\bar{\sigma}_2 = 0.0040 \pm 0.0023$

chooses to return to a previous location, the visitation probability P_r correlates with rank r , expressed as Eq. (2). Alternatively, he visits a new location with the probability $P_{new} = 1 - P_{ret} = 1 - \sum P_r$. In what follows, we show that the individual mobility model incorporating preferential return and exploration is sufficient to explain a series of scaling properties.

3.3 Theoretical analysis

Our VEPR model has two parameters, λ and σ , which control the user's probability of returning to a previously visited location and determine the tendency of the user to explore a new location. We aim to obtain the analytic forms of the three characterizing quantities mentioned above.

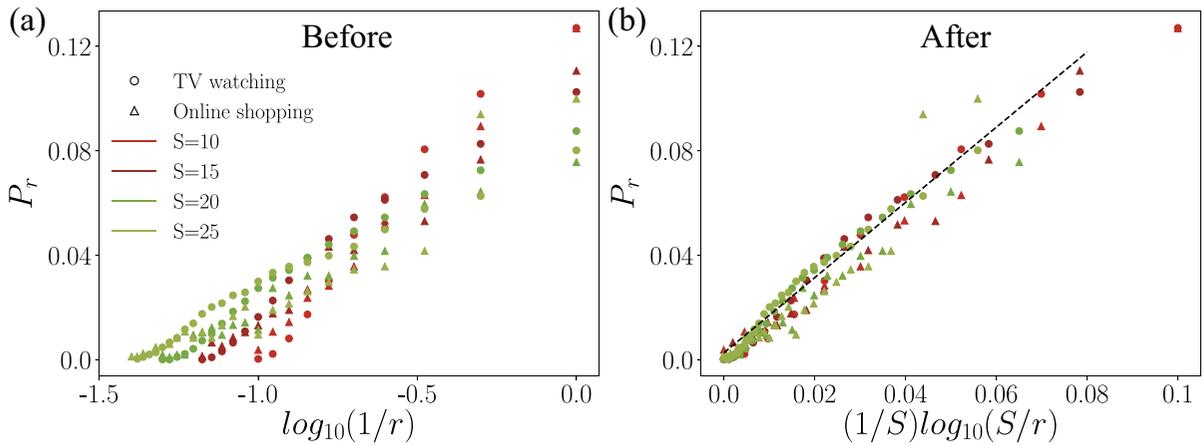


Fig. 5 Comparison before and after rescaling. **a** Before rescaling, the slope and intercept vary with S . **b** After rescaling, all P_r and $\frac{1}{S} \log_{10} \frac{S}{r}$ fall near the straight line with a slope of 1.44

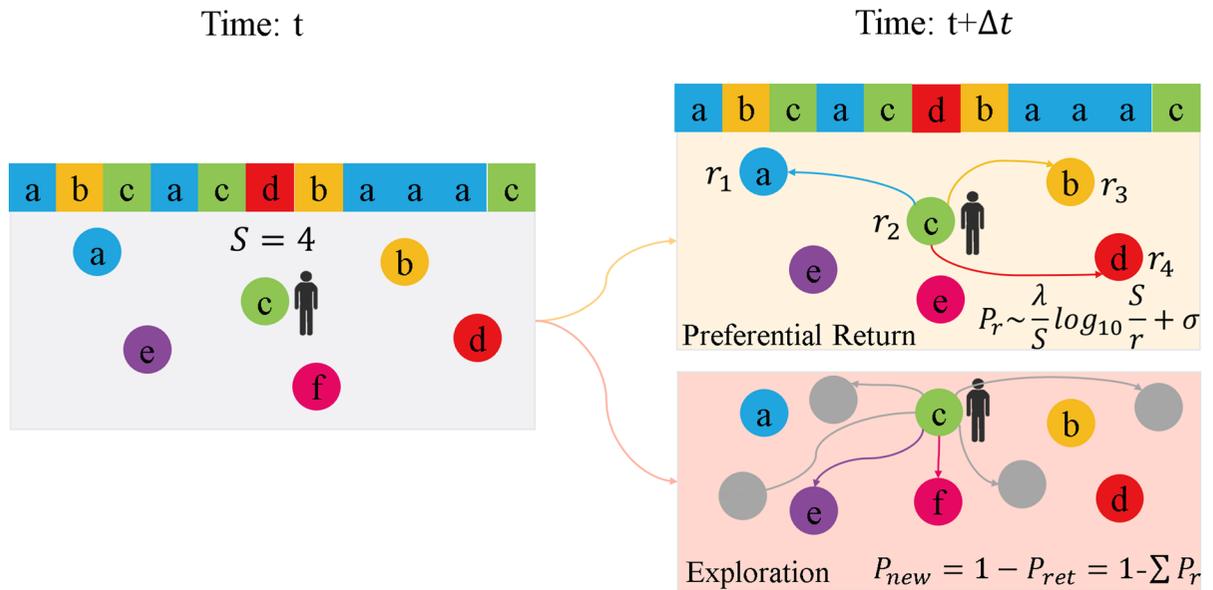


Fig. 6 Schematic model illustration. A typical trajectory of an individual visiting the 4 locations denoted by letters $a - e$ with different colors is indicated: $a \rightarrow b \rightarrow c \rightarrow a \rightarrow c \rightarrow d \rightarrow b \rightarrow a \rightarrow a \rightarrow a \rightarrow c$. At time t , the user has visited $S = 4$ locations. At time $t + \Delta t$ (with Δt drawn

from $P(\Delta t) \propto \Delta t^{-1-\beta}$, the user can either return to a previously visited location with probability $P_r \approx \frac{\lambda}{S} \log_{10} \frac{S}{r} + \sigma$ or visit a new location chosen from locations not visited with probability $P_{new} = 1 - P_{ret}$ ($P_{ret} = \sum_{r=1}^S P_r$)

The probability of exploring a new location P_{new} . Obviously, $P_{new} = 1 - P_{ret}$. P_{ret} is the sum of probabilities of returning to the previously visited locations. In general, we find (see Supplementary S4)

$$\begin{aligned}
 P_{new} &= 1 - \sum_{r=1}^S P_r = 1 - \sum_{r=1}^S \left(\frac{\lambda}{S} \log_{10} \frac{S}{r} + \sigma \right) \\
 &= 1 - \left(\frac{\lambda}{S} \log_{10} \frac{S^S}{S!} + \sigma S \right) \approx \frac{4}{3} S^{-\frac{1}{2}}. \tag{3}
 \end{aligned}$$

The number of distinct sites $S(t)$. We note that the probability that an individual moves to a new location is approximately $\frac{4}{3}S^{-\frac{1}{2}}$, that is, $\frac{dS}{dn} \approx \frac{4}{3}S^{-\frac{1}{2}} \propto S^{-\frac{1}{2}}$, predicting $S \propto n^{\frac{2}{3}}$, where n is the number of discrete moves the individual made up to time t . Given the heavy-tailed waiting time distribution $P(\Delta t) \propto \Delta t^{-1-\beta}$, time t scales with the number of moves n as $t \propto n^{\frac{1}{\beta}}$ [12], indicating that $S(t)$ follows $S(t) \propto t^\mu$ with

$$\mu = \frac{2}{3}\beta. \quad (4)$$

Since $\beta_1 = 0.78$, $\beta_2 = 0.88$, we have $\mu_1 = 0.52$, $\mu_2 = 0.57$, identical to the exponents of empirical data.

The visitation frequency f_r . We introduce an intermediate quantity m_r , the number of visits to the r_{th} location. Note that at each step, the probability that m_r increases is the probability that the r_{th} site is returned, that is, $\frac{dm_r}{dn} = P_r$, where P_r , as in Eq. (2). Because P_r discussed here is for one site and σ is a small quantity close to 0, we further approximate $\frac{dm_r}{dn} = P_r \approx \frac{\lambda}{S} \log_{10} \frac{S}{r}$. In addition, $\frac{dS}{dn} \approx \frac{4}{3}S^{-\frac{1}{2}}$, and we obtain

$$\frac{dm_r(S)}{dS} \approx \frac{3\lambda}{4\ln 10} S^{-\frac{1}{2}} \ln \frac{S}{r}. \quad (5)$$

Its solution is $m_r(S) \approx \frac{3\lambda}{2\ln 10} (S^{\frac{1}{2}} \ln \frac{S}{r} - 2S^{\frac{1}{2}})$. As the visitation frequency $f_r = \frac{m_r}{n} \approx \frac{m_r(S)}{\frac{4}{3}S^{-\frac{1}{2}}}$, after simplification and approximation, we have (see Supplementary S4)

$$f_r \propto r^{-\frac{3\lambda}{2\ln 10}}. \quad (6)$$

Eq. (6) predicts $\zeta = \frac{3\lambda}{2\ln 10}$, so $\zeta_1 = \zeta_2 = \frac{3 \times 1.44}{2\ln 10} = 0.94$, which is in excellent agreement with the empirical value.

Note that the exponents β of the waiting time $P(\Delta t)$ are different in the two datasets. The exponents μ of $S(t)$, related to t , are also different. However, $\mu = \frac{2}{3}\beta$ holds for both datasets. Moreover, the scaling properties of P_{new} and f_r are identical. It reveals that for different virtual movements, the selection mechanism behind each individual is the same, but due to different waiting time in viewing, browsing, etc., β and μ vary with scenes. However, the underpinning is unified.

3.4 Numerical simulation

We numerically simulate our VEPR model to verify the scaling laws of three characterizing quantities. Note that the waiting time distribution $P(\Delta t)$ is determined

by the specific system, and we take it as a premise and set $\beta = 0.78$ according to the TV watching data. We use the setting of 20000 trajectories in a space of 300 sites, that is, $N = 20000$ and $M = 300$. λ is the most critical parameter in our model, and we set $\lambda = 1.44$ for all S according to the empirical results.

The results, which are shown in Fig. 7, indicate that the three characterizing quantities are consistent with the empirical results. For further quantitative analysis, we conduct the independent sample T test for each pair of empirical values and simulated values of $S(t)$, P_{new} , and f_r respectively based on the 30,000 viewing records and 20,000 simulation trajectories. An empirical value is colored green if the P value is less than 0.05 and is red otherwise. More than 80% of point pairs pass the T test, proving our simulation's effectiveness (see S6 for detailed T test results). Moreover, we note that there is a slight deviation between the simulation and the empirical results when S/r is small or large. When S/r is small, only a few points can be used for regression. When S/r is large, few people watch/browse many sites in one trajectory, so the sample size is small. In both cases, the parameter λ is biased from $\bar{\lambda}$ (also reflected in Fig. 4d). Therefore, for small or large S/r , the simulation results obtained by substituting $\bar{\lambda}$ for the true λ are biased. In addition, the number of distinct sites S visited by individuals is much smaller than $M = 300$, showing that each step of each individual is based on the principles of our model, without overflow, which also conforms to the virtual space that has almost no space size restriction. Thus far, our VEPR model has been validated in simulation and theory.

4 Discussion

Movements in virtual space have attracted increasing academic attention. We explore virtual mobility dynamics based on two online datasets of different fields. We find striking consistency in the two virtual movements by measuring the scaling properties and summarize a universal choice mechanism behind virtual travel from empirical data. We also develop an individual mobility model called the Virtual Exploration and Preferential Return (VEPR) model. Compared with previous studies, we comprehensively discuss the scaling characteristics across datasets and find a more con-

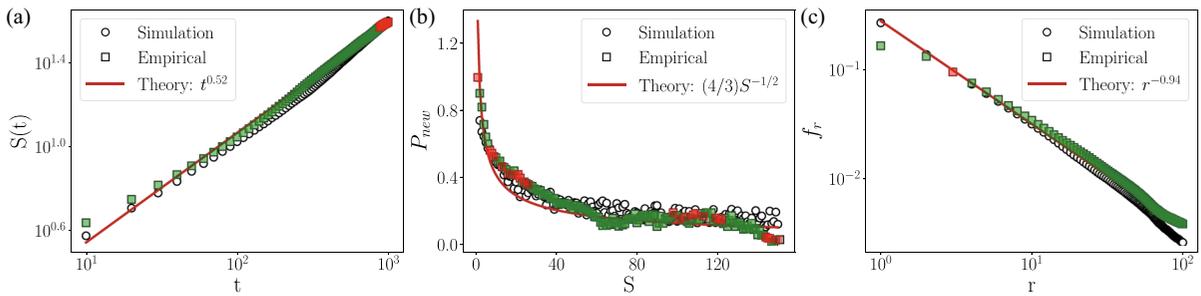


Fig. 7 Comparison between simulation and theoretical predictions of the VEPR model. The number of trajectories $N = 20000$, the number of locations $M = 300$, the exponent of waiting time $\beta = 0.78$, and the parameter $\lambda = 1.44$. The circles represent the simulation results, the squares represent the empirical results, and the red line represents theoretical value. The independent sample t tests are conducted for each pair of simulated and empirical val-

ues. A square is colored green if the P value is less than 0.05 and is red otherwise. Most squares are green, proving that the empirical results are consistent with the simulation results at a 95% confidence level. **a** The number of distinct locations $S(t)$. **b** The probability of exploring a new location P_{new} . **c** The visitation frequency f_r .

cise behavior rule, contributing to revealing the unified underpinning of human mobility.

Our results can be summarized as follows. (i) We find that the scaling properties of virtual mobility are jointly determined by the unified mechanism and system features. The characteristics of the probability of exploring a new location, P_{new} , are the same for TV watching and online shopping. The exponents of visitation frequency f_r are also the same. The equation $\mu = \frac{2}{3}\beta$ holds for both datasets, where β is the waiting time distribution's exponent and μ is the exponent of the number of distinct locations $S(t)$. (ii) We summarize a specific rule that governs individuals' behavior from the empirical data. In virtual space, the memory effect is $P_r \propto \log_{10} \frac{1}{r}$. Compared with $P_r \propto \frac{1}{r}$ in the real world, the visitation probability is more uniform, which is in agreement with the smaller travel restriction in virtual space. (iii) We obtain a more general choice mechanism by rescaling the coefficients to be independent of S . The equation is $P_r \approx \frac{\lambda}{3} \log_{10} \frac{S}{r} + \sigma$, which can reflect the return rule and facilitate theoretical derivation. (iv) We incorporate the rule under the EPR framework to establish our VEPR model. The three scaling properties can be reproduced in theory and simulation, which proves the model's validity. In conclusion, we extract the rule of virtual travel from empirical data, and our individual mobility model based on the rule reproduces the scaling properties in theory and simulation. At the same time, the virtual mobility's features can also be well reflected by the mechanism of our model. The same macro scaling properties and micro behavior mechanism behind watching TV and online shopping

inspire further exploration of whether there is a unified underpinning behind other virtual movements such as web browsing, online communication, and social interactions. Additionally, our work also contributes to solving practical problems. It is feasible to establish the mobility model for each individual according to his/her λ to predict his/her digital trajectory, which can be applied to make personalized recommendation in practice.

While we treat all users' trajectories equally for regression to obtain the rule governing return, users are heterogeneous in behavior. The model incorporating memory heterogeneity may reproduce much richer statistical properties. In addition, it is worthwhile to obtain each user's λ by regression to understand his mobility style. For example, a large λ indicates that the individual prefers to return to previous locations, while a small λ represents the fact that he is an explorer. This key parameter has potential applications for current important issues such as traffic optimization and online recommendation. Generally, it is of profound significance to extend the model considering heterogeneity at both individual and group levels.

Author contributions MW: Software, Validation, Visualization, Writing—original draft. AZ: Conceptualization, Methodology, Writing—review&editing. XC: Conceptualization, Methodology, Validation, Writing—review&editing.

Funding This work was supported by the National Natural Science Foundation of China (Grant Nos. 11775020 and 11675001).

Data availability 1. Dataset D1: This anonymized dataset represents 3 months of viewing records from 30 thousand users in 2015. The dataset cannot be disclosed for confidentiality reasons. 2. Dataset D2: This dataset contains behavior data for 60 thousand users from a large multicategory online store in October 2019. The dataset is free and available at <https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store>.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

- Brockmann, D., Hufnagel, L., Geisel, T.: The scaling laws of human travel. *Nature* **439**, 462–465 (2006)
- Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L.: Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008)
- Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., Mascolo, C.: A tale of many cities: universal patterns in human urban mobility. *PLOS ONE* **7**, e37027 (2012)
- Yan, X.-Y., Han, X.-P., Wang, B.-H., Zhou, T.: Diversity of individual mobility patterns and emergence of aggregated scaling laws. *Sci. Rep.* **3**, 2678 (2013)
- Eubank, S., Guclu, H., Anil Kumar, V., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N.: Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184 (2004)
- Toroczkai, Z., Guclu, H.: Proximity networks and epidemics. *Phys A: Stat. Mech. Appl.* **378**, 68–75 (2007)
- Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D., Tatem, A.J.: Dynamic population mapping using mobile phone data. *Proc. Nat. Acad. Sci.* **111**, 15888–15893 (2014)
- Berlingerio, M., Calabrese, F., Lorenzo, G. D., Nair, R., Pinelli, F., Sbodio, M. L.: Allaboard: a system for exploring urban mobility and optimizing public transport using cellphone data. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, pp. 663–666 (2013)
- Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabasi, A.-L.: Human mobility, social ties, and link prediction. In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1100–1108 (2011)
- Miritello, G., Moro, E., Lara, R.: Dynamical strength of social ties in information spreading. *Phys. Rev. E* **83**, 045102 (2011)
- Blondel, V.D., Decuyper, A., Krings, G.: A survey of results on mobile phone datasets analysis. *EPJ Data Sci.* **4**, 10 (2015)
- Song, C., Koren, T., Wang, P., Barabási, A.-L.: Modelling the scaling properties of human mobility. *Nat. Phys.* **6**, 818–823 (2010)
- Yan, X.-Y., Wang, W.-X., Gao, Z.-Y., Lai, Y.-C.: Universal model of individual and population mobility on diverse spatial scales. *Nat. Commun.* **8**, 1639 (2017)
- Kang, C., Liu, Y., Guo, D., Qin, K.: A generalized radiation model for human mobility: spatial scale, searching direction and trip constraint. *PLOS ONE* **10**, e0143500 (2015)
- Szell, M., Sinatra, R., Petri, G., Thurner, S., Latora, V.: Understanding mobility in a social petri dish. *Sci. Rep.* **2**, 457 (2012)
- Lin, M., Hsu, W.-J., Lee, Z. Q.: Modeling high predictability and scaling laws of human mobility. In: *2013 IEEE 14th International Conference on Mobile Data Management*, Volume 02, volume 2, IEEE, pp. 125–130 (2013)
- Lee, K., Hong, S., Kim, S. J., Rhee, I., Chong, S.: Slaw: A new mobility model for human walks. In: *IEEE INFOCOM 2009*, IEEE, pp. 855–863 (2009)
- Han, X.-P., Wang, B.-H.: Impacts of distance and memory in the emergence of scaling mobility pattern of human. *Phys. Proc.* **3**, 1907–1911 (2010)
- Hu, Y., Zhang, J., Huan, D., Di, Z.: Toward a general understanding of the scaling laws in human and animal mobility. *Europhys. Lett.* **96**, 38006 (2011)
- Moro, E., Calacci, D., Dong, X., Pentland, A.: Mobility patterns are associated with experienced income segregation in large us cities. *Nat. Commun.* **12**, 4633 (2021)
- Zhao, Z.-D., Huang, Z.-G., Huang, L., Liu, H., Lai, Y.-C.: Scaling and correlation of human movements in cyberspace and physical space. *Phys. Rev. E* **90**, 050802 (2014)
- Zhao, Y.-M., Zeng, A., Yan, X.-Y., Wang, W.-X., Lai, Y.-C.: Unified underpinning of human mobility in the real world and cyberspace. *New J. Phys.* **18**, 053025 (2016)
- Deville, P., Song, C., Eagle, N., Blondel, V.D., Barabási, A.-L., Wang, D.: Scaling identity connects human mobility and social interactions. *Proc. Nat. Acad. Sci.* **113**, 7047–7052 (2016)
- Hu, T., Xia, Y., Luo, J.: To return or to explore: Modelling human mobility and dynamics in cyberspace. In: *The World Wide Web Conference*, pp. 705–716 (2019)
- De Nadai, M., Cardoso, A., Lima, A., Lepri, B., Oliver, N.: Strategies and limitations in app usage and human mobility. *Sci. Rep.* **9**, 10935 (2019)
- Arthur, R., Williams, H.T.: Scaling laws in geo-located twitter data. *PLOS ONE* **14**, e0218454 (2019)
- Jiang, S., Fiore, G. A., Yang, Y., Ferreira Jr, J., Frazzoli, E., González, M. C.: A review of urban computing for mobile phone traces: current methods, challenges and opportunities. In: *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, pp. 1–9 (2013)
- Zhao, C., Zeng, A., Yeung, C.H.: Characteristics of human mobility patterns revealed by high-frequency cell-phone position data. *EPJ Data Sci.* **10**, 5 (2021)
- Cai, H., Zhan, X., Zhu, J., Jia, X., Chiu, A.S., Xu, M.: Understanding taxi travel patterns. *Physica A: Stat. Mech. Appl.* **457**, 590–597 (2016)
- Metzler, R., Klafter, J.: The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* **339**, 1–77 (2000)
- Yuste, S.B., Klafter, J., Lindenberg, K.: Number of distinct sites visited by a subdiffusive random walker. *Phys. Rev. E* **77**, 032101 (2008)

32. Viswanathan, G.M., Buldyrev, S.V., Havlin, S., Da Luz, M., Raposo, E., Stanley, H.E.: Optimizing the success of random searches. *Nature* **401**, 911–914 (1999)
33. Zhao, Z.-D., Cai, S.-M., Lu, Y.: Non-markovian character in human mobility: online and offline, chaos: an interdisciplinary. *J. Nonlin. Sci.* **25**, 063106 (2015)
34. Crang, M.: *Cyberspace as the new public domain*. Johns Hopkins University Press, Woodrow Wilson International Center for Scholars (2010)
35. Chen, Y., Li, X., Zhang, R., Huang, Z.-G., Lai, Y.-C.: Instantaneous success and influence promotion in cyberspace-how do they occur? *Physica A: Stat. Mech. Appl.* **556**, 124725 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.