Cascading Walks Model for Human Mobility Patterns

Xiao-Pu Han[1,2]*, Xiang-Wen Wang[3,4], Xiao-Yong Yan[2,5,6,7], Bing-Hong Wang[4]

1 Institute of Information Economy, Hangzhou Normal University, Hangzhou 311121, China, 2 Alibaba Research Center for Complexity Sciences, Hangzhou Normal University, Hangzhou 311121, China, 3 Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0435, USA, 4 Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China, 5 School of Systems Science, Beijing Normal University, Beijing 100875, China, 6 Department of Transportation Engineering, Shijiazhuang Tiedao University, Shijiazhuang 050043, P. R. China, 7 Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 610051, China

* xp@hznu.edu.cn

Abstract

Background

Uncovering the mechanism behind the scaling laws and series of anomalies in human trajectories is of fundamental significance in understanding many spatio-temporal phenomena. Recently, several models, e.g. the explorations-returns model (Song et al., 2010) and the radiation model for intercity travels (Simini et al., 2012), have been proposed to study the origin of these anomalies and the prediction of human movements. However, an agent-based model that could reproduce most of empirical observations without priori is still lacking.

Methodology/Principal Findings

In this paper, considering the empirical findings on the correlations of move-lengths and staying time in human trips, we propose a simple model which is mainly based on the cascading processes to capture the human mobility patterns. In this model, each long-range movement activates series of shorter movements that are organized by the law of localized explorations and preferential returns in prescribed region.

Conclusions/Significance

Based on the numerical simulations and analytical studies, we show more than five statistical characters that are well consistent with the empirical observations, including several types of scaling anomalies and the ultraslow diffusion properties, implying the cascading processes associated with the localized exploration and preferential returns are indeed a key in the understanding of human mobility activities. Moreover, the model shows both of the diverse individual mobility and aggregated scaling displacements, bridging the micro and macro patterns in human mobility. In summary, our model successfully explains most of empirical findings and provides deeper understandings on the emergence of human mobility patterns.

Introduction

In recent years, benefited from the development of location tracking technologies, the studies on the statistics of human daily mobility rise to be a new research hot point. Researchers have analyzed varieties of datasets which could reflect human mobility (e.g. dollar-bill tracking, mobile phone roaming, global positioning system, recording of taxi and bus passengers, et al). These empirical studies show two major quantitative features of human mobility patterns: i). the scaling anomalies in several statistics, which mainly include the aggregated jump-size (or say move-length or displacement) of each travel [1–8], waiting time [2, 6], visitation frequency of each distinct location [2, 6], and the growth of the total number of visited locations [2, 6]. ii). Ultraslow diffusion property, which is observed in the slow growth of the mean square displacement (MSD) and radius of gyration [1–3, 6]. Beyond that, several other anomalies, such as high regularity and predictability [9, 10] and the limitation of traffic system [5], have also been observed in human mobility. These features are not only against the traditional view of human mobility based on random walks but also sharply different from the pure Lévy-flight nature. Consequently, researchers have to review the discussions on many social and economic dynamics that are deeply affected by human mobility, such as the spreading of pathogens and information [11–16], the emergence of collective behaviors [17], the planning of urban sub-domains and traffic systems [18], and so on.

Moreover, with the progress on this issue, even some recent formed consensus are also challenged by the new findings. For instance, the aggregated distribution of long-range move-length shows widely existed scaling properties, however, this conclusion can not be extend to be the universal law of human mobility [19], since human individuals' movements show high diversity [20] and the pattern of urban trips usually appear to be exponential-like [21–24].

A key issue is how to uncover the origin of these properties. Recent studies have proposed several approaches and dynamical models. The simplest type is the descriptive model, including the continuous-time random-walk (CTRW), which can generate power-law-like displacement distribution and show slow diffusion [1], and the self-similar least action walk (SLAW) [4]. Many possible origins have been proposed in explaining the scaling anomalies. The first type is the aggregated effect on population to mimic the macroscopic features of human mobility. Such as the radiation model that can reproduce much realistic inter-city mobility patterns [25], and the model based on maximum entropy theory under Maxwell-Boltzmann statistics which aggregates the scaling law from diverse individual movements [20]. In the second type, the spatial heterogeneity of population density and urban environment was discussed as a factor that contributes to the emergence of human mobility patterns [23, 26, 27]. Beyond these qualities, recently, the hierarchical nature of human mobility is empirically observed [7] and is considered as one of the origin of the scaling move-length [7, 28]. For the ultraslow diffusion, the model reported in Ref. [29] indicates that it seems to be the result of the high-frequency trips between homes and working locations. And some recent studies focus on the effect of social interactions and propose models based on the co-evolution between agents' mobility and social networks, creating scaling properties on both social networks and mobility patterns [30]. Recently, the model proposed by Song et al [6] attracts much attention. It considers two generic mechanisms: the explorations of new locations and the preferential returns of former visited locations. In combination of some previous results obtained from empirical analysis, this model provides several statistical patterns consistent with the empirical data, indicating that the two mechanisms play fateful role in human mobility. Unfortunately, due to the dependence on the previous empirical results, this model does not give a sufficient explanation on the characteristic key of human mobility patterns: the scaling law on jump-size. Overall, a model that catches

the key in human mobility and reproduces most of empirical observations (scaling laws and ultraslow diffusion) from few basic assumptions is still lacking.

In the present paper, considering the empirical findings of the correlations of displacements [31], based on the cascading-like process and combined with the localized exploration and preferential returns, we propose an agent-based model to explain the underlying mechanisms behind human daily travels. We find that, this minimum model can quantitatively well rebuild almost all the empirical findings and cover most of the results of previous modeling studies, implying the cascading-like process is indeed important in understanding human mobility patterns.

Results

The model

One of the noticeable features of our daily long-range travels is that, we usually prefer to take some short-range movements around the aiming locations before returning to the starting point. For example, when we travel from our hometown to another city, generally speaking, we need to find a hotel, then we will take some activities for the purpose of the travel (e.g., working, meeting, or tours) at certain locations near the city, and at last we return to hometown after all objectives have been completed. Each activity we take during the travel will create one or several additional short-range movements in the region of the city, that is to say, the long-range travel from our hometown to the aiming city activates a series of shorter movements in the aiming city. Similarly, such kind of activations can also be observed in many middle-range or short-range movements, for instance, the commuter's trips vs. the movements in workplaces, and the trips from home to marketplace vs. the strolls during shopping. Associating with the activation by long-range movements, an explicit feature of human mobility emerges: a series of activating processes perform most transitions from the relatively longer travels to the shorter trips, and each of them can initiate a series of shorter trips near the aiming location of the longer travel. These serial initiations of movements are named as "Cascading Walks Process".

This cascading walks process in human mobility have been observed by recent empirical studies based on the database of GPS trajectories [31], in which the evidences of cascading walks process that the positive correlation between the displacements of two consecutive movements, positively relates to the scaling property of human mobility patterns, implying the cascading walks process might play an important role in the emergence of human scaling mobility patterns. Also, previous studies on the temporal patterns of human activities indicated a similar cascading-like process would be relevant to the emergence of burst [32], as well as the long-term correlations (can be created by the cascading processes) [33, 34]. Overall, combining both the daily experiences and empirical studies above, the cascading walks process is regarded as the basic mechanism of our model which will be elucidated later.

The initiation of a series of shorter movements after a longer travel is the basic dynamics of cascading walks process, and thereby in the present model, the longer travel and initiated shorter movements can be distinguished into two layers: the longer travel stays in the higher layer, the shorter movements are in the lower one, and the corresponding end locations also can be set into different layers. An illustration for the algorithm of the cascading walks in the present model is shown in Fig 1. Associating with Fig 1, detailed rules will be shown below.

i). The localized cascading process. It's the basic rule of the present model. We assume that the walker moves to an i-th layer location A (the movement I in Fig 1). The long-range movement I is in the i-th layer and initiates several $(i + 1)$-th layer shorter movements in the region of A (the red dashed circle), which are marked as II (A $\rightarrow$ B) and VI (B $\rightarrow$ E), as shown in Fig 1. The two $(i + 1)$-th layer locations B and E are randomly chosen in the region of A. The
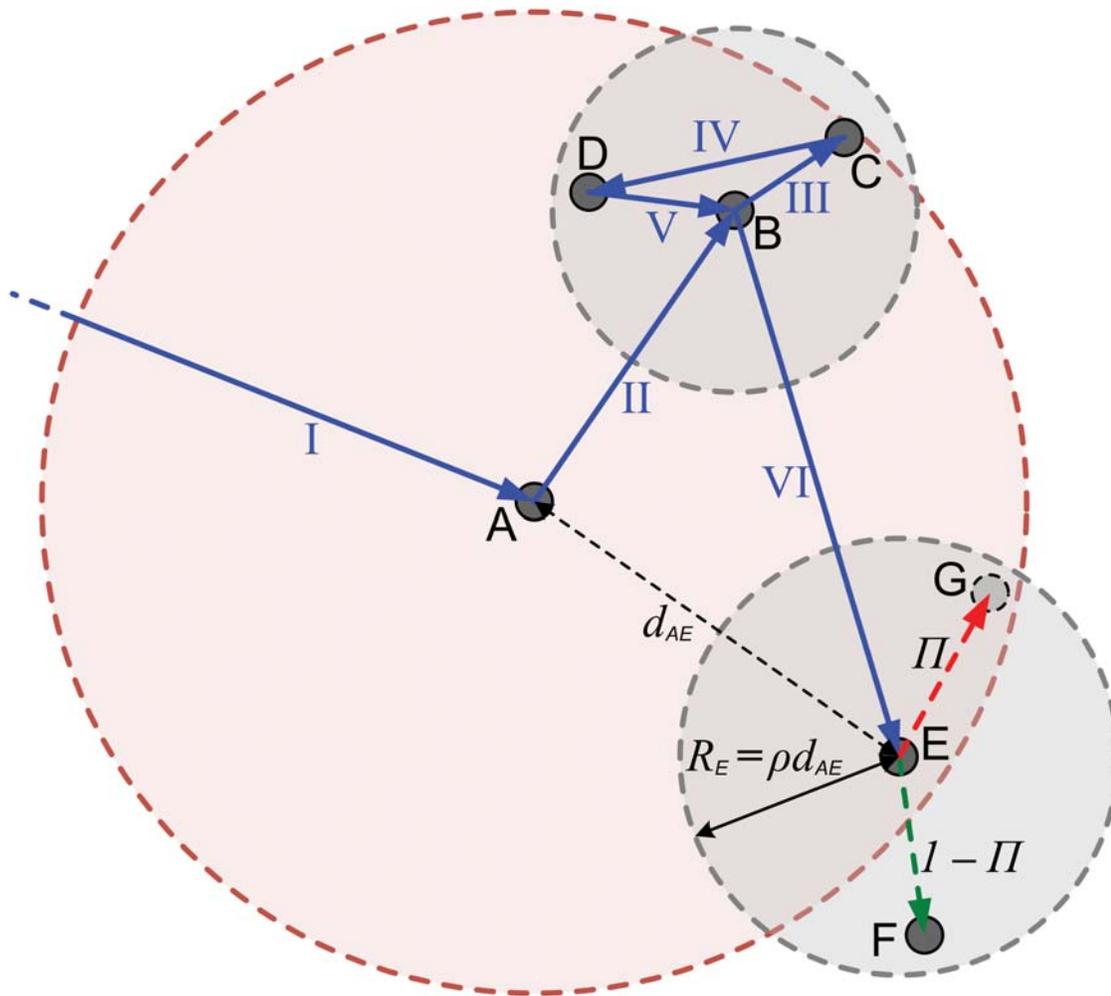
Fig 1. Illustration of the cascading walks process in the present model. It shows several consecutive movements of the walker, and the dashed circles represent the regions in different layers.

movements II and VI respectively initiate several $(i + 2)$-th layer short trips in the region of B and E (the two gray dashed circles, and they are called the sub-region of A), for instance, the movements III, IV and V in the region of B. These $(i + 2)$-th layer trips could again initiate $(i + 3)$-th layer movements (it's not shown in Fig 1). The radius $R's$ of the region B/E, are respectively proportional to the distances from A to B/E: $R_B = \rho d_{AB}$ for location B, as well as $R_E = \rho d_{AE}$ for E, where ρ is a parameter satisfying $0 < \rho < 1$. Here it is allowed that parts of the region of B/E are outside the region of A. And during the term that the walker moves in the region of location A, say, every visited locations besides location A are set into the layer lower than location A, until the exhaustion of the staying time of A (it will be defined in the next rule), as well as other locations (B, C, · · ·).

ii). The staying time is correlated with the traveling distance. For the location B, say, the total time of all these lower-layer short movements which are directly or indirectly initiated by the movement II is τ_B, which means τ_B is the time interval from the beginning of movement III to the end of movement V, or say, the staying time of B. We assume that the staying time τ_B is

positively correlated with the distance from the up-layer location A to the location B,

$$\tau_B = f_P(d_{AB}^\gamma). \tag{1}$$

It means that τ_B is a random number which follows a Poisson distribution f_P with an average value of d_{AB}^γ, as well as $\tau_E = f_t(d_{AE}^\gamma)$ for the location E, and similarly for other locations, for instance, $\tau_C = f_P(d_{BC}^\gamma)$, where γ is another main parameter of the model, denoting how the staying time is correlated with the distance between up-layer locations. According to the empirical findings, γ is set to be the positive and not greater than 1, namely, the relationship of Eq (1) is mainly sub-linear. In the simulations, we set a rule that when the staying time runs out, the walker must return to the center location of the current region to prepare the departure to another region. For instance, when $t = t_B + \tau_B - 1$, the walker moving in the region of B must return to B, no matter whether the staying time in the current region runs out or not. This assumption is realistic and supported by our empirical analysis results of the volunteers' GPS records (see Section 0.1 of *Materials and Methods*), in which indeed show the above power functional correlation on the staying time in the relevant region and the moving displacement in the region, and the fitting value of γ is less than 1.

iii). Localized exploration and preferential return in prescribed region. During the period that the walker moves inside the region of a location (the $(i + 1)$-th layer location E, say), as shown in Fig 1, each initiation of a new $(i + 2)$-th layer movement will be of the probability

$$\Pi = \frac{\lambda}{\lambda + \sum_j k_j}. \tag{2}$$

The walker can visit a new $(i + 2)$-th layer location (G, say) from its present location, where G is randomly chosen in the region of E, the parameter λ controls the preference of explorations, and $\Sigma_j k_j$ denotes the total number of visitations of all the $(i + 2)$-th layer locations in current $(i + 1)$-th layer region except the current location. And with probability $1 - \Pi$, the walker can return to a former visited $(i + 2)$-th layer location (F, say), chosen with a probability that is proportional to its former visitation number k. In other words, in this region and this layer, the location that are visited more frequently are more possible to be visited again. Note that, the $(i + 1)$-th layer location E is also treated as one of these $(i + 2)$-th layer locations in the region, if E is not the current location, as well as A for the the $(i + 1)$-th layer locations in its region. The walks in other layers, e.g. A $\rightarrow$ B and B $\rightarrow$ E, also obey the above algorithms. Compared with the algorithm of Song's model [6], the explorations and preferential returns are localized in the current region and current layer.

Modeling Results

At the initial time $t = 0$ in simulations, the walker starts traveling from the initial location with coordinate (0, 0) at the top layer. The radius of the region of the initial location, which also is the maximum radius for all the regions, is fixed to be $R_0 = 10^4$. In the following discussions, if not specially mentioned, the value of λ is set to be the default value 0.5. All the statistical results that will be discussed below are obtained from the simulations with 5×10^4 time steps after a relaxation with the same time steps. What's more, instead of using the absolute value of displacement, we use the relative displacement $\Delta x_t = \frac{|\vec{X}_{t+1} - \vec{X}_t|}{R_0}$, where $\vec{X}_t$ and $\vec{X}_{t+1}$ are respectively the coordinates of the agent at time step t and $t + 1$.

The patterns of trajectories generated by the present model are very similar to our intuitive experience in daily life. Fig 2 shows the typical trajectory patterns under different parameter settings. The trajectories in different scales have significant self-similar property, and large ρ,
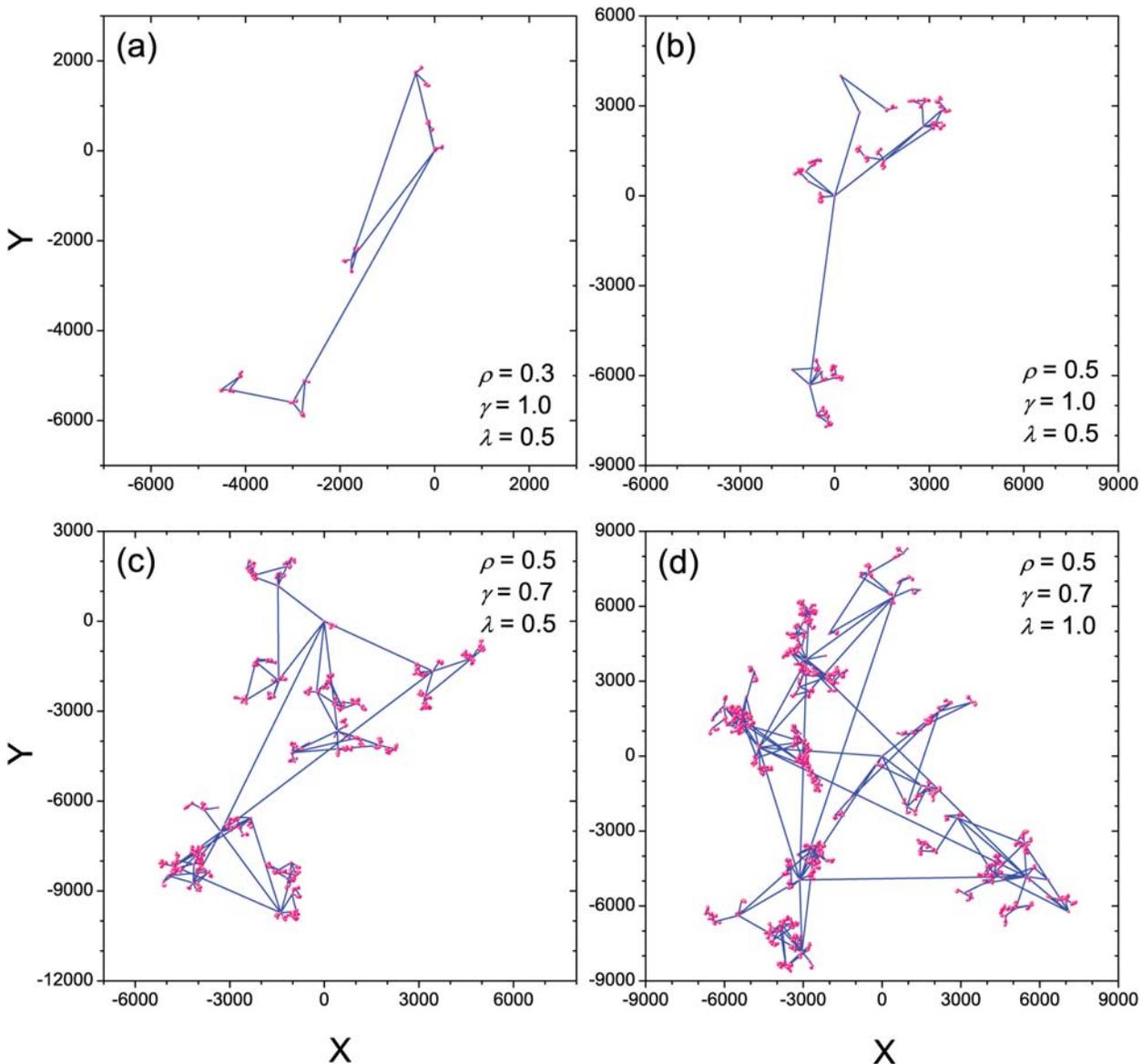
Fig 2. Typical trajectories of the model. Illustrations of the trajectories of 5×10^4 consecutive travels generated by the model with four typical parameter settings of ρ, γ and λ.

small γ along with large λ can generate more complicated trajectory patterns. Most importantly, we find that the present model can create most of the properties observed in empirical studies, including both the scaling anomalies and ultraslow diffusion. In the following subsections, we will discuss them one by one.

i) Displacement distributions.

The scaling statistics of move-lengths is one of the most noticeable differences between real-world human mobility patterns and the random walk nature, and has been observed in many animals movements [35–37]. Empirical studies have provided lots of evidences to support a widely accepted viewpoint that the displacement distributions of human travels at the population level are power-law-like [1, 2], no matter the arguments at the individual level. Now,
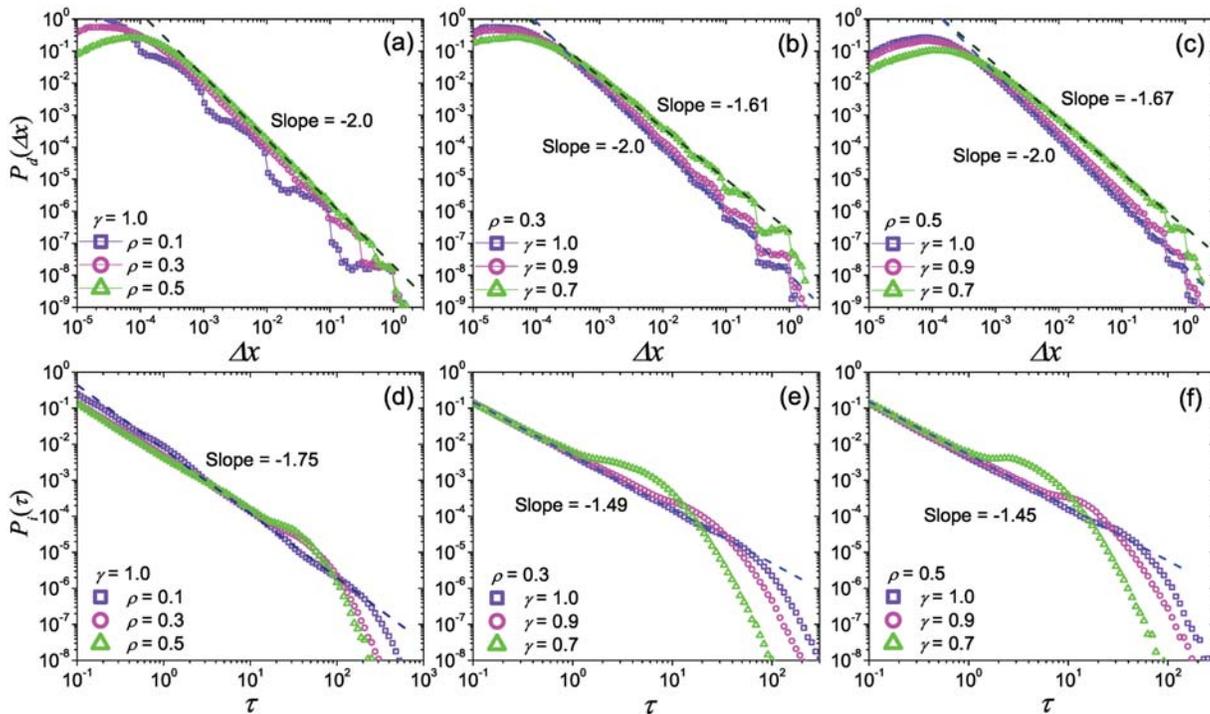
Fig 3. The displacement distributions and the inter-visit time distributions. Panels (a), (b) and (c) are respectively the displacement distributions generated by the present model with different parameters ρ and γ, as well as panels (d), (e) and (f) for the inter-visit time distributions. The result is averaged by 10^3 independent runs. The dark blue dashed lines and dark green dashed lines respectively corresponds to the fitting lines of the data signed by light blue squares and light green rectangles.

doi:10.1371/journal.pone.0124800.g003

understanding the origin of the scaling displacement distribution becomes an important task for modeling researches. In present model, the numerical simulations indicate that the displacement distribution with a power-law tail can emerge in all parameter space (Fig 3(a), 3(b) and 3(c)). The exponent α of the power-law tail is mainly affected by γ, and is insensitive to other parameters. The value of α is 2.0 when $\gamma = 1.0$, and will reduce along the reduction of γ (Fig 3(b) and 3(c)), covering the range of many previous empirical studies and our findings on GPS records (see Section 0.1 of *Materials and Methods*). When the value of ρ is small, even though the fluctuations on the curves of displacement distribution are obvious, it does not damage the scaling property in the tail (Fig 3(a), 3(b) and 3(c)). Moreover, we estimated the average value of ρ and γ from empirical analysis on Geo-life GPS dataset, and then run the model on this empirical-based parameter settings ($\rho = 0.66$, $\gamma = 0.54$). The curve of simulation results is almost parallel to the empirical one (see Section 0.1 of *Materials and Methods*). More detailed discussion can be found in Section 0.1 of *Materials and Methods*.

Simulations also show that the parameter λ doesn't affect the displacement distribution at the aggregated level at all. Moreover, due to the locations is randomly chosen in region of upper layer location, the displacement distribution for the movements in single layer is homogeneous. In the real-world human mobility, the use of vehicle usually follow the travel length and thus it shows obvious hierarchy, for instance, taxi or bus for urban trips, and airplane for long-range travels. Recent studies indicated that human travels using single type of vehicle (e.g. taxi, et al) and short-distance commuting usually do not show scaling property [22, 38]. Our modeling results for single layer mobility successfully match the empirical findings. The
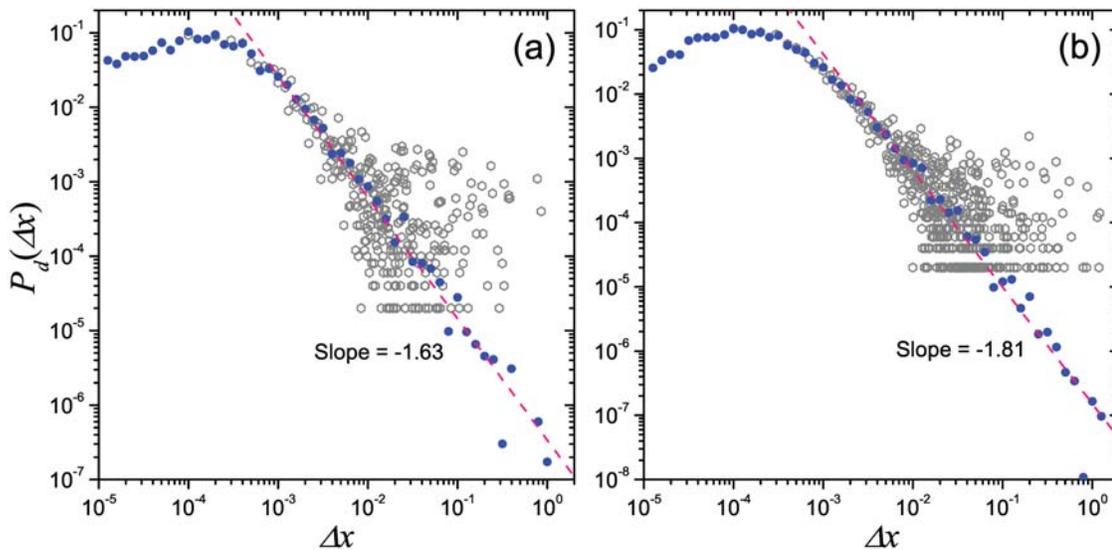
Fig 4. Two examples of the displacement distributions of single walker in single run. Panels (a) and (b) respectively correspond to the simulation shown in Fig 2(c) and 2(d), and blue dots and grey hexagons respectively show the log-binning distribution and its original form, and the pink dashed lines shows the fitting power law curves.

doi:10.1371/journal.pone.0124800.g004

analytical result (see Section 0.2 of *Materials and Methods*) indicates that, the exponent α only depends on γ with the expression of $\alpha = 1 + \gamma$, and the scaling displacement distribution essentially emerges from the localized cascading process, which determines the hierarchical organization of locations. It plays an important role in the sub-linear relationship between the staying time τ in a certain region and the corresponding traveled displacement Δx.

Note that, the results in Fig 3 are obtained by an average of a large number of simulations. In our simulations each run is completely independent with other runs, therefore we can consider each run as a simulation for an agent, accordingly $P_d(\Delta x)$ in Fig 3(a), 3(b) and 3(c) actually shows the aggregated patterns of displacement distributions for multiple agents, which corresponds to most of empirical studies. What does the real-world travel patterns at the individual level look like is still in controversy. Recent studies reported some evidences that, human mobility shows rich diversity at the individual level, and the displacement distribution of some users would show obvious deviation from the power law [20, 29]. Our empirical analysis on Geo-life GPS dataset also provide similar properties (see Section 0.1 of *Materials and Methods*). The diversity in individual-level mobility is also observed in the present model. We plot the displacement distributions generated by single runs. Fig 4 shows two typical examples of the displacement distribution of single run, in which the original data points in tail are scattered in a large region. The log-binning curves still looks close to a power law, however, due to many of the repeated trips, the probabilities of few long-range displacements usually are 10^3 or more times larger than the expected values on fitting power law. And also, the displacements of these repeated long-range movements for different runs (corresponds to different individuals) are usually very different. In this case, it is not surprising that a small sample deviating from power law can be abstracted from this system with intrinsical scaling property. Possibly, it is the origin of diversity on human mobility patterns at the individual level.

ii) Inter-visit time distributions.

The inter-visit time of a location is the time interval between two consecutive visits of the same location. Several earlier researches, for example, the power-law-like inter-event time distributions of library borrowing [39], implies that the inter-visiting time distribution of many locations could have non-Poisson properties. Recently, several direct observations based on cell phone and online society datasets pointed out that the inter-visiting time distribution indeed has obvious fat-tail property and usually is effected by users' life cycles [2, 8].

Power-law-like inter-visit time distributions surprisingly emerge in our model with slopes between -1 and -2 (Fig 3(d), 3(e) and 3(f)) under different parameter settings. The curves of the inter-visit time distribution $P_i(\tau)$ vary as γ decrease, from power-law functions to bimodal forms with power-law heads. The fitting exponent ζ of the head of $P_i(\tau)$ is close to 1.5 and somewhat negatively correlate to the value of ρ (Fig 3(d), 3(e) and 3(f)), but not sensitive to γ. For large ρ, a tiny peak appears in the tail of $P_i(\tau)$ and moves to the front when γ is small. Similar to the displacement distributions, λ almost doesn't affect the inter-visit time distribution.

iii) Visitation frequency and the number of visited locations

Empirical studies of human mobility patterns have indicated that, a few locations, which usually are working places and supermarkets, are visited much more frequently than other places. This heterogeneity is described by the power-law-like visitation frequency distributions in Zipf's ranking plot with a Zipf's exponent $\xi \approx 1.2$ [2, 6]. According to the relationship $\xi' = 1 + \xi^{-1}$ between the Zipf's exponent ξ and ξ', where ξ' is the exponent of the correspondingly power-law probability density function (PDF), we can get a exponent ξ' of 1.8 in the PDF of visitation frequency distribution.

This scaling visitation frequency is reproduced by the present model. The curves in Fig 5(a)–5(d) illustrate the power-law-like PDF $P_v(k)$ of visitation frequency for locations under different parameter settings. $P_v(k)$ is very sensitive to the value of ρ: The fitting value of the PDF exponent ξ' reduces from 1.8 to 1.1 along the growth of ρ from 0.1 to 0.5. The analysis in Section 0.3 of *Materials and Methods* indicates that in a extreme case without the cascading effect, ξ' is 1.0, implying that larger ξ' possibly originates from the cascading effect. Moreover, the preference on explorations λ also strongly affect $P_v(k)$: the more explorations, the larger value of ξ' (Fig 5(d)).

The growth of the number of visited locations $S(t)$ is generally expected to be in a scaling form $S(t) \sim t^\mu$. $\mu = 1$ stands for the pure Lévy flights, and its value depends on the exponent of the staying time distribution, where staying time is the time that a walker spend at one location in CRTW [9]. Empirical studies have indicated that, in real-world $S(t)$ obeys a scaling law with an exponent of $\mu \approx 0.6$, which is smaller than the prediction in CRTW [2, 9], indicating an ultraslow growth of the number of visited locations and a growing returning probability versus the visited locations. The growth of $S(t)$ in the present model also obeys a power-law function with an exponent $\mu < 1$ (Fig 5(e)–5(h)). Similar to the situation of visitation frequency distribution, the fitting exponent μ of $S(t)$ is mainly affected by the value of ρ, where smaller ρ corresponds to slower growth of $S(t)$ (small μ), indicating that the walkers are more likely to return to the former visited locations for smaller ρ. The reason is that, smaller ρ limits the movements and explorations of new locations in sub-layer regions, and the walker has to frequently return to former visited locations in upper layer. The effect of γ and λ is weak. An extreme situation without the cascading effect shows a logarithmic-growing $S(t)$ (see Section 0.3 of *Materials and Methods*), which results that $S(t)$ with large ρ can generate logarithmic sections (Fig 5(e)). Thus we infer that the cascading effect deeply changes $S(t)$ from the logarithmic form into the power-law form.

iv) Radius of gyration and MSD.

The properties of human spreading are mainly shown in the growth of radius of gyration r_g and MSD versus time. For the discrete time model, the radius of gyration of a walker in the
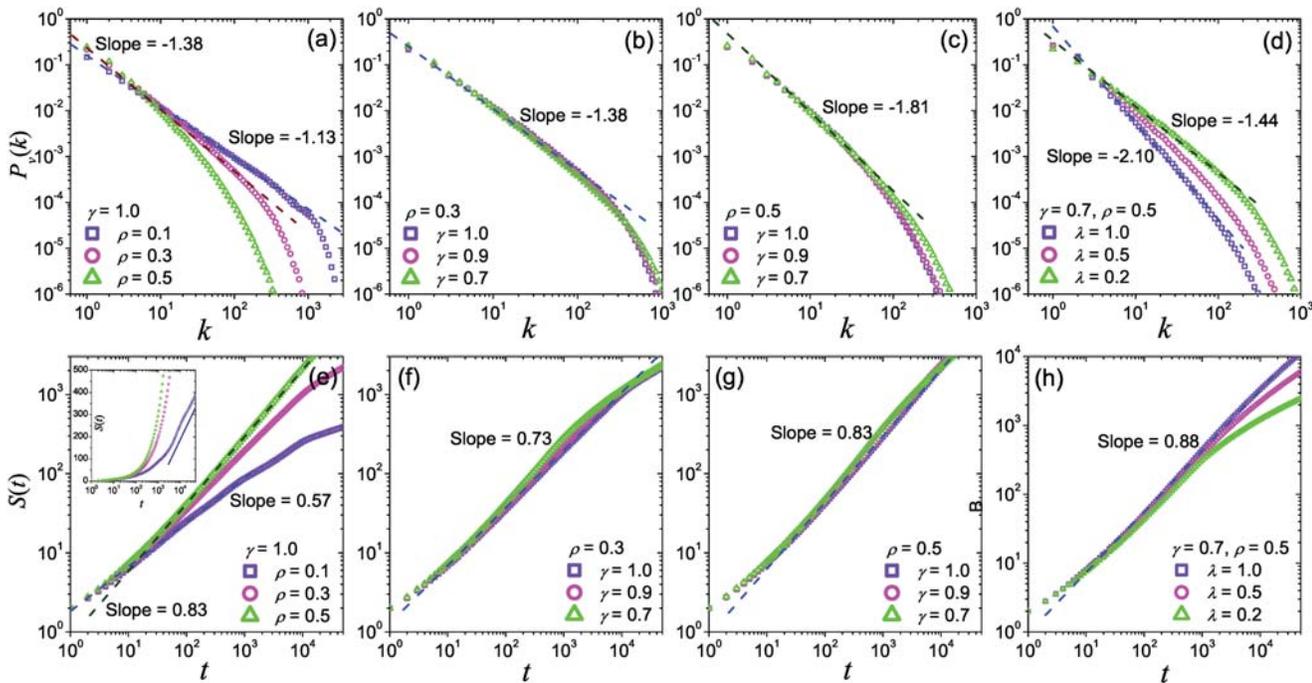
Fig 5. Visiting frequency distributions and the growths of the number of visited locations. Panels (a), (b), (c) and (d) are respectively the visiting frequency distributions generated by the present model with different parameters ρ and γ, as well as panels (e), (f), (g) and (h) with the growth of the number of visited locations. The result is averaged by 10^3 independent runs. The dark blue dashed lines and dark green dashed lines respectively corresponds to the fitting lines of the data signed by light blue squares and light green rectangles, and the blue line in the inset of (e) shows the logarithmical section.

doi:10.1371/journal.pone.0124800.g005

time period $(0, t)$ is defined as:

$$r_g(t) = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(\vec{x}_i - \vec{x}_m)^2}, \tag{3}$$

where $\vec{x}_i$ represents the position of each visited location, and $\vec{x}_m$ is the center of the trajectory during the period [2]. Empirical studies noted that the growth of the radius of gyration of human $r_g(t)$ is close to the logarithmic form rather than the power function predicted by the CRTW model and Lévy flight, indicating an ultraslow spreading in human long-range travels [2, 6]. We also calculate the $r_g(t)$ of GPS carriers using Geo-life GPS datasets and find that r_g obviously deviates from power function for the t larger than 0.5 day (see Section 0.1 of *Materials and Methods*). This abnormal property can also be reproduced by the present model. Generally, the modeling $r_g(t)$ shows a gradual deviation from the initial power functional form with exponent about 0.8 (close to our empirical finding discussed in Section 0.1 of *Materials and Methods*). When $\gamma < 1$, $r_g(t)$ shows a partly logarithmic-like growth in the region $t > 100$, as shown in Fig 6(c), which generally is in agreement with some empirical findings on the mobility of mobile users [2, 6].

The MSD $\langle \Delta x^2 \rangle$ at time t is defined as:

$$\langle \Delta x^2(t) \rangle = \langle (\vec{x}_t - \vec{x}_0)^2 \rangle \tag{4}$$

where $\vec{x}_0$ provides the coordinates of initial position and $\langle \cdot \rangle$ means the average over total running time. Previous studies indicated that, $\langle \Delta x^2(t) \rangle \sim t$ for Brownian random walks and
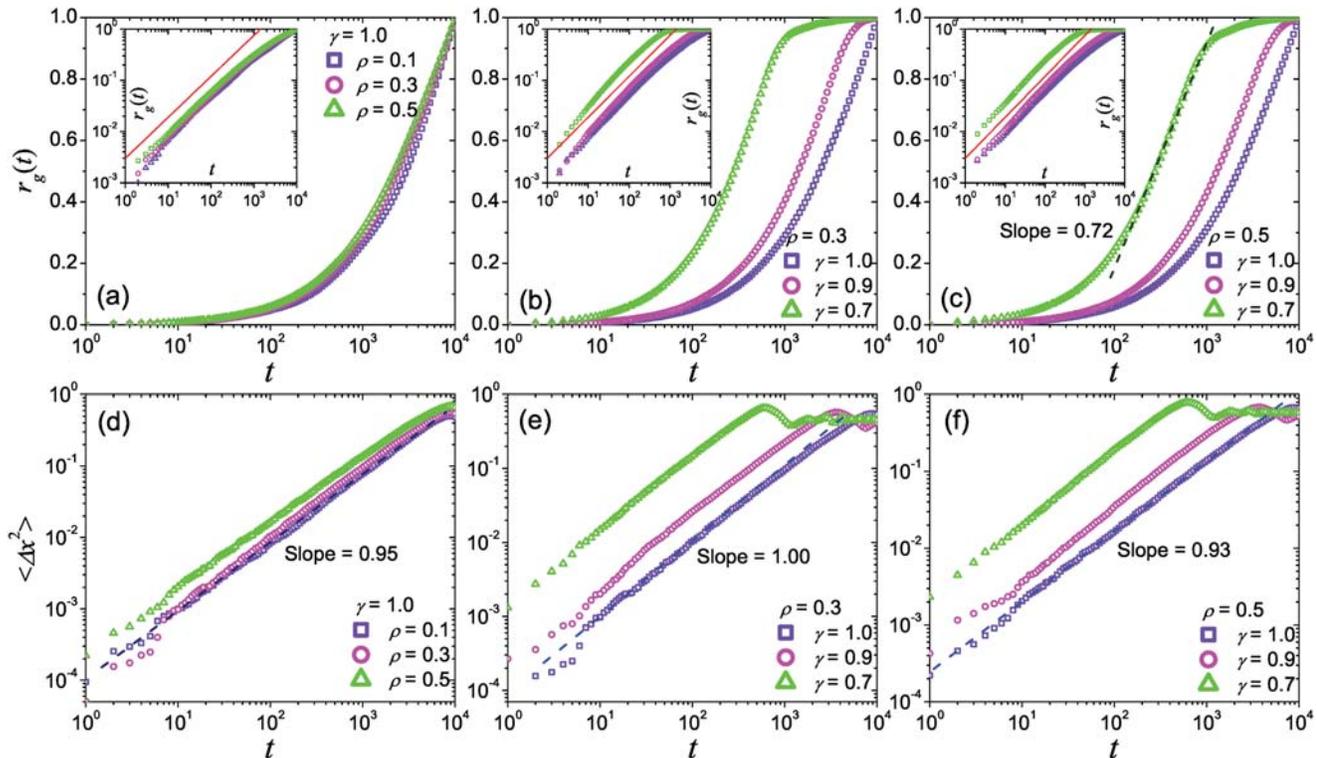
Fig 6. Gyration radius and MSD of the model. Panels (a), (b) and (c) are respectively the growth of radius of gyration generated by the present model with different parameters ρ and γ, as well as panels (d), (e) and (f) for the MSD. The result is averaged by 10^3 independent runs. The insets in panels (a), (b) and (c) show the corresponding $r_g(t)$ curves in log-log plots. The dark blue dashed lines and dark green dashed lines respectively corresponds to the fitting lines of the data signed by light blue squares and light green rectangles, and the red line in each inset indicates the power function with exponent 0.80.

doi:10.1371/journal.pone.0124800.g006

$\langle \Delta x^2(t) \rangle \sim t^{1/(\alpha-1)}$ for pure Lévy flight, and the MSD of real-world human mobility is generally slower than the prediction of Lévy flight [3, 6].

The curves of MSD generated with different parameter settings are shown in Fig 6(d)–6(f). The fitting exponents are much close to 1.0, which is generally close to the empirical findings based on GPS carriers [3], and also can in agreement with our empirical findings on Geo-life GPS datasets (see Section 0.1 of *Materials and Methods* and Fig 7(d)). For the cases $\gamma < 1$, due to the fitting exponent α of the displacement distribution $P_d(\Delta x)$ is less than 2.0, the growth of MSD in our model is slower than the prediction by the pure Lévy flight and is more similar to the empirical findings. Here, λ also does not show obvious effect on both r_g and MSD.

These results indicate that the ultraslow diffusions mainly emerge under the condition of the sub-linear relationship ($\gamma < 1$). In this situation, the time that the walker spends moving within the regions of lower layer locations is relatively short, and therefore the walker will present more frequent returns to the upper layer locations and consequently results in the ultraslow diffusions.

Discussions

The basic assumption of the model is the cascading process in both the spatial and temporal correlations, which is supported by several recent empirical observations [7, 31]. For another assumption, the localized preferential returns and explorations, compared with Song et al's model [6], its expression is linear, and its effect is localized. Apart from these two assumptions,
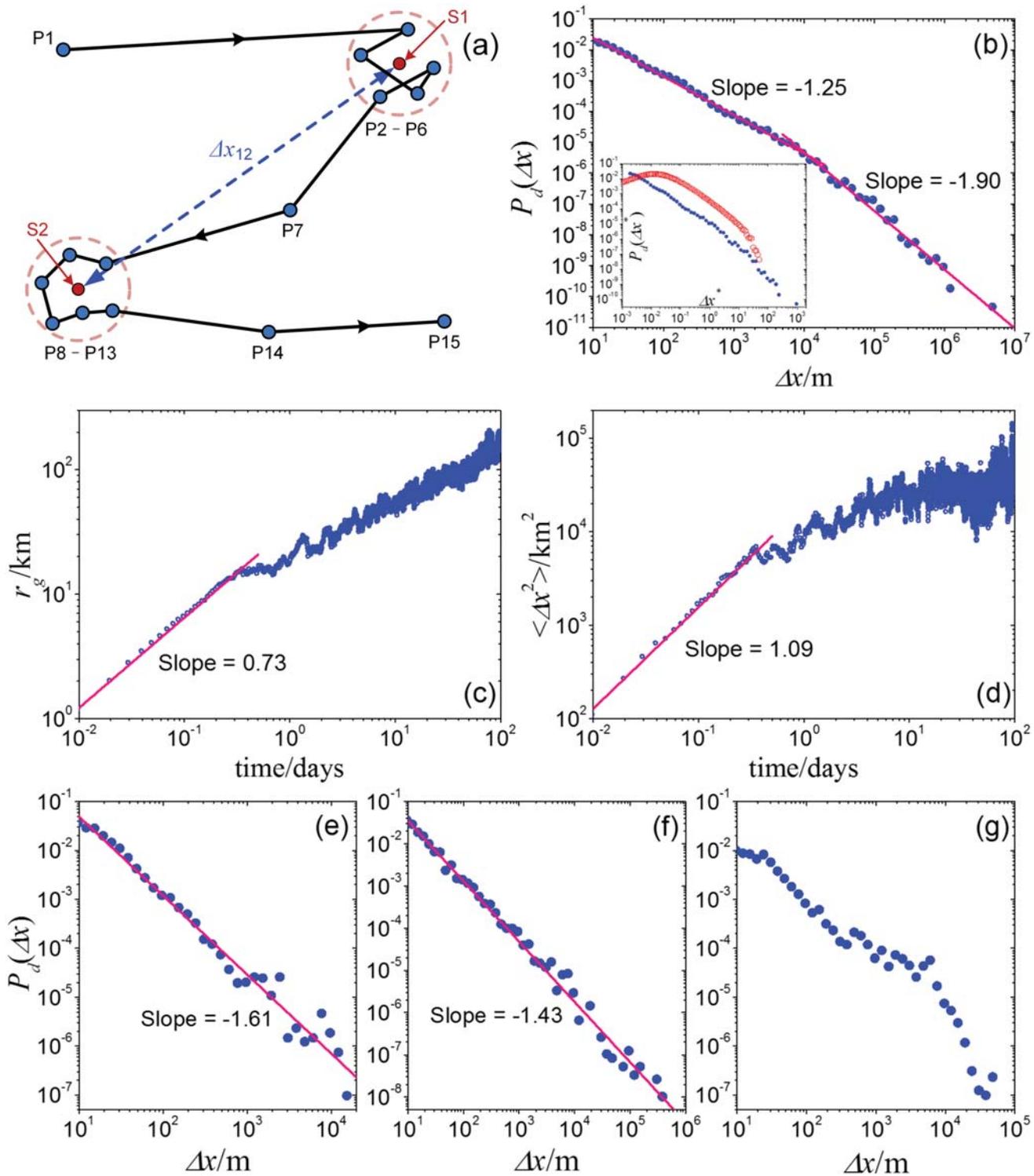
Fig 7. Empirical results of the analysis on GPS records. (a) Illustration of the distinguishing on effective staying positions. (b) The log-binning aggregated displacement distribution $P_d(\Delta x)$, and the inset shows the displacement distributions $P_d(\Delta x^*)$ of the unified displacement $\Delta x^* = \Delta x/\langle\Delta x\rangle$ for the empirical results (the blue dots) and the modeling results under parameter settings $\rho = 0.66$, $\gamma = 0.54$, and $\lambda = 0.5$ (the red hexagons). Panels (c) and (d) respectively show the gyration radius and MSD of the 32 users in aggregated level, and the unit respectively are kilometer (km) and square kilometer (km^2). Panels (e), (f) and (g) show the log-binning displacement distributions of three of the 32 users, and the unit of displacement is meter (m).

this model does not introduce any other spatio-temporal heterogeneities, implying that the scaling properties in long-range human mobility can emerge even without the effect of environmental heterogeneity (e.g. population density, urban environments, et al) [23, 26].

This minimum model reproduces most statistics observed in the empirical analysis, including the scaling anomalies in move-length, inter-visit time and visitation frequency, the slow growth of the number of visited locations, radius of gyration and MSD. Since the scaling displacements is the most noticeable feature in real-world human mobility, the present model successfully creates it without introducing any prior scaling properties. The analytical result of this model indicates that the cascade-like process is the key of the emergence of scaling move-lengths. And also, our model indicates the important role of the sub-linear relationship between the staying time and the displacement in dynamics.

Moreover, the present model provides an unified explanation for the differences on empirical findings in different levels and bridges several controversial issues. First, it shows both the aggregated scaling law and diversity of individual mobility, in agreement with the recent empirical findings across the macroscopic and microcosmic domains [20], indicating a possible mode for the emergence of the aggregated scaling properties of large number of individuals with diverse mobility patterns. Secondly, the result that the mobility in single layer do not show the scaling jump-length, is similar to the real-world situations with single type of vehicle (e.g. bus, taxi, et al) and the short-distance commuting [22, 38], successfully explaining both the scaling anomalies on move-length in global level and the homogeneous properties in urban mobility and single-vehicle trips [20–23, 27]. In addition, the scaling visitation frequency in the model implies the existence of dominant trips, in agreement with peoples' daily experiences and empirical findings [2, 6, 20].

In summary, the present model shows that, only based on the two basic mechanisms, the cascading walks process and the localized preferential returns and explorations, are enough to explain most of the statistics for human mobility in different levels. This conclusion would be very helpful in the understanding of the underlying mechanisms driving human mobility patterns and valuable for the prediction of human daily trips [40] and further applications.

Materials and Methods

0.1 Empirical analysis and findings

To test the assumption that the staying time τ in a region power-functionally depends on the displacement that the walker moves into that region (see Eq (1)), we empirically analyze the real-world human mobility patterns.

We analyze the same data set that were discussed in Ref. [31]. The dataset provides the GPS records from the Microsoft Research Asia in Geo-life Project [41–43] and available online (http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/). It contains the GPS records of handheld equipments and smart phones over three years (April, 2007—Sep., 2010) of 165 volunteers. In the dataset, the trajectory records of each user are of several files. Each of the trajectory files consists of a continuous sequence of tracking points, and each tracking point provides information on the latitude, longitude, and altitude of the position of the GPS holder, along with the corresponding recording time. The total recording time for different individuals is different, and ranges from several weeks to several years.

Due to GPS records can not directly show the positions that users really have stayed in, firstly we distinguish the effective staying positions from the records. Same to the method used in Ref. [31], we set the resolutions of distinguishing the staying positions to be 10 meters for the displacement which is the critical spatial resolution of a handhold GPS equipment, and 120 seconds for the time which is close to the interval of traffic signals. Consider a trajectory

labelled by $\{P_1, P_2, \cdots, P_N\}$, where a continuous sub-sequence $\{P_j, \cdots, P_k\}$ $(1 \leq j \leq k \leq N)$ satisfies the following two conditions: the distances between two consecutive track points are less than 10 meters, and the total time length of the sub-sequence $\{T_j, \cdots, T_k\}$ is larger than 120 seconds. The average position of the sub-sequence is recorded as an effective staying position. As illustrated in Fig 7(a), the average position S1 of track points from P2 to P6 are considered as an effective staying point, as all the geographical distances from P2 to P6 are no more than $10m$ and $T_6 - T_2 < 120s$. Similarly, the track points from P8 to P13 is distinguished to be another effective staying position S2. The straight-line distance between S1 to S2 is set as the user's displacement for the movement from S1 to S2.

The analysis for users' mobility patterns requires long-time continuous movement records. However, most of the files in this dataset only contain the records of a few minutes or hours, and we usually can not obtain enough effective stay positions to achieve good statistical patterns of user's mobility. To get long-time continuous records, we therefore abandon all the files where the recording time is less than 6 hours, and we are left with 927 files from 100 users. Using the above distinguishing algorithms, we distinguish all the effective staying positions and get 17296 effective movements in total. Fig 7(b) shows the log-binning displacement distribution of these movements, in which the tail can be well fitted by a power law with slope -1.90, indicating the scaling property at the aggregated level. The average ratio $\langle \rho_* \rangle$ between each two consecutive displacements (the shorter to the longer) is 0.44, and the real-world average value of ρ is estimated as 0.66 ($\rho_* = 2\rho/3$).

Nonetheless, in the analysis at the individual level, the number of effective staying positions from many of the 100 users is still too small to extract its patterns. We thus remain with the data of 32 users with a number of effective staying positions that is more than 200. Here we one by one analyze the files from the same user, and the results of the user are aggregated from all of his/her files. The aggregated radius of gyration and MSD of the 32 users are shown in Fig 7(c) and 7(d). The growth of radius of gyration obeys power function (the fitting slope is 0.73) within the time scale less than half-day and then trends to slow, which is slightly different from the logarithmic-like type in some previous reports [2]. MSD also shows similar growth pattern, and the fitting exponent of its front power functional region is 1.08. We calculate the displacement distributions for each of the 32 users and find that most of them can be well fitted by a power law in log bins, and the fitting power-law slopes generally are between -1 and -2, as the two examples shown in Fig 7(e) and 7(f). However, some of them show obvious deviation from strict power-law distribution, as the typical example shown in Fig 7(g), denoting the diversity on human mobility patterns at the individual level.

Since the assumption of Eq (1) is one of the cores of our model, the dependence between the staying time τ and the corresponding displacement Δx is empirically analyzed to test this assumption, where τ is the time interval between the moment that the user moves into a small region with radius $\rho\Delta x$ and the moment it moves out, and Δx is the displacement that the user takes when moving into that region. We plot all the data points $(\tau, \Delta x)$ of the 32 users' movements and calculate the distribution of τ versus Δx, as shown in Fig 8. For different ρ, the correlations are all obviously positive. And its log-binning curves show that its averaged effect generally is indeed power-functional-like and sub-linear, and the fitting exponent γ slightly grows from 0.58 to 0.82 along the reduction of ρ from 1/2 to 1/8. That is why we introduce the assumption of Eq (1) and set $\gamma \leq 1$ in our model. Considering the estimated average value of ρ is 0.66, the corresponding fitting γ is about 0.54. Under this parameter settings, the displacement distribution of our model generally is parallel to the empirical curve (see the inset of Fig 7(b)), even though the size of our model limits the stretching on the tail, indicating our model can well fit the empirical findings on displacement distribution.
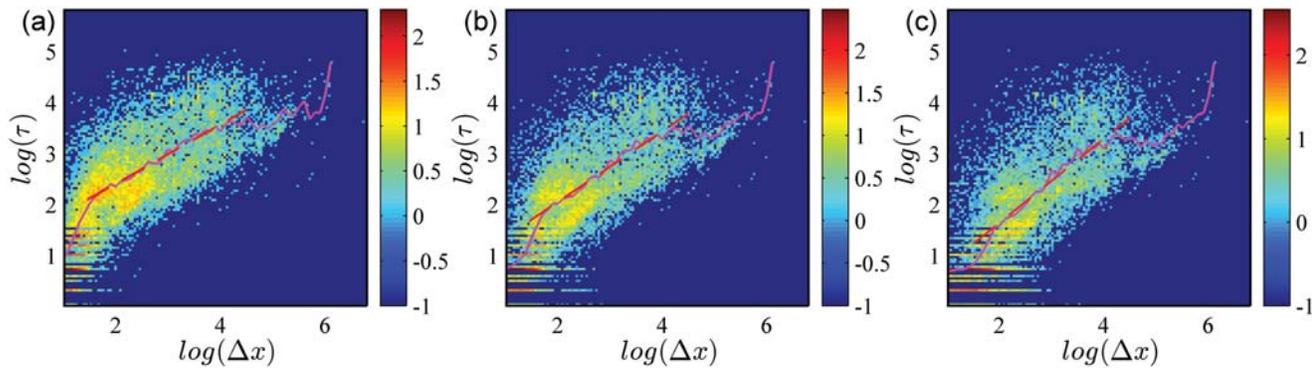
Fig 8. The heat map of the staying time τ in the region with radius ρd vs. the displacement Δx that the user travels into the region. Here $\rho = 0.500$ (a), $\rho = 0.250$ (b) and $\rho = 0.125$ (c). The color is proportional to the logarithm of the heat (count) on each data point. The magenta curves provide the log-binning result of the relationship τ vs. Δx, and the red dashed lines are the fitting power functions of the log-binning curves, where fitting slopes $\gamma = 0.58 \pm 0.01$ for (a), $\gamma = 0.69 \pm 0.02$ for (b) and $\gamma = 0.82 \pm 0.03$ for (c).

0.2 Analytical results of the displacement distributions

In the present model, according to the basic rules of the model: $R = \rho d_c$, $\tau = f_p(d_c^\gamma)$, and assuming $\bar{\Delta x}_1 = 1$ for the movements in first layer, the averaged displacement in the i-th layer movements and the corresponding staying time can be derived immediately:

$$\overline{\Delta x_i} = \rho_*^{i-1}, \tag{5}$$

and

$$\bar{\tau}_i = \overline{\Delta x}_i^{\,\gamma} = \rho_*^{(i-1)\gamma}, \tag{6}$$

where ρ_* is the ratio of the averaged radius of the $(i + 1)$-th layer and i-th layer regions: $\rho_* = \bar{R}_{i+1}/\bar{R}_i$. Obviously, $\rho_* = 2\rho/3$, since the center of the $(i + 1)$-th layer is randomly chosen in the i-th layer region.

From Eq (2), averagely, each ith-layer movements can activate $\overline{\tau}_i/\overline{\tau_{i+1}}$ of $(i + 1)$-th layer trips. On average, we can have

$$\Omega_{i+1} = \Omega_i \overline{\tau}_i / \overline{\tau_{i+1}} = \rho_*^{-\gamma} \Omega_i, \tag{7}$$

where Ω_i denotes the summation of the probability of all the i-th layer movements, as well as Ω_{i+1} for the $(i + 1)$-th layer. The expression of Ω_i is:

$$\Omega_i = \int_0^{\Delta x_{max}} p_i(\Delta x) d(\Delta x), \tag{8}$$

where $\Delta x_{max} = 2\rho^{i-1}$ denotes the maximum possible length of the movements in the i-th layer, and $p_i(\Delta x)$ presents the probability that an i-th layer movement is of a displacement Δx, and Ω_i satisfies the normalized condition $\Sigma_i \Omega_i = 1$. From Eqs (7) and (8), we can immediately have:

$$\int_0^{2\rho^i} p_{i+1}(\Delta x) d(\Delta x) = \rho_*^{-\gamma} \int_0^{2\rho^{i-1}} p_i(\Delta x) d(\Delta x). \tag{9}$$

Considering $\overline{\Delta x}_{i+1} = \rho_* \overline{\Delta x}_i$, and assuming that $p_i(\Delta x)$ and $p_{i+1}(\Delta x)$ have similar forms, the $p_i(\Delta x)$ and $p_{i+1}(\Delta x)$ will have following relationship:

$$p_{i+1}(\Delta x) = C p_i\left(\frac{\Delta x}{\rho_*}\right) \tag{10}$$

where the constant C can be obtained approximately by Eqs (9) and (10):

$$C \approx \rho_*^{-(1+\gamma)} \tag{11}$$

The probability of a displacement with a relative length Δx is the summation of $p(\Delta x)$ over different layers:

$$P_d(\Delta x) = \sum_{i \leq j} p_i(\Delta x) \tag{12}$$

where j is the maximum layer number that satisfies $\Delta x \leq 2\rho^{j-1}$, since the maximum possible length of the j-th layer movements is $2\rho^{j-1}$.

Also due to the center of the j-th layer is randomly chosen in the $(j-1)$-th layer region, the distribution $p_j(\Delta x)$ is in homogeneous type, and thus we can assume that the peak of $p_j(\Delta x)$ is much closer to the averaged value $\overline{\Delta x}_j$. And considering Eq (10), we have:

$$p_i(\overline{\Delta x}_j)\big|_{i<j} < \rho_*^{(1+\gamma)} p_j(\overline{\Delta x}_j) \tag{13}$$

We thus can have:

$$p_j(\overline{\Delta x}_j) < P_d(\overline{\Delta x}_j) < \frac{1 - \rho_*^{j(1+\gamma)}}{1 - \rho_*^{(1+\gamma)}} p_j(\overline{\Delta x}_j) \tag{14}$$

$P_d(\overline{\Delta x}_j)$ therefore is approximately proportional to $p_j(\overline{\Delta x}_j)$. Combining with Eq (10), and $\overline{\Delta x}_j = \rho_*^{j-1}$, we can get:

$$P_d(\overline{\Delta x}_j) \approx \rho_*^{-(1+\gamma)} P_d\left(\frac{\overline{\Delta x}_j}{\rho_*}\right) \tag{15}$$

We therefore obtain $P_d(\overline{\Delta x}_j) \propto \overline{\Delta x}_j^{-\alpha}$, where the power law exponent $\alpha = \gamma + 1$.

$P_d(\overline{\Delta x})$ for the $\overline{\Delta x}$ of different layers construct the basic power-law-like form of the displacement distribution $P_d(\Delta x)$, therefore $P_d(\Delta x)$ generally obeys a power law form with the exponent $\alpha = 1 + \gamma$, even though it has slight fluctuation in the condition with small ρ. The analytical result is in agreement with the numerical simulations shown in Fig 3. Here the exponent α only depends on γ, nevertheless, it is in the condition where $0 < \rho_* < 1$, indicating that the cascading process in the model is of the basic requirement in the origin for the power-law-like displacement distribution.

0.3 Analysis of the visitation frequency and the number of visited locations in a situation without the cascading effect

An extreme situation is the model without the cascading effect ($\rho^* = 0$). In this case, all the movements and locations are in the same layer, and Eq (2) would be the only rule that drives the model. The evolution of the visitation frequency distribution $P'_v(k')$ can be derived directly:

$$k P'_v(k') = (k'-1) P'_v(k'-1), \tag{16}$$

Note that, here k' represents the total number of visitation of a location from the initial time $t = 0$. We obtain:

$$P'_v(k') \propto k'^{-1}. \tag{17}$$

Considering the initial process from $t = 0$ to $t = t_0$, for $t > t_0$, we have: $P'_v(k_0 + k) \propto (k_0 + k)^{-1}$ and $P'_v(k_0) \propto k_0^{-1}$, where k_0 and k respectively denote the total number of visitation of a location before and after t_0. So we can have $P'_v(k) \propto k^{-1}$, which is the visitation frequency distribution in this extreme situation.

For the growth of the number of visited locations $S'(t)$, in this extreme situation, according to the definition of $S'(t)$ and $P'_v(k)$, we have

$$S'(t) = \sum_k P'_v(k), \tag{18}$$

and

$$t = \sum_k [k P'_v(k)]. \tag{19}$$

Substituting $P'_v(k) \propto k^{-1}$ into the two equations above, we can easily obtain that $S'(t) \propto \ln t$.

In brief, removing the cascading effect, the visitation frequency distribution of the model will obey a power law form with the exponent of -1, and the growth of the number of visited locations will show a logarithmic relationship over time.

Acknowledgments

Author Contributions

Conceived and designed the experiments: XPH. Performed the experiments: XPH. Analyzed the data: XPH XWW. Contributed reagents/materials/analysis tools: XPH XWW XYY BHW. Wrote the paper: XPH XWW.

References

1. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. Nature 439:462–465. doi: 10.1038/nature04292 PMID: 16437114

2. González MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. Nature 453:779–782. doi: 10.1038/nature06958 PMID: 18528393

3. Rhee I, Shin M, Hong S, Lee K, Chong S (2008) On the Levy-walk nature of human mobility: Do humans walk like monkeys? IEEE Conference on Computer Communications 2008, pp. 924.

4. Lee K, Hong S, Kim SJ, Rhee I, Chong S (2009) SLAW: A New Mobility Model for Human Walks. IEEE Conference on Computer Communications 2009, pp. 855.

5. Jiang B, Yin J, Zhao S (2009) Characterizing the human mobility pattern in a large street network. Phys Rev E 80: 021136. doi: 10.1103/PhysRevE.80.021136

6. Song C, Koren T, Wang P, Barabási A-L (2010) Modelling the scaling properties of human mobility. Nat Phys 6:818–823. doi: 10.1038/nphys1760

7. Jia T, Jiang B, Carling K, Bolin M, Ban Y (2012) An empirical study on human mobility and its agent-based modeling. J Stat Mech P11024. doi: 10.1088/1742-5468/2012/11/P11024

8. Szell M, Sinatra R, Petri G, Thurner S, Latora V (2012) Understanding mobility in a social petri dish. Sci Rep 2: 457. doi: 10.1038/srep00457 PMID: 22708055

9. Sinatra R, Szell M (2014) Entropy and the predictability of online life. Entropy 16:543–556. doi: 10.3390/e16010543

10. Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of Predictability in Human Mobility. Science 327:1018–1021. doi: 10.1126/science.1177170 PMID: 20167789

11. Belik V, Geisel T, Brockmann D (2011) Natural human mobility patterns and spatial spread of infectious diseases. Phys Rev X 1: 011001.

12. Balcan D, Vespignani A (2011) Phase transitions in contagion processes mediated by recurrent mobility patterns. Nat Phys 7:581–586. doi: 10.1038/nphys1944 PMID: 21799702

13. Wang L, Li X, Zhang Y-Q, Zhang Y, Zhang K (2011) Evolution of scaling emergence in large-scale spatial epidemic spreading. PLOS ONE 6: e21197. doi: 10.1371/journal.pone.0021197 PMID: 21747932

14. Ni S, Weng W (2009) Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks. Phys Rev E 79: 016111. doi: 10.1103/PhysRevE.79.016111

15. Wang P, González MC, Hidalgo CA, Barabási A-L (2009) Understanding the Spreading Patterns of Mobile Phone Viruses. Science 324:1071–1076. doi: 10.1126/science.1167053 PMID: 19342553

16. Zhao Z-D, Liu Y, Tang M (2012) Epidemic variability in hierarchical geographical networks with human activity patterns. Chaos 22: 023150. doi: 10.1063/1.4730750 PMID: 22757557

17. Liu C, Du W-B, Wang W-X (2014) Particle Swarm Optimization with Scale-Free Interactions. PLOS ONE 9: e97822. doi: 10.1371/journal.pone.0097822 PMID: 24859007

18. Horner MW, O'Kelly MES (2001) Embedding economies of scale concepts for hub networks design. J Transp Geogr 9:255–265. doi: 10.1016/S0966-6923(01)00019-9

19. Petrovskii S, Mashanova A, Jansen VAA (2011) Variation in individual walking behavior creates the impression of a Lévy flight. Proc Natl Acad Sci USA 108:8704–8707. doi: 10.1073/pnas.1015208108 PMID: 21555579

20. Yan X-Y, Han X-P, Wang B-H, Zhou T (2013) Diversity of individual mobility patterns and emergence of aggregated scaling laws. Sci Rep 3: 2678. doi: 10.1038/srep02678 PMID: 24045416

21. Bazzani A, Giorgini B, Rambaldi S, Gallotti R, Giovannini L (2010) Statistical laws in urban mobility from microscopic GPS data in the area of Florence. J Stat Mech P05001.

22. Liang X, Zheng X, Lv W, Zhu T, Xu K (2012) The scaling of human mobility by taxis is exponential. Physica A 391:2135–2144. doi: 10.1016/j.physa.2011.11.035

23. Noulas A, Scellato S, Lambiotte R, Pontil M, Mascolo C (2012) Tale of Many Cities: Universal Patterns in Human Urban Mobility, PLOS ONE 7: e37027. doi: 10.1371/journal.pone.0037027 PMID: 22666339

24. Peng C, Jin X, Wong K-C, Shi M, Liò P (2012) Collective Human Mobility Pattern from Taxi Trips in Urban Area, PLOS ONE 7; e34487. doi: 10.1371/journal.pone.0034487 PMID: 22529917

25. Simini F, González MC, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns, Nature 484:96–100. doi: 10.1038/nature10856 PMID: 22367540

26. Veneziano D, González MC, Trip Length Distribution Under Multiplicative Spatial Models of Supply and Demand: Theory and Sensitivity Analysis, arxiv: 1101.3719.

27. Gallotti R, Bazzani A, Rambaldi S (2012) Towards a statistical physics of human mobility. Int J Mod Phys C 23: 2150061.

28. Han X-P, Hao Q, Wang B-H, Zhou T (2011) Origin of the scaling law in human mobility: Hierarchy of traffic systems. Phys Rev E 83: 036117. doi: 10.1103/PhysRevE.83.036117

29. Yan X-Y, Han X-P, Zhou T, Wang B-H (2011) Exact Solution of Gyration Radius of Individual's Trajectory for a Simplified Human Regular Mobility Model. Chin Phys Lett 28: 120506. doi: 10.1088/0256-307X/28/12/120506

30. Grabowicz PA, Ramasco JJ, Gonçalves B, Eguíluz VM (2014) Entangling Mobility and Interactions in Social Media. PLOS ONE 9: e92196. doi: 10.1371/journal.pone.0092196 PMID: 24651657

31. Wang X-W, Han X-P, Wang B-H (2014) Correlations and scaling laws in human mobility. PLOS ONE 9: e84954. doi: 10.1371/journal.pone.0084954 PMID: 24454769

32. Malmgren RD, Stouffer DB, Motter AE, Amaral LAN (2008) A Poissonian explanation for heavy tails in e-mail communication, Proc. Natl. Acad. Sci. USA. 105:18153–18158. doi: 10.1073/pnas.0800332105 PMID: 19017788

33. Rybski D, Buldyrev SV, Havlin S, Liljeros F, Makse HA (2009) Scaling laws of human interaction activity. Proc Natl Acad Sci USA 106:12640–12645. doi: 10.1073/pnas.0902667106 PMID: 19617555

34. Rybski D, Buldyrev SV, Havlin S, Liljeros F, Makse HA (2012) Communication activity in a social network: relation between long-term correlations and inter-event clustering. Sci Rep 2: 560. doi: 10.1038/srep00560 PMID: 22876339

35. Bartumeus F, Peters F, Pueyo S, Marrasé S, Catalan J (2003) Helical Lévy walks: Adjusting searching statistics to resource availability in microzooplankton. Proc Natl Acad Sci USA 100:12771–12775. doi: 10.1073/pnas.2137243100 PMID: 14566048

36. Ramos-Fernández G, Mateos JL, Miramontes O, Cocho G, Larralde H, Ayala-Orozco B (2004) Lévy Walk Patterns in the Foraging Movements of Spider Monkeys. Behav Ecol Sociobiol 55:223–230. doi: 10.1007/s00265-003-0700-6

37. Sims DW, Southall EJ, Humphries NE, Hays GC, Bradshaw CJA, Pitchford JW, et al. (2008) Scaling laws of marine predator search behaviour. Nature 451:1098–1102. doi: 10.1038/nature06518 PMID: 18305542

38. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignania A (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci USA 106:21484–21489. doi: 10.1073/pnas.0906910106 PMID: 20018697

39. Vázquez A, Oliveira JG, Dezsö Z, Goh KI, Kondor I, Barabási A-L (2006) Modeling bursts and heavy tails in human dynamics. Phys Rev E 73: 036127. doi: 10.1103/PhysRevE.73.036127

40. Domenico MD, Lima A, Musolesi M (2013) Interdependence and Predictability of Human Mobility and Social Interactions. Pervasive and Mobile Computing 9:798–807. doi: 10.1016/j.pmcj.2013.07.008

41. Zheng Y, Zhang L, Xie X, Ma W-Y (2009) Mining interesting locations and travel sequences from GPS trajectories. WWW'09, pp. 791–800.

42. Zheng Y, Li Q, Chen Y, Xie X, Ma W-Y (2008) Understanding Mobility Based on GPS Data. UbiComp'08, pp. 312–321.

43. Zheng Z, Xie X, Ma W-Y (2010) GeoLife: A Collaborative Social Networking Service among User, location and trajectory. IEEE Data Eng Bull, pp. 32–40.