

Correlations and Scaling Laws in Human Mobility

Xiang-Wen Wang^{1,2,3}, Xiao-Pu Han^{1*}, Bing-Hong Wang^{3,4,5}

1 Alibaba Research Center for Complexity Sciences, Hangzhou Normal University, Hangzhou, China, **2** Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, United States of America, **3** Department of Modern Physics, University of Science and Technology of China, Hefei, China, **4** College of Physics and Electronic Information Engineering, Wenzhou University, Wenzhou, China, **5** The Research Center for Complex System Science, University of Shanghai for Science and Technology, Shanghai, China

Abstract

Background: In recent years, several path-breaking findings on human mobility patterns point out a novel issue which is of important theoretical significance and great application prospects. The empirical analysis of the data which can reflect the real-world human mobility provides the basic cognition and verification of the theoretical models and predictive results on human mobility. One of the most noticeable findings in previous studies on human mobility is the wide-spread scaling anomalies, e.g. the power-law-like displacement distributions. Understanding the origin of these scaling anomalies is of central importance to this issue and therefore is the focus of our discussion.

Methodology/Principal Findings: In this paper, we empirically analyze the real-world human movements which are based on GPS records, and observe rich scaling properties in the temporal-spatial patterns as well as an abnormal transition in the speed-displacement patterns together with an evidence to the real-world traffic jams. In addition, we notice that the displacements at the population level show a significant positive correlation, indicating a cascading-like nature in human movements. Furthermore, our analysis at the individual level finds that the displacement distributions of users with stronger correlations usually are closer to the power law, suggesting a correlation between the positive correlation of the displacement series and the form of an individual's displacement distribution.

Conclusions/Significance: These empirical findings make connections between the two basic properties of human mobility, the scaling anomalies on displacement distributions and the positive correlations on displacement series, implying the cascading-like dynamics which is exhibited by the positive correlations would cause the emergence of scaling properties on human mobility patterns. Our findings would inspire further researches on mechanisms and predictions of human mobility.

Citation: Wang X-W, Han X-P, Wang B-H (2014) Correlations and Scaling Laws in Human Mobility. PLoS ONE 9(1): e84954. doi:10.1371/journal.pone.0084954

Editor: Yamir Moreno, University of Zaragoza, Spain

Received: August 23, 2013; **Accepted:** November 20, 2013; **Published:** January 13, 2014

Copyright: © 2014 Wang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the National Natural Science Foundation of China Grants Nos. 11205040, 11105024, 70971089 and 10635040. XPH acknowledges the Zhejiang Provincial Natural Science Foundation of China (Grant No. LY12A05003), the start-up foundation and Pandeng project of Hangzhou Normal University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: xp@hznu.edu.cn

Introduction

The statistical patterns of human daily movements directly affect the physical contacts between humans and thus deeply impact on the dynamics of many social systems. The understanding of real-world human mobility patterns would be very helpful for the advancements of many aspects of social dynamics, such as epidemics spreading [1–4], the designing of traffic systems [5], or localized recommendations [6,7]. Since the pioneering work of Brockmann et al (2006) [8], the temporal-spatial statistical properties in human movements have become a new issue in complex sciences and have attracted much attention in recent years.

The most dramatic discovery in the statistical patterns of human mobility is the existence of wide-spread scaling properties [8–10]. The first one is the power-law-like displacement distribution, which has been empirically observed not only in many analyses of real-world human movements [8,9,11] but also in the study of virtual world of online-games [12]. This result sharply differs with the traditional understanding based on random walks, and reveals long-range correlations in human travels and social interactions.

Other scaling properties include the staying time distributions which denote that humans usually stay in a few locations for quite long periods of time [9], and the visitation frequency distributions are dominated by a few locations that are usually much more frequently visited [9,10], and so on.

Many other abnormal properties are also found in human mobility patterns, including ultra-slow diffusion [8,9], anisotropism [9], high predictability [13], and the limitation of roads [14]. These discoveries reveal abnormal features in real-world human mobility, in stark contrast to the traditional understandings based on the hypothesis of random-walk-like human mobility or on that of Lévy flights with the same scaling displacement distributions.

However, these findings are still facing several controversies. Due to the limitation of original data, most previous works are at the population level, and a direct analysis of individuals is rarely seen. Recently, Yan, et al. [15] reported the diversity in individual-level mobilities and found out that most of the individuals' displacement distributions do not obey the scaling law. Moreover, several recent researches indicated that the move length in human urban trips or the travels by a single type of transportation do not obey well a power law [16–18]. These controversies require the

confirmation from a more in-depth empirical analysis of human mobility patterns.

Recent studies also proposed many models to explain the underlying mechanisms that drives the emergence of these anomalies in human mobility. Generally, the basic dynamics of previous modeling works can be divided into the following classes: i) The descriptive models: Lévy flights [11], Self-similar least action walk (SLAW) [19], and Continuous-time random walks [8]; ii) The exploration of new locations and the preferential return to visited places [10]; iii) The effect of hierarchical traffic systems [20]; iv) the effect of few dominant trips [21]; v) The spatial heterogeneity of population density or the geographic locations [18,22]; vi) The radiation model proposed by Simini et. al. [23], which can reproduce many mobility patterns at the global level; vii) The aggregation of individuals without scaling properties [15]. These models can reproduce parts of the empirical findings. Nevertheless, it is difficult to identify common rules from these models, and thus it remains controversial what drives the emergence of these abnormal properties in human mobility. It would therefore be helpful if the empirical analysis can identify characteristic factors affecting the emergence of these anomalies.

In this paper, based on the empirical analysis of GPS data sets, we report one of the characteristic factors that is relevant to the scaling displacement distributions: the correlation of the series of displacement. We first show the aggregated temporal-spatial properties at the population level (Section II) and then we analyze the correlation of the aggregated series of displacements (Section III). Finally, we discuss the diversity in individuals' mobility patterns and the relationship between the correlation of the series of displacements and the scaling properties of displacement distributions (Section IV). We show that the correlation is indeed a tool that allows to investigate the underlying mechanisms from the empirical data.

Results

The scaling properties at the population level

The data set in our analysis contains records from 165 volunteers that have been gathered over three years (April, 2007 – Sep., 2010). The GPS trajectories result from the Microsoft Research Asia in Geo-life Project [24–26]. More details can be found in *Materials and Methods*.

We determine the effective staying positions from the dataset using a resolution of 10 meters in space and 120 seconds in time. Fig. 1 illustrates for a case of two staying positions, S1 and S2, that are obtained from a sequence of GPS records. Details of our approach will be discussed in *Materials and Methods*. The geographical distance between two consecutive staying positions, e.g. S1 and S2 in Fig. 1, is defined as the displacement of travel. The staying time in each staying position is defined as the time interval between the first and last GPS records in the given staying position.

Using the above method, we obtain 927 trajectories with recording times longer than 6 hours that contain 19376 effective staying points. The total staying time is 4463 hours, and the total displacement is 95472.33 kilometers. From each of these trajectories, we can obtain a sequence that contains the staying positions, displacements and staying times.

We combine the displacements and staying times in all 927 files to calculate the displacement distribution and the staying time distribution at the population level. After log-binning, the displacement distribution $P(d)$ generally obeys the following power-law function with two different regimes (Fig.2 (a)):

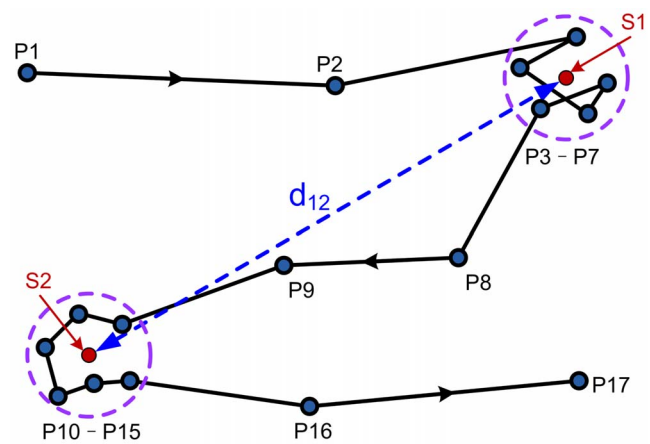


Figure 1. Illustration of the distinguishing on effective staying positions. P1–P17 represent 17 track points recorded by a GPS equipment from which we obtain two staying points S1 and S2. The displacement of travel is defined as the distance between the centers of the two staying points.

doi:10.1371/journal.pone.0084954.g001

$$P(d) \sim \begin{cases} d^{-1.25}, & (d < 6.5\text{km}), \\ d^{-1.90}, & (d \geq 6.5\text{km}). \end{cases} \quad (1)$$

This power-law displacement distribution indicates that the typical behavior consists of many short-range trips and few long-range travels. This conclusion is in substantial agreement with the conclusions of several previous findings [9,10]. The transition at $d \simeq 6.5\text{km}$ is related to the average extend of the urban district of cities, indicating the difference between human urban movements and intercity travels. This difference may be due to the convenience of urban movements and the dominant high-frequency movements between few positions (such as home and working places) [15,21].

A similar scaling property is also observed in the staying time distribution $P(t_s)$ at the population level, which can be well fitted by a power-law function with an exponent -1.98 (Fig. 2 (b)), indicating that humans usually stay in few positions for quite a long time. This result is also close to previous findings based on other data sets [9,11,16,27].

The distribution $P(t_e)$ of the elapsed time t_e that individuals spend on the way from an effective staying position to the next one has also been studied. As shown in Fig. 2 (c), $P(t_e)$ shows a strange behavior where two power-law-like sections are separated by an unusual bump when $10^4 < t_e < 2 \times 10^4$ seconds. It seems that this bump results from the traffic jams. This result is somewhat different to the previous findings in urban taxi data [17].

Moreover, we calculate the average speed $v_i = d_i/t_{ei}$ for every user i , and plot each pair (v_i, d_i) on the plane to get the pattern of the relationship between speed and displacement. We surprisingly find that v vs. d generally obeys two-section scaling form, in which the first section ($d < 10^3$ meters) is almost linear (slope ≈ 1.0), whereas another part ($d > 10^3$ meters) is sub-linear (slope ≈ 0.5), as shown in Fig. 3. The point of transition $d \approx 10^3$ meters and $v \approx 1$ m/s, could relate to the length and speed of walking, therefore the two sections would correspond to the travel by foot or by automobile, with humans preferring a trip by automobile (bus, car, etc.) for distances longer than 1 kilometer.

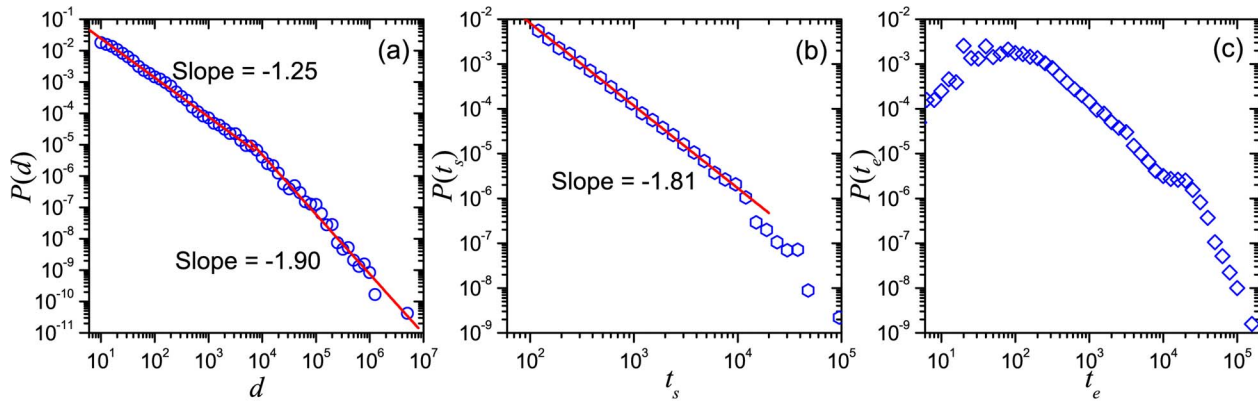


Figure 2. Aggregated mobility patterns. (a) The aggregated displacement distribution $P(d)$, (b) the staying time distribution $P(t_s)$, (c) and the elapsed time distribution $P(t_e)$ in log-log plots.
doi:10.1371/journal.pone.0084954.g002

In addition, some movements have ultraslow speed, as indicated by the dashed line in Fig. 3. The corresponding displacements of these ultrashow movements are generally between 10^2 meters and 10^3 meters, and the corresponding elapsed time is mainly in the range from 10^4 seconds to 2×10^4 seconds, corresponding to the bump in $P(t_e)$ displayed in Fig. 2, possibly indicating displacements hampered by traffic jams.

Correlations of displacements at the population level

For each sequence of displacements of individuals, the correlation between two consecutive displacements reflects the trends and causal relationship in human travels. To get the pattern of the correlation, we plot each of the data points (d_i, d_{i+1}) and calculate the density of these data points. Here d_i and d_{i+1} denote the i -th and the $(i+1)$ -th displacement in the series \vec{d} . As shown in Fig. 4(a), most of the data points (d_i, d_{i+1}) accumulate close to the diagonal line $d_{i+1} = d_i$, corresponding to a positive correlation.

We also plot the pattern using the related displacement $\bar{d} = d/d_*$, where d_* is the average displacement of the user. We first calculate the average displacement of each user and then obtain the sequences of \bar{d} from each file. Fig. 4(b) shows the

pattern of the density of the data points $(\bar{d}_i, \bar{d}_{i+1})$, where the positive correlation is much clearer.

Taking into account the heterogeneous d , we use the rank-based correlation coefficient named *Kendall's Tau* to quantify the strength of this correlation. We first set $\vec{d}_i = \{d_1, d_2, \dots, d_i, \dots, d_{N-1}\}$, and $\vec{d}_{i+1} = \{d_2, d_3, \dots, d_{i+1}, \dots, d_N\}$ for every series, where N is the total number of the displacements in the corresponding file. The detailed introduction of Kendall's Tau can be found in the second section of *Materials and Methods*. The value of the Kendall's Tau $\tau_K = 0.424$ for the series \vec{d}_i and \vec{d}_{i+1} , and the confidence interval with 95% significance level is 0.015. For the related displacement series $\vec{\bar{d}}_i$ and $\vec{\bar{d}}_{i+1}$, $\tau_K = 0.435$ which represents a significant positive correlation.

This remarkable positive correlation shows that a trip can have effect on the next one: if the current displacement is long, the next one has a high probability to be only slightly different. The change in displacement is usually gradual. This gradual change agrees with our daily experience. For example, if we travel to another city, we first need to find a hotel in the city. The movement from our city to the target hotel generally is a long travel (the length may be several hundred miles). In the next several days, we might leave the hotel to visit some places around the city (generally tens of miles). During each trip, our visit will contain many short moves (usually less than one mile). A direct trip from our city to the place in the target city rarely appears.

Furthermore, to investigate the long-term correlations in human mobility, we calculate the Kendall's Tau τ_K of the series \vec{d}_i and \vec{d}_{i+m} ($m = 1, 2, \dots$), and find that the function τ_K vs. m shows a remarkable slow decay, which can be well fitted by a power-law function with an exponent -0.276 (Fig. 5(a)), implying that the effect of previous movements can continue a very long time. To ensure it, we plot the Pearson correlation coefficient R_m between $\log(\vec{d}_i)$ and $\log(\vec{d}_{i+m})$. It does obey a power-law decay with a slope $\gamma = -0.367$ (Fig. 5(b)). The value $H = 1 + \gamma/2 \approx 0.82$ is the well-known Hurst exponent that denotes the long-term correlations in the fluctuation of the series [28]. Using the method of detrended fluctuation analysis (DFA) [29] (see the third section in *Materials and Methods*), we obtain a similar Hurst exponent value of $H = 0.87$, indicating a strong long-term correlation among the displacement series.

Two other correlations have also been studied, namely the correlation among the series of staying times, and the correlation between the staying time and the displacement. The series of

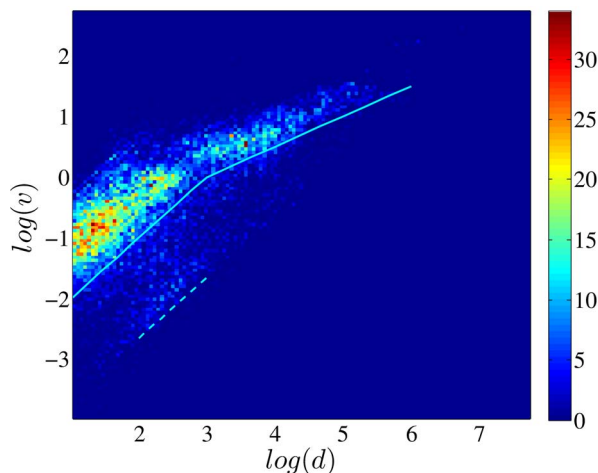


Figure 3. The relationship between the average speed v and the displacement d . The slopes of the upper lines are 1.0 and 0.5 respectively, whereas the slope of the lower dashed line is 1.0.
doi:10.1371/journal.pone.0084954.g003

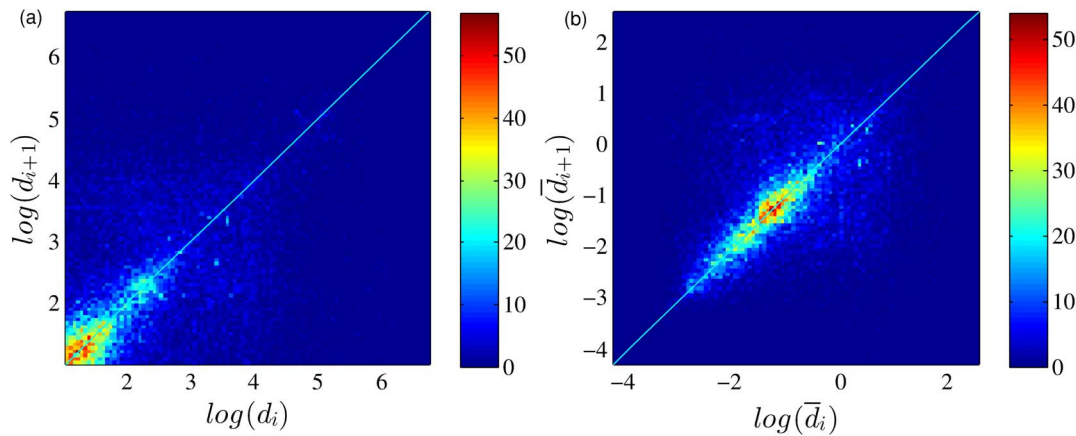


Figure 4. Aggregated displacement correlations d_i vs. d_{i+1} (a) and \bar{d}_i vs. \bar{d}_{i+1} (b) in scatter plot. The figure shows a high density of points near the diagonal line $d_{i+1} = d_i$.
doi:10.1371/journal.pone.0084954.g004

staying times show only a weak positive correlation (its $\tau_K = 0.120$ with the confidence interval 0.015), and the staying times and displacements are almost independent (its $\tau_K = 5.30 \times 10^{-3}$).

Mobility patterns at the individual level

The above discussions showed the scaling patterns and positive correlation of human movements at the population level. Nevertheless, since the above results are aggregated over all individuals, we can not directly conclude that the movements of each individual also exhibit the same properties. Actually, power-law-like displacement distribution at the population level can even be observed in a system where all the individuals' movements are Poissonian [15,30]. Because of the lack of direct evidence, it remains controversial whether the scaling mobility patterns are universal at the individual level. Recently, Yan et al. reported the diversity of human mobility patterns at the individual level and that many individuals' displacement distributions usually are dominated by some frequently-appearing mobilities [15]. Due to the limitations in the original data sets of Yan's work, this conclusion still needs to be confirmed by more in-depth empirical studies based on datasets with higher resolution.

Among the 100 remaining users, we choose the users who had more than 200 effective staying positions to study their mobility

patterns at the individual level. We choose 200 to be the minimum number of effective staying points in order to obtain efficient statistical patterns. By doing this, 32 effective individuals with 698 files and 15189 staying positions are chosen. The number of effective staying positions n_s and the number of displacements n_d of each of the 32 users are listed in Table 1.

Plotting the displacement distribution $P(d)$ and correlation patterns (d_i, d_{i+1}) for each of the 32 users, we remark that users with stronger positive correlation seem to have usually a displacement distribution that is closer to a power law. The trajectories, displacement distributions and correlation patterns (d_i, d_{i+1}) of two typical users are shown in Fig. 6. User No. 9 has many long-range movements, and his/her displacement distribution obeys well a power law. Significant positive correlation is also observed. In contrast, the displacement distribution of user No. 22 is bimodal-like, and the correlation is also not obvious.

The positive correlation reflects a gradually changing nature of human displacements. Previous studies in the temporal patterns have found that this gradually changing process, or say the cascading effect, is of close relevance to the emergence of burstiness in human activities [31], as well as the long-term persistences [32,33]. Our results seem to indicate that the positive

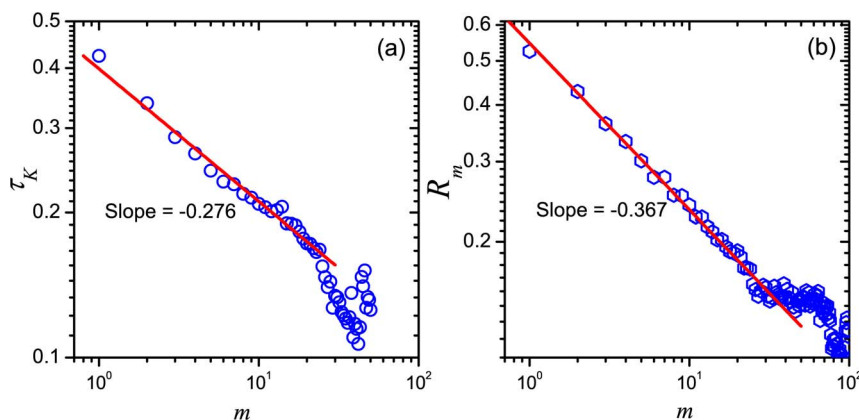


Figure 5. Long-term correlations in human mobility. (a) The decay of Kendall's Tau τ_K between \bar{d}_i and \bar{d}_{i+m} as a function of the interval m . (b) Pearson correlation coefficient R_m between $\log(\bar{d}_i)$ and $\log(\bar{d}_{i+m})$ as a function of the interval m .
doi:10.1371/journal.pone.0084954.g005

Table 1. Information and fitting parameters of the 32 individuals, where n_s and n_d are the number of effective staying positions and displacements of each user, d_* and d_{max} are the average displacement and the maximum displacement of the user, and α is the fitting exponent of $P(d)$ using the estimated lower bound x_{min} .

user ID	n_s	n_d	d_*/m	d_{max}/m	τ_K	$\Delta\tau$	R_d	H	p_{KS}	x_{min}	α	p'_{KS}	R'_d	R'_t
1	424	407	10980	879844	0.231	0.085	0.360	0.83	4.37×10^{-3}	49.9	1.496	1.21×10^{-17}	-0.978	-0.989
2	278	257	12180	258183	0.360	0.108	0.500	0.85	3.98×10^{-1}	3203.3	1.604	1.05×10^{-7}	-0.983	-0.980
3	747	716	14311	5261960	0.312	0.065	0.421	0.80	6.42×10^{-1}	404.9	1.469	2.22×10^{-26}	-0.989	-0.993
4	232	220	7301	257901	0.390	0.113	0.515	0.66	6.76×10^{-1}	4477.9	2.084	3.97×10^{-13}	-0.968	-0.991
6	323	310	1373	40117	0.327	0.099	0.392	0.79	2.41×10^{-1}	97.6	1.547	1.13×10^{-5}	-0.989	-0.974
8	237	228	905	21326	0.300	0.111	0.325	0.55	3.89×10^{-1}	152.5	1.649	2.91×10^{-4}	-0.980	-0.991
9	1036	996	3161	490551	0.368	0.056	0.504	0.84	1.23×10^{-1}	9.9	1.426	1.21×10^{-1}	-0.994	-0.993
10	563	541	2022	158233	0.367	0.075	0.457	0.86	5.18×10^{-1}	98.6	1.543	5.64×10^{-7}	-0.992	-0.997
12	299	290	1086	21866	0.288	0.100	0.348	0.51	9.51×10^{-1}	97.6	1.752	7.32×10^{-18}	-0.966	-0.993
15	240	218	47632	526428	0.187	0.116	0.227	0.72	3.43×10^{-1}	35114.0	2.065	9.23×10^{-13}	-0.976	-0.980
22	1050	986	3055	60918	0.230	0.055	0.344	0.69	4.72×10^{-1}	5601.9	3.232	1.10×10^{-24}	-0.964	-0.972
26	2702	2650	370	16074	0.376	0.034	0.546	0.85	3.62×10^{-1}	15.2	1.656	3.07×10^{-10}	-0.987	-0.996
27	297	287	10732	639964	0.380	0.104	0.546	1.00	1.44×10^{-1}	35.4	1.335	1.71×10^{-3}	-0.985	-0.986
28	729	681	5271	103931	0.326	0.068	0.403	0.72	1.74×10^{-7}	10.1	1.277	1.82×10^{-7}	-0.985	-0.983
29	296	279	14295	867024	0.188	0.103	0.217	0.70	8.50×10^{-1}	132.7	1.512	2.92×10^{-17}	-0.984	-0.986
34	243	233	1682	34774	0.443	0.110	0.554	0.99	8.35×10^{-1}	264.9	1.692	3.16×10^{-5}	-0.975	-0.967
37	243	237	731	25306	0.231	0.110	0.307	0.78	9.61×10^{-1}	21.7	1.692	2.88×10^{-5}	-0.982	-0.983
39	731	673	1138	6481	0.153	0.068	0.244	0.53	6.37×10^{-21}	10.0	1.316	6.03×10^{-21}	-0.898	-0.959
40	379	365	1234	18140	0.518	0.093	0.655	0.88	9.56×10^{-1}	2282.1	2.344	1.32×10^{-11}	-0.963	-0.996
41	234	224	1533	33011	0.373	0.114	0.471	0.81	1.06×10^{-1}	116.9	1.498	4.49×10^{-5}	-0.986	-0.983
42	290	278	982	37454	0.363	0.101	0.446	0.79	7.77×10^{-1}	225.8	1.706	2.24×10^{-4}	-0.984	-0.990
43	382	357	4419	72620	0.369	0.090	0.438	0.91	1.39×10^{-3}	10.7	1.293	1.15×10^{-3}	-0.987	-0.989
44	361	354	212	9105	0.333	0.091	0.544	0.88	4.25×10^{-1}	20.8	1.827	6.84×10^{-5}	-0.979	-0.982
46	215	204	2206	19166	0.260	0.119	0.333	0.66	2.58×10^{-1}	73.2	1.443	1.44×10^{-7}	-0.970	-0.989
52	258	248	1497	39341	0.242	0.106	0.308	0.54	3.73×10^{-1}	867.4	2.023	2.44×10^{-8}	-0.981	-0.992
54	823	796	8716	1159507	0.442	0.064	0.578	0.82	1.04×10^{-1}	9.9	1.408	1.02×10^{-1}	-0.992	-0.993
78	319	306	4432	151888	0.416	0.096	0.509	0.79	8.21×10^{-1}	17.5	1.427	2.63×10^{-2}	-0.990	-0.993
79	549	530	1038	78978	0.444	0.076	0.568	0.81	1.72×10^{-1}	12.7	1.490	7.41×10^{-3}	-0.991	-0.985
116	272	261	567	13700	0.390	0.105	0.412	0.82	8.14×10^{-1}	10.9	1.525	7.65×10^{-1}	-0.986	-0.986
123	247	234	6133	255818	0.437	0.114	0.576	0.83	9.76×10^{-1}	3999.2	2.024	3.58×10^{-13}	-0.970	-0.994
134	360	340	9839	624928	0.348	0.094	0.416	0.86	1.01×10^{-1}	47.5	1.339	5.37×10^{-10}	-0.985	-0.984
137	528	483	21151	1291274	0.374	0.079	0.479	0.81	6.66×10^{-3}	92.6	1.304	8.13×10^{-7}	-0.991	-0.997

The definition of other parameters can be found in the main text.
doi:10.1371/journal.pone.0084954.t001

correlation in the displacements is related to the scaling properties in human mobility patterns.

To prove this hypothesis, we need to test the relationships between the strength of the correlation and the form of the users' displacement distributions.

Using the method introduced above, we first calculate the Kendall's Tau of the series \vec{d}_i and \vec{d}_{i+1} for each user, as shown in Table 1. Although all τ_K of the 32 users are positive, the value varies in a wide range from 0.2 to 0.5, showing a great diversity in the correlation. More than 2/3 of all users (23/32) have the Kendall's Tau $\tau_K > 0.3$ and exhibit significant positive correlation.

The correlation coefficients R_d of each user's displacement series are also calculated. Due to the heterogeneous displacements,

the logarithm of displacement $\ln d$ is used here, so R_d is defined as:

$$R_d = \frac{\langle (\log(d_i) - \log(d_*))(\log(d_{i+1}) - \log(d_*)) \rangle}{\sigma^2}, \quad (2)$$

where d_* is the average displacement of the user and σ is the variance of the displacement series \vec{d} . The values of R_d for all 32 users are shown in Table 1. All of them are higher than 0.5, showing strong positive correlation in agreement with the above results for the method of Kendall's Tau. And also, to quantify the long-term correlations, we calculate the Hurst exponent H of the series $\log \vec{d}$ of each user using DFA and find $H > 0.5$ for all of

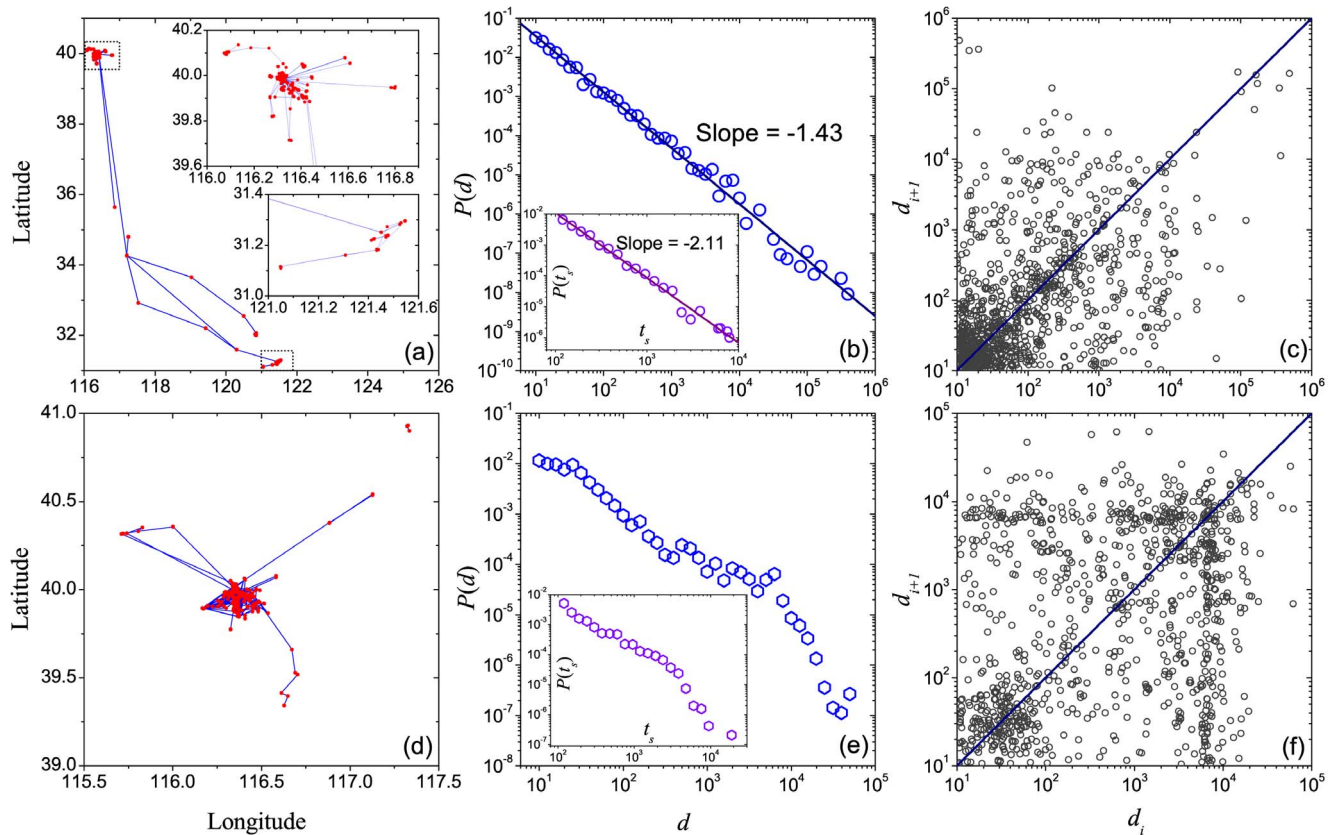


Figure 6. Mobility patterns for two typical users. Trajectories (a, d), displacement distributions $P(d)$ (b, e), staying time distributions $P(t_s)$ (insets in (b, e)) and correlation patterns (c, f) of two typical individuals (upper and lower three panels for individual No. 9 and No. 22 respectively). doi:10.1371/journal.pone.0084954.g006

them (Table 1), showing significant long-term persistence on displacements.

To check whether individual-level displacement distributions exhibit a power-law form, we plot these distributions and find that most of them seem to be power-law-like after log-binning. Here the Kolmogorov-Smirnov Test (KS Test) [34] is used to test the power-law fits of these empirical data points. After estimating and setting a lower bound x_{min} in the dataset, KS test will return confidence probability p_{KS} . Generally speaking, the bigger p_{KS} is, the better the fit is. Table 1 shows p_{KS} of the log-binning displacement distribution for each user, in which most of them have $p_{KS} > 0.1$ and have a well-fitted power-law-like section.

However, several users have very large estimated values for x_{min} in the KS test, showing that the power-law-like section only covers a small range in the tail of $P(d)$. We therefore fix x_{min} to 10 meters to test if $P(d)$ can be well fitted by a power law in all of the range. This yields the confidence probability p'_{KS} . Unfortunately, in only a few users is the requirement $p'_{KS} > 1$ fulfilled, as shown in Table 1, indicating that for most of these users a power law is observed over only a small range.

To quantify the differences between $P(d)$ and a strict power law, one can also directly linear fit the data points of $P(d)$ under a double-logarithmic coordinate system to get the Pearson correlation coefficient R'_d between the fitting curve and $P(d)$ [35]. The better fitting corresponds to smaller negative values of R'_d due to the decaying power-law function, and $R_d = -1$ relates to a strict power law. As shown in Table 1, all users' R'_d are less than -0.9 .

Now we have five quantities for each individual, τ_K , R_a and H are the ones for the correlations of user's displacements, p_{KS} and

R'_d are the ones for the quality of the power-law fitting. We plot six relationships of these quantities and respectively calculate their Kendall's Tau values, as shown in Fig. 7. Most of these correlations are significant, and in supporting of our previous guess that the scaling mobility patterns usually correspond to higher correlation of displacements. This result implies that the cascading-like processes play an important role in the emergence of the scaling properties in human movements.

However, unlike the previous findings in human communications [32,33], the long-term correlations of move-lengths look independent of the power-law exponents of $P(d)$ (The Kendall's Tau between α and H is -0.129 with 95% significance level confidence interval 0.262).

Similarly, we calculate the Pearson correlation coefficient R'_t between the staying time distribution $P(t_s)$ and power law fits for each individual, as shown in Table 1. However, R'_t does not show significant correlations with R'_d and p'_{KS} (Kendall's Taus respectively are 0.214 and -0.048 for the confidence interval 0.262), and weak negative correlations with R_a and τ_K (Kendall's Taus respectively are -0.266 and -0.262 with the confidence interval 0.262). Combining these results with the observation that the staying time does not correlate with the displacement at the population level, we infer that the effect of the dynamics on the staying time is rather unrelated to that on the displacement.

At last, we compare the empirical correlations of the displacement series with the ones that are generated by typical models. Several models that are based on the random walks on either hierarchical or self-similar organization, e.g. SLAW [19] and the hierarchical-traffic-system model (HTS model for short)

[20], can create displacement series with inherent positive correlation and long-term correlations, which are mainly caused by the cascading-like process since each movement can activate a series of movements with similar changing trends on displacements. A more detailed discussion can be found in the fourth section in *Materials and Methods*. These models partially explained the origin of the correlations in human mobility, nevertheless, the explanation is not complete. How to understand the long-term correlations is still of heightened interest in the future studies.

Discussion

By analyzing the dataset of GPS carriers, we observe the scaling temporal-spatial properties in the aggregated human movements as well as individual-level diversities. The displacement distribution at the population level is well-fitted by a power law. However, the individuals' mobility shows much diversity: some of them display common scaling properties, but others are irregular, in agreement with several recent studies [15].

Our most remarkable finding is the significant positive correlation of the series of displacements both at the population level and at the individual level, showing that the gradually changing nature is wide-spread in human mobility. We surprisingly find that the strength of the correlation for each individual is significantly related to their displacement distribution: the individuals with stronger displacement correlation have a higher probability to possess a power-law-like displacement distribution. This result is confirmed by four types of correlations (Fig. 7) and implies that the cascading-like dynamics is an important mechanism in the emergence of scaling properties of human mobility. Although the total number of samples in our analysis is not very big, this result is still highly believable, as most of the correlations/correlations well pass the test with 95% significant level and support each other.

We notice that the displacements and staying times are largely independent both at the population level and at the individual level, indicating that the mechanisms that drive the emergences of their scaling laws are also independent. This result is helpful for

the modeling, as it indicates that we can divide the empirical findings into several classes that may have similar dynamics according to their correlations, and then can be modeled independently.

Finally, the speed-displacement pattern shows the abnormal transition from a linear to a sub-linear relationship (Fig. 3), which may indicate the change of transportation from walks to automobile and the average longest walking distance in daily life. In addition, the impact on human mobility patterns due to traffic jams are observed here.

In summary, we find that the positive correlation of the series of displacements that describes the cascading-like movements, is a characteristic factor that is relevant to the underlying mechanisms of the scaling of mobility patterns from the empirical analysis. Our findings and the methods used provide some useful insights for further empirical and modeling studies of human mobility patterns.

Materials and Methods

Dataset descriptions and the judgement of effective stay positions

The data used in this study has been provided by the Microsoft Geo-life project and contains over 2 years of GPS trajectories (from April 2007 to August 2009) of 165 individuals. The dataset is available at the website: <http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/>. The GPS data was collected by different GPS handheld equipments or GPS phones. In most of them, the interval of recording time ranges from 2 to 5 seconds. The dataset includes more than 10,000 trajectories, the total recording distance is more than 1 million kilometers, and the total recording time is more than 48,000 hours. The trajectories are widely distributed in the world, covering more than 30 cities in China, and several cities in North America, Europe, South-east Asia, etc. The movements recorded by the dataset include not only trips to work or home, but also many daily-life activities, such as shopping, sightseeing, dining, hiking, and cycling, etc. The recording time for different

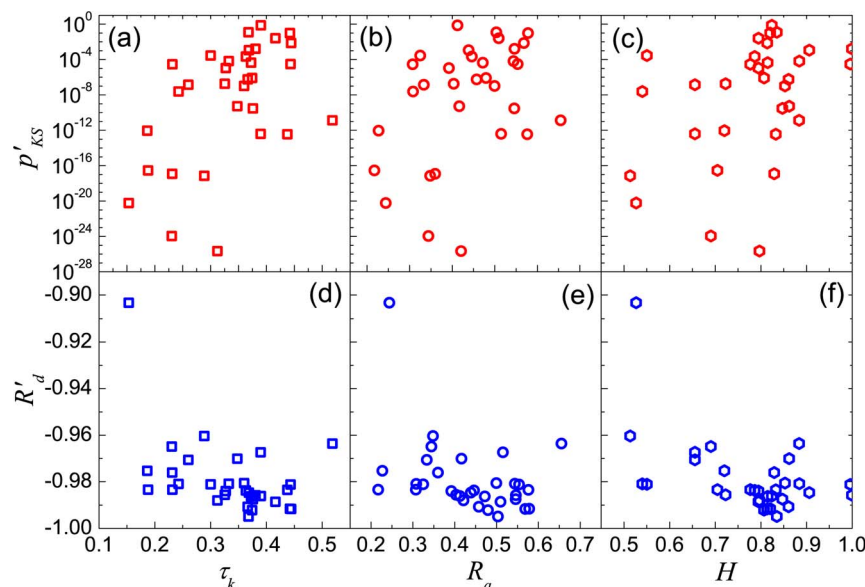


Figure 7. Patterns of six correlations. (a) p'_{KS} vs. τ_k , (b) p'_{KS} vs. R_a , (c) p'_{KS} vs. H , (d) R'_d vs. τ_k , (e) R'_d vs. R_a , (f) R'_d vs. H . Kendall's Tau of these correlations respectively are (a) 0.367, (b) 0.246, (c) 0.278, (d) -0.274, (e) -0.286, (f) -0.254, with 95% significance level confidence interval 0.262. doi:10.1371/journal.pone.0084954.g007

individuals is different, and ranges from several weeks to several years. A trajectory file consists of a sequence of the records of trajectory points, and each record provides information on the latitude, longitude, and altitude of the position of the GPS holder, and the corresponding recording time.

The dataset is composed by a series of geographic locations with corresponding time recordings ordered by the time sequence. They can not directly show the positions that users really have stayed in, so first of all we should distinguish the effective stay positions from the record. We set the resolutions for distinguishing of staying positions to 10 meters for the displacement which is the critical spatial resolution of a handheld GPS equipment, and 120 seconds for the time which is the interval of traffic signals.

Consider a trajectory labelled by $\{P_1, P_2, \dots, P_N\}$, where a continuous sub-sequence $\{P_j, \dots, P_k\}$ (where $1 \leq j \leq k \leq N$) 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 $\{T_j, \dots, T_k\}$ is larger than 120 seconds. The average position of the sub-sequence is recorded as an effective stay position, and $t_s = T_k - T_j$ is the staying time of the stay position. As illustrated in Fig. 1, the average position S1 of track points from P3 to P7 are considered as an effective stay point, as all the geographical distances from P3 to P7 are no more than 10 m and $T_7 - T_3 < 120s$. The same holds true for S2 for the track points from P10 to P15. The straight-line distance between S1 to S2 is set as the user's displacement for the movement from S1 to S2.

Most of the files in the dataset only contain the records of a few minutes or hours. Since the critical staying time in each stay position is set as 120 seconds, we usually can not obtain enough effective stay positions to achieve good patterns of user's mobility, and 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 algorithms, we distinguish the effective stay positions of each of the 100 users from the 927 files, which are used in our analysis at the population level. An example of the extraction of the effective staying positions is shown in Fig. 8. In comparison with the original trajectory (the left panel of Fig. 8), all the non-effective stay points are filtered out (the right panel of Fig. 8).

However, in our empirical analysis at the individual level, the number of effective stay positions of more than half of the 100 users is too small to extract its patterns. We thus remain with the data of 32 users with a number of effective stay positions that is larger than 200. Notice that we analyze the files of a same user one by one, and the statistical patterns of the user are aggregated from all of his/her files.

Kendall's Tau

In our empirical analysis, the displacements of the users are very heterogeneous, covering several orders of magnitude. Thus classical measurements like the Pearson coefficient are not suitable in analyzing the correlation of these displacements. We therefore use the rank-based correlation coefficient named *Kendall's Tau*. For two series $\vec{x} = \{x_1, x_2, \dots, x_m\}$ and $\vec{y} = \{y_1, y_2, \dots, y_m\}$, the Kendall's Tau is defined as [36]

$$\tau_K = \frac{2}{m(m-1)} \sum_{i < j} \text{sgn}[(x_i - x_j)(y_i - y_j)], \quad (3)$$

where $\text{sgn}(x)$ is the signum function, which equals +1 if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$. τ_K ranges from +1 (exactly the same ordering of \vec{x} and \vec{y}) to -1 (reverse ordering of \vec{x} and \vec{y}), and two

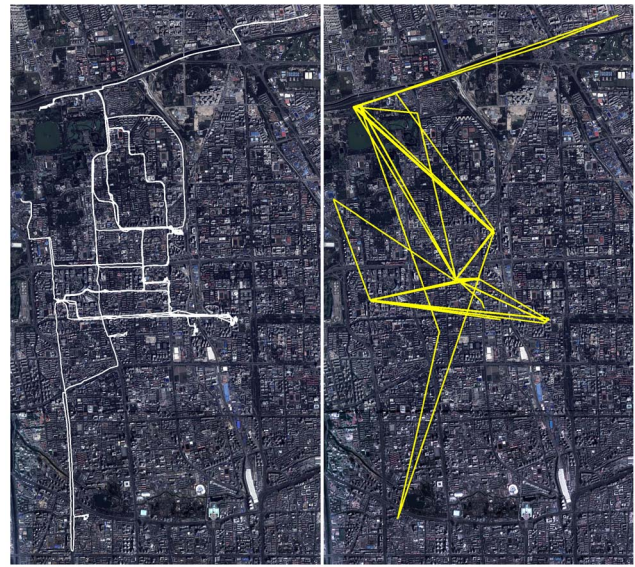


Figure 8. An example of the distinguishing of effective staying positions. The figure left shows original trajectory of one GPS carrier. The figure right shows the effective staying positions connected by lines in order, where each vertex represents an effective staying position.

doi:10.1371/journal.pone.0084954.g008

uncorrelated series have $\tau_K \approx 0$. Obviously, as τ_K is calculated based on the order of the elements in two series, the magnitudes of differences on the value of the elements do not impact τ_K .

Detrended fluctuation analysis

The detrended fluctuation analysis (DFA) is a method proposed to evaluate the self-affinity of a time series in stochastic processes. It was first developed by Peng, et al. [29], and is helpful to reveal the extent of long-term correlations of a time series. Using the DFA method, the Hurst exponent can be derived through the following procedures.

i) Considering a time series $\{d_t, t \in \mathbb{N}\}$, we first need to calculate the integration $D(t)$ of the time series,

$$D(t) = \sum_{i=1}^t (d_i - \langle d_t \rangle). \quad (4)$$

where $\langle \cdot \rangle$ means an average over all t 's.

ii) Then divide $D(t)$ into mutually disjoint boxes of size Δt .

iii) In each box, using the least square method, we can get a n -order polynomial fit $D_{fit}(t)$, which is called the n -order trend. The residual series, in which the trend has been eliminated, can be derived by applying a subtraction.

$$Y(t) = D(t) - D_{fit}(t). \quad (5)$$

iv) Calculate the mean square error of each box over the size Δt after eliminating the trend.

$$E(\Delta t)^2 = \sum_{t=1}^{\Delta t} Y(t)^2, \quad (6)$$

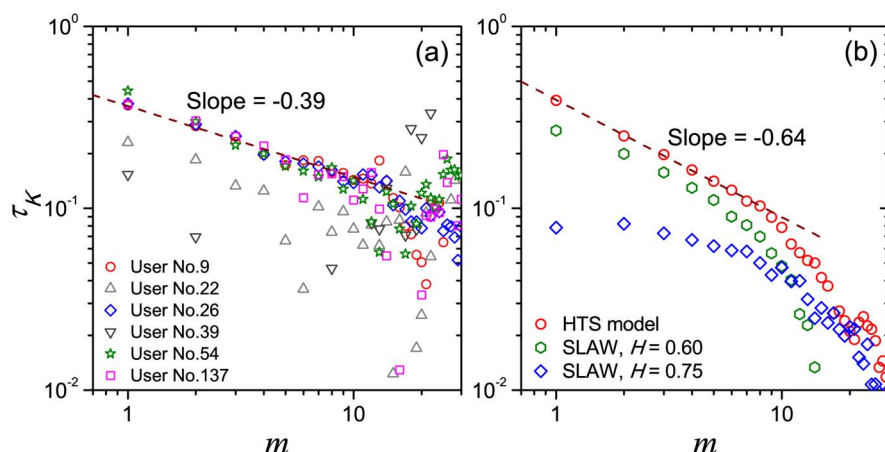


Figure 9. The decay of Kendall's Tau τ_K between \vec{d}_i and \vec{d}_{i+m} versus the interval m for several typical users (a) and modeling series (b). The wine dashed line in panel (a) is the fitting curve of user No. 26. The red, green and blue data points in panel (b) are respectively the results of HTS model, and SLAW model with different Hurst exponent H . Each data point in the modeling results is an average of ten runs of simulations. doi:10.1371/journal.pone.0084954.g009

v) Calculate the root-mean-square deviation, or say fluctuation, over different (Δt) 's.

$$F(\Delta t) = \sqrt{\frac{1}{\Delta t} E(\Delta t)^2}. \quad (7)$$

vi) If the time series $\{d_i\}$ satisfies a power-law distribution, the quality $F(\Delta t)$ will also follow a power-law increasing function,

$$F(\Delta t) \sim (\Delta t)^H, \quad (8)$$

where H is the Hurst exponent that we want to calculate. Here $H=0.5$ represents the time series are completely uncorrelated, and $0.5 < H < 1.0$ indicates the time series are of long-term correlation.

Long-term correlations in individual level and the comparison with modeling series

Our findings have shown that there exists significant positive and long-term correlations among the displacements of human movements at the aggregated level. Meanwhile, at the individual level, as shown in Fig. 9(a), the curves of Kendall's Tau τ_K between \vec{d}_i and \vec{d}_{i+m} versus the interval m for different users are diverse. The curves τ_K vs. m generally are closer to the power-law form for the users with higher $\tau_K(m=1)$ (e.g. users Nos. 9, 26, 54 and 137), and fluctuate dramatically for the users with lower $\tau_K(m=1)$ (e.g. users Nos. 22 and 39). On the whole, long-term correlations of displacement series are also widely existed at the individual level.

Researchers have proposed a series of agent-based models to explain the origin of human mobility patterns. We can compare the modeling displacement series with our empirical findings to test these agent-based models. Since the scaling property of displacement distribution is the first thing that would be taken into consideration when modeling, our testing therefore will mainly focus on the comparison of the correlations among displacement series.

One of the simplest models is the pure Lévy flight, which describes the random walk completely with power-law distributed move-lengths (displacements). Obviously, its expected correlation

between two successive move-lengths is zero. Similarly, the CTRW model [8] which considers both the random staying times and the random displacements does not introduce any correlation into displacements. Moreover, Song's model [10], which is based on two hypotheses: the exploration of new locations and the preferential return to former visited locations, can only satisfy parts of the empirical findings on correlations. The correlation between two successive displacements is positive ($\tau_K > 0.2$). However, the τ_K between \vec{d}_i and \vec{d}_{i+m} is close to zero when $m \geq 2$, which greatly deviates from the empirical findings.

Several models can create correlation patterns more similar to the empirical results. One is SLAW [19], which describes that an agent preferentially moves to the nearby point in a fractal landscape with self-similarly distributed points. Here the preferential movement is expressed by the probability $p_i = d_i^{-\alpha} / \sum d^{-\alpha}$ in which the i -th point is chosen to be the next staying point, and d is the distance between the point that the agent currently locates and the i -th point. The fractal landscape is controlled by the Hurst exponent H that defines the distribution of the distance x between any two immediate neighboring points $p(x) \sim x^{-\beta}$, and β negatively depends on H . As shown in Fig. 9(b), for small interval m , when H is lower, stronger correlations between movements appear, but the decay of τ_K is quicken to be the exponential form. Another model is the hierarchical-traffic-system model (HTS model) reported in Ref. [20]. The basic rules and results of the model can be briefly introduced as follows: Firstly we create a hierarchical geographic network on a two-dimensional plane. In the plane, K top-layer nodes, $K(M-1)$ 2nd-layer nodes, \dots , $KM^{n-2}(M-1)$ n th-layer ($n > 2$) nodes, \dots , and $KM^{N-2}(M-1)$ N th-layer nodes are randomly distributed on the plane. Each node is then connected to its nearest up-layer node. For the n th-layer node, its weight is $w_n = r^{N-n}$, where $r > 1$, represents that the upper layer nodes have more attraction for agents. After the construction of the hierarchical network, an agent randomly walks on it. The probability for an agent to move to a neighboring city is proportional to its weight. Obviously, due to the hierarchical organization, the probability that walkers directly move from a top-layer node to a bottom layer node is small. Since the long range movements only exist between two higher layer nodes, the displacements of the agents are gradual changed and has inherent positive correlation. τ_K of the displacement series is 0.39, very

close to our empirical results. In addition, τ_K of the modeling series also shows the power-law-like decay when m increases, as shown in Fig. 9(b), however, the decay is quicker than empirical results. Ref. [37] also mentioned a model based on the hierarchical structure on purpose cluster graph. Although we can not test it directly for its result is partially depending on the empirical data, it would be reasonable to take the view that its correlation patterns would be similar to the HTS model because of its hierarchical hypothesis.

References

- Belik V, Geisel T, Brockmann D (2011) Natural human mobility patterns and spatial spread of infectious diseases. *Phys. Rev. X* 1: 011001.
- Balcan D, Vespignani A (2011) Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nat. Phys.* 7: 581–586.
- Wang L, Li X, Zhang YQ, Zhang Y, Zhang K (2011) Evolution of scaling emergence in large-scale spatial epidemic spreading. *PLoS ONE* 6: e21197.
- Ni S, Weng W (2009) Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks. *Phys. Rev. E* 79: 016111.
- Horner MW, O’Kelly MES (2001) Embedding economics of scale concepts for hub networks design. *J. Transp. Geogr.* 9: 255–265.
- Clements M, Serdyukov P, de Vries AP, Reinders MJY, Personalised travel recommendation based on location co-occurrence. arXiv: 1106.5213. Available: <http://arxiv.org/abs/1106.5213>. Accessed 2011 Jun 26.
- Scellato S, Noulas A, Mascolo C (2011) Exploiting place features in link prediction on location based social networks. *KDD’11* pp. 1046–1054.
- Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439: 462–465.
- González MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. *Nature* 453: 779–782.
- Song C, Koren T, Wang P, Barabási AL (2010) Modelling the scaling properties of human mobility. *Nat. Phys.* 6: 818–823.
- Rhee I, Shin M, Hong S, Lee K, Chong S (2008) On the Levy-walk nature of human mobility. *IEEE/ACM Transactions on Networking* 19: 630–643.
- Szell M, Sinatra R, Petri G, Thurner S, Latora V (2012) Understanding mobility in a social petri dish. *Sci. Rep.* 2: 457.
- Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of Predictability in Human Mobility. *Science* 327: 1018–1021.
- Jiang B, Yin J, Zhao S (2009) Characterizing the human mobility pattern in a large street network. *Phys. Rev. E* 80: 021136.
- 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.
- 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.* 2010: P05001.
- 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.
- 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.
- Lee K, Hong S, Kim SJ, Rhee I, Chong S (2009) SLAW: A New Mobility Model for Human Walks. *IEEE Conference on Computer Communications*. pp. 855.
- 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.
- 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.
- Veneziano D, González MC, Trip Length Distribution Under Multiplicative Spatial Models of Supply and Demand: Theory and Sensitivity Analysis. arXiv: 1101.3719. Available: <http://arxiv.org/abs/1101.3719>. Accessed 2010 Nov 19.
- Simini F, González MC, Maritan A, Barabási AL (2012) A universal model for mobility and migration patterns. *Nature* 484: 96–100.
- Zheng Y, Zhang L, Xie X, Ma W-Y (2009) Mining interesting locations and travel sequences from GPS trajectories. *WWW’09* pp. 791–800.
- Zheng Y, Li Q, Chen Y, Xie X, Ma W-Y (2008) Understanding Mobility Based on GPS Data. *UbiComp’08* pp. 312–321.
- 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.
- Gallotti R, Bazzani A, Rambaldi S (2012) Towards a statistical physics of human mobility. *Int. J. Mod. Phys. C* 23: 2150061.
- Feder J (1988) *Fractals: Physics of Solids and Liquids*. New York: Plenum Press.
- Peng C-K, Buldyrev SV, Havlin S, Simons M, Stanley HE, et al. (1994) Mosaic organization of DNA nucleotides. *Phys. Rev. E* 49: 1685.
- 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.
- Malmgren RD, Stouffer DB, Motter AE, Amaral LAN (2008) A Poissonian explanation for heavy tails in e-mail communication. *Natl. Acad. Sci. USA*. 105: 18153–18158.
- 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.
- 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.
- Clauset A, Shalizi CR, Newman MEJ (2009) Power-Law Distributions in Empirical Data. *SIAM Rev.* 51: 661–703.
- Rodgers JL, Nicewander WA (1988) Thirteen ways to look at the correlation coefficient. *Am. Stat.* 42: 59–66.
- Kendall M (1938) A New Measure of Rank Correlation. *Biometrika* 30: 81–93.
- 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.* 2012: P11024.

Acknowledgments

We thank Dr. Michel Pleimling and Linjun Li for their helpful discussions on this paper.

Author Contributions

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