nature physics

SUPPLEMENTARY INFORMATION

DOI: 10.1038/NPHYS1760

Modeling The Scaling Properties of Human Mobility

Supplementary Material

Chaoming Song, Tal Koren, Pu Wang, Albert-László Barabási

DOI: 10.1038/NPHYS1760

CONTENTS

S1. DATA DESCRIPTION

The research was based on two datasets, collected for billing purposes and anonymized by a European mobile phone company.

1. Dataset D_1: This anonymized data set represents 1 year of call patterns from 3 million anonymized mobile phone users in 2008. The data contains the routing tower location each time a user initiates or receives a call or a text message. From this information, a user's trajectory may be reconstructed.

2. Dataset D_2: Some mobile services, such as pollen and traffic forecasts, rely on the approximate knowledge of a customer's location at all times. For customers voluntarily enrolled in such services the date, time and the closest tower coordinates are recorded on a regular basis, independent of phone usage. We were provided with the anonymized records of 1,000 such users, whose coordinates were recorded every hour for two weeks.

S2. ESTIMATING THE EXPONENTS α AND β

Jump size Δr and *waiting time* Δt are two quantities that play a key role in our efforts to characterize individual mobility patterns. The first measures the distances over which the users travel and the second counts the time a user spends at different locations.

In order to quantify them, we need to accurately measure the jump size distribution $P(\Delta r)$ and waiting time distribution $P(\Delta t)$ in our datasets. It was previously known that these distributions are well approximated by $P(\Delta r) \sim \Delta r^{-(1+\alpha)}$ and $P(\Delta t) \sim \Delta t^{-(1+\beta)}$, i.e. $P(\Delta r)$ and $P(\Delta t)$ are fat tailed [5, 6]. In this section, we show how to obtain the exponents α and β in our datasets.

It is important to notice that communication patterns are highly heterogeneous [1], i.e. the call patterns is bursty, which leads to biased sampling of the D_1 dataset. To correct for these sampling biases, we call the *interevent time* ΔT the time elapsed between two consecutive communication records (phone calls and SMS, sent or received) for the same user [2–4].

For modeling purposes we will also need to measure the $P(\Delta t)$ *waiting time* distribution, where Δt represents the time interval a user spends in the same location, before changing

DOI: 10.1038/NPHYS1760

its location. Below we will first investigate the interevent time ΔT characterizing communication patterns and then discuss the methodologies to extract the waiting time Δt and the appropriate mobility information from the dataset D_1 and compare the results with the measurements obtained for the hourly recorded D_2 dataset.

A. Interevent time distribution

To characterize the individual communication activity, we measured the interevent time distribution $\mathrm{P}(\Delta T)$ for different user groups selected based on their call frequency f (the average daily number of calls). The probability distribution $P(\Delta T) \sim \Delta T^{-1}$ has a fat tail, consistent with the queuing model predictions [2].

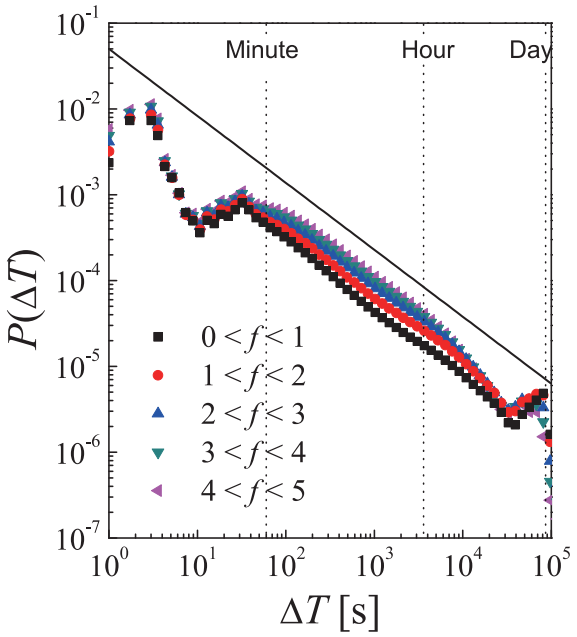

FIG. S1: Call interevent time distribution $P(\Delta T)$. Different symbols indicate the measurements done over groups of users with different call frequency f (number of calls per day). The solid line corresponds to ΔT^{-1}.

Next we will show that after proper correction the sampling heterogeneity in the dataeset D_1 does not affect our results.

B. Jump size distribution

To measure the jump size distribution $P(\Delta r)$ in the D_1 dataset, we need to correct the bias from the widely varying interevent times that characterize the calling pattern of each user. Thus, we filtered the events, accepting displacements Δr only for events that separated

by the same time interval $T \pm 0.05T$. Figure S2a shows $P(\Delta r)$ for the D_1 dataset with $T = 1$ hour, and compare the result with the measurement obtained from the unbiased D_2 dataset. The similar trends followed by the data extracted from the two datasets confirms that $P(\Delta r)$ is not driven by the statistics of the call activity, supporting the validity of our filtering methodology for the D_1 dataset. Furthermore, we find (Fig. S2) that $P(\Delta r)$ is independent of both the call frequency f and the threshold time T. The jump size distribution $P(\Delta r)$ displays a fat tailed distribution $\Delta r^{-1-\alpha}$ with $\alpha = 0.55 \pm 0.05$ and a sharp cutoff at $\Delta r \approx 100$ km. The value of α is consistent with the jump size exponent $\alpha \approx 0.59 \pm 0.02$ reported by Brockmann *et al.* [5] by measuring the jump size distribution between consecutive sightings of bank notes.

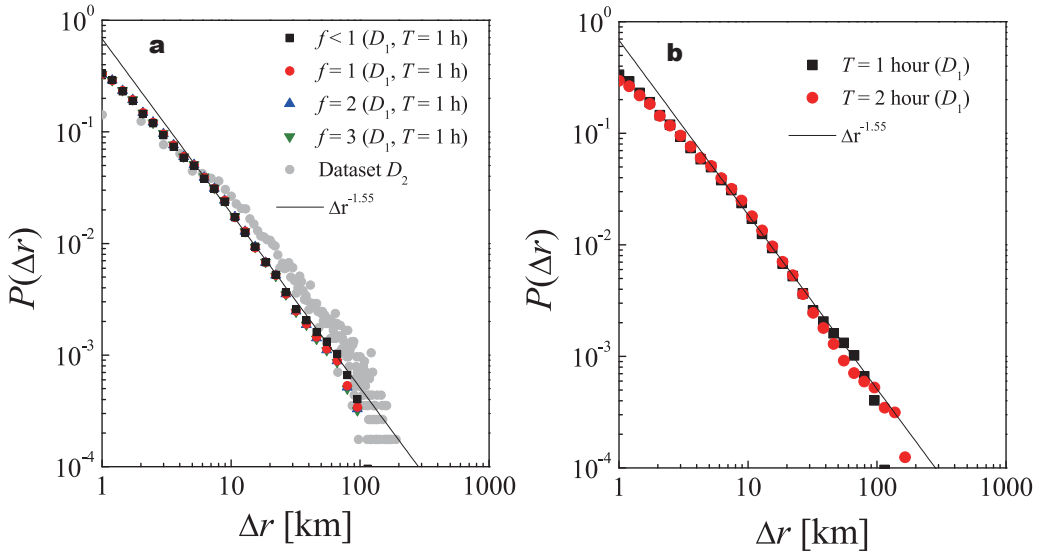

FIG. S2: The jump size distribution, $P(\Delta r)$, for (a) the D_1 dataset with $T = 1$ hour and different call frequency f, and the hourly recorded D_2 dataset and (b) D_1 dataset with $T = 1$ hour and $T = 2$ hour. The straight lines represent $\Delta r^{-1-\alpha}$ with $\alpha = 0.55$.

C. Waiting time distribution

To measure the waiting time distribution $P(\Delta t)$ in the D_1 dataset, we first discreterize the time series with a time unit T. Figure S3 offers a schematic description of the procedure, where each grey block with a number represents a tower ID from where a call was placed in that time interval and white blocks represent time interval with unknown locations (no call records during these time intervals).

DOI: 10.1038/NPHYS1760

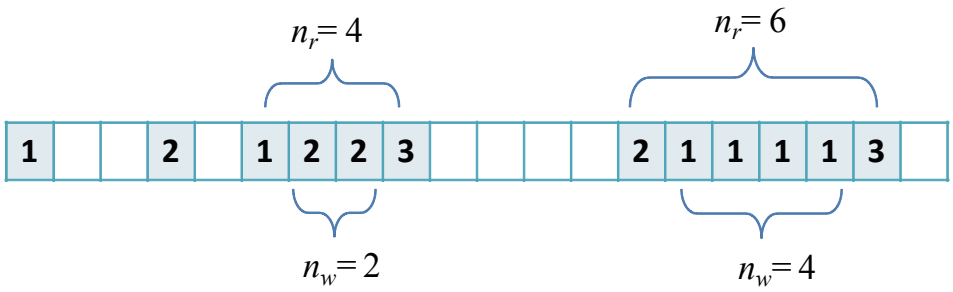

FIG. S3: A schematic discrete calling and mobility record in the D_1 dataset. Each white block indicates a T long time interval when we do not have location recorded for the user. The grey block represents time interval with a phone call and the number identifies the specific "towers" (locations) from where the call was placed.

To obtain evidence that a user did not change location for $\Delta t \approx n_w T$ waiting time, we need at least $n_r = n_w + 2$ calling records where all towers except the first and last ones are same (Fig. S3).

The probability that a user spends n_w-intervals at the same tower is

$$P_{measure}(n_w) = P_w(n_w) P_{sample}(n_r|n_w), \tag{S1}$$

where $P_w(n_w)$ is the actual probability that a user did not change location for n_w time intervals and $P_{sample}(n_r|n_w)$ is the sampling distribution with n_r known records given a n_w-long staying event. If the correlation between the phone activity and the waiting times is neglected (i.e. it is small), we have $P_{sample}(n_r|n_w) \approx P_{sample}(n_r)$, where $P_{sample}(n_r)$ is the probability that a user makes n_r calls continuously. Thus, we have

$$P_w(n_w) = P_{measure}(n_w)/P_{sample}(n_r). \tag{S2}$$

By measuring $P_{measure}(n_w)$ and $P_{sample}(n_r)$ directly from the data, we can estimate the waiting time distribution as

$$P(\Delta t = n_w T) = T^{-1} P_w(n_w). \tag{S3}$$

To test the validity of our methodology, we generated a random time series over 10 different locations with waiting time distribution $P(\Delta t) \sim \Delta t^{-(1+\beta)}$ with $\beta = 0.8$. The

phone activity pattern is simulated via the queuing model of Ref. [2] with the parameters $L = 2$ and $p = 0.8$. As we show in Fig. S4a, our methodology captures the proper waiting time distribution.

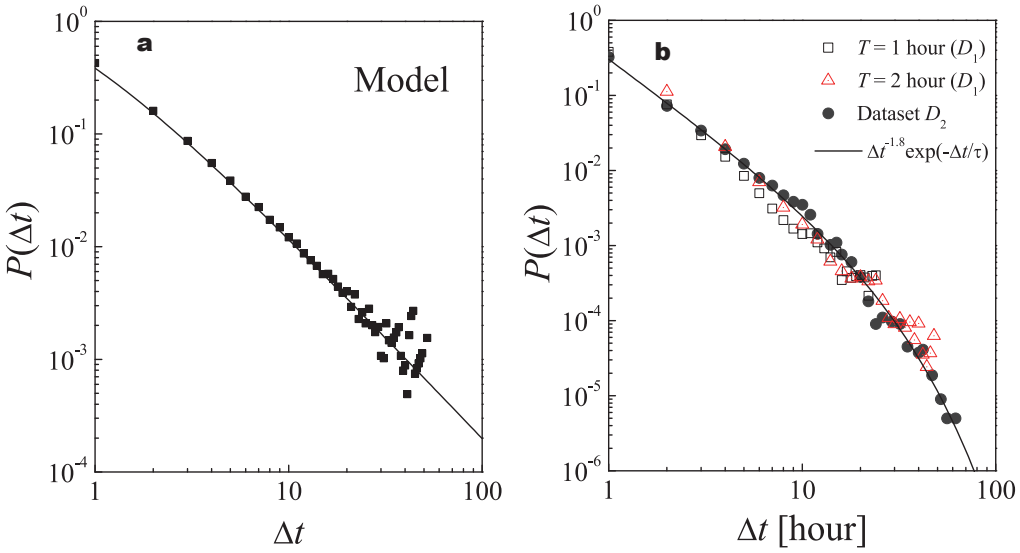

FIG. S4: The waiting time distribution, $P(\Delta t)$, for (a) a model system with biased sampling generated by the queuing model of Ref. [2], where the straight line presents the input distribution, and (b) the D_1 dataset with $T = 1$ hour and $T = 2$ hour, in comparison with the hourly recorded D_2 dataset. The solid line represents $\Delta t^{-1-\beta}\exp(-\Delta t/\tau)$ with $\beta = 0.8$ and $\tau \approx 17$ hours.

We applied our methodology to the D_1 dataset for $T = 1$ hour and $T = 2$ hours. We compared the reconstructed waiting time distribution in D_1 dataset with the unbiased D_2 dataset, finding good agreement (Fig. S4b). The $P(\Delta t)$ is best fitted by $\Delta t^{-1-\beta}\exp(-\Delta t/\tau)$ with $\beta = 0.8 \pm 0.1$ and cutoff time $\tau \approx 17$ hours. Note, unlike the jump size exponent α, the waiting time exponent β is a bit greater than the one measured from bank notes movements [5], indicating that the waiting time of bank notes is more heterogenous than the waiting time of individuals. Indeed, a bank note jumps only when the owner travels, thus we expect the exponent α to be the same for both individual mobility and the spreads of bank notes. On the other hand, bank notes can stay in the same locations much longer than individuals because they could be simply placed in a drawer or car, which offers extra time heterogeneity for the bank note mobility pattern.

DOI: 10.1038/NPHYS1760

S3. BRIEF REVIEW OF THE RANDOM WALK MODELS PERTINENT TO HUMAN MOBILITY

A. Lévy flight

For a Lévy flight it is assumed that a particle or an individual moves in randomly chosen directions, where the step sizes Δr are independent with each other, following a heavy tailed distribution

$$P(\Delta r) \simeq \frac{1}{\Delta r^{\alpha+1}}, \quad 0 < \alpha < 2. \tag{S4}$$

B. Continuous time random walk (CTRW)

In general, the jump sizes Δr and the waiting times Δt between consecutive jumps can be chosen from a well behaved probability density function, $\psi(\Delta r, \Delta t)$.

In the simplest case the jumps sizes and waiting times are assumed to be uncorrelated,

$$\psi(\Delta r, \Delta t) = P(\Delta r)P(\Delta t), \tag{S5}$$

where $P(\Delta r)$ follows Eq. (S4) and

$$P(\Delta t) \simeq \frac{1}{\Delta t^{\beta+1}}, \quad 0 < \beta < 1 \tag{S6}$$

is also fat tailed. That is, it is assumed that the user moves in a series of successive step, where both the step lengths and times between steps (i.e waiting times) are taken from heavy tailed distribution and are independent of each other.

S4. ANALYTICAL DERIVATIONS

A. Scaling relation between number of jumps n and the travel time/distance

We denote the Laplace transform of $P(\Delta t)$ with $\psi(u)$. For a power law waiting time distribution

$$\psi(u) \sim 1 - (Au)^{\beta}, \quad 0 < \beta < 1, \quad u \to 0 \tag{S7}$$

where A is normalization constant. The probability $\chi_n(t)$ that a user has taken exactly n steps in time t is a multiple convolution over $\psi(t)$ expressed via the relation between their Laplace transforms [7],

$$\mathcal{L}\chi_n(t) \equiv \chi_n(u) = \big[\psi(u)\big]^n \big[1-\psi(u)\big]/u. \tag{S8}$$

$\chi_n(t)$ introduces the relation between number of steps and time,

$$\big\langle n\big\rangle(u) = \sum_{n=0}^{\infty} n\chi_n(u), \tag{S9}$$

which implies that

$$\big\langle n\big\rangle(u) = \sum_{n=0}^{\infty} n\big[\psi(u)\big]^n \big[1-\psi(u)\big]/u. \tag{S10}$$

Clearly,

$$\big\langle n\big\rangle(u) = \frac{\big[1-\psi(u)\big]}{u} \sum_{n=0}^{\infty} n\big[\psi(u)\big]^n, \tag{S11}$$

which gives

$$\big\langle n\big\rangle(u) = \frac{\big[1-\psi(u)\big]}{u}\psi(u) \sum_{n=0}^{\infty} n\big[\psi(u)\big]^{n-1}, \tag{S12}$$

and

$$\big\langle n\big\rangle(u) = \frac{\big[1-\psi(u)\big]}{u}\psi(u)\frac{\partial}{\partial\psi(u)} \sum_{n=0}^{\infty} \big[\psi(u)\big]^n, \tag{S13}$$

$$\big\langle n\big\rangle(u) = \frac{\big[1-\psi(u)\big]}{u}\frac{\psi(u)}{(1-\psi(u))^2}. \tag{S14}$$

Therefore, we obtain

$$\big\langle n\big\rangle(u) = \frac{\psi(u)}{u(1-\psi(u))}. \tag{S15}$$

Substituting the Laplace transform

DOI: 10.1038/NPHYS1760

$$\langle n \rangle(u) = \frac{1-(Au)^{\beta}}{u(1-[1-(Au)^{\beta}])}, \tag{S16}$$

and taking $u \to 0$ gives

$$\langle n \rangle(u) \sim \frac{1}{u^{\beta+1}}. \tag{S17}$$

Finally,

$$\langle n \rangle(t) \equiv n \sim t^{\beta} \tag{S18}$$

or

$$t \sim n^{1/\beta}. \tag{S19}$$

Similarity we can derive the scaling relationship between travel distance r and number of jumps n as

$$r \sim n^{1/\alpha}. \tag{S20}$$

B. The the number of distinct sites visited (S)

In this the section we derive the asymptotic scaling law of the the mean number of distinct sites visited S vs. time t.

Based on Eq. (3), the average changes of S at the n-th steps is proportional to $S^{-\gamma}$, leading under continuous approximation to the differential equation

$$\frac{dS}{dn} = p_{\text{new}} = \rho S^{-\gamma}. \tag{S21}$$

Its solution is simply

$$S = (1+\gamma)(\rho n)^{1/(1+\gamma)}, \tag{S22}$$

which offers the dependence of S on the total number of steps n.

By combining Eq. (S22) with the scaling relationship between the time t and step n (derived in Eq. (S19)), we find

$$S(t) \sim t^{\beta/(1+\gamma)}. \tag{S23}$$

C. Visitation frequency and Zipf's law

For a trajectory with n jumps and S distinct locations, we rank all visited locations $L_1, L_2, \ldots, L_S$ in the order of discovery time (i.e. the time of the first visit). It is easy to see from the definition that

$$S(n_i) = i, \tag{S24}$$

where n_i is the jumps number at which location i was discovered.

The visitation frequency f_i of the i-th location is given by

$$f_i = \frac{m_i}{\sum_{i=1}^{S} m_i}, \tag{S25}$$

where m_i is the number of visits to the i-th location. Thanks to the preferential return (Eq. (4)), the sooner a new location is discovered the more visitations it has in the long run. Thus, the order $L_1, L_2, \ldots, L_S$ is same as the frequency-ranking, or

$$k(L_i) = i, \tag{S26}$$

where $k(L)$ denotes the frequency-based rank of location L. Combining Eq. (S24) and Eq. (S26), we find

$$k(L_i) = S(n_i) \sim n_i^{1/(1+\gamma)}, \tag{S27}$$

or

$$n_i \sim k(L_i)^{1+\gamma}. \tag{S28}$$

To derive the relationship between k and f, we notice that the number of visitations m_i at the i-th location follows

$$\frac{dm_i}{dn} = \Pi(m_i)(1 - p_{\text{new}}), \tag{S29}$$

where p_{new} and $\Pi(m_i)$ are given by Eqs. (3-4) respectively. It is also useful to include the initial condition

$$m_i(n_i) = 1, \tag{S30}$$

which indicates that the number of visits is one for a location when it was first visited.

We next solve the Eq. (S29) with the initial condition (S30) for two different classes of γ values.

- Case 1: $\gamma > 0$

DOI: 10.1038/NPHYS1760

In this case $\Pi(m_i) = 1 - \rho S^{-\gamma}$ approaches 1 for large S (or equivalently large t, as S increases with t, see Eq. (S23)).

Thus the asymptotic form of Eq. (S29) reduces to

$$\frac{dm_i}{dn} = \Pi(m_i) = \frac{m_i}{\sum_{i=1}^{S} m_i}. \tag{S31}$$

We find that

$$\sum_{i=1}^{S} \frac{dm_i}{dn} = \sum_{i=1}^{S} \left(\frac{m_i}{\sum_{i=1}^{S} m_i} \right) = 1 \tag{S32}$$

or simply

$$\sum_{i=1}^{S} m_i = n, \tag{S33}$$

which represents a "mass conservation", i.e. the total number of visits should agree with the total number of steps at each location.

By substituting Eq. (S33) into Eq. (S31), we obtain

$$\frac{dm_i}{dn} = \frac{m_i}{n}. \tag{S34}$$

The solution of above differential equation is simply $m_i = C_i n$, where C_i is an arbitrary constant. The initial condition Eq. (S30) requires $m_i(n_i) = C_i n_i = 1$, or $C_i = 1/n_i$. Thus, we find

$$m_i = n/n_i \tag{S35}$$

By substituting Eq. (S28) into Eq. (S35), we find

$$f_i \sim n_i^{-1} \sim k^{-(1+\gamma)}. \tag{S36}$$

- Case 2: $\gamma = 0$

  In this case $\Pi(m_i) = 1 - \rho$ thus Eq. (S29) reduces to

$$\frac{dm_i}{dn} = (1-\rho)\Pi(m_i) = (1-\rho)\frac{m_i}{\sum_{i=1}^{S} m_i}. \tag{S37}$$

Because of $S = \rho n$ (Eq. S22), we have

$$d \ln m_i = (1-\rho)\frac{dn}{\sum_{i=1}^{\rho n} m_i}, \tag{S38}$$

thus $m_i = C_i g(n)$ where $g(n)$ is function of n and $C_i = 1/g(n_i)$ because of $m_i(n_i) = 1$ (Eq. S30), which gives

$$m_i = \frac{g(n)}{g(n_i)}. \tag{S39}$$

Moreover, Eq. (S24) suggests $S(n_i) = \rho n_i = i$, or $n_i = i/\rho$. Together with Eq. (S39), we obtain

$$m_i = \frac{g(n)}{g(i/\rho)}. \tag{S40}$$

Applying a continuous approximation we obtain

$$\sum_{i=1}^{S} m_i = \sum_{i=1}^{\rho n} \frac{g(n)}{g(i/\rho)} \approx g(n) \int_1^{\rho n} \frac{1}{g(i/\rho)} di = g(n)\rho \int_{1/\rho}^{n} \frac{1}{g(x)} dx. \tag{S41}$$

After substitute Eq. (S41) into Eq. (S37) we obtain

$$\frac{dg(n)}{dn} = \frac{1-\rho}{\rho}\left[\int_{1/\rho}^{n} \frac{1}{g(x)} dx\right]^{-1}. \tag{S42}$$

The equation above has asymptotic solution $g(n) \sim n^{1-\rho}$. Thus,

$$m_i = (n/n_i)^{1-\rho}. \tag{S43}$$

By substitute Eq. (S43) into Eq. (S28), finding

$$f_i \sim n_i^{-(1-\rho)} \sim k^{-(1-\rho)}. \tag{S44}$$

Together with (S36) and (S44), we obtain

$$f_k \sim k^{-\zeta}, \tag{S45}$$

DOI: 10.1038/NPHYS1760

where

$$\zeta = \begin{cases} 1+\gamma, & \gamma > 0 \\ 1-\rho, & \gamma = 0 \end{cases} \tag{S46}$$

D. Mean Square Displacement (MSD)

To estimate the scaling relationship between time and distance, we first introduce $P(l|i)$, the probability that the i-th location is l steps away from the starting point of the individual trajectory.

For a trajectory with n jumps and S distinct locations, the walker can jump to the latest S-th location through any of the previous locations, suggesting the following recurrence equation

$$P(l|S) = \sum_{i=l-1}^{S-1} P(l-1|i) f_S(i), \tag{S47}$$

with the initial condition

$$P(1|1) = 1, \tag{S48}$$

where $1 \leq l \leq S$ and f_i^S is the probability of visitation of the i-th location given that the total number of locations visited previously is $S-1$.

Furthermore, from Eq. (S47), it is easy to see that $P(l|S)$ satisfies the following conservation law

$$\sum_{l=1}^{S} P(l|S) = 1. \tag{S49}$$

As we showed in the previous section, the visitation frequency follows a power law, which implies that

$$f_i^S \approx \frac{\zeta - 1}{1 - S^{1-\zeta}} i^{-\zeta} \tag{S50}$$

for $\zeta > 1$ or $\zeta < 1$. If $\zeta = 1$ we have $f_i^S = i^{-1}/\ln S$, and for large enough n, Eq. (S47) can be approximated as,

$$P'(l|x) = \int_l^S P(l-1|j) f_i^S di = \frac{1}{x} \int_{a_l}^{x} P'(l-1|x')dx', \tag{S51}$$

where $x \equiv (1 - S^{1-\zeta})/(\zeta - 1) \geq (1 - l^{1-\zeta})/(\zeta - 1) \equiv a_l \geq 0$ and $P'(l|x) = P(l|S)$. We set

$p_l(y) = l!xP'(l|x)$ with $y = \ln x \geq \ln a_l \equiv b_l$, thus we have

$$p_l(y) = l \int_{b_l}^{y} p_{l-1}(y')dy', \tag{S52}$$

or

$$\frac{dp_l(y)}{dy} = lp_{l-1}(y), \tag{S53}$$

with certain boundary condition, where

$$P(l|S) = \exp(-y)\frac{p_l(y)}{l!}, \tag{S54}$$

and

$$y = \ln\left(\frac{1 - S^{1-\zeta}}{\zeta - 1}\right). \tag{S55}$$

Equation (S53) indicates that $p_l(y)$ is an Appell sequence [8], characterized by the generating function

$$g(z)e^{zy} = \sum_{l=0}^{\infty} \frac{p_l(y)}{l!} z^l, \tag{S56}$$

where $g(z)$ are power series in z.

We can rewrite Eq. (S49) for sufficiently large S as

$$1 = \sum_{l=1}^{S} P(l|S) \approx e^{-y} \sum_{l=0}^{\infty} \frac{p_l(y)}{l!} = g(1), \tag{S57}$$

which indicates $g(1) = 1$.

To calculate $\langle l^a \rangle (S)$ with arbitrary exponent $a > 0$, we notice that,

$$\langle l^a \rangle (S) = \sum_{l=1}^{S} l^a P(l|S) \approx e^{-y} \sum_{l=0}^{\infty} l^a \frac{p_l(y)}{l!} \approx e^{-y} \frac{d^a}{dz^a} \sum_{l=0}^{\infty} \frac{p_l(y)}{l!} z^l \Big|_{z=1} = e^{-y} \frac{d^a}{dz^a}[g(z)e^{zy}]\Big|_{z=1} \approx y^a + O(y^{a-1}). \tag{S58}$$

Applying the scaling relation in Eq. (S20) between Δr and l, we obtain

$$\left\langle \Delta r^2 \right\rangle^{\alpha/2} \sim \ln\left(\frac{1 - S^{1-\zeta}}{\zeta - 1}\right) + O(1). \tag{S59}$$

DOI: 10.1038/NPHYS1760

Using the power law relation between S and t, we finally have

$$\left\langle \Delta r^2 \right\rangle (t) \sim \begin{cases} (\ln t)^{2/\alpha} & \zeta < 1 \\ (\ln \ln t)^{2/\alpha} & \zeta = 1 \\ const & \zeta > 0 \end{cases} . \tag{S60}$$

S5. THE TAIL OF $P(r_g)$

For a trajectory $\{\vec{r}_i\}$ with $1 \leq i \leq n$ where n is the total number of steps and the vector $\vec{r}_i \in R^2$ represents the position at step i, we define the center of mass $\vec{r}_{CM} \equiv \frac{1}{n}\sum_{i=1}^{n} \vec{r}_i$ and the radius of gyration as

$$r_g \equiv \sqrt{\frac{1}{n}\sum_{i=1}^{n} |\vec{r}_i - r_{CM}|^2} \tag{S61}$$

It has been proven that [9]

$$r_g = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{i-1} |\vec{r}_i - \vec{r}_j|^2}. \tag{S62}$$

Define functions $P(x)$ and $f(x)$ are *asymptotically equivalent* if there exist constants $x_0 > 0$ and $c_2 \geq c_1 > 0$ that for every $x > x_0$, $P(x)$ is bounded as $c_1 f(x) \leq P(x) \leq c_2 f(x)$, denoting as $P(x) \sim f(x)$,

In this section, we will show if the jump size distribution $P(\Delta r) \sim \Delta r^{-(1+\alpha)}$ is fat tailed, then $P(r_g) \sim r_g^{-(1+\alpha)}$ with the same α exponent from $P(\Delta r)$.

We first prove following Lemma,

Lemma S5.1 *For two random variables x and y where y satisfies $c_1 x \leq y \leq c_2 x$ with $0 < c_1 \leq c_2$. If the probability density function $P(x) \sim x^{-(1+\alpha)}$ ($\alpha > 0$) then $P(y) \sim y^{-(1+\alpha)}$.*

Proof $P(x) \sim x^{-(1+\alpha)}$ thus $\exists x_0 > 0$ and $0 < d_1 \leq d_2$, $\forall x > x_0$, $d_1 x^{-(1+\alpha)} \leq P(x) < d_2 x^{-(1+\alpha)}$, or $(d_1/\alpha)x^{-\alpha} \leq C(x) \leq (d_2/\alpha)x^{-\alpha}$, where $C(x) \equiv \int_x^\infty P(x)dx$ is the cumulative distribution. Thus, $\forall x > x_0/c_1 \geq x_0/c_2$, we have $(d_1/\alpha)(c_2 x)^{-\alpha} \leq C(c_2 x) \leq C(y) \leq C(c_1 x) \leq (d_2/\alpha)(c_1 x)^{-\alpha}$, or, $(d_1/\alpha)(c_2 y/c_1)^{-\alpha} \leq C(y) \leq (d_2/\alpha)(c_1 y/c_2)^{-\alpha}$. Thus $\forall y > y_0 \equiv x_0(c_2/c_1)$, we have $[d_1(c_2/c_1)^{-\alpha}]\, y^{-(1+\alpha)} \leq P(y) \leq [d_2(c_2/c_1)^{\alpha}]\, y^{-(1+\alpha)}$.

The Lemma S5.1 basically claims that a random variable y follows $P(y) \sim y^{-(1+\alpha)}$ $(0 < \alpha < 2)$ if it is bounded by another random variable x satisfying $P(x) \sim x^{-(1+\alpha)}$.

We next show

Lemma S5.2 *if there are n independent random variables $x_1, \ldots, x_n$ following same probability density function $P(x) \sim x^{-(1+\alpha)}$ $(0 < \alpha < 2)$, then $P(x_M) \sim x^{-(1+\alpha)}$, where x_M is the maximum one among them.*

To prove the this, we notice that the sum $x_S \equiv \sum_{i=1}^{n} x_i$ follows $P(x_S) \sim x_S^{-(1+\alpha)}$ (see [10] and Section S3A). Since $x_S/n \leq x_M \leq x_S$, the result follows from Lemma S5.1.

Finally, we switches to the mobility models. Assume an individual had n number of discrete moves. Since in our model the user could either explore a new location or return to previous ones, we denote the total number of exploring steps with l and the rest $n - l$ steps associate with the returning process. In the Lévy flight or CTRW models, l is simply equal to n.

The exploring process generates a series of jumps $\{\Delta r_1, \ldots, \Delta r_i, \ldots, \Delta r_l\}$, which follow $P(\Delta r_i) \sim \Delta r_i^{-(1+\alpha)}$ $(0 < \alpha < 2)$ and are independent with other each. Let's define $\Delta r_M = \max\{\Delta r_i\}$, the largest jumps among l-jumps. Lemma S5.2 tells us $P(\Delta r_M) \sim \Delta r_M^{-(1+\alpha)}$.

From Eq. (S62), we have

$$r_g = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{i-1}|\vec{r}_i - \vec{r}_j|^2} \geq \frac{1}{n}\Delta r_M. \tag{S63}$$

On the other hand, since for any i and j, $|\vec{r}_i - \vec{r}_j| \leq n\Delta r_M$, it is easy to see that $r_g \leq \sqrt{(n-1)/2}\Delta r_M$. Thus

$$\frac{1}{n}\Delta r_M \leq r_g \leq \sqrt{(n-1)/2}\Delta r_M. \tag{S64}$$

By applying Lemma S5.1, we find that

$$P(r_g) \sim r_g^{-(1+\alpha)}. \tag{S65}$$

It is worth noticing that this proof and Eq. S65 also applies to other models with fat-tailed jump size distribution (Lévy flight , CTRW, etc.).

DOI: 10.1038/NPHYS1760

S6. FINITE SIZE EFFECTS

As we mentioned in the main text, the saturation of the MSD could also be rooted in the finite size of the country (the finite number of towers a user can visit). To investigate whether the ultra-slow diffusive process of human mobility is mainly rooted in finite size effects, we simulate both Brownian motion and truncated LF on a finite lattice with 10,000 sites, which is approximately the order of magnitude of the total number of cell towers in our system. The jump size distribution of truncated LF has $\alpha = 0.55 \pm 0.05$ and a cutoff at $\Delta r \sim 10\%$ of the maximum size, consistent with real measurements. In Fig. S5, we plot MSD vs. S (the number of distinct locations visited) for the real data, Brownian motion and truncated LF, respectively. In order to compare with the real data directly, we only focus on the range where S is less than 100. We find that the Brownian motion doesn't saturate before $S < 100$ and more interestingly, the truncated LF shows a clear saturation for $S > 50$. However, both models predict a power law relationship between MSD and S before MSD reaches saturation, which is inconsistent with the real data, which shows a double-logarithmic slow-down. The prediction of the IM model does fit, however, the real data very well. Thus it is important to understand how saturation is reached, as it allows us to distinguish different models. Incorporating the finiteness of the area in the IM model might speed up the saturation further, which unfortunately we cannot tell from the current 1-year-long dataset.

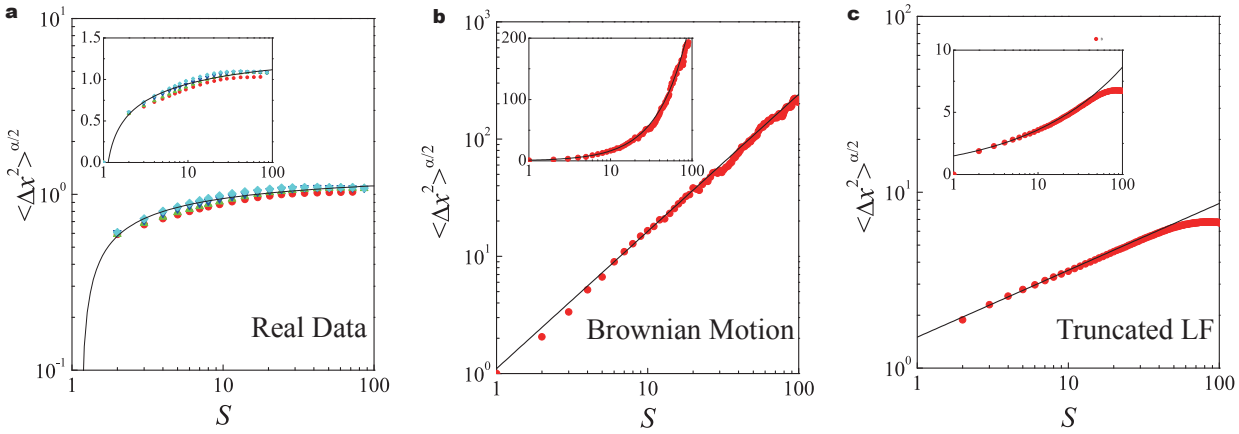

FIG. S5: The MSD vs. S for (a) real data with different r_g, where the black curve represents the prediction of IM model, (b) Brownian motion and (c) truncated LF, where $P(\Delta r)$ follows $\Delta r^{-(1+\alpha)}$ with $\alpha = 0.55$. Both (b) and (c) are confined within a 100×100 lattice. The cutoff of truncated LF is equal to 10, same as 10% of lattice size.

S7. TEMPORAL CORRELATIONS

To explore the short range temporal order, we used the D_2 dataset and measured the jump probability $\Pi_i(t)$ to location i at time t. We ranked the locations based on the visitation frequencies and found that the ratio Π_i/f_i for different time, as shown in Fig. S6. Indeed, an individual tends to return his/her mostly visited locations during night and explore low-ranked locations during daylight and weekends. However, the derivation is relatively small. If we coarse-grained the time-dependent return process over short time period, say, several days, in Fig. 3b we find $\langle \Pi_i(t) \rangle = f_i$, covering the preferential return (4), where $\langle\rangle$ is averaged over a two weeks time period. This means that the short time correlation is smoothed and the remaining long-time correlation is mainly rooted in the preferential return process.

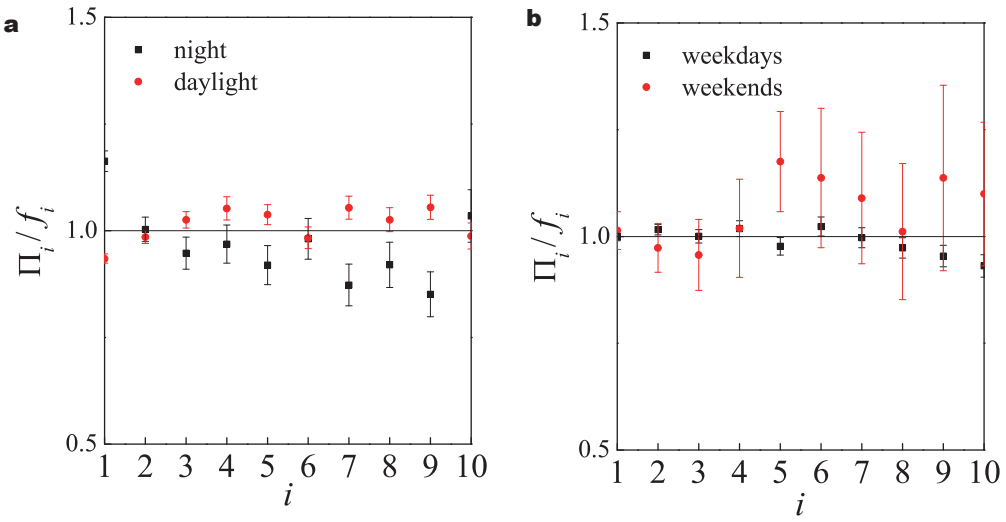

FIG. S6: Preferential return process explored by plotting Π_i/f_i vs. the frequency-ranking of site i for (a) the daylight and night, and (b) weekdays and weekends.

[1] Onnela, J.-P., Saramaki, J., Hyvonen, J., Szabo, G., Lazer, D., Kaski, K., Kertesz, K. and Barabasi A.L. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences of the United States of America* 104, 7332-7336 (2007).

[2] Barabasi, A.L. The origin of bursts and heavy tails in human dynamics. *Nature* 435, 207-211 (2005).

DOI: 10.1038/NPHYS1760

[3] Vazquez, A., Oliveira, J.G., Dezso, Z., Goh,K.-I, Kondor, I. and Barabasi, A.-L. Modeling bursts and heavy tails in human dynamics. *Phys. Rev. E* 73, 036127 (2006). 17

[4] Hidalgo, C. and Barabasi, A.L. Inter-event time of uncorrelated and seasonal systems. *Physica A* 369, 877-883 (2007).

[5] D. Brockmann, L. Hunagel and T. Geisel. The scaling laws of human travel. *Nature* 439, 462-465 (2006).

[6] M. C. González, C. A. Hidalgo and A.-L. Barabási. Understanding individual human mobility patterns. *Nature* 453, 779-782 (2008).

[7] S.B. Yuste, J. Klafter, and K. Lindenberg. Number of distinct sites visited by a subdiffusive random walk. *Phys. Rev. E* 77, 032101 (2008).

[8] Steven Roman and Gian-Carlo Rota, "The Umbral Calculus", *Advances in Mathematics* 27, 95 - 188, (1978).

[9] Hughes, B.D. "Random Walks and Random Environments". (Oxford University Press, USA, 1995).

[10] Feller, W. An Introduction to Probability Theory and Its Applications, Volume 2. (Wiley, 1971)