nature physics

ARTICLES

PUBLISHED ONLINE: 12 SEPTEMBER 2010 | DOI: 10.1038/NPHYS1760

Modelling the scaling properties of human mobility

Chaoming Song[1,2†], Tal Koren[1,2†], Pu Wang[1,2†] and Albert-László Barabási[1,2,3★]

Individual human trajectories are characterized by fat-tailed distributions of jump sizes and waiting times, suggesting the relevance of continuous-time random-walk (CTRW) models for human mobility. However, human traces are barely random. Given the importance of human mobility, from epidemic modelling to traffic prediction and urban planning, we need quantitative models that can account for the statistical characteristics of individual human trajectories. Here we use empirical data on human mobility, captured by mobile-phone traces, to show that the predictions of the CTRW models are in systematic conflict with the empirical results. We introduce two principles that govern human trajectories, allowing us to build a statistically self-consistent microscopic model for individual human mobility. The model accounts for the empirically observed scaling laws, but also allows us to analytically predict most of the pertinent scaling exponents.

Uncovering the statistical patterns that characterize the trajectories humans follow during their daily activity is not only a major intellectual challenge, but also of importance for public health[1–5], city planning[6–8], traffic engineering[9,10] and economic forecasting[11]. For example, quantifiable models of human mobility are indispensable for predicting the spread of biological pathogens[1–5] or mobile-phone viruses[12].

In the past few years the availability of mobile-phone records, global-positioning-system data and other data sets capturing aspects of human mobility have given a new empirically driven momentum to the subject. Although the available data sets significantly differ in their reach and resolution, the results seem to agree on a number of quantitative characteristics of human mobility. For example, both dollar-bill tracking[13] and mobile-phone data[14] indicate that the aggregated jump-size (Δr) and waiting-time (Δt) distributions characterizing human trajectories are fat-tailed, that is, $P(\Delta r) \sim |\Delta r|^{-1-\alpha}$ and $P(\Delta t) \sim |\Delta t|^{-1-\beta}$ with $0 < \alpha \leq 2$ and $0 < \beta \leq 1$, where Δr denotes the distances covered by an individual between consecutive sightings and Δt is the time spent by an individual at the same location. These findings suggest that human trajectories are best described as Lévy flights or CTRWs, a much studied modelling framework in the random-walk community[13,15–20].

The purpose of the present Article is to show, using a series of direct measurements, that human trajectories do follow several highly reproducible scaling laws. Yet, many of these laws are either not explained by the CTRW model, or they are in direct contradiction with the CTRW predictions, indicating the lack of a modelling framework capable of capturing the basic features of human mobility. To explain the origin of the observed scaling laws, we introduce two principles that govern human mobility, serving as the starting point of a statistically acceptable microscopic model for individual human motion. We show that the model can account for the empirically observed scaling laws and allows us to analytically predict the pertinent scaling exponents.

Scaling anomalies

We used two data sets to uncover the patterns characterizing individual mobility. The first data set (D_1) captures for a one-year period the time-resolved trajectories of three million anonymized mobile-phone users. Each time a user initiated or received a phone call the tower that routed the communication was recorded for billing purposes. Thus, the user's location was recorded with the resolution that is determined by the local tower density. The reception area of a tower varies from as little as a few hundred metres in metropolitan areas to a few kilometres in rural regions, controlling our uncertainty about the user's precise location. However, as here we focus on the asymptotic scaling properties of human trajectories, these short distance uncertainties are not expected to affect our results (see Supplementary Section S2). The second data set (D_2) uses the anonymized location record of 1,000 users who signed up for a location-based service; thus, their location was recorded every hour for a two-week period. As a first step we calculated the displacement at hourly intervals, finding $P(\Delta r) \sim |\Delta r|^{-1-\alpha}$ with $\alpha = 0.55 \pm 0.05$ and an expected cutoff at $\Delta r \sim 100$ km, corresponding to the distance people could reasonably cover in an hour. We used the D_2 data set to measure $P(\Delta t)$, where the waiting time Δt is defined as the time a user spent at one location. We find that $P(\Delta t)$ follows $P(\Delta t) \sim |\Delta t|^{-1-\beta}$ with $\beta = 0.8 \pm 0.1$ and a cutoff of $\Delta t = 17$ h, probably capturing the typical awake period of an individual. Taken together, the fat-tailed natures of $P(\Delta r)$ and $P(\Delta t)$ suggest that humans follow a CTRW during their daily mobility. Next we discuss three empirical observations that indicate that human trajectories follow reproducible scaling laws, but also illustrate the shortcoming of the CTRW model in capturing the observed scaling properties.

(A) The number of distinct locations $S(t)$ visited by a randomly moving object is expected to follow[21–23]

$$S(t) \sim t^{\mu} \tag{1}$$

where $\mu = 1$ for Lévy flights[24] and $\mu = \beta$ for CTRW. Interestingly, our measurements indicate that for humans $\mu = 0.6 \pm 0.02$ (see Fig. 1a), smaller than the CTRW prediction of $\beta = 0.8 \pm 0.1$. The fact that $\mu < 1$ indicates a slow-down at large timescales, a decreasing tendency of the user to visit previously unvisited locations.

[1]Center for Complex Network Research, Department of Physics, Biology and Computer Science, Northeastern University, Boston, Massachusetts 02115, USA, [2]Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA, [3]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Massachusetts 02115, USA. †These authors contributed equally to this work. *e-mail: alb@neu.edu.

© 2010 Macmillan Publishers Limited. All rights reserved.

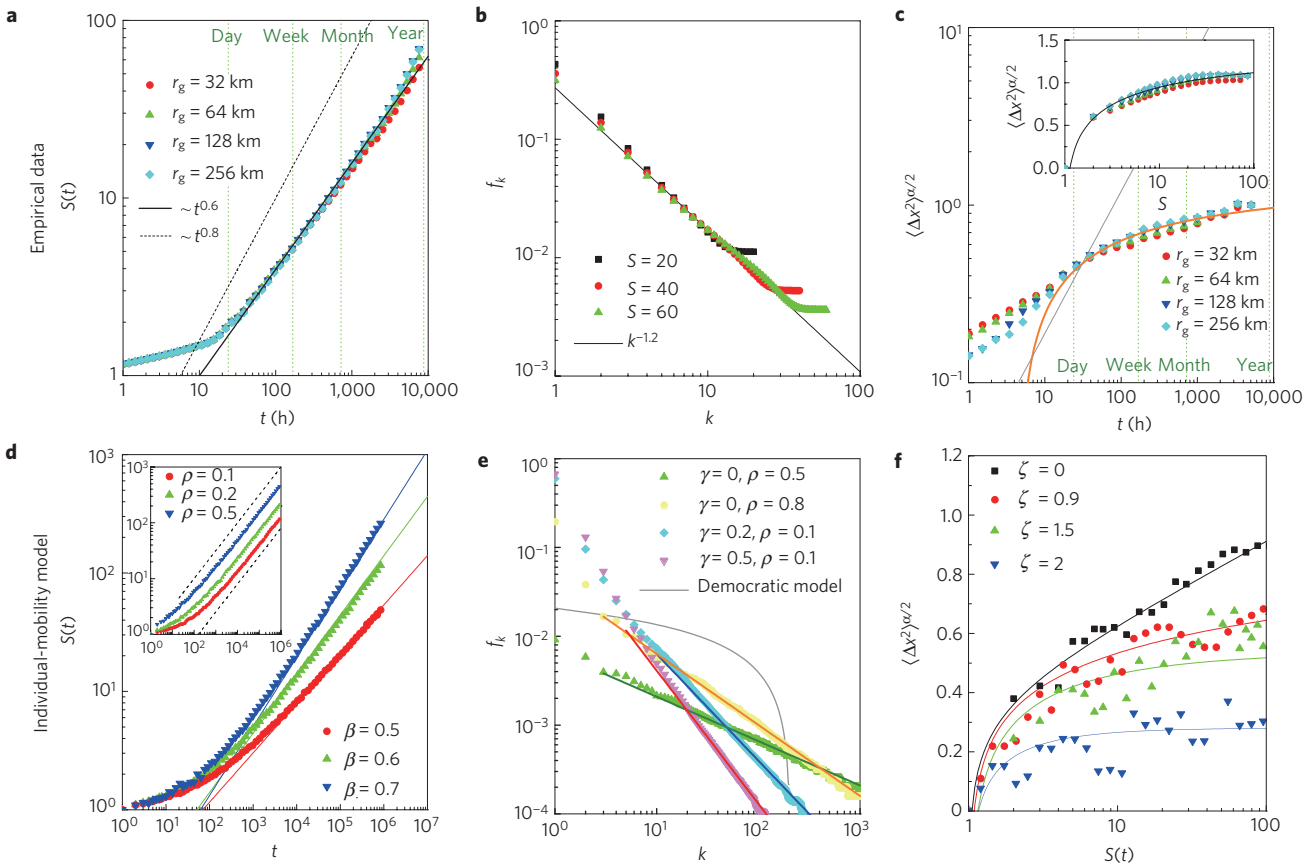

Figure 1 | Empirical results versus the predictions of the individual-mobility model. a, The number of visited distinct locations $S(t)$ versus time for different r_g groups, indicating that $S(t)$ grows as t^{μ}, with $\mu \approx 0.6 \pm 0.02$ (straight line). The dashed line represents the prediction of CTRW $S(t) \sim t^{\beta}$ with $\beta \approx 0.8$. **b**, Zipf's plot showing the visitation frequency f_k of the kth most visited location of a user for different S values. The empirical data is well approximated by $f_k \sim k^{-\xi}$, where $\xi \approx 1.2 \pm 0.1$. **c**, Time evolution of the MSD ($\langle \Delta x^2 \rangle^{\alpha/2}$) in a log–log scale for user groups with different radii of gyration r_g, where the MSDs are normalized by their value at $t = 1$ year. The orange curve represents the analytical prediction for the asymptotic behaviour. The grey line represents the analytical prediction of CTRW, $\langle \Delta x^2 \rangle^{\alpha/2} \sim t^{\beta}$ with $\beta = 0.8$. Inset: The normalized MSD versus S for different r_g groups, where the black curve represents the analytical prediction. **d**, The number of visited distinct locations $S(t)$ versus time in a log–log plot, as predicted by the individual-mobility model with different β values ($\alpha = 0.5$, $\gamma = 0.2$ and $\rho = 0.1$). The straight lines represent the analytical prediction. Inset: The individual-mobility model prediction for different ρ values ($\alpha = 0.5$, $\beta = 0.6$ and $\gamma = 0.2$). **e**, Zipf's plot showing the visitation frequency f_k in the individual-mobility model with different γ and ρ values ($\alpha = 0.5$, $\beta = 0.6$). The straight lines show the analytical prediction. The grey line corresponds to a democratic model in which preferential attachment is absent. **f**, The MSD ($\langle \Delta x^2 \rangle^{\alpha/2}$) versus the number of visited distinct locations $S(t)$ in a log–linear scale. The symbols correspond to simulations with different ξ values, where $\xi = 0$ corresponds to the democratic model. The solid lines represent the analytical prediction.

(B) Visitation frequency: the probability f of a user to visit a given location is expected to be asymptotically ($t \to \infty$) uniform everywhere ($f \sim$ const.) for both Lévy flights and CTRWs. In contrast, the visitation patterns of humans is rather uneven, so that the frequency f of the kth most visited location follows Zipf's law[14]

$$f_k \sim k^{-\zeta} \tag{2}$$

where $\zeta \approx 1.2 \pm 0.1$ (see Fig. 1b). This suggests that the visitation frequency distribution follows $P(f) \sim f^{-(1+1/\zeta)}$.

(C) Ultraslow diffusion: the CTRW model predicts that the mean square displacement (MSD) asymptotically follows $\langle \Delta x^2(t) \rangle \sim t^{\nu}$ with $\nu = 2\beta/\alpha \approx 3.1$. As both $P(\Delta r)$ and $P(\Delta t)$ have cutoffs, asymptotically the MSD should converge to a Brownian behaviour with $\nu = 1$. However, this convergence is too slow[25] to be relevant in our observational time frame. Either way, CTRW predicts that the longer we follow a human trajectory, the further it will drift from its initial position. Yet, humans have a tendency to return home on a daily basis, suggesting that simple diffusive processes, which are not recurrent in two dimensions, do not offer a suitable description of human mobility. Indeed, our measurements indicate an ultraslow diffusive process, in which the MSD seems to follow a slower than logarithmic growth (see Fig. 1c and ref. 14). Such ultraslow growth of the MSD is rare in diffusion, having been observed before only in a few disordered systems, from glasses (for example the Sinai model[26]) to polymers[27] and iterated maps[28].

On one hand, the findings summarized in A–C indicate that individual human mobility does follow reproducible scaling laws, the origins of which remain to be uncovered. Yet, they also document systematic deviations from the predictions of the Lévy-flight- or CTRW-based null models. The main purpose of this Article is to offer a model that not only explains the origin of anomalies A–C, but also leads to a self-consistent statistical model of individual human mobility.

Generic mechanisms and individual mobility model

As we build our model, we will take for granted the observations that the jump-size $P(\Delta r)$ and the waiting-time $P(\Delta t)$ distributions characterizing individual human trajectories are heavy-tailed, a phenomenon addressed by a series of models[29–33]. Yet, $P(\Delta r)$ and $P(\Delta t)$ alone are not sufficient to explain scaling laws A–C. We

© 2010 Macmillan Publishers Limited. All rights reserved.

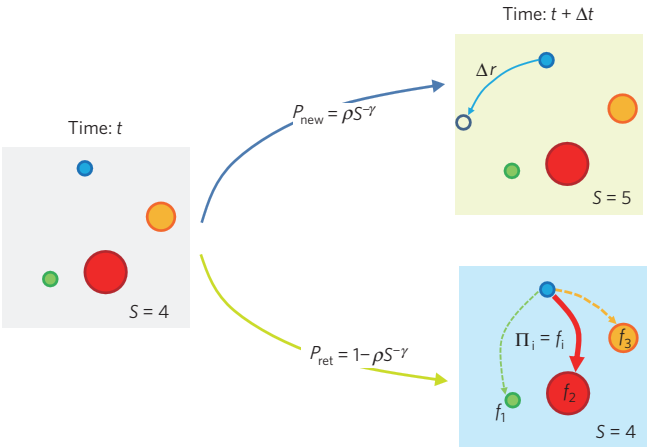

Figure 2 | Schematic description of the individual-mobility model. Starting at time t from the configuration shown in the left panel, indicating that the user visited previously $S = 4$ locations with frequency f_i that is proportional to the size of circles drawn at each location, at time $t + \Delta t$ (with Δt drawn from the $P(\Delta t)$ fat-tailed distribution) the user can either visit a new location at distance Δr from his/her present location, where Δr is chosen from the $P(\Delta r)$ fat-tailed distribution (exploration; upper panel), or return to a previously visited location with probability $P_{\text{ret}} = 1 - \rho S^{-\gamma}$, where the next location will be chosen with probability $\Pi_i = f_i$ (preferential return; lower panel).

propose that the main reason for this discrepancy is that two generic mechanisms, exploration and preferential return, both unique to human mobility, are missing from the traditional random-walk (Lévy-flight or CTRW) models. (i) Exploration: random-walk models assume that the next diffusive step is independent of the previously visited locations. In contrast, the scaling law (1) indicates that the tendency to explore additional locations decreases with time. Indeed, the longer we observe a person's trajectory, the harder it is to find locations in the vicinity of their home/workplace that they have not yet visited. (ii) Preferential return: in contrast with the random-walk-based models for which the visitation probability is random and uniform in space, humans show significant propensity to return to the locations they visited frequently before, such as their home or workplace.

In what follows we present an individual-mobility model that incorporates ingredients (1) and (2), showing that they are sufficient to explain anomalies A–C. The model, intended to describe the trajectory of an individual, assumes that at time $t = 0$ the individual is at some preferred location (see Fig. 2). After a waiting time Δt chosen from the $P(\Delta t)$ distribution, the individual will change his/her location. We assume that the individual has two choices. (i) Exploration: with probability

$$P_{\text{new}} = \rho S^{-\gamma} \tag{3}$$

the individual moves to a new location (different from the S locations he/she visited before). The distance Δr that he/she covers during this exploratory jump is chosen from the $P(\Delta r)$ distribution and his/her direction is selected to be random. As the individual moves to this new position, the number of previously visited locations increases from S to $S + 1$. (ii) Preferential return: with the complementary probability $P_{\text{ret}} = 1 - \rho S^{-\gamma}$ the individual returns to one of the S previously visited locations. In this case, the probability Π_i to visit location i is chosen to be proportional to the number of visits the user previously had to that location. That is, we assume that

$$\Pi_i = f_i \tag{4}$$

an assumption known as preferential attachment or cumulative advantage in network and social science[34–38].

Model predictions

The individual-mobility model has two parameters, $0 < \rho \leq 1$ and $\gamma \geq 0$, both of which control the user's tendency to explore a new location during his/her next move versus returning to a previously visited location. The numerical values of these two parameters will be determined later from the empirical data.

To explain the origin of the anomaly A, we note that in the individual-mobility model the probability that an individual moves to a new location is proportional to $S^{-\gamma}$, that is, $\mathrm{d}S/\mathrm{d}n \propto S^{-\gamma}$, predicting $S \sim n^{1/(1+\gamma)}$, where n is the total number of discrete moves the individual had up to time t. For a fat-tailed waiting-time distribution $P(\Delta t) \sim |\Delta t|^{-1-\beta}$ the time t scales with the number of jumps n as $t \sim n^{1/\beta}$ (Supplementary Section S4A), showing that $S(t)$ follows (1) with the exponent

$$\mu = \beta/(1+\gamma) \tag{5}$$

To verify the validity of this prediction for the individual-mobility model, in Fig. 1d we calculated $S(t)$ numerically for different values of β, finding that the asymptotic scaling exponent of $S(t)$ follows equation (5). Therefore, we predict that $\mu \leq \beta$, in line with the empirical data.

To account for anomaly B we notice that m_i, the number of visits to location i, increases as $\mathrm{d}m_i/\mathrm{d}n = \Pi_i(1 - P_{\text{new}})$, where $\Pi_i = f_i = m_i/\Sigma_i m_i(n)$ is the probability to return to the location i during step n. When $\gamma > 0$, in the limit of $S(t) \to \infty$ the probability to explore a new location is negligible compared with the return visits; thus, asymptotically we have $\mathrm{d}m_i/\mathrm{d}n = m_i/\Sigma_i m_i(n)$. As $\Sigma_i m_i(n) = n$, we obtain $m_i(n) = n/n_i$, where n_i denotes the jump during which location i was first visited, at which moment $m_i(n_i) = 1$. Owing to preferential return (4), the earlier a location is visited, the more it is visited later. Thus, the ranking k_i for location i coincides with the order in which it was first visited, that is, $k_i = S(n_i) \sim n_i^{1/(1+\gamma)}$. As the visitation frequency f_i is proportional to $m_i(n) = n/n_i$, we have $f_k \sim k^{-\zeta}$ with the exponent $\zeta = 1 + \gamma$. In general, we find (Supplementary Section S4C)

$$\zeta = \begin{cases} 1+\gamma, & \gamma > 0 \\ 1-\rho, & \gamma = 0 \end{cases} \tag{6}$$

Note that for $\gamma = 0$ and $\rho = 1$ equation (6) predicts $\zeta = 0$, indicating, as expected, that the visitation becomes homogeneous in the CTRW limit.

To test the validity of prediction (6), in Fig. 1e we measured f_k for different values of γ and ρ in the individual-mobility model, finding that the numerically observed scaling behaviour is in agreement with equation (6). The model also suggests that the observed Zipf's law is rooted mainly in the preferential return (4). Indeed, if we calculate f_k for a 'democratic' model in which Π_i is independent of f_i (that is, the individual visits the previously visited locations with the same likelihood), Zipf's law vanishes (see Fig. 1e).

To understand the origin of anomaly C, we note that the number of jumps to new locations l relates to the displacement Δr as $\Delta r \sim l^{1/\alpha}$ (Supplementary Section S4A), suggesting that $\langle \Delta r^2 \rangle \sim \langle l^{2/\alpha} \rangle = \sum_{l=1}^{n} l^{2/\alpha} P(l|S)$, where $P(l|S)$ is the probability that the Sth location is l steps away from the starting point of the individual. Note that l is different from n, as l counts only moves that result in a jump to some new location, so $l \leq n$. $P(l|S)$ follows the recurrent equation $P(l|S) = \sum_{i=1}^{S} P(l-1|k) f_k^S$, where each term within the sum represents a jump from the kth location to the Sth location. Here f_i^s is the probability of visitation of the kth location given that the total number of locations visited previously is $S - 1$, and is well approximated by Zipf's law (2)

© 2010 Macmillan Publishers Limited. All rights reserved.

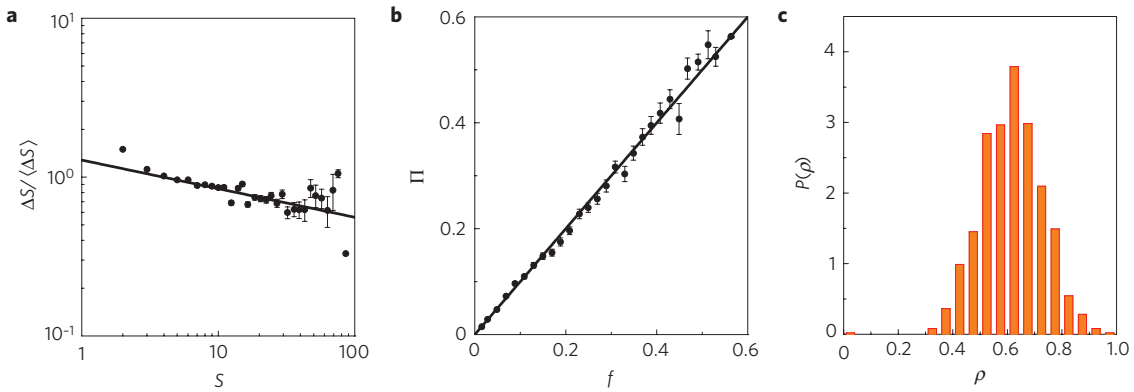

Figure 3 | Testing the hypotheses behind the individual-mobility model. a, The changes in the number of distinct locations $\Delta S/\langle\Delta S\rangle$ visited by the users versus the number of distinct locations S visited previously, indicating that $P_{\text{new}} \sim \Delta S \sim S^{-\gamma}$ with $\gamma = 0.21 \pm 0.02$. The error bars correspond to the standard error of the mean. **b**, The probability to return to a previously visited location Π as a function of the previous frequency of visitation f, offering evidence for preferential attachment $\Pi = f$. The error bars correspond to the standard error of the mean. **c**, The probability density function $P(\rho)$ for the user group, finding that $P(\rho)$, where ρ is the model parameter (see Fig. 2), approximately follows a normal distribution with mean $\rho \approx 0.6$.

$f_k^s \approx (\zeta - 1)/(1 - S^{1-\zeta})k^{-\zeta}$. By applying this approximation (see Supplementary Section S4D), we find

$$\langle \Delta x^2 \rangle^{\alpha/2} \sim \log\left(\frac{1 - S^{1-\zeta}}{\zeta - 1}\right) + \text{const} \qquad (7)$$

which relates the MSD to the number of distinct locations S visited by the user. In Fig. 1c,f, we plot the MSD versus S for different values of ζ, finding that equation (7) agrees well with the numerical results. Furthermore, in the inset of Fig. 1c we plot the prediction (7) against the empirical data, again with excellent agreement.

To determine the explicit time dependence of MSD, we use Equations (1) and (5) to predict three possible scaling regimes for large t (or S). (a) For $\zeta < 1$ ($\gamma = 0$ and $\rho < 1$), $S^{1-\zeta}$ diverges and thus the MSD grows with time as $(\log t)^{2/\alpha}$. In this regime the visitation frequency $P(f) \sim f^{-(1+1/\zeta)}$ has an exponent greater than two and thus the diffusion is dominated by the infrequently visited locations. (b) For $\zeta = 1$, we have $(\zeta - 1)/(1 - S^{1-\zeta}) = \log(S)$, and thus the MSD $\sim \log(\log(t))^{2/\alpha}$. (c) For $\zeta > 1$, $S^{1-\zeta}$ approaches zero for large S and thus the MSD is expected to saturate. This is because $P(f)$ decays with an exponent that has a value of less than two and thus the individual's motion is dominated by his/her most visited location.

To compare further the individual-mobility model with the empirical data, we first need to test the validity of hypotheses (i) and (ii) as formulated by equations (3) and (4). To do this, we used the D_2 data set and measured the rate at which the users visit new locations (Fig. 3a). We find that $P_{\text{new}} \propto S^{-\gamma}$ with $\gamma = 0.21 \pm 0.02$, a result that not only confirms the validity of hypothesis (i), but also provides the numerical value of the exponent γ. To test the validity of hypothesis (ii), in Fig. 3b we plot the probability Π that a user returns to a previous location as a function of the previous visitation frequency f of this location. The plot indicates that $\Pi = f$, confirming the validity of preferential return (4). Furthermore, the ratio between P_{new} and $S^{-\gamma}$ measures the parameter ρ for each user, allowing us to plot $P(\rho)$ for our user group, finding that $P(\rho)$ approximately follows a normal distribution with the mean at $\langle\rho\rangle \approx 0.6$. Note that the scaling properties of the individual-mobility model do not depend on ρ for $\gamma > 0$; thus, although the precise value of ρ is important to parameterize the model, it does not affect the scaling laws discussed in A–C.

Finally, to match our model-based prediction with the empirical data, it is necessary to inspect the relationship between the exponents characterizing the model and those observed for real human mobility. In this respect our starting point is the set of three independently determined exponents $\alpha = 0.55 \pm 0.05$, $\beta = 0.8 \pm 0.1$ and $\gamma = 0.21 \pm 0.02$. Equation (5) predicts $\mu = 0.67 \pm 0.07$, in agreement within the error bar with the empirical value $\mu = 0.6 \pm 0.02$ (Fig. 1a). Furthermore, equation (6) predicts $\zeta = 1.21 \pm 0.02$, again in excellent agreement with the empirical value $\zeta = 1.2 \pm 0.1$. Moreover, equations (5) and (6) predict the existence of the scaling relationship $\beta = \mu\zeta$ between three empirical exponents (valid for $\zeta > 1$), which is again consistent with the empirical data (Fig. 1b). Finally, for anomaly C the analytical predictions offer three different scaling regimes, determined by the value of ζ. The empirical data indicate the Zipf exponent $\zeta = 1.2 \pm 0.1$, for which we predict a saturation in the MSD. To understand how the system reaches saturation, we expand ζ around $\zeta = 1$, finding that in the transient regime the MSD scales as $(\log\log(t))^{2/\alpha}$. This prediction is valid only if $S(t) < \exp(1/(\zeta - 1)) \approx 148$, which is true for 89% of the users that have $S < 148$ over the one-year period. This prediction is consistent with the empirical data shown in Fig. 1c, documenting a slower than logarithmic growth of the MSD. Note that in the empirical data we do not observe the predicted saturation of the MSD, potentially owing to the finite time frame (1 year) used in the study. To estimate the saturation time, we extrapolated equation (1) up to $S(t_s) = 148$ for the empirical data, predicting that the saturation can be reached only after $t_s = 5$ years, beyond our data horizon. Note that the MSD saturation could also be rooted in the finite size of the country (that is, the finite number of towers a user can visit). Yet, scaling arguments (see Supplementary S6) indicate that the saturation time for S because of the finite size effects is approximately 10,000 years. Therefore, the saturation of MSD is rooted in the need to return to the most visited locations, forcing the new locations explored by a user to be close to the user's most visited location.

Another key empirical observation is population heterogeneity: the radius of gyration r_g of the trajectory of different individuals is found to follow a fat-tailed distribution[14] (see Fig. 4a). As we show in Fig. 4c, our model can reproduce this feature as well, indicating that the fat-tailed $P(r_g)$ is a consequence of the inherent fluctuations present within the model and it is rooted in the $P(\Delta r)$ distribution. Indeed, in Supplementary Section S5 we show that the tails of $P(r_g)$ and $P(\Delta r)$ are expected to share the same exponent $1 + \alpha$. Furthermore, we find that the model reproduces not only the $P(r_g)$ distribution, but the ultraslow growth of r_g as well (Fig. 4d), in agreement with the empirical data (Fig. 4b).

It is important to note that in contrast with the traditional random-walk, Lévy-flight or CTRW models, our model is dynamically quenched. That is, after an individual explores a new location, he/she will have an increasing tendency to return to it in the future, generating a recurrent and relatively stable mobility pattern for each individual. In principle one could also consider a

© 2010 Macmillan Publishers Limited. All rights reserved.

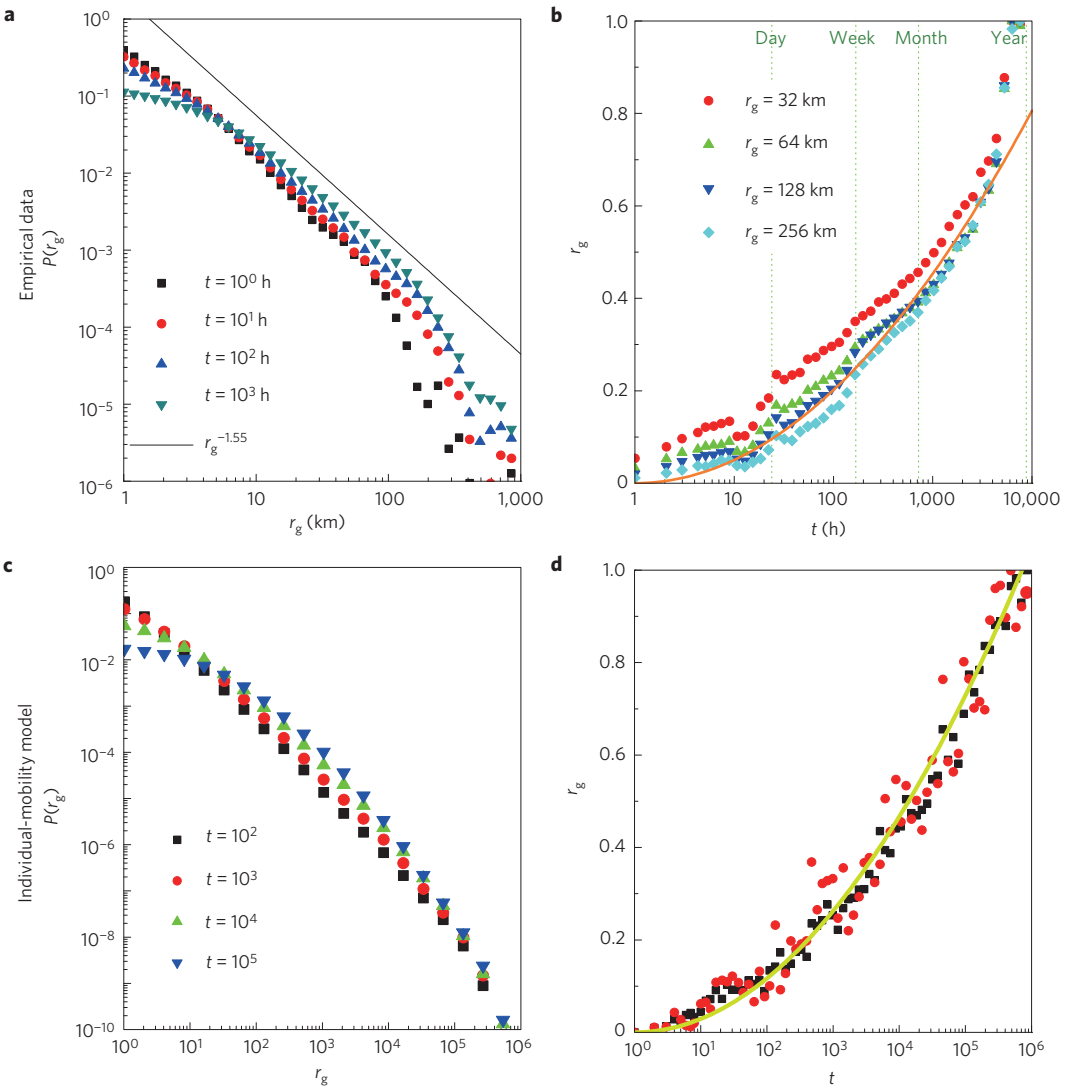

Figure 4 | Population heterogeneity and ultraslow growth of the radius of gyration. a, The $P(r_g)$ distribution of the radius of gyration r_g for the mobile-phone users at different moments of time. The straight line, shown as a guide to the eye, represents a power-law decay with the exponent $1+\alpha \approx 1.55$. **b**, The time evolution of the radius of gyration $r_g(t)$ for the mobile-phone users, where $r_g(t)$ is normalized by its maximal value $r_g(T)$ at time $T = 1$ year. The orange curve is a guide to the eye, following $(\log t)^2$. **c**, The distribution of the radius of gyration $P(r_g)$ for the individual-mobility model using $\alpha = 0.75$, $\beta = 0.6$, $\gamma = 0.2$ and $\rho = 0.1$, the values found to be of direct relevance to human mobility. **d**, The time evolution of the radius of gyration $r_g(t)$ for the individual-mobility model with the same parameters as in **c**, where $r_g(t)$ is normalized by its value $r_g(T)$ at $T = 10^6$. The green curve is a guide to the eye, following $(\log t)^2$.

model that assigns a quenched visitation variable to each site. Our approach not only avoids the need to parameterize such a model, but also achieves the trajectory selection dynamically, through its self-quenching character.

We also note that the model is designed to capture the long-term spatial and temporal scaling patterns; thus, in its present from it does not reproduce the short-term temporal order and correlations potentially present in individual mobility. Our choice to focus the asymptotic properties is driven not only by theoretical arguments (we aim to reproduce the universal and not the transient patterns), but also by practical considerations: many human-mobility-driven processes, from epidemic spreading to city planning, are driven by the asymptotic characteristics of human mobility. To achieve a better short-range temporal fidelity, we need to incorporate the periodic modulations that are known to characterize human mobility (there is a 24 hour and 7 day periodicity in human mobility; individuals are less likely to change locations during the night and are quite mobile in the morning and late afternoon, see Supplementary S7) as well potential correlations in spatial mobility (that is, if location B is between locations A and C in space, the likely order of visitation will be A → B → C or C → B → A). These correlations further constrain the human trajectories, being partly responsible for the high degree of predictability characterizing individual mobility patterns[39]. Finally, we note that the dynamical quenching and the recurrent behaviour are unique to human trajectories, and do not restrict banknote diffusion or foraging behaviour[13,33,40]. As such, our model represents an improvement over the CTRW/Lévy-flight models, as it is adapted to capture the specifics of human mobility. By reproducing the basic scaling laws characterizing human trajectories, the present model offers a conceptual framework that has the flexibility to absorb future extensions, potentially improving the temporal fidelity of its short-term dynamics as well.

Methods

MSD. Owing to the significant population heterogeneity of $P(r_g)$, MSD averaged over all users diverges and thus is ill-defined. We therefore grouped users on

© 2010 Macmillan Publishers Limited. All rights reserved.

the basis of their r_g, guaranteeing that the MSDs defined within groups do not diverge. We calibrate $t = 0$ to correspond to the moment when an individual leaves his/her most visited location. To compare the empirical measure with the analytical prediction (7) directly, the displacement is counted only when the individual explores a new site. This methodology does not affect the scaling behaviour of regular random-walk-based models.

Number of distinct locations (S). Similar to the MSD, $S(t)$ is the average over different r_g groups. Given the large population in our data set, the standard error of the mean is quite small and the error bar in Fig. 1a is smaller than the symbols. The error bar in μ mainly comes from fitting of the power law, where the correlation coefficient $r = 0.998$ and the p value associated with the fit is less than 10^{-6}.

Received 4 May 2010; accepted 26 July 2010; published online 12 September 2010

References

1. Vespignani, A. Predicting the behavior of techno-social systems. *Science* **325,** 425–428 (2009).
2. Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. Predictability and epidemic pathways in global outbreaks of infectious diseases: The SARS case study. *BMC Med.* **5,** 34 (2007).
3. Balcan, D. *et al.* Seasonal transmission potential and activity peaks of the new influenza A(H1N1): A Monte Carlo likelihood analysis based on human mobility. *BMC Med.* **7,** 45 (2009).
4. Eubank, S. *et al.* Modelling disease outbreaks in realistic urban social networks. *Nature* **429,** 180–184 (2004).
5. Toroczkai, Z. & Guclu, H. Proximity networks and epidemics. *J. Phys. A* **378,** 68–75 (2007).
6. Makse, H. A., Havlin, S. & Stanley, H. E. Modelling urban growth patterns. *Nature* **377,** 608–612 (1995).
7. Hufnagel, L., Brockmann, D. & Geisel, T. Forecast and control of epidemics in a globalized worlds. *Proc. Natl Acad. Sci. USA* **101,** 15124–15129 (2004).
8. Rozenfeld, H. D. *et al.* Laws of population growth. *Proc. Natl Acad. Sci. USA* **105,** 18702–18707 (2008).
9. Krings, G., Calabrese, F., Ratti, C. & Blondel, V. D. Urban gravity: A model for inter-city telecommunication flows. *J. Stat. Mech.-Theor. Exp.* L07003 (2009).
10. Ratti, C. & Richens, P. Raster analysis of urban form. *Environ. Plan. B* **31,** 297–309 (2004).
11. Gabaix, X., Gopikrishnan, P., Plerou, V. & Stanley, H. E. A theory of power-law distributions in financial market fluctuations. *Nature* **423,** 267–270 (2003).
12. Wang, P., González, M. C., Hidalgo, C. A. & Barabási, A. L. Understanding the spreading patterns of mobile phone viruses. *Science* **324,** 1071–1076 (2009).
13. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439,** 462–465 (2006).
14. González, M. C., Hidalgo, C. A. & Barabási, A. L. Understanding individual human mobility patterns. *Nature* **453,** 779–782 (2008).
15. Montroll, E. W. & Weiss, G. H. Random walks on lattices II. *J. Math. Phys.* **6,** 167–181 (1965).
16. Weiss, G. H. *Aspects and Applications of the Random Walk* (North-Holland, 1994).
17. Metzler, R. & Klafter, J. The random walk's guide to anomalous diffusion: A fractional dynamic approach. *Phys. Rep.* **339,** 1–77 (2000).
18. Ben-Avraham, D. & Havlin, S. *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge Univ. Press, 2000).
19. Redner, S. *A Guide to First-Passage Processes* (Cambridge Univ. Press, 2001).
20. Ciliberti, S., Caldarelli, G., De los Rios, P., Pietronero, L. & Zhang, Y. C. Discretized diffusion processes. *Phys. Rev. Lett.* **85,** 4848–4851 (2000).
21. Larralde, H., Trunfio, P., Havlin, S., Stanley, H. E. & Weiss, G. H. Territory covered by N diffusing particles. *Nature* **355,** 423–426 (1992).
22. Larralde, H., Trunfio, P., Havlin, S., Stanley, H. E. & Weiss, G. H. Number of distinct sites visited by N random walkers. *Phys. Rev. A* **45,** 7128–7138 (1992).
23. Yuste, S. B., Klafter, J. & Lindenberg, K. Number of distinct sites visited by a subdiffusive random walker. *Phys. Rev. E* **77,** 032101 (2008).
24. Gillis, J. E. & Weiss, G. H. Expected number of distinct sites visited by a random walk with an infinite variance. *J. Math. Phys.* **11,** 1307–1312 (1970).
25. Mantegna, R. N. & Stanley, H. E. Stochastic process with ultraslow convergence to a Gaussian: The truncated Levy flight. *Phys. Rev. Lett.* **73,** 2946–2949 (1994).
26. Sinai, Y. G. The limiting behavior of a one-dimensional random walk in random medium. *Theor. Probl. Appl.* **27,** 256–268 (1982).
27. Schiessel, H., Sokolov, I. M. & Blumen, A. Dynamics of a polyampholyte hooked around an obstacle. *Phys. Rev. E* **56,** R2390–R2393 (1997).
28. Drager, J. & Klafter, J. Strong anomaly in diffusion generated by iterated maps. *Phys. Rev. Lett.* **84,** 5998–6001 (2000).
29. Viswanathan, G. M. *et al.* Optimizing the success of random searches. *Nature* **401,** 911–914 (1999).
30. Santos, M. C. *et al.* Origin of power-law distributions in deterministic walks: The influence of landscape geometry. *Phys. Rev. E* **75,** 061114 (2007).
31. Lomholt, M. A., Koren, T., Metzler, R. & Klafter, J. Levy strategies in intermittent search processes are advantageous. *Proc. Natl Acad. Sci. USA* **105,** 11055–11059 (2008).
32. Raposo, E. P., Buldyrev, S. V., da Luz, M. G. E., Viswanathan, G. M. & Stanley, H. E. Levy flights and random searches. *J. Phys. A* **42,** 434003 (2009).
33. Viswanathan, G. M., Raposo, E. P. & da Luz, M. G. E. Levy flights and superdiffusion in the context of biological encounters and random searches. *Phys. Life Rev.* **5,** 133–150 (2008).
34. Eggenberger, F. & Polya, G. Uber die Statistik verketteter vorgange. *Z. Angew. Math. Mech.* **1,** 279–289 (1923).
35. Yule, G. U. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Phil. Trans. R. Soc. Lond. B* **213,** 21–87 (1925).
36. Simon, H. A. On a class of skew distribution functions. *Biometrika* **42,** 425–440 (1955).
37. Price, D. d. S. A general theory of bibliometric and other cumulative advantage processes. *J. Am. Soc. Inform. Sci.* **27,** 292–306 (1976).
38. Barabási, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286,** 509–512 (1999).
39. Song, C., Qu, Z., Blumm, N. & Barabási, A. L. Limits of predictability in human mobility. *Science* **327,** 1018–1021 (2010).
40. Brown, C. T., Liebovitch, L. S. & Glendon, R. Lévy flights in Dobe Ju/'hoansi foraging patterns. *Hum. Ecol.* **35,** 129–138 (2007).

Acknowledgements

We thank M. Gonzales, D. Wang, J. Bagrow and Z. Qu for discussions and comments on the manuscript. This work was supported by the James S. McDonnell Foundation 21st Century Initiative in Studying Complex Systems; NSF within the Information Technology Research (DMR-0426737), and IIS-0513650 programmes; the Defense Threat Reduction Agency Award HDTRA1-08-1-0027 and the Network Science Collaborative Technology Alliance sponsored by the US Army Research Laboratory under Agreement Number W911NF-09-2-0053.

Author contributions

C.S., T.K. and A-L.B. conceived and executed the research; C.S. and T.K. ran the numerical simulations. C.S., T.K. and P.W. analysed the empirical data.

Additional information

The authors declare no competing financial interests. Supplementary information accompanies this paper on www.nature.com/naturephysics. Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions. Correspondence and requests for materials should be addressed to A-L.B.

© 2010 Macmillan Publishers Limited. All rights reserved.