Understanding individual human mobility patterns

Marta C. González[1], César A. Hidalgo[1,2] & Albert-László Barabási[1,2,3]

Despite their importance for urban planning[1], traffic forecasting[2] and the spread of biological[3–5] and mobile viruses[6], our understanding of the basic laws governing human motion remains limited owing to the lack of tools to monitor the time-resolved location of individuals. Here we study the trajectory of 100,000 anonymized mobile phone users whose position is tracked for a six-month period. We find that, in contrast with the random trajectories predicted by the prevailing Lévy flight and random walk models[7], human trajectories show a high degree of temporal and spatial regularity, each individual being characterized by a time-independent characteristic travel distance and a significant probability to return to a few highly frequented locations. After correcting for differences in travel distances and the inherent anisotropy of each trajectory, the individual travel patterns collapse into a single spatial probability distribution, indicating that, despite the diversity of their travel history, humans follow simple reproducible patterns. This inherent similarity in travel patterns could impact all phenomena driven by human mobility, from epidemic prevention to emergency response, urban planning and agent-based modelling.

Given the many unknown factors that influence a population's mobility patterns, ranging from means of transportation to job- and family-imposed restrictions and priorities, human trajectories are often approximated with various random walk or diffusion models[7,8]. Indeed, early measurements on albatrosses[9], followed by more recent data on monkeys and marine predators[10,11], suggested that animal trajectory is approximated by a Lévy flight[12,13]—a random walk for which step size Δr follows a power-law distribution $P(\Delta r) \sim \Delta r^{-(1+\beta)}$, where the displacement exponent $\beta < 2$. Although the Lévy statistics for some animals require further study[14], this finding has been generalized to humans[7], documenting that the distribution of distances between consecutive sightings of nearly half-a-million bank notes is fat-tailed. Given that money is carried by individuals, bank note dispersal is a proxy for human movement, suggesting that human trajectories are best modelled as a continuous-time random walk with fat-tailed displacements and waiting-time distributions[7]. A particle following a Lévy flight has a significant probability to travel very long distances in a single step[12,13], which seems to be consistent with human travel patterns: most of the time we travel only over short distances, between home and work, whereas occasionally we take longer trips.

Each consecutive sighting of a bank note reflects the composite motion of two or more individuals who owned the bill between two reported sightings. Thus, it is not clear whether the observed distribution reflects the motion of individual users or some previously unknown convolution between population-based heterogeneities and individual human trajectories. Contrary to bank notes, mobile phones are carried by the same individual during his/her daily routine, offering the best proxy to capture individual human trajectories[15–19].

We used two data sets to explore the mobility pattern of individuals. The first (D_1) consisted of the mobility patterns recorded over a six-month period for 100,000 individuals selected randomly from a sample of more than 6 million anonymized mobile phone users. Each time a user initiated or received a call or a text message, the location of the tower routeing the communication was recorded, allowing us to reconstruct the user's time-resolved trajectory (Fig. 1a, b). The time between consecutive calls followed a 'bursty' pattern[20] (see Supplementary Fig. 1), indicating that although most consecutive calls are placed soon after a previous call, occasionally there are long periods without any call activity. To make sure that the obtained results were not affected by the irregular call pattern, we also studied a data set (D_2) that captured the location of 206 mobile phone users, recorded every two hours for an entire week. In both data sets, the spatial resolution was determined by the local density of the more than 10^4 mobile towers, registering movement only when the user moved between areas serviced by different towers. The average service area of each tower was approximately $3\,km^2$, and over 30% of the towers covered an area of $1\,km^2$ or less.

To explore the statistical properties of the population's mobility patterns, we measured the distance between user's positions at consecutive calls, capturing 16,264,308 displacements for the D_1 and 10,407 displacements for the D_2 data set. We found that the distribution of displacements over all users is well approximated by a truncated power-law:

$$P(\Delta r) = (\Delta r + \Delta r_0)^{-\beta} \exp(-\Delta r / \kappa) \qquad (1)$$

with exponent $\beta = 1.75 \pm 0.15$ (mean $\pm$ standard deviation), $\Delta r_0 = 1.5\,km$ and cutoff values $\kappa|_{D_1} = 400\,km$ and $\kappa|_{D_2} = 80\,km$ (Fig. 1c, see the Supplementary Information for statistical validation). Note that the observed scaling exponent is not far from $\beta = 1.59$ observed in ref. 7 for bank note dispersal, suggesting that the two distributions may capture the same fundamental mechanism driving human mobility patterns.

Equation (1) suggests that human motion follows a truncated Lévy flight[7]. However, the observed shape of $P(\Delta r)$ could be explained by three distinct hypotheses: first, each individual follows a Lévy trajectory with jump size distribution given by equation (1) (hypothesis A); second, the observed distribution captures a population-based heterogeneity, corresponding to the inherent differences between individuals (hypothesis B); and third, a population-based heterogeneity coexists with individual Lévy trajectories (hypothesis C); hence, equation (1) represents a convolution of hypotheses A and B.

To distinguish between hypotheses A, B and C, we calculated the radius of gyration for each user (see Supplementary Information), interpreted as the characteristic distance travelled by user a when observed up to time t (Fig. 1b). Next, we determined the radius of gyration distribution $P(r_g)$ by calculating r_g for all users in samples D_1 and D_2, finding that they also can be approximated with a truncated power-law:

$$P(r_g) = \left(r_g + r_g^0 \right)^{-\beta_r} \exp(-r_g / \kappa) \qquad (2)$$

[1]Center for Complex Network Research and Department of Physics, Biology and Computer Science, Northeastern University, Boston, Massachusetts 02115, USA. [2]Center for Complex Network Research and Department of Physics and Computer Science, University of Notre Dame, Notre Dame, Indiana 46556, USA. [3]Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA.

with $r_g^0 = 5.8$ km, $\beta_r = 1.65 \pm 0.15$ and $\kappa = 350$ km (Fig. 1d, see Supplementary Information for statistical validation). Lévy flights are characterized by a high degree of intrinsic heterogeneity, raising the possibility that equation (2) could emerge from an ensemble of identical agents, each following a Lévy trajectory. Therefore, we determined $P(r_g)$ for an ensemble of agents following a random walk (RW), Lévy flight (LF) or truncated Lévy flight (TLF) (Fig. 1d)[8,12,13]. We found that an ensemble of Lévy agents display a significant degree of heterogeneity in r_g; however, this was not sufficient to explain the truncated power-law distribution $P(r_g)$ exhibited by the mobile phone users. Taken together, Fig. 1c and d suggest that the difference in the range of typical mobility patterns of individuals (r_g) has a strong impact on the truncated Lévy behaviour seen in equation (1), ruling out hypothesis A.

If individual trajectories are described by an LF or TLF, then the radius of gyration should increase with time as $r_g(t) \sim t^{3/(2+\beta)}$ (ref. 21), whereas, for an RW, $r_g(t) \sim t^{1/2}$; that is, the longer we observe a user, the higher the chance that she/he will travel to areas not visited before. To check the validity of these predictions, we measured the time dependence of the radius of gyration for users whose gyration radius would be considered small ($r_g(T) \leq 3$ km), medium ($20 < r_g(T) \leq 30$ km) or large ($r_g(T) > 100$ km) at the end of our observation period ($T = 6$ months). The results indicate that

the time dependence of the average radius of gyration of mobile phone users is better approximated by a logarithmic increase, not only a manifestly slower dependence than the one predicted by a power law but also one that may appear similar to a saturation process (Fig. 2a and Supplementary Fig. 4).

In Fig. 2b, we chose users with similar asymptotic $r_g(T)$ after $T = 6$ months, and measured the jump size distribution $P(\Delta r|r_g)$ for each group. As the inset of Fig. 2b shows, users with small r_g travel mostly over small distances, whereas those with large r_g tend to display a combination of many small and a few larger jump sizes. Once we rescaled the distributions with r_g (Fig. 2b), we found that the data collapsed into a single curve, suggesting that a single jump size distribution characterizes all users, independent of their r_g. This indicates that $P(\Delta r|r_g) \sim r_g^{-\alpha} F(\Delta r/r_g)$, where $\alpha \approx 1.2 \pm 0.1$ and $F(x)$ is an r_g-independent function with asymptotic behaviour, that is, $F(x) \sim x^{-\alpha}$ for $x < 1$ and $F(x)$ rapidly decreases for $x \gg 1$. Therefore, the travel patterns of individual users may be approximated by a Lévy flight up to a distance characterized by r_g. Most important, however, is the fact that the individual trajectories are bounded beyond r_g; thus, large displacements, which are the source of the distinct and anomalous nature of Lévy flights, are statistically absent. To understand the relationship between the different exponents, we note that the measured probability distributions are related
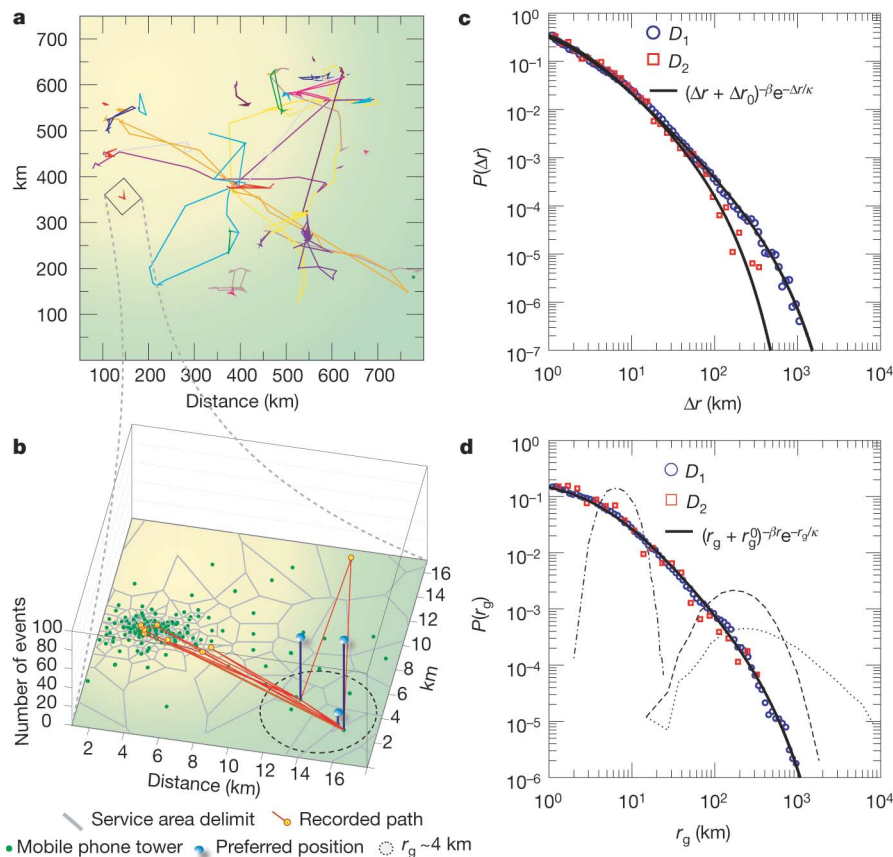

Figure 1 | Basic human mobility patterns. a, Week-long trajectory of 40 mobile phone users indicates that most individuals travel only over short distances, but a few regularly move over hundreds of kilometres. **b,** The detailed trajectory of a single user. The different phone towers are shown as green dots, and the Voronoi lattice in grey marks the approximate reception area of each tower. The data set studied by us records only the identity of the closest tower to a mobile user; thus, we can not identify the position of a user within a Voronoi cell. The trajectory of the user shown in **b** is constructed from 186 two-hourly reports, during which the user visited a total of 12 different locations (tower vicinities). Among these, the user is found on 96 and 67 occasions in the two most preferred locations; the frequency of visits for each location is shown as a vertical bar. The circle represents the radius of gyration centred in the trajectory's centre of mass. **c,** Probability density function $P(\Delta r)$ of travel distances obtained for the two studied data sets D_1 and D_2. The solid line indicates a truncated power law for which the parameters are provided in the text (see equation (1)). **d,** The distribution $P(r_g)$ of the radius of gyration measured for the users, where $r_g(T)$ was measured after $T = 6$ months of observation. The solid line represents a similar truncated power-law fit (see equation (2)). The dotted, dashed and dot-dashed curves show $P(r_g)$ obtained from the standard null models (RW, LF and TLF, respectively), where for the TLF we used the same step size distribution as the one measured for the mobile phone users.

by $P(\Delta r) = \int_0^\infty P(\Delta r | r_g) P(r_g) dr_g$, which suggests (see Supplementary Information) that up to the leading order we have $\beta = \beta_r + \alpha - 1$, consistent, within error bars, with the measured exponents. This indicates that the observed jump size distribution $P(\Delta r)$ is in fact the convolution between the statistics of individual trajectories $P(\Delta r_g | r_g)$ and the population heterogeneity $P(r_g)$, consistent with hypothesis C.

To uncover the mechanism stabilizing r_g, we measured the return probability for each individual $F_{pt}(t)$ (first passage time probability)[21,22], defined as the probability that a user returns to the position where he/she was first observed after t hours (Fig. 2c). For a two-dimensional random walk, $F_{pt}(t)$ should follow $\sim 1/(t \ln^2(t))$ (ref. 21). In contrast, we found that the return probability is characterized by several peaks at 24 h, 48 h and 72 h, capturing a strong tendency of humans to return to locations they visited before, describing the recurrence and temporal periodicity inherent to human mobility[23,24].

To explore if individuals return to the same location over and over, we ranked each location on the basis of the number of times an individual was recorded in its vicinity, such that a location with $L = 3$ represents the third-most-visited location for the selected individual. We find that the probability of finding a user at a location with a given rank L is well approximated by $P(L) \sim 1/L$, independent of the number of locations visited by the user (Fig. 2d). Therefore, people devote most of their time to a few locations, although spending their remaining time in 5 to 50 places, visited with diminished regularity. Therefore, the observed logarithmic saturation of $r_g(t)$ is rooted in the high degree of regularity in the daily travel patterns of individuals, captured by the high return probabilities (Fig. 2b) to a few highly frequented locations (Fig. 2d).

An important quantity for modelling human mobility patterns is the probability density function $\Phi_a(x, y)$ to find an individual a in a given position (x, y). As it is evident from Fig. 1b, individuals live and travel in different regions, yet each user can be assigned to a well defined area, defined by home and workplace, where she or he can be found most of the time. We can compare the trajectories of different users by diagonalizing each trajectory's inertia tensor, providing the probability of finding a user in a given position (see Fig. 3a) in the user's intrinsic reference frame (see Supplementary Information for the details). A striking feature of $\Phi(x, y)$ is its prominent spatial anisotropy in this intrinsic reference frame (note the different scales in Fig. 3a); we find that the larger an individual's r_g, the more pronounced is this anisotropy. To quantify this effect, we defined the anisotropy ratio $S \equiv \sigma_y/\sigma_x$, where σ_x and σ_y represent the standard deviation of the trajectory measured in the user's intrinsic reference frame (see Supplementary Information). We found that S decreases monotonically with r_g (Fig. 3c), being well approximated with $S \sim r_g^{-\eta}$ for $\eta \approx 0.12$. Given the small value of the scaling exponent, other functional forms may offer an equally good fit; thus, mechanistic models are required to identify if this represents a true scaling law or only a reasonable approximation to the data.

To compare the trajectories of different users, we removed the individual anisotropies, rescaling each user trajectory with its respective σ_x and σ_y. The rescaled $\tilde{\Phi}(x/\sigma_x, y/\sigma_y)$ distribution (Fig. 3b) is similar for groups of users with considerably different r_g, that is, after the anisotropy and the r_g dependence are removed all individuals seem to follow the same universal $\tilde{\Phi}(\tilde{x}, \tilde{y})$ probability distribution. This is particularly evident in Fig. 3d, where we show the cross section of $\tilde{\Phi}(x/\sigma_x, 0)$ for the three groups of users, finding that apart from the noise in the data the curves are indistinguishable.

Taken together, our results suggest that the Lévy statistics observed in bank note measurements capture a convolution of the population heterogeneity shown in equation (2) and the motion of individual users. Individuals display significant regularity, because they return to a few highly frequented locations, such as home or work. This regularity does not apply to the bank notes: a bill always follows the trajectory of its current owner; that is, dollar bills diffuse, but humans do not.

The fact that individual trajectories are characterized by the same r_g-independent two-dimensional probability distribution $\tilde{\Phi}(x/\sigma_x, y/\sigma_y)$ suggests that key statistical characteristics of individual trajectories are largely indistinguishable after rescaling. Therefore, our results establish the basic ingredients of realistic agent-based models, requiring us to place users in number proportional with the population density of a given region and assign each user an r_g taken from the observed $P(r_g)$ distribution. Using the predicted anisotropic rescaling, combined with the density function $\tilde{\Phi}(x, y)$, the shape of which is provided as Supplementary Table 1, we can obtain the likelihood of finding a user in any location. Given the known correlations between spatial proximity and social links, our results could help quantify the role of space in network development and evolution[25–29] and improve our understanding of diffusion processes[8,30].
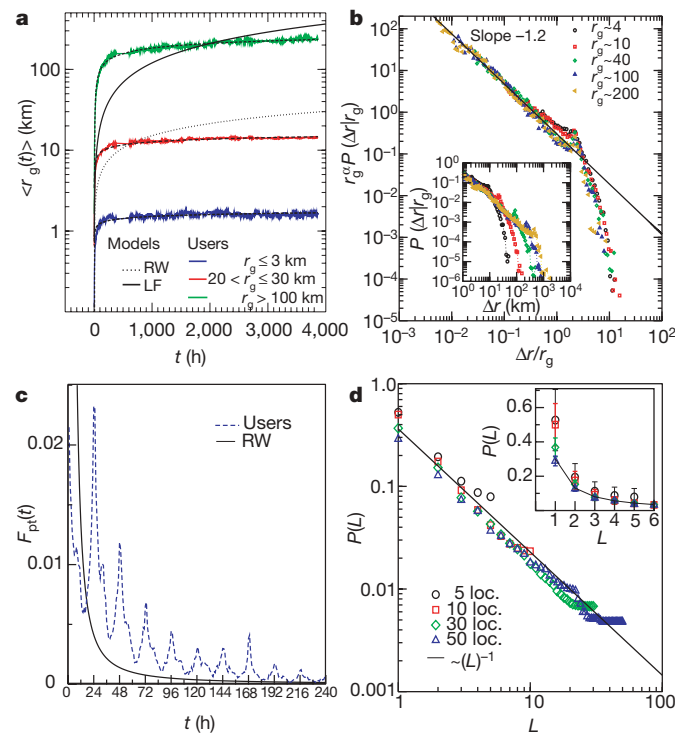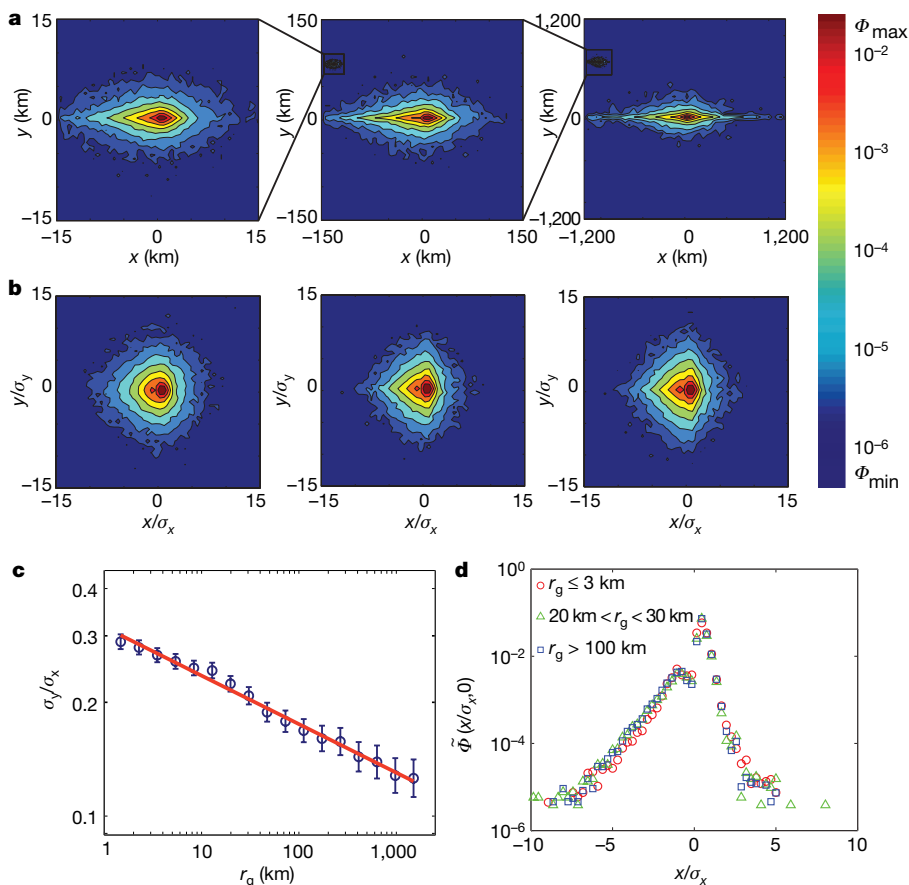
Figure 2 | The bounded nature of human trajectories. a, Radius of gyration $\langle r_g(t) \rangle$ versus time for mobile phone users separated into three groups according to their final $r_g(T)$, where $T = 6$ months. The black curves correspond to the analytical predictions for the random walk models, increasing with time as $\langle r_g(t) \rangle |_{LF,TLF} \sim t^{3/2 + \beta}$ (solid curve) and $\langle r_g(t) \rangle |_{RW} \sim t^{0.5}$ (dotted curve). The dashed curves corresponding to a logarithmic fit of the form $A + B \ln(t)$, where A and B are time-independent coefficients that depend on r_g. **b**, Probability density function of individual travel distances $P(\Delta r | r_g)$ for users with $r_g = 4$, 10, 40, 100 and 200 km. As the inset shows, each group displays a quite different $P(\Delta r | r_g)$ distribution. After rescaling the distance and the distribution with r_g (main panel), the different curves collapse. The solid line (power law) is shown as a guide to the eye. **c**, Return probability distribution, $F_{pt}(t)$. The prominent peaks capture the tendency of humans to return regularly to the locations they visited before, in contrast with the smooth asymptotic behaviour $\sim 1/(t \ln(t)^2)$ (solid line) predicted for random walks. **d**, A Zipf plot showing the frequency of visiting different locations (loc.). The symbols correspond to users that have been observed to visit $n_L = 5$, 10, 30 and 50 different locations. Denoting with (L), the rank of the location listed in the order of the visit frequency, the data are well approximated by $R(L) \sim L^{-1}$. The inset is the same plot in linear scale, illustrating that 40% of the time individuals are found at their first two preferred locations; bars indicate the standard error.

Figure 3 | The shape of human trajectories.
a, The probability density function $\Phi(x,y)$ of finding a mobile phone user in a location (x,y) in the user's intrinsic reference frame (see Supplementary Information for details). The three plots, from left to right, were generated for 10,000 users with: $r_g \leq 3$, $20 < r_g \leq 30$ and $r_g > 100$ km. The trajectories become more anisotropic as r_g increases. **b,** After scaling each position with σ_x and σ_y, the resulting $\tilde{\Phi}(x/\sigma_x, y/\sigma_y)$ has approximately the same shape for each group. **c,** The change in the shape of $\Phi(x,y)$ can be quantified calculating the isotropy ratio $S \equiv \sigma_y/\sigma_x$ as a function of r_g, which decreases as $S \sim r_g^{-0.12}$ (solid line). Error bars represent the standard error. **d,** $\tilde{\Phi}(x/\sigma_x, 0)$ representing the x-axis cross-section of the rescaled distribution $\tilde{\Phi}(x/\sigma_x, y/\sigma_y)$ shown in **b.**

1. Horner, M. W. & O'Kelly, M. E. S Embedding economies of scale concepts for hub networks design. *J. Transp. Geogr.* **9**, 255–265 (2001).
2. Kitamura, R., Chen, C., Pendyala, R. M. & Narayaran, R. Micro-simulation of daily activity-travel patterns for travel demand forecasting. *Transportation* **27**, 25–51 (2000).
3. Colizza, V., Barrat, A., Barthélémy, M., Valleron, A.-J. & Vespignani, A. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Medicine* **4**, 95–110 (2007).
4. Eubank, S. *et al.* Controlling epidemics in realistic urban social networks. *Nature* **429**, 180–184 (2004).
5. Hufnagel, L., Brockmann, D. & Geisel, T. Forecast and control of epidemics in a globalized world. *Proc. Natl Acad. Sci. USA* **101**, 15124–15129 (2004).
6. Kleinberg, J. The wireless epidemic. *Nature* **449**, 287–288 (2007).
7. Brockmann, D. D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
8. Havlin, S. & Ben-Avraham, D. Diffusion in disordered media. *Adv. Phys.* **51**, 187–292 (2002).
9. Viswanathan, G. M. *et al.* Lévy flight search patterns of wandering albatrosses. *Nature* **381**, 413–415 (1996).
10. Ramos-Fernandez, G. *et al.* Lévy walk patterns in the foraging movements of spider monkeys (*Ateles geoffroyi*). *Behav. Ecol. Sociobiol.* **273**, 1743–1750 (2004).
11. Sims, D. W. *et al.* Scaling laws of marine predator search behaviour. *Nature* **451**, 1098–1102 (2008).
12. Klafter, J., Shlesinger, M. F. & Zumofen, G. Beyond brownian motion. *Phys. Today* **49**, 33–39 (1996).
13. Mantegna, R. N. & Stanley, H. E. Stochastic process with ultraslow convergence to a gaussian: the truncated Lévy flight. *Phys. Rev. Lett.* **73**, 2946–2949 (1994).
14. Edwards, A. M. *et al.* Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature* **449**, 1044–1049 (2007).
15. Sohn, T. *et al.* in *Proc. 8th Int. Conf. UbiComp 2006* 212–224 (Springer, Berlin, 2006).
16. Onnela, J.-P. *et al.* Structure and tie strengths in mobile communication networks. *Proc. Natl Acad. Sci. USA* **104**, 7332–7336 (2007).
17. González, M. C. & Barabási, A.-L. Complex networks: from data to models. *Nature Physics* **3**, 224–225 (2007).
18. Palla, G., Barabási, A.-L. & Vicsek, T. Quantifying social group evolution. *Nature* **446**, 664–667 (2007).
19. Hidalgo, C. A. & Rodriguez-Sickert, C. The dynamics of a mobile phone network. *Physica A* **387**, 3017–3024 (2008).
20. Barabási, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211 (2005).
21. Redner, S. *A Guide to First-Passage Processes* (Cambridge Univ. Press, Cambridge, UK, 2001).
22. Condamin, S., Bénichou, O., Tejedor, V. & Klafter, J. First-passage times in complex scale-invariant media. *Nature* **450**, 77–80 (2007).
23. Schlich, R. & Axhausen, K. W. Habitual travel behaviour: evidence from a six-week travel diary. *Transportation* **30**, 13–36 (2003).
24. Eagle, N. & Pentland, A. Eigenbehaviours: identifying structure in routine. *Behav. Ecol. Sociobiol.* (in the press).
25. Yook, S.-H., Jeong, H. & Barabási, A. L. Modeling the Internet's large-scale topology. *Proc. Natl Acad. Sci. USA* **99**, 13382–13386 (2002).
26. Caldarelli, G. *Scale-Free Networks: Complex Webs in Nature and Technology.* (Oxford Univ. Press, New York, 2007).
27. Dorogovtsev, S. N. & Mendes, J. F. F. *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Univ. Press, New York, 2003).
28. Song, C. M., Havlin, S. & Makse, H. A. Self-similarity of complex networks. *Nature* **433**, 392–395 (2005).
29. González, M. C., Lind, P. G. & Herrmann, H. J. A system of mobile agents to model social networks. *Phys. Rev. Lett.* **96**, 088702 (2006).
30. Cecconi, F., Marsili, M., Banavar, J. R. & Maritan, A. Diffusion, peer pressure, and tailed distributions. *Phys. Rev. Lett.* **89**, 088102 (2002).