

ARTICLE

Received 20 Jun 2016 | Accepted 15 Jul 2016 | Published 30 Aug 2016

DOI: 10.1038/ncomms12600

OPEN

# A stochastic model of randomly accelerated walkers for human mobility

Riccardo Gallotti<sup>1</sup>, Armando Bazzani<sup>2,3</sup>, Sandro Rambaldi<sup>2,3</sup> & Marc Barthelemy<sup>1,4</sup>

Recent studies of human mobility largely focus on displacements patterns and power law fits of empirical long-tailed distributions of distances are usually associated to scale-free superdiffusive random walks called Lévy flights. However, drawing conclusions about a complex system from a fit, without any further knowledge of the underlying dynamics, might lead to erroneous interpretations. Here we show, on the basis of a data set describing the trajectories of 780,000 private vehicles in Italy, that the Lévy flight model cannot explain the behaviour of travel times and speeds. We therefore introduce a class of accelerated random walks, validated by empirical observations, where the velocity changes due to acceleration kicks at random times. Combining this mechanism with an exponentially decaying distribution of travel times leads to a short-tailed distribution of distances which could indeed be mistaken with a truncated power law. These results illustrate the limits of purely descriptive models and provide a mechanistic view of mobility.

<sup>1</sup>Institut de Physique Théorique, CEA, CNRS-URA 2306, F-91191 Gif-sur-Yvette, France. <sup>2</sup>Department of Physics and Astronomy, University of Bologna, Viale Berti Pichat 6/2, 40126 Bologna, Italy. <sup>3</sup>INFN Bologna Section, 40126 Bologna, Italy. <sup>4</sup>CAMS (CNRS/EHESS) 190-198, avenue de France, 75244 Paris, France. Correspondence and requests for materials should be addressed to M.B. (email: marc.barthelemy@cea.fr).

Understanding individual mobility has important implications for traffic forecasting<sup>1</sup>, epidemics spreading<sup>2,3</sup> or the evolution of cities<sup>4-6</sup>. With the development of Information and Communication Technologies<sup>7</sup>, the investigations' focus shifted from the traditional travel diary surveys<sup>8-10</sup> to several new data sources. In particular, it became possible to follow individual trajectories from mobile phone calls<sup>11-13</sup>, location-sharing services<sup>14-16</sup> and microblogging<sup>17</sup>, or directly extracted from public transport ticketing system<sup>10,18</sup>, global positioning system (GPS) tracks of taxis<sup>10,19-23</sup>, private cars<sup>24-26</sup> or single individuals<sup>27,28</sup>. For most data sources, the spatial position  $r$  is the most reliable quantity. This information can be used for studying two different aspects of human mobility: how far and where we are moving. The second question is far more complex than the first and can be approached with several different tools, from aggregated origin-destination matrices<sup>29</sup> for mobility prediction<sup>30</sup> or land use analysis<sup>31</sup> to individual mobility networks and patterns suitable for describing the natural tendency to return frequently to a few locations (such as homes, offices and so on)<sup>11,12,24-26,32</sup>. On the other hand, the first question, generally characterized by the distribution  $P(\Delta r)$  of the individuals' displacements  $\Delta r$  across all users, although apparently simple is still far from being completely understood. Indeed, even if the study of the distribution  $P(\Delta r)$  has become a trademark for recent works on human mobility, there are still no consensus about the functional form of this distribution. At a large scale (national or inter-urban), one may observe a long tail behaviour<sup>9,11,12,14,15,17,20,22,33</sup> characterized by a power law decay for long displacements. At a smaller (urban) scale, the distribution seems to have an exponential tail<sup>10,13,19,21,22,24,25</sup>. The unclear nature of this probability distribution makes its interpretation difficult and dependent on the data set used, the scale and possible empirical and fitting errors<sup>34</sup>. It is therefore necessary to obtain data as clean as possible and to propose a model that can be tested against empirical results. So far, essentially power law fits were used and led the authors to draw conclusions about the nature and mechanisms of the mobility, but this way of proceeding could actually lead to erroneous conclusions. Similar unresolved controversy also exists in the study of animal's foraging movements: relying only on a fit of the empirical data, the same distribution can be understood in different ways, leading to contrasting conclusions on the nature of the underlying process<sup>35-38</sup>.

Remarkably enough, what appears to be under-evaluated in the study of human mobility is the relevance of travel itself. Human travelling behaviour can in general be described as a sequence of rest times of duration  $\tau$  and jumps  $\Delta r$  in space<sup>12</sup>. These two processes need to be separated for modelling human mobility, since costs are in general associated to trips while a positive utility can be associated to activities performed during stops<sup>1</sup>. However, proposed models usually neglect the role of travel time and the moving velocity and assume instantaneous jumps. This is essentially a consequence of the limitations inherent to data sources: phone calls or social networks capture the spatial character of individuals' movements<sup>39</sup>, but are limited by sampling rates or by the bursty nature of human communications<sup>40,41</sup> and are thus not suitable for an exhaustive temporal description of human mobility.

In this paper, we show that the observed truncated power laws in the jump size distribution can be the consequence of simple processes such as random walks with random velocities<sup>42</sup>. We test this model over a large GPS database describing the mobility of 780,000 private vehicles in Italy, where travels and pauses can be easily separated, as the transition is identified by the moment when the engine is turned on or off (but we introduce a lower threshold of 5 min in the elapsed time, to distinguish real stops from accidentally switched off of the engine during a trip). This allows us to evaluate accurately not only the displacements  $\Delta r$ , but also travel times  $t$ , speeds  $v$  and rest times  $\tau$ .

## Results

**The current empirical view.** Several studies suggested that the displacements' distribution  $P(\Delta r)$  has a fat tail, and power law fits display a wide range of exponent values depending on the data set and the fitting form used (see Table 1). We note that, strictly speaking, the distribution cannot be scale-free since displacements are always limited in space<sup>43</sup>. Thus, human movements could possibly be identified as truncated Lévy flights only<sup>11,44</sup>. In contrast, displacements at the urban scale consistently display an exponential tail<sup>19,24,25</sup>. Short tails also emerge when studying the distribution  $P(t)$  of travel times  $t$  of individual trajectories originating in different cities. For private cars' mobility, as in the data set studied here, we observe that  $P(t)$  is indeed characterized by an exponential decay  $P(t) \propto e^{-t/\tau}$  as in refs 9,19,28,45 (see

**Table 1 | Parameter values for the fit of the displacement distribution with a truncated power law found in previous studies.**

Data source	Trajectories	$\beta$	$\kappa$	$\Delta r_0$
Dollar bills <sup>33</sup>	464 K	1.59	$\infty$	0
Mobile phones <sup>11</sup>	100 K	1.75	400 km	1.5 km
Mobile phones <sup>11</sup>	206	1.75	80 km	1.5 km
Mobile phones <sup>12</sup>	3 M	1.55	100 km	0
Location sharing <sup>14</sup>	220 K	1.88	$\infty$	0
GPS tracks <sup>27</sup>	101	[1.16, 1.82]	$\infty$	0
Location sharing <sup>15</sup>	900 K	1.50	$\infty$	2.87 km
Location sharing <sup>15</sup>	900 K	4.67	$\infty$	18.42 km
Taxis <sup>19</sup>	12 K	0	4.29 km	—
Taxis <sup>20</sup>	7 K	1.2	10 km	0.31 km
Mobile phones <sup>13</sup>	3 M	0	[2, 5.8] km	—
Travel diaries <sup>9</sup>	230	1.05	50 km	0
Tweets <sup>17</sup>	13 M	1.62	$\infty$	0
Location sharing <sup>16</sup>	521 K	0	300 km	—
Taxis <sup>21</sup>	34 K	0	[2, 4.6] km	—
Taxis <sup>23</sup>	1,100	[0.50, 1.17]	[4.5, 6.5] km	0

This list includes studies on different data sources and spatial or temporal scales. Only fits consistent with the function  $P(\Delta r) = (\Delta r + \Delta r_0)^{-\beta} \exp(-\frac{\Delta r}{\kappa})$  proposed in ref. 11 are presented here. The case  $\kappa = \infty$  is associated to non-truncated power laws, while  $\beta = 0$  to exponential distributions. When  $\Delta r_0 = 0$  this parameter was omitted in the fit, and we set  $\beta = 0$  when this value is not defined. Further studies propose: (i) a polynomial form close to an exponential behaviour for private cars<sup>24</sup>; (ii) two different behaviours for urban and inter-urban trajectories for cars and taxis<sup>22,25</sup>; and (iii) a lognormal distribution for individual GPS tracks<sup>28</sup>.

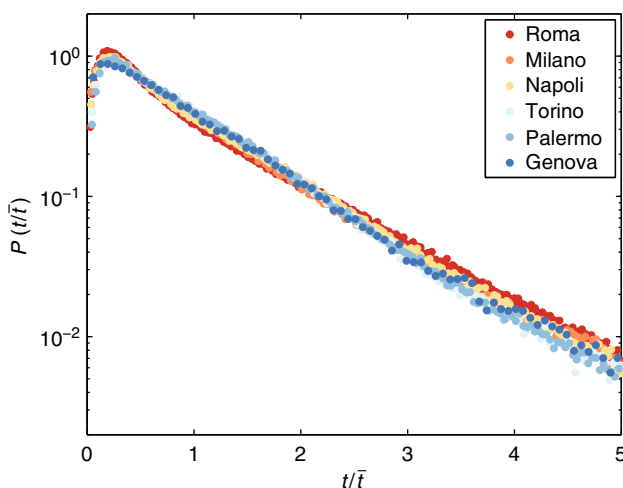
Fig. 1), where  $\bar{t}$  is the average travel time that may vary among cities (see Supplementary Note 1 and Supplementary Fig. 1). For public transportation, we also observe a rapidly decreasing tail for the travel times between metro stations (see ref. 18 and Supplementary Fig. 2). A short tail is expected for these distributions, since the total daily travel time spent in public<sup>46</sup> or private transportation<sup>45</sup> has an exponential tail (see Supplementary Note 1 and Supplementary Fig. 3 for the analysis of the distribution  $P(\tau)$  of rest times  $\tau$ ).

We have therefore ambiguous empirical results and purely descriptive models, leading to a very unclear view of human mobility. Here we investigate the statistics of displacements starting from simple assumptions about the dynamics that governs mobility. In contrast with empirical and descriptive approaches, we start by modelling this mechanism and show that our predictions are consistent with data. To model mobility, we must understand the relationship between the duration of a trip and its average velocity. Velocity is a natural quantity for describing the mobility, and classical traffic modelling focuses on predicting its average value in freeways or in cities with different vehicle's densities<sup>47</sup>. The relevance of velocity appears to be underestimated in recent studies even when the hierarchy of transportation networks is suggested to be at the origin of the fat tail<sup>48</sup>. As discussed above, this omission is a consequence of data sources limitations. To go beyond descriptive approaches, it is necessary to obtain richer data such as the large GPS database used here.

**Random uncorrelated accelerations.** A first remark is that apparent truncated power laws can result from interrupting at random times simple processes<sup>49</sup> such as random walks in the velocity space. To illustrate the problem with this simple approach, we consider the evolution of the velocity  $v$  described by a Brownian motion with diffusion coefficient  $D$

$$\dot{v} = \sqrt{D}\xi(\theta) \tag{1}$$

where  $\theta$  is time and  $\xi$  is a white noise. This random acceleration model has been the subject of many theoretical studies (see ref. 50 and references therein) and provides here an interesting null model. We define the displacement for a given time time  $t$  as  $\Delta r(t) = |\int_0^t v d\theta| = \bar{v}t$  where  $\bar{v} > 0$  is the average velocity. To



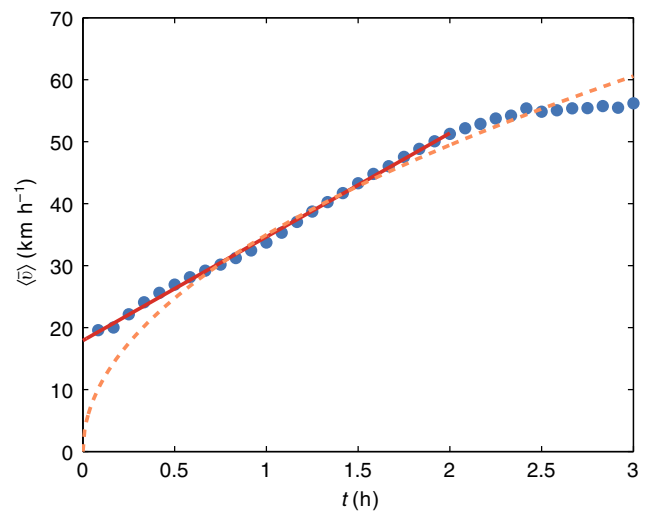
**Figure 1 | Travel time distribution in the six largest Italian cities.** The probability distribution  $P(t)$  is well fitted by an exponential function. As demonstrated by the data collapse shown here, differences among cities are encoded in a single parameter: the average travel time  $\bar{t}$ .

compute the displacement distribution, we use the fact that  $\bar{v}$  is a Gaussian variable and that the travel time distribution is approximatively exponential  $P(t) \sim e^{-t/\bar{t}}$ . Using a saddle-point approximation (see Supplementary methods), we obtain a stretched exponential distribution for large displacements

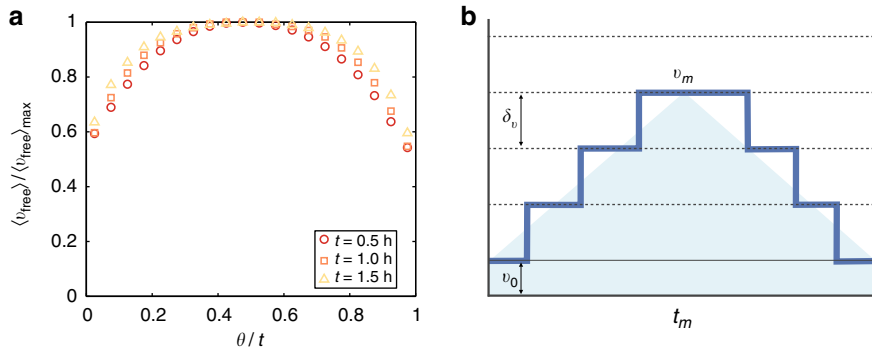
$$P(\Delta r) \propto \Delta r^{-\gamma} \exp(-C\Delta r^\delta) \tag{2}$$

with  $\gamma = 3/4$ ,  $\delta = 1/2$  and  $C$  a free parameter. We show in Supplementary Fig. 4(a) the best fit with  $C = 0.49 \text{ km}^{-0.5}$ . At this stage, this model offers an already reasonable description of the empirical pattern usually described by a truncated power law with three parameters<sup>11</sup> which results here from the combination of a random walk and a random duration model. In addition, this simple random model (equation (1)) implies a relation<sup>51</sup> between the travel time  $t$  and the average velocity modulus of the form  $\langle |\bar{v}| \rangle \propto t^{1/2}$ . However, for both private cars (see Fig. 2) and public transportation (see Supplementary Fig. 5(a) and Supplementary Note 2), we find that average speed grows linearly with the travel time. This important empirical observation invalidates our first simple model (which might still be relevant for describing different mobility patterns, such those of animals<sup>52</sup>) and one has to understand the origin of this uniform acceleration. Our idea here is that this behaviour results from the optimal use of the hierarchical nature of the transportation networks for longer trips. Indeed, it is likely that faster transportation modes or faster roads are used more frequently for longer trajectories<sup>25,53</sup>. In the following, we propose a stochastic model based on this idea, which correctly predicts both travel speed and displacement distributions.

**Random acceleration kicks.** In contrast with the assumption underlying equation (2), accelerations are not uncorrelated: in Fig. 3a, we indeed observe that an average trip can be separated into a first half where one uses progressively faster roads and a second half with deceleration until the velocity reaches back the base speed at time  $t$ . We use this simple representation in a stochastic model (see the schematic in Fig. 3b) and make the following assumptions. The transportation network is modelled by  $n$  layers  $L_n$ ,



**Figure 2 | Acceleration of private transportation.** Empirical average speeds (blue dots) versus the duration of the trip for our GPS data set. The red solid line represents a constant average acceleration  $\langle \bar{v} \rangle = v_0 + at$  with  $v_0 = 17.9 \text{ km h}^{-1}$  and  $a = 16.7 \text{ km h}^{-2}$ . For  $t > 2 \text{ h}$  the speed reaches a saturation at  $\langle \bar{v} \rangle \approx 55 \text{ km h}^{-1}$  imposed by the finite number of layers (see Methods section and Supplementary information). The dashed line represents the best fit with  $\langle |\bar{v}| \rangle \propto t^{1/2}$  (as predicted by the simple random model equation (1) with uncorrelated accelerations).



**Figure 3 | Empirical evidence and schematic representation of the model.** (a) We compute for different trips the empirical average free-flow speeds  $\langle v_{\text{free}} \rangle$  for the used roads, measured in the absence of congestion (see Methods section). We then aggregate the trips according to their duration  $t$ , and we plot the evolution of the relative average free-flow speed  $\langle v_{\text{free}} \rangle / \langle v_{\text{free}} \rangle_{\text{max}}$ , where  $\langle v_{\text{free}} \rangle_{\text{max}}$  is the maximal average free-flow speed reached at  $\theta \approx t/2$ . We observe symmetrical curves with a progressive increase of velocity in the first half of the trip, a short flat central part followed by a deceleration. The curves do not collapse, since longer trips have a larger fraction of their trajectory spent at a maximal speed. The maximal speed  $\langle v_{\text{free}} \rangle_{\text{max}}$  is reached at the middle of the trip and increases from  $79 \text{ km h}^{-1}$  for trips with a duration  $t = 0.5$  h to  $106 \text{ km h}^{-1}$  for a duration of  $t = 1.5$  h. (b) Schematic representation of the random acceleration model (equation (3)). Each trajectory starts with an acceleration phase, where the speed increases by constant kicks equal to  $\delta v$ . These kicks happen at random times with a uniform probability per unit time  $p$ . When approaching the destination, there is a deceleration phase with random kicks of  $-\delta v$  (with the same probability  $p$ ). The symmetry of the problem implies that the maximal speed  $v_m$  is reached on average at  $t_m = t/2$  and depends on  $\delta v$  and  $p$ :  $v_m = v_0 + k(t_m)\delta v$ , where  $k(\theta)$  is the number of kicks after a time  $\theta$  and follows a Poisson distribution of average  $\langle k(\theta) \rangle = p\theta$ . The average speed  $\bar{v}$  of the trajectory is evaluated by estimating the shaded area and leads to  $\bar{v} = (v_0 + v_m)/2$ .

corresponding to different travel speeds  $v_n$  (we neglect saturation effects, see Methods section and Supplementary methods). The speed differences between layers are taken constant and equal to  $\delta v$ , and the speed on layer  $L_k$  is then  $v_k = v_0 + k\delta v$ . An individual starts then her trip of duration  $t$  in the layer  $L_0$  with base speed  $v_0$ , and we assume that there are two phases in a trip, acceleration and deceleration, of roughly the same duration. In both the ascending and descending phases, we define a Poissonian process where all individuals have the same probability per unit time  $p$  to jump to the successive layer and to change their speed.

Within this model, we can estimate the maximal speed  $v_m = v_0 + k(t_m)\delta v$ , where  $k(t_m)$  is the number of jumps at mid trajectory  $t_m = t/2$ , and get approximatively the average speed  $\bar{v} \approx (v_0 + v_m)/2$ . Since the process is Poissonian, we have  $\langle k(t_m) \rangle = pt_m$  and the average speed is given by

$$\langle \bar{v}(t) \rangle = v_0 + \frac{p\delta v}{4}t \tag{3}$$

where the brackets denote the average over the Poisson variable  $k$ . The average speed thus grows linearly with  $t$ , in agreement with empirical observations. Remarkably enough, this model allows us to predict also the shape of the conditional probability distribution  $P(\bar{v}|t)$ . Indeed, the number of jumps  $k$  is distributed following the Poisson distribution  $P(k) = \frac{e^{-\lambda} \lambda^k}{k!}$  with  $\lambda = pt/2$ . Using the Gamma function as the natural analytic continuation of the factorial  $k! = \Gamma(1 + k)$ , we obtain the distribution

$$P(\bar{v}|t) = \frac{1}{\delta v'} \frac{\exp(-p't + \frac{\bar{v}-v_0}{\delta v'} \log(p't))}{\Gamma(1 + \frac{\bar{v}-v_0}{\delta v'})} \tag{4}$$

where  $p' = p/2$  and  $\delta v' = \delta v/2$  are free parameters fitted using empirical speeds (see Fig. 4). The shape of the displacement distribution  $P(\Delta r)$  can then be computed as a superimposition of Poisson distributions (see Supplementary Fig. 4(b)) and is given by

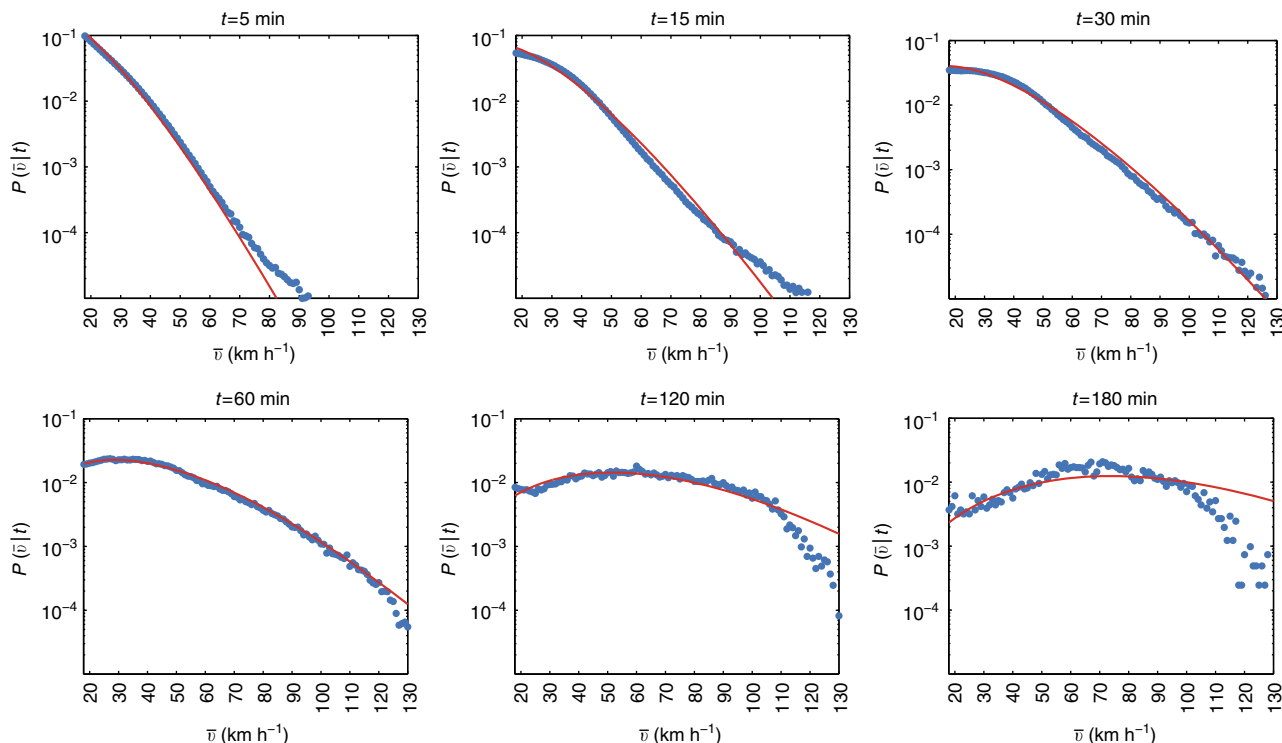
$$\begin{aligned} P(\Delta r) &= \int_0^\infty dt \int_{v_1}^{v_2} d\bar{v} \delta(\Delta r - \bar{v}t) P(t) P(\bar{v}|t) \\ &= \int_0^\infty \frac{dt}{t} \frac{\exp(- (p' + \frac{1}{t})t + \log(p't) \frac{\Delta r/t - v_0}{\delta v'})}{\delta v' \Gamma(1 + \frac{\Delta r/t - v_0}{\delta v'})} \end{aligned} \tag{5}$$

where  $\delta(x)$  is the Dirac delta function. This equation (5) is our main analytical result and its exact form cannot be exactly

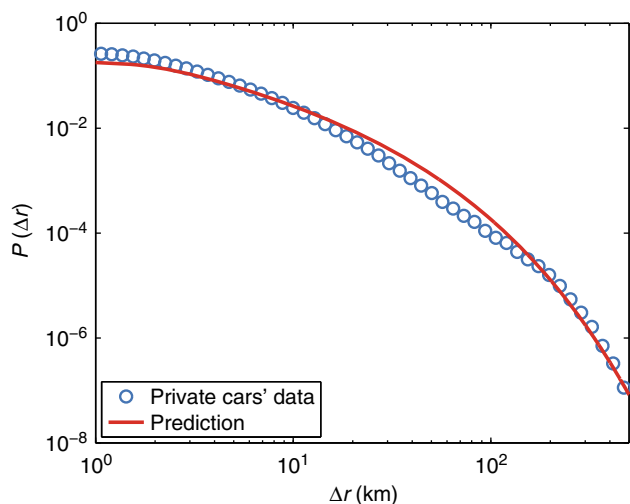
computed, but the limiting behaviour for large  $\Delta r$  is again equation (2) (see Supplementary methods). In particular, this distribution is not fat-tailed, in clear contrast with Lévy flights which have divergent moments and are governed by large fluctuations. Therefore, all phenomena associated to Lévy flights such as superdiffusion, for example, are not expected from our model. In Fig. 5, we compare the empirical  $P(\Delta r)$  with our prediction and show that the proposed random acceleration model (together with the exponential behaviour for  $P(t)$ ) is in excellent agreement with data.

### Discussion

The identification of an approximate power law behaviour of a complex system is rarely scientifically useful by itself<sup>43</sup> but needs to be model-informed. Since multiple competing models can explain the same pattern<sup>40,41</sup>, it even risks to swamp future research with years of replicating the same, and possibly wrong, pattern analysis (see Table 1). Proposing models with simple, reasonable assumptions and processes can help in identifying the fundamental constituents of the problem, and provide predictions that can be tested and trigger future research and improvements. In the case of human mobility, the random acceleration model proposed here allows for a deeper quantitative understanding, leads to predictions in excellent agreement with data, and brings evidence that the long-standing interpretation with Lévy flights is incorrect. The central idea of our model is the existence of a hierarchical organization of transportation layers with different velocities. The main ingredients that describe the variability of  $P(\Delta r)$  among cities (see Supplementary Note 3 and Supplementary Fig. 6) are different travel times (Supplementary Fig. 1), the base speed, the speed gap between layers, and the effective acceleration (Supplementary Fig. 7), which is proportional to the jumping rate to a different layer. An important assumption in this model is that these quantities are constant, but this is not necessarily true. Deviations observed in empirical distributions at the urban level suggest the need for more elaborate velocity models, able to include spatial and temporal inhomogeneities of transportation systems by allowing  $v_0$ ,  $p$  and  $\delta v$  to vary across cities or during the day.



**Figure 4 | Conditional distribution  $P(\bar{v}|t)$  of the average speeds for different travel times.** We fit simultaneously the curves  $P(\bar{v}|t)$  given by equation (4) in the interval  $[v_0, v_{\max}]$ , with  $v_0 = 17.9 \text{ km h}^{-1}$  and  $v_{\max} = 130 \text{ km h}^{-1}$  and for  $t = 5, 10, 15, 20, \dots, 180 \text{ min}$  (see Methods section). Each plot refers to a different trip durations ( $t = 5, 15, 30, 60, 120, 180 \text{ min}$ ). The dots are the empirical data whereas the solid line is the fit obtained using equation (4). The best fit value of the parameters are  $p' = 1.06$  jumps per hour,  $\delta v' = 20.9 \text{ km h}^{-1}$  (and  $v_0 = 17.9 \text{ km h}^{-1}$ ). We therefore get  $\delta v \approx 40 \text{ km h}^{-1}$  for the speed gap, in excellent agreement with the progression of the most common speed limits in Italy:  $50 \text{ km h}^{-1}$  (urban),  $90 \text{ km h}^{-1}$  (extra-urban) and  $130 \text{ km h}^{-1}$  (highways). These results suggest that a multilayer hierarchical transportation infrastructure can explain the constant acceleration observed in both public and private transportation.



**Figure 5 | Displacements distribution.** We show (circles) the aggregated empirical distribution  $P(\Delta r)$  for all the  $\approx 780,000$  cars in the data set. The solid line is the model prediction (and not an *a posteriori* best fit) based on our model equation (5) with: (i)  $v_0 = 17.9 \text{ km h}^{-1}$  estimated as the intercept of the fit in Fig. 2; (ii)  $p' = 1.06$  jumps per hour and  $\delta v' = 20.9 \text{ km h}^{-1}$  estimated from the fit of  $P(\bar{v}|t)$  in Fig. 4; and (iii)  $\bar{t} = 0.30 \text{ h}$  coming from the average of the travel times for all selected trips (see Methods section). This remarkable prediction has a quality comparable with the commonly used direct fits by a truncated power law with three free parameters (see Supplementary Fig. 4).

**Methods**

**Private transportation data.** We compute spatial displacements and travel times for private car transportation from a database of GPS measures sampling the trajectories of private vehicles in the whole Italy during the month of May 2011. The data are collected for insurance reasons by a device installed on the vehicles and are mainly related to private vehicles, since taxi cabs or delivery companies use their own GPS systems and do not contribute to it. A small percentage of vehicles belongs to private companies and are used for professional reasons. This database includes  $\approx 2\%$  of the vehicles registered in Italy, containing a total of 128,363,000 trips performed by 779,000 vehicles. Records contain information about the engine starts and stops, and travel times include the time spent looking for parking. We introduce a lower threshold of 5 min in the elapsed time to define the end of a trip and to distinguish real stops related to an activity. When the quality of the satellite signal is good, we have an average spatial accuracy of the order of 10 m, but during a trip, the error fluctuations may increase up to 30 m or more<sup>54</sup>. The error in temporal resolution is negligible.

We have applied correction and filtering procedures to exclude from our analysis data affected by systematic errors. Approximately 10% of the data were discarded for this reason. When the engine is switched on or when the vehicle is parked inside a building, there are errors due to signal loss. In such cases, we use the redundant information given by the previous stopping point to correct 20% of the data. When the engine was off for  $< 30 \text{ s}$ , the subsequent trajectory is always considered as a continuation of the same trip, except if the vehicle is going back towards the origin of the previous trajectory segment.

For privacy reasons, the drivers' city of residence is unknown. Therefore, it has been necessary to associate to each car an urban area using available information. We do that by identifying a driver as living in a certain city if the most part of its parking time is spent in the corresponding municipality area. For each driver we then consider all trips both within and outside this 'residence' urban area.

**Measuring the free-flow speed.** Each trajectory is sampled at a spatial scale at most equal to 2 km or, in highways, at a time scale of 30 s. Using trajectory reconstructions<sup>55,56</sup>, we estimate for each road the average speed for all the trajectories passing through that road. Using the distribution of the travel speed, the free-flow speed of a road is defined as the speed corresponding to the 85% of percentile. As a matter of fact, the free-flow speed corresponds to trajectories

carried out during the night and the choice of 85% of percentile takes into account the individual heterogeneity: in many cases there are individuals driving at a speed much larger than the average and that are neglected in the computation of an average free-flow speed.

In Fig. 3a, we have computed the free-flow velocity for the used roads in each trajectory and for different travel time durations ( $0.5 \pm 0.1$  h,  $1.0 \pm 0.1$  h and  $1.5 \pm 0.1$  h).

**Details of the model.** The model proposed here assumes that urban mobility is performed at a baseline speed of  $v_0 \approx 18 \text{ km h}^{-1}$ . Very short trips hardly reach the base speed  $v_0$  that we impose in our model. For this reason, in all the measures proposed in this paper, we considered only trips longer than 1 km and with duration larger than 5 min.

The upper bound of  $130 \text{ km h}^{-1}$  limits the number of jumps to  $k = (130 \text{ km h}^{-1} - v_0)/\delta v \approx 3$ . We have a jumping rate of order  $p' \approx 1$  jump per hour and we expect significant deviations from our prediction for trips longer than 4 h. For such long trips, one cannot in principle approximate the area below the step function in Fig. 3b with a triangle. Indeed, the Poisson fit in Fig. 4 overestimates the frequency of trips faster than  $110 \text{ km h}^{-1}$  for  $t = 120$  and 180 min. In Supplementary methods, we show that if the acceleration kicks are limited by a finite number of layers, the speed grows linearly for small times and saturates to a finite limiting speed (see Supplementary Fig. 8). This saturation can be clearly observed at national level (see Fig. 2) and for five out of six of the cities represented in Supplementary Fig. 7.

**Fitting the model parameters.** The parameters are computed according to the following procedure. First, we estimate the value of  $v_0$  from a linear fit of the curve  $\langle \bar{v} \rangle(t)$  (Fig. 2 and Supplementary Fig. 7) in the interval  $[0, 2]$  h. Then, the study of the surface generated by the curve  $P(\bar{v}|t)$  for different durations  $t$  is limited to the speed interval  $[v_0, 130]$  (in  $\text{km h}^{-1}$ ) and the time interval  $[2.5, 182.5]$  min. The values of speeds are binned at integer values in  $\text{km h}^{-1}$ , whereas trip durations are binned at intervals of 5 min (the centres of the bins fall at  $t = 5, 10, 15, \dots, 180$  min). Finally, we estimate the parameters  $p'$  and  $\delta v'$  of equation (4) by minimizing the sum of the square errors for all curves (for all  $t$  bins) simultaneously. In Fig. 4 we show the fits of six sections of the surface. The last parameter needed is the average travel time  $\bar{t}$  and has been measured separately for each city (see Supplementary Fig. 1). To predict consistently the curve  $P(\Delta r)$ , the value  $\bar{t} = 0.30$  h used in Fig. 5 and Supplementary Fig. 6 refers to the trips considered only ( $\Delta r > 1$  km and  $t > 5$  min).

**Data availability.** Anonymized displacement lengths and travel times for the private cars' data are available from the authors upon request.

## References

- Axhausen, K. W. & Gärling, T. Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transp. Rev.* **12**, 323–341 (1992).
- Colizza, V., Barrat, A., Barthelemy, M., Valleron, A. J. & Vespignani, A. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Med.* **4**, 95 (2007).
- Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl Acad. Sci. USA* **106**, 21459–21460 (2009).
- Makse, H. A., Havlin, S. & Stanley, H. E. Modelling urban growth patterns. *Nature* **377**, 608–612 (1995).
- Bettencourt, L. M., Lobo, J., Helbing, D., Kuhnert, C. & West, G. B. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl Acad. Sci. USA* **104**, 7301–7306 (2007).
- Louf, R. & Barthelemy, M. How congestion shapes cities: from mobility patterns to scaling. *Sci. Rep.* **4**, 5561 (2014).
- Vespignani, A. Modelling dynamical processes in complex socio-technical systems. *Nat. Phys.* **8**, 32–39 (2012).
- Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfuser, G. & Haupt, T. Observing the rhythms of daily life: a six-week travel diary. *Transportation* **29**, 95–124 (2002).
- Yan, X.-Y., Han, X.-P., Wang, B.-H. & Zhou, T. Diversity of individual mobility patterns and emergence of aggregated scaling laws. *Sci. Rep.* **3**, 2678 (2013).
- Liang, X., Zhao, J., Dong, L. & Xu, K. Unraveling the origin of exponential law in intra-urban human mobility. *Sci. Rep.* **3**, 2983 (2013).
- González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
- Song, C., Koren, T., Wang, P. & Barabási, A.-L. Modelling the scaling properties of human mobility. *Nat. Phys.* **6**, 818–823 (2010).
- Kang, C., Ma, X., Tong, D. & Liu, Y. Intra-urban human mobility patterns: an urban morphology perspective. *Physica A* **391**, 1702–1717 (2012).
- Cheng, Z., Caverlee, J., Lee, K. & Sui, D. Z. Exploring millions of footprints in location sharing services. *ICWSM* **2011**, 81–88 (2011).
- Noulas, A., Scellato, S., Lambiotte, R., Pontil, M. & Mascolo, C. A tale of many cities: universal patterns in human urban mobility. *PLoS ONE* **7**, e37027 (2012).
- Liu, Y., Sui, Z., Kang, C. & Gao, Y. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS ONE* **9**, e86026 (2014).
- Hawelka, B. *et al.* Geo-located twitter as proxy for global mobility patterns. *Cartogr. Geogr. Inf. Sci.* **41**, 260–271 (2014).
- Roth, C., Kang, S. M., Batty, M. & Barthelemy, M. Structure of urban movements: polycentric activity and entangled hierarchical flows. *PLoS ONE* **6**, e15923 (2011).
- Liang, X., Zheng, X., Lv, W., Zhu, T. & Xu, K. The scaling of human mobility by taxis is exponential. *Physica A* **391**, 2135 (2012).
- Liu, Y., Kang, C., Gao, S., Xiao, Y. & Tian, Y. Understanding intra-urban trip patterns from taxi trajectory data. *J. Geogr. Syst.* **14.4**, 463–483 (2012).
- Wang, W., Pan, L., Yuan, N., Zhang, S. & Liu, D. A comparative analysis of intra-city human mobility by taxi. *Physica A* **420**, 134–147 (2015).
- Liu, H., Chen, Y.-H. & Liha, J.-S. Crossover from exponential to power-law scaling for human mobility pattern in urban, suburban and rural areas. *Eur. Phys. J. B* **88**, 117 (2015).
- Tang, J., Liu, F., Wang, Y. & Wang, H. Uncovering urban human mobility from large scale taxi GPS data. *Physica A* **438**, 140–153 (2015).
- Bazzani, A., Giorgini, B., Rambaldi, S., Gallotti, R. & Giovannini, L. Statistical laws in urban mobility from microscopic GPS data in the area of Florence. *J. Stat. Mech.* **2010**, P05001 (2010).
- Gallotti, R., Bazzani, A. & Rambaldi, S. Toward a statistical physics of human mobility. *Int. J. Mod. Phys. C* **23**, 1250061 (2012).
- Gallotti, R., Bazzani, A., Degli Esposti, M. & Rambaldi, S. Entropic measures of individual mobility patterns. *J. Stat. Mech.* **2013**, P10022 (2013).
- Rhee, I., Shin, M., Hong, S., Lee, K. & Kim, S. On the levy-walk nature of human mobility. *ACM Trans. Network.* **19**, 630–643 (2011).
- Zhao, K., Musolesi, M., Hui, P., Rao, W. & Tarkoma, S. Explaining the power-law distribution of human mobility through transportation modality decomposition. *Sci. Rep.* **5**, 9136 (2015).
- Sagarra, O., Szell, M., Santi, P., Diaz-Guilera, A. & Ratti, C. Supersampling and network reconstruction of urban mobility. *PLoS ONE* **10**, e0134508 (2015).
- Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96 (2012).
- Louail, T. *et al.* From mobile phone data to the spatial structure of cities. *Sci. Rep.* **4**, 5276 (2014).
- Gärbling, T. & Axhausen, K. W. Introduction: habitual travel choice. *Transportation* **30**, 1–11 (2003).
- Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
- Clauaset, A., Shalizi, C. R. & Newman, M. E. Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703 (2009).
- Edwards, A. M. *et al.* Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature* **449**, 1044–1048 (2007).
- Edwards, A. M. Overturning conclusions of Lévy flight movement patterns by fishing boats and foraging animals. *Ecology* **92**, 1247–1257 (2011).
- Petrovskii, S., Mashanova, A. & Jansen, V. A. A. Variation in individual walking behaviour creates the impression of a Lévy flight. *Proc. Natl Acad. Sci. USA* **108**, 8704–8707 (2011).
- Jansen, V. A. A., Mashanova, A. & Petrovskii, S. Comment on Lévy walks evolve through interaction between movement and environmental complexity. *Science* **335**, 918 (2012).
- Lenormand, M. *et al.* Cross-checking different sources of mobility information. *PLoS ONE* **9**, e105184 (2014).
- Barabási, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211 (2005).
- Malmgren, R. D., Stouffer, D. B., Motter, A. E. & Amaral, L. A. N. A Poissonian explanation for heavy tails in e-mail communication. *Proc. Natl Acad. Sci. USA* **105**, 18153–18158 (2008).
- Zaburdaev, V., Schmiedeberg, M. & Stark, H. Random walks with random velocities. *Phys. Rev. E* **78**, 011119 (2008).
- Stumpf, M. P. H. & Porter, M. A. Critical truths about power laws. *Science* **335**, 665–666 (2012).
- Mantegna, R. N. & Stanley, H. E. Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight. *Phys. Rev. Lett.* **73**, 2946 (1994).
- Gallotti, R., Bazzani, A. & Rambaldi, S. Understanding the variability of daily travel-time expenditures using GPS trajectory data. *EPJ Data Sci.* **4**, 1–14 (2015).
- Kölbl, R. & Helbing, D. Energy laws in human travel behaviour. *New J. Phys.* **5**, 48.1–48.12 (2003).
- Helbing, D. Traffic and related self-driven many-particle systems. *Rev. Mod. Phys.* **73**, 1067–1141 (2001).

48. Han, X.-P., Hao, Q., Wang, B.-H. & Zhou, T. Origin of the scaling law in human mobility: hierarchy of traffic systems. *Phys. Rev. E* **83**, 036117 (2011).
49. Reed, W. J. & Hughes, B. D. From gene families and genera to incomes and internet file sizes: why power laws are so common in nature. *Phys. Rev. E* **66**, 067103 (2002).
50. Burkhardt, T. W. in *First-Passage Phenomena and their Applications*. (eds Metzler, R., Oshanin, G. & Redner, S. Chapter 2 (World Scientific, 2014).
51. Gallotti, R. *Statistical Physics and Modeling of Human Mobility*. PhD thesis University of Bologna 79–80, (2013).
52. Viswanathan, G. M., da Luz, M. G. E., Raposo, E. P. & Stanley, H. E. *The Physics of Foraging: An Introduction to Random Searches and Biological Encounters* (Cambridge University Press, (2011).
53. Gallotti, R. & Barthelemy, M. Anatomy and efficiency of urban multimodal mobility. *Sci. Rep.* **4**, 6911 (2014).
54. Bazzani, A. *et al.* in *Proceedings of the Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, 1455–1459 (2011).
55. Giovannini, L. *A Novel Map-Matching Procedure for Low-Sampling GPS Data with Applications to Traffic Flow Analysis*. PhD thesis University of Bologna, (2011).
56. Bazzani, A., Giorgini, B., Giovannini, L., Gallotti, R. & Rambaldi, S. in *MIPRO, 2011 Proceedings of the 34th International Convention*, 1615–1618 (2011).

### Acknowledgements

We thank P. Krapivsky, M. Lenormand and R. Louf for useful discussions. We thank Octo Telematics S.p.A. for providing the GPS database.

### Author contributions

R.G. and M.B. designed and performed research. A.B. and S.R. obtained the data set and performed the data pre-elaboration. R.G., A.B. and M.B. wrote the paper. R.G. prepared the figures.

### Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Gallotti, R. *et al.* A stochastic model of randomly accelerated walkers for human mobility. *Nat. Commun.* **7**:12600 doi: 10.1038/ncomms12600 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016