nature human behaviour

LETTERS

https://doi.org/10.1038/s41562-018-0364-x

Evidence for a conserved quantity in human mobility

Laura Alessandretti[1,2], Piotr Sapiezynski[2], Vedran Sekara[2,3], Sune Lehmann[2,4*] and Andrea Baronchelli[1*]

Recent seminal works on human mobility have shown that individuals constantly exploit a small set of repeatedly visited locations[1–3]. A concurrent study has emphasized the explorative nature of human behaviour, showing that the number of visited places grows steadily over time[4–7]. How to reconcile these seemingly contradicting facts remains an open question. Here, we analyse high-resolution multi-year traces of ~40,000 individuals from 4 datasets and show that this tension vanishes when the long-term evolution of mobility patterns is considered. We reveal that mobility patterns evolve significantly yet smoothly, and that the number of familiar locations an individual visits at any point is a conserved quantity with a typical size of ~25. We use this finding to improve state-of-the-art modelling of human mobility[4,8]. Furthermore, shifting the attention from aggregated quantities to individual behaviour, we show that the size of an individual's set of preferred locations correlates with their number of social interactions. This result suggests a connection between the conserved quantity we identify, which as we show cannot be understood purely on the basis of time constraints, and the 'Dunbar number'[9,10] describing a cognitive upper limit to an individual's number of social relations. We anticipate that our work will spark further research linking the study of human mobility and the cognitive and behavioural sciences.

There is disagreement between the current scientific understanding of human mobility as highly predictable and stable over time[1,4,5], and the fact that individual lives are constantly evolving due to changing needs and circumstances[11]. The role of cultural, social and legal constraints on the space–time fixity of daily activities has long been recognized[2,12,13]. Recent studies based on the analysis of human digital traces, including mobile phone records[14,15], online location-based social networks[16–20] and global positioning system (GPS) location data of vehicles[21–26] have shown that individuals universally exhibit a markedly regular pattern characterized by few locations or points of interest[27,28], where they return regularly[6,29] and predictably[4]. However, the observed regularity mainly concerns human activities taking place at the daily[28,30,31] or weekly[14,15,17] time scales, such as commuting between home and the office[14,15,32,33], pursuing habitual leisure activities, and socializing with established friends and acquaintances[16]. Thus, while the role played by slowly occurring changes on the evolution of individuals' social relationships has been widely investigated[34–41], their effects on human mobility behaviour are not well understood and not included in most available models[4,8,42–47].

Here, we investigate individuals' routines across months and years. We reveal how individuals balance the trade-off between the exploitation of familiar places and the exploration of new opportunities. We point out that predictions of state-of-the-art models can be significantly improved if a finite memory is assigned to individuals, and we show that individuals' exploration–exploitation behaviours in the social and spatial domain are correlated.

Our study is based on the analysis of ~40,000 high-resolution mobility trajectories of 2 samples of individuals measured for at least 12 months (see Supplementary Table 1): users of the Lifelog mobile application (Lifelog), traced over 19 months, and participants in a longitudinal experiment, the Copenhagen Networks Study (CNS)[48], spanning 24 months. The results were corroborated with data from two other experiments with fixed-rate temporal sampling but lower spatial resolution and sample size (see Supplementary Table 1): the Mobile Data Challenge dataset from the Lausanne Data Collection Campaign (MDC), which lasted for 19 months[49,50], and the Reality Mining dataset (RM)[51,52], spanning 10 months. Our datasets rely on different types of location data and collection methods (see Methods, Supplementary Note 1.1 and Supplementary Figs. 1–6), but share the high spatial resolution and temporal sampling necessary to capture mobility patterns beyond highly regular ones, such as home–work commuting[53].

All the datasets considered display statistical properties consistent with those reported in previous studies focusing on larger samples but shorter timescales[4,5] (see Supplementary Note 1.2 and Supplementary Figs. 7–9), and their temporal resolution and duration make them ideal for investigating the evolution of individual geo-spatial behaviours on longer timescales. Moreover, three of the datasets considered (CNS, MDC and RM) also include information on individuals' interactions across multiple social channels (phone call, SMS and Facebook), allowing us to connect individuals' spatial and social behaviours across long timescales. Two of the datasets (CNS and RM), collectively describing ~2% of the individuals analysed in this study, consist of the trajectories of university students (both CNS and RM) and faculty members (RM only). These subjects are homogeneous with respect to sociodemographic indicators affecting mobility behaviour[54], and their displacements are constrained by a similar academic schedule. Notwithstanding this possible source of bias, all results presented below hold for the four considered datasets.

Our first finding is that individuals' sets of visited locations grow with characteristic sublinear exponent. When initiating a transition from one place to another, individuals may either choose to return to a previously visited place, or explore a new location. To

---

[1]Department of Mathematics, City, University of London, London, United Kingdom. [2]Technical University of Denmark, Lyngby, Denmark. [3]Sony Mobile Communications, Lund, Sweden. [4]Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark. *e-mail: sljo@dtu.dk; Andrea.Baronchelli.1@city.ac.uk

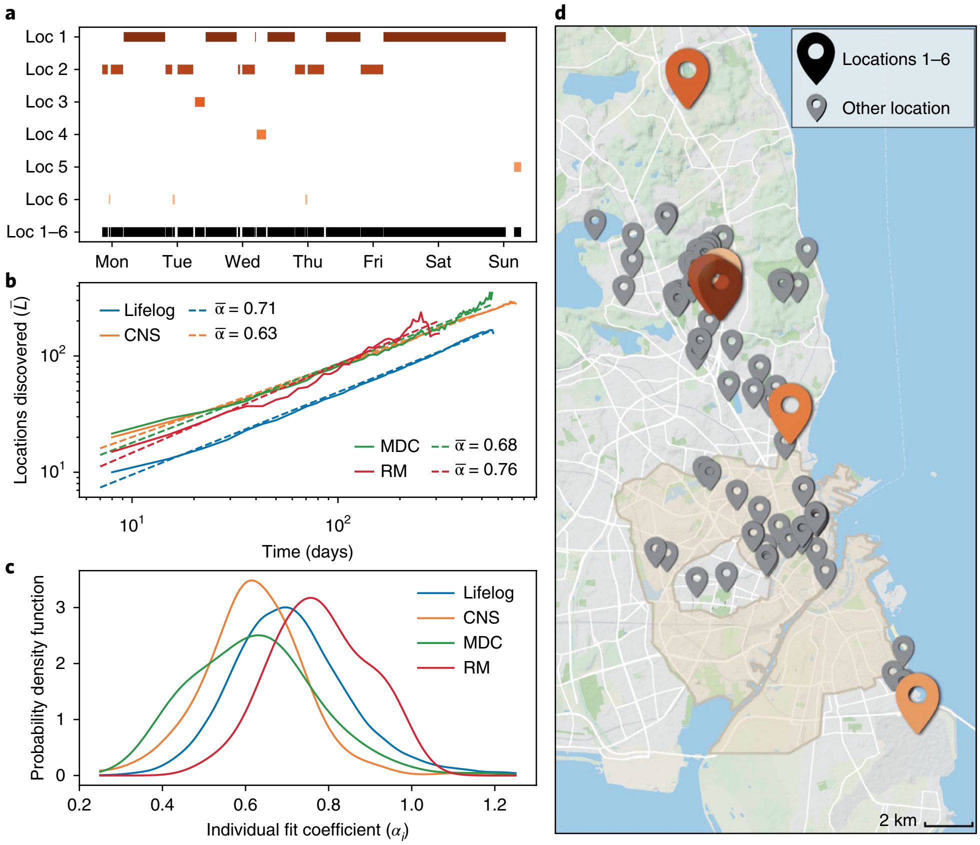

Fig. 1 | Activity set and exploration of new locations. **a**, An example of an individual's mobility trace. The visiting temporal patterns of the six most visited locations (Loc 1, ..., Loc 6) are shown, along with a black trace including all visits to these six locations (Loc 1–6). **b**, Total number of discovered locations L in time. The solid lines show the average across users for each dataset, while the dashed lines show a power-law fitting function with exponent α. **c**, Probability density functions of individuals' power-law fit coefficients for each of the datasets, with peaks around their average values. **d**, Example of an individual's activity set. Locations were plotted as pins on a map, after random noise was added to protect the user's privacy. The six most visited locations are displayed as larger pins using the same colours as in **a**. The light orange area shows the city of Copenhagen. Map data © 2018 Google.

characterize this exploration–exploitation trade-off, we represent individual geo-spatial trajectories as sequences of locations, where 'locations' are defined as places where participants in the study stopped for more than 10 min (Fig. 1a; see also Supplementary Note 1.1). Examples of CNS locations after pre-processing included places offering commercial activities, metro stations, classrooms and other areas within the University campus (see Supplementary Fig. 6). Despite the differences in data spatial resolution, the number of unique locations visited weekly is comparable among all four datasets (see Supplementary Table 2).

A central question concerning the long-term exploration behaviour of the individuals is whether an individual's set of known locations continuously expands or saturates over time. We find that the total number of unique locations $L_i(t)$ an individual i has discovered up to time t grows as $L_i \propto t^{\alpha_i}$ (Fig. 1b), and that individuals' exploration is homogeneous across the populations studied, with α_i peaked around $\overline{\alpha}$ (Lifelog: $\overline{\alpha} = 0.71$; CNS: $\overline{\alpha} = 0.63$; MDC: $\overline{\alpha} = 0.68$; RM: $\overline{\alpha} = 0.76$) (Fig. 1c). This sublinear growth occurs regardless of how the locations are defined, when in time the measurement starts and individuals' ages (see Supplementary Figs. 19–21). This behaviour is a characteristic signature of Heaps' law[55], and consistent with findings from previous studies focusing on shorter time scales[4].

While continually exploring new places, individuals allocate most of their time among a small subset of all visited locations (see Fig. 1d and Supplementary Fig. 10), in agreement with previous research on human mobility behaviour[4,6,29] and time-geography[3,12,56–58]. Hence, at any point in time, each individual is characterized by an activity set containing the locations they visit, as a result of their daily activities[57,59]. This is defined to capture important locations visited multiple times even if each visit is only short[29,60], and is closely related to the concept of 'activity space', which is widely used in geography[59]. Operationally, we define it as the set $\mathrm{AS}_i(t) = \{\ell_1, \ell_2, \ldots, \ell_k, \ldots \ell_C\}$ of locations ℓ_k that individual i visited at least twice and where they spent on average more than 10 min wk^{-1} during a time-window of 20 consecutive weeks preceding time t. The results presented below are robust with respect to variations of this definition, such as changes in the time-window size or definition of a location (Supplementary Note 1.3, Supplementary Figs. 11, 13 and 19, and Supplementary Tables 3, 4 and 6).

Thus, individuals continually explore new places yet they are loyal to a limited number of familiar ones forming their activity set. But how does discovery of new places affect an individual's activity set? We find that the average probability $\overline{P}$ that a newly discovered location will become part of the activity set stabilizes at P^* (CNS: $P^* = 15\%$; Lifelog: $P^* = 7\%$; MDC: $P^* = 15\%$; RM: $P^* = 20\%$) over the long term, indicating that individuals' activity sets are inherently unstable and new locations are continually added. However, over time, individuals may also cease to visit locations that are part of their activity set. The balance between newly added and dismissed familiar locations is captured by the temporal evolution of the activity set, which we characterize by the location capacity and net gain. We define location capacity C_i as the number of an individual's

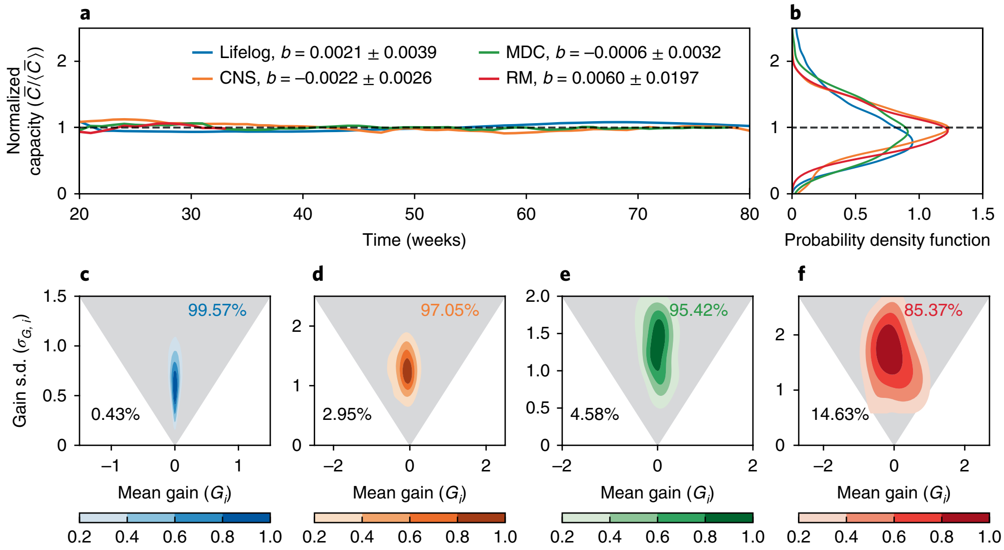

Fig. 2 | Conserved size of evolving activity sets. a, Evolution of the average normalized capacity for the four datasets considered. The dashed black line corresponds to constant capacity. The error on the angular coefficient b of a linear fit, reported in the key, shows that the fit is compatible with a constant line. **b**, Probability density function of individuals' average capacity for the four datasets considered. **c–f**, Gain s.d. $\sigma_{G,i}$ versus the mean gain $\langle G_i\rangle$ for the Lifelog (**c**), CNS (**d**), MDC (**e**) and RM (**f**) datasets. Lines representing cumulative probabilities are obtained through a kernel density estimation from the data. The grey area corresponds to individuals for whom $|\langle G_i\rangle| < \sigma_{G,i}$; that is, whose average gain is compatible with zero. It contains 99.57% (Lifelog), 97.05% (CNS), 95.42% (MDC) and 85.37% (RM) of the population.

familiar locations; that is, the activity set size, at any given moment. The net gain G_i is defined as the difference between the number of locations that are respectively added (A_i) and removed (D_i) at a specific time; hence, $G_i = A_i - D_i$. Figure 2a shows the evolution of the average capacity $\overline{C}$ for the populations considered, normalized to account for the effects due to different data collection methods (see Supplementary Note 1.1).

We find that $\overline{C}$ is constant in time. Using a linear fit of the form $\overline{C} = a + b \times t$, where b represents the slope and a is the intercept, we find that b is not significantly different from 0 (Lifelog: $b = 0.0021 \pm 0.0039$; CNS: $b = -0.0022 \pm 0.0026$; MDC: $b = -0.0006 \pm 0.0032$; RM: $b = 0.0060 \pm 0.0197$). Analogously, a power-law fit of the form $\overline{C(t)} \propto t^{\beta}$ yields β consistent with 0 (Lifelog: $9\times10^{-4} \pm 3\times10^{-2}$; CNS: $-2\times10^{-3} \pm 6\times10^{-2}$; MDC: $-2\times10^{-4} \pm 5\times10^{-3}$; RM: $2\times10^{-3} \pm 3\times10^{-1}$). As a further control, we performed a multiple hypothesis test with false discovery rate correction to compare the averages of the capacity distribution at different times (see Supplementary Table 3). We find no evidence for rejecting the hypothesis that the average capacity does not change in time. Additionally, we find that, for the CNS and Lifelog datasets, the radius of gyration[5] of the activity set—a measure of its spatial extent—is, on average, constant in time (see Supplementary Fig. 31) under the two tests above. Thus, despite individual activity sets evolving over time, the average location capacity is a conserved quantity.

Conservation of the average location capacity may result from either (1) each individual maintaining a stable number of familiar locations over time or (2) a substantial heterogeneity of the populations considered, with certain individuals shrinking their activity sets and others expanding theirs. We tested the two hypotheses by measuring the individual average net gain across time $\langle G_i\rangle$ and its s.d. $\sigma_{G,i}$. If a participant's average gain is closer than one s.d. to 0; hence, $|\langle G_i\rangle|/\sigma_{G,i} < 1$, the net gain is consistent with $\langle G_i\rangle = 0$. If this is true for the majority of individuals, the location capacity is conserved at the individual level and hypothesis (1) holds. In contrast, if $|\langle G_i\rangle|/\sigma_{G,i} \geq 1$, the individual capacity must either increase or decrease in time, supporting hypothesis (2). We find that hypothesis (1) holds for most individuals (Lifelog: 99.57%; CNS: 97.05%; MDC: 95.42%; RM: 85.37%) (Fig. 2c–f; see also Supplementary Table 4). For the large majority of each population, the average net gain of familiar locations added or removed to the activity set at any point is not significantly different from zero; hence, their individual capacity is conserved. Also, we find that the individual capacity has low variability, with the ratio between the average individual capacity and its s.d. $\langle C_i\rangle / \sigma_{C,i}$ typically limited below 30% (Lifelog: 30%; CNS: 28%; MDC: 27%; RM: 14%), demonstrating that fluctuations of the capacity are relatively small. Further evidence suggesting that individual location capacity is conserved is provided in Supplementary Note 1.5 and Supplementary Figs. 33–35.

These results indicate that each individual is characterized by a fixed-size but evolving activity set of familiar locations. We find that the typical size of the activity set saturates at ~25 for increasingly larger values of the time-window defining the activity set (see Supplementary Fig. 12). This value is consistent across all four samples, provided that the differences in time coverage are taken into account by rescaling. Individuals' values are homogeneously distributed around the sample mean (Fig. 2b; see also Supplementary Fig. 14). Previous analyses identified two distinct classes of individuals: 'returners', whose characteristic travelled distance is dominated by movements between few important locations, and 'explorers', characterized by a larger number of places[6]. We observe that 'explorers' typically have higher location capacity than 'returners' (see Supplementary Figs. 8, 9 and 32).

To interpret the information contained in the measured value of the location capacity, we randomized the temporal sequences of locations in two ways, preserving individual routines only up to the daily level. After breaking individual time series into modules of one day in length, we: (1) randomized individual time series, preserving the module/day units (local randomization); or (2) created new sequences by assembling modules extracted randomly by the whole set of individual traces (global randomization; see Supplementary Fig. 22). Due to the absence of temporal correlations, the capacity is also constant in time for the randomized datasets. However, the capacity of the random sets is significantly

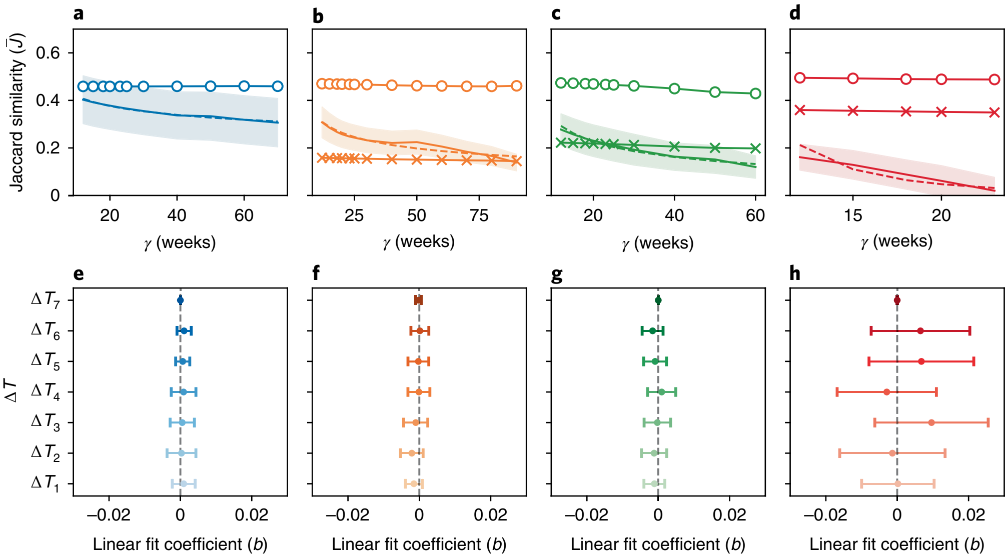

Fig. 3 | Evolution of activity sets and conservation of time allocation. a–d, Average Jaccard similarity $\bar{J}$ between the weekly activity sets measured at t and $t+\gamma$ as a function of γ for data (solid lines), the globally randomized series (lines with crosses) and the locally randomized series (lines with circles). Shaded areas correspond to the 50% interquartile range. Dashed lines correspond to power-law fits $\bar{J} \sim \gamma^{\lambda}$, with $\lambda = -0.15$ for the Lifelog data (**a**), $\lambda = -0.31$ for CNS (**b**), $\lambda = -0.49$ for MDC (**c**) and $\lambda = -3.00$ for RM (**d**). The results were obtained setting a sliding window of length $w = 10$ weeks. The anonymization procedure applied by SONY Mobile before supplying the data makes it impossible to perform the global randomization on the Lifelog trajectories. **e–h**, Linear fit coefficients of the average capacity versus time for several categories of location ΔT are consistent with zero within errors. The intervals considered are: $\Delta T_1 = 10$–30 min wk^{-1}, $\Delta T_2 = 30$–60 min wk^{-1}, $\Delta T_3 = 1$–6 h wk^{-1}, $\Delta T_4 = 6$–12 h wk^{-1}, $\Delta T_5 = 12$–24 h wk^{-1}, $\Delta T_6 = 24$–48 h wk^{-1} and $\Delta T_7 = 48$–168 h wk^{-1}. The results are shown for the Lifelog (**e**), CNS (**f**), MDC (**g**) and RM (**h**) datasets.

higher than in the real-time series for both randomizations under the Kolmogorov–Smirnov test (see Supplementary Table 5), implying that the observed value in real data is not a simple consequence of time constraints. Instead, the fixed capacity is an inherent property of human behaviour.

The time evolution of the activity set supports this finding. We measured the turnover of familiar locations using the Jaccard similarity $J_i(t,\gamma)$ between the weekly activity set at t and at $t+\gamma$ (see Fig. 3a–d). Despite seasonality effects (see Supplementary Figs. 15 and 16), which imply fluctuations around a typical behaviour, J_i does not depend on the initial point, but only on the waiting time γ, and we can consider $J_i(\gamma)$ independently of t (see Supplementary Fig. 17). We find that the average similarity decreases as a power law $\bar{J} \propto \gamma^{\lambda}$ with the coefficient λ significantly different from 0 (Lifelog: $\lambda = -0.15$; CNS: $\lambda = -0.31$; MDC: $\lambda = -0.49$; RM: $\lambda = -3.00$; see Supplementary Fig. 18). Furthermore, the centre of mass of the activity set changes position across time (see Supplementary Fig. 30). In contrast, for the randomized sequences, the Jaccard similarity is constant in time as familiar locations are never abandoned ($\bar{J} \propto \gamma^{0}$). This confirms that individual activity sets change continually, and individual routines evolve gradually in time.

To characterize the structure of the activity set, we investigated how individuals allocate time among different location classes, defined on the basis of their average visit duration. We considered intervals ΔT, with ΔT ranging from 10–30 min wk^{-1} (the time it takes to visit a bus stop or grocery shop) to 48–168 h wk^{-1} (such as for home locations). For each of these location classes, we computed the evolution of the capacity $C_i^{\Delta T}$ and the gain $G_i^{\Delta T}$, and tested the hypothesis $G_i^{\Delta T} = 0$, as above. We find that, although the activity set subsets are continuously evolving (see Supplementary Table 7), $C_i^{\Delta T}$ is conserved for each ΔT (Fig. 3e–h; see also Supplementary Figs. 24–27 and Supplementary Table 6), indicating that the number of places where individuals spend a range of time ΔT does not change over time. This result holds independent of the choice of specific ΔT, and implies that the individual capacity $C_i = \sum C_i^{\Delta T}$, where both C_i and each $C_i^{\Delta T}$ are conserved across time. Thus, both location capacity and time allocation are conserved quantities.

Our results have consequences for the modelling of human mobility. The renown exploration and preferential return model[4,8] describes agents who, when not exploring a new location, return to a previously visited place selected with a probability proportional to the number of former visits. Another state-of-the-art model introduces a mechanism assigning higher return probability to recently visited locations[61]. These models reproduce some of the empirical observations described above, including the conservation of the location capacity (Fig. 2), but fail to describe the time evolution of the activity set (Fig. 3). To overcome this limitation, we start from the observation that the exploitation probability for a location is time dependent[61,62], and endow the agents with a finite memory M so that the probability of returning to a location is based on the number of visits occurred in the past M days. The model including this simple modification qualitatively reproduces all the observations, including the long-term evolution of the activity set (see Supplementary Note 1.4 and Supplementary Fig. 28).

Finally, we analysed the connection between the social and spatial domain. Empirical observations suggest that there are upper limits to the size of an individual's social circle—the so-called Dunbar number[9,10,38,39], due to cognitive constraints[9], and it has been hypothesized that the geography of one's activity set is proportional to one's social network geography[63]. Motivated by these observations, we tested the hypothesis of a correlation between individuals' location capacity and the size of their social circle, as measured by the people contacted by phone (see Fig. 4) and Facebook (see Supplementary Fig. 36) over a period of 20 weeks. We find that a significant positive correlation exists (see Fig. 4). Furthermore, for the CNS dataset, we are able to show that both quantities correlate with the individual personality trait of extraversion[64], which tends to be manifested in outgoing, talkative and energetic behaviour[65] (see Supplementary Fig. 29; Pearson correlation $\rho = 0.22$, two-tailed $P < 10^{-9}$ for location capacity versus extraversion; $\rho = 0.40$, two-tailed $P < 10^{-28}$ for size of

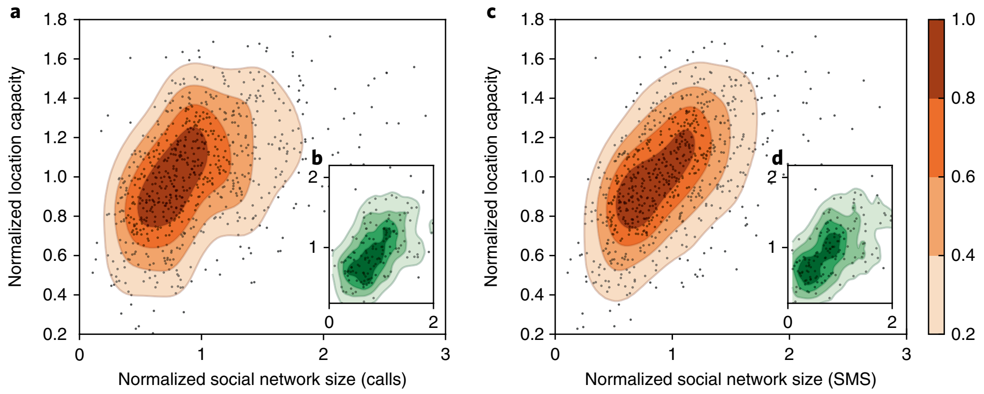

Fig. 4 | Correlation between location capacity and social network size. a–d, Values of individuals' average normalized location capacity versus their normalized social network size computed from phone call interactions (**a** and **b**) and SMS interactions (**c** and **d**) (see Supplementary Fig. 36 for results obtained using Facebook interactions). Data points are indicated by black dots. Coloured filled areas correspond to cumulative probabilities estimated via Gaussian kernel density estimations for visualization purposes. Results are shown for the CNS (**a** and **c**) and MDC (**b** and **d**) datasets. The values of the Pearson correlation coefficient are 0.31 (**a**), 0.48 (**b**), 0.52 (**c**), 0.54 (**d**) (two-tailed $P < 10^{-13}$ in all cases). Social network size is normalized to the population average value.

social network versus extraversion[66]). We consider that these observations call for further analyses on the connections between human social and spatial behaviour.

In summary, we have shown that the number of locations an individual visits regularly is conserved over time, even while individual routines are unstable in the long term because of the continual exploration of new locations. This individual location capacity is peaked around a typical value of ~25 locations across the population, and significantly (typically, at least 30%) smaller than would be expected if only time constraints were at play (see Supplementary Table 5 and Supplementary Fig. 23).

The location capacity is hierarchically structured, indicating that individual time allocation for categories of places is also conserved. These results have allowed us to improve existing models of human mobility, which are unable to fully account for long-term instabilities and fixed-capacity effects.

Taken together, these findings shed new light on the underlying dynamics shaping human mobility, with potential impact for a better understanding of phenomena such as urban development and epidemic spreading.

Extending our scope beyond mobility, we have shown that individuals' location capacity is correlated with the size of their social circles. In this respect, it is interesting to note that fixed-size effects in the social domain[9,10,38,39] have been directly related to human cognitive abilities[9]. We anticipate that our results will stimulate new research exploring this connection.

Methods

Data description. *RM dataset.* The RM project was conducted from 2004–2005 at the Massachusetts Institute of Technology (MIT) Media Laboratory. It measured 94 subjects using mobile phones over the course of 9 months. Of these 94 subjects, 68 were colleagues working in the same building on campus (90% graduate students and 10% staff), while the remaining 26 subjects were incoming students at the university's business school[51]. An application installed on users' phones continuously logged location data from cell tower IDs at fixed-rate sampling. The study was approved by the MIT Committee on the Use of Humans as Experimental Subjects. Subjects were provided with detailed information about the type of information captured, and provided informed consent[67].

CNS dataset. The CNS experiment took place between September 2013 and September 2015[48] and involved 851 Technical University of Denmark students (~22% female and ~78% male), typically aged between 19 and 21 years old. Participants' position over time was estimated combining their smart phones' WiFi and GPS data using the method described in ref. [68] (see also Supplementary Note 1.1 and Supplementary Fig. 6). The location estimation error was below 50 m in 95% of the cases. Data on participants' calls and SMS activity were also collected as part of the experiment. Individuals' background information was obtained through a survey with 310 questions, including the Big Five Inventory[69], measuring five broad domains of human personality traits (openness, conscientiousness, extraversion, agreeableness and neuroticism). Data collection was approved by the Danish Data Protection Agency. All participants provided informed consent.

Lifelog dataset. The Lifelog dataset consists of anonymized GPS location data for ~36,000 users of the Lifelog app between 2015 and 2016. Lifelog users are geo-localized across the world (see Supplementary Note 1.1 and Supplementary Fig. 4), and are aged between 18 and 65 years old, with an average age of 36 years. About one-third of users are female. Data are not collected with a fixed time interval. Instead, the app receives updates when there is a change is the motion-state of the device (if the accelerometer registers a change; see Supplementary Fig. 5). Location estimation error is below 100 m for 93% of data points. To preserve privacy, GPS traces were pre-processed (internally at SONY Mobile) to infer stop-locations using the method described in ref. [70]. This method is built on the idea that a stop corresponds to a temporal sequence of locations within a maximal distance $d_{\max}$ from each other. The results presented are for $d_{\max} = 30$ m. Data collection for the Sony dataset has been approved by the Sony Mobile Logging Board, and informed consent was obtained for all study participants according to the Sony Mobile Application Terms of Service and the Sony Mobile Privacy Policy.

MDC dataset. Data were collected by the MDC between October 2009 and March 2011. The campaign involved a heterogeneous sample of ~185 volunteers with mixed backgrounds from the Lake Geneva region (Switzerland), who were allocated smart phones[50]. In this work, we used Global System for Mobile communications data since they have a higher sampling frequency than the GPS data collected from the same experiment. Following Nokia's privacy policy, individuals participating in the study provided informed consent[50]. The Lausanne Data Collection Campaign involves 62% male and 38% female participants, where the age range of 22–33 years accounts for roughly two-thirds of the population[71].

Code availability. The code used to generate the results of this paper is available from the corresponding authors upon request.

Data availability. The CNS data are not publicly available due to privacy considerations, including European Union regulations and Danish Data Protection Agency rules. Due to the data security of participants, data cannot be shared freely, but are available to researchers who meet the criteria for access to confidential data, sign a confidentiality agreement and agree to work under supervision in Copenhagen. Please direct your queries to S.L. (sljo@dtu.dk).

The MDC data are available from the Idiap Research Institute, but restrictions apply to the availability of these data, which were used under licence for the current study, and so are not publicly available. Data are, however, available from the Idiap Research Institute to eligible institutions upon reasonable request (https://www.idiap.ch/dataset/mdc/download).

Raw data are not publicly available to preserve users' privacy under the Sony Mobile Privacy Policy. Derived data supporting the findings of this study are available from the corresponding authors upon request.

The RM dataset is available from the MIT Human Dynamics Lab (http://realitycommons.media.mit.edu/realitymining4.html).

Received: 21 September 2016; Accepted: 14 May 2018;

References

1. Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).
2. Schwanen, T., Kwan, M.-P. & Ren, F. How fixed is fixed? Gendered rigidity of space–time constraints and geographies of everyday activities. *Geoforum* **39**, 2109–2121 (2008).
3. Golledge, R. G. *Spatial Behavior: A Geographic Perspective* (Guilford Press, London & New York, NY, 1997).
4. Song, C., Koren, T., Wang, P. & Barabási, A.-L. Modelling the scaling properties of human mobility. *Nat. Phys.* **6**, 818–823 (2010).
5. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
6. Pappalardo, L. et al. Returners and explorers dichotomy in human mobility. *Nat. Commun.* **6**, 8166 (2015).
7. Alessandretti, L., Sapiezynski, P., Lehmann, S. & Baronchelli, A. Multi-scale spatio-temporal analysis of human mobility. *PLoS ONE* **12**, e0171686 (2017).
8. Jiang, S. et al. The TimeGeo modeling framework for urban motility without travel surveys. *Proc. Natl Acad. Sci. USA* **113**, E5370–E5378 (2016).
9. Dunbar, R. I. Coevolution of neocortical size, group size and language in humans. *Behav. Brain Sci.* **16**, 681–694 (1993).
10. Gonçalves, B., Perra, N. & Vespignani, A. Modeling users' activity on twitter networks: validation of Dunbar's number. *PLoS ONE* **6**, e22656 (2011).
11. Sarason, I. G., Johnson, J. H. & Siegel, J. M. Assessing the impact of life changes: development of the life experiences survey. *J. Consult. Clin. Psychol.* **46**, 932–946 (1978).
12. Hägerstraand, T. What about people in regional science? *Pap. Reg. Sci.* **24**, 7–24 (1970).
13. Burns, L. D. *Transportation, Temporal, and Spatial Components of Accessibility* (Lexington Books, Lexington, 1980).
14. Csáji, B. C. et al. Exploring the mobility of mobile phone users. *Phys. A Stat. Mech. Appl.* **392**, 1459–1473 (2013).
15. Sevtsuk, A. & Ratti, C. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *J. Urban Technol.* **17**, 41–60 (2010).
16. Cho, E., Myers, S. A. & Leskovec, J. Friendship and mobility: user movement in location-based social networks. In *Proc. 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1082–1090 (ACM, 2011).
17. Cheng, Z., Caverlee, J., Lee, K. & Sui, D. Z. Exploring millions of footprints in location sharing services. In *Proc. 5th International AAAI Conference on Weblogs and Social Media* 81–88 (AAAI, 2011).
18. Brown, C., Lathia, N., Mascolo, C., Noulas, A. & Blondel, V. Group colocation behavior in technological social networks. *PLoS ONE* **9**, e105816 (2014).
19. 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).
20. Bapierre, H., Jesdabodi, C. & Groh, G. Mobile homophily and social location prediction. Preprint at https://arxiv.org/abs/1506.07763 (2015).
21. Giannotti, F. et al. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *VLDB J.* **20**, 695–719 (2011).
22. Scellato, S., Musolesi, M., Mascolo, C., Latora, V. & Campbell, A. T. in *Pervasive Computing. Pervasive 2011. Lecture Notes in Computer Science* Vol. 6696 (eds Lyons K., Hightower J. & Huang E. M.) 152–169 (Springer, Berlin & Heidelberg, 2011).
23. Liang, X., Zheng, X., Lv, W., Zhu, T. & Xu, K. The scaling of human mobility by taxis is exponential. *Phys. A Stat. Mech. Appl.* **391**, 2135–2144 (2012).
24. Gallotti, R., Bazzani, A. & Rambaldi, S. Towards a statistical physics of human mobility. *Int. J. Modern Phys. C* **23**, 1250061 (2012).
25. 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. Theory Exp.* **2010**, P05001 (2010).
26. Jiang, B., Yin, J. & Zhao, S. Characterizing the human mobility pattern in a large street network. *Phys. Rev. E* **80**, 021136 (2009).
27. Mülligann, C., Janowicz, K., Ye, M. & Lee, W.-C. Analyzing the spatial–semantic interaction of points of interest in volunteered geographic information. In *Proc. International Conference on Spatial Information Theory* 350–370 (Springer, 2011).
28. Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R. & Ratti, C. Activity-aware map: identifying human daily activity pattern using mobile phone data. In *Int. Workshop on Human Behavior Understanding* 14–25 (Springer, 2010).
29. Isaacman, S. et al. Identifying important places in people's lives from cellular network data. In *Pervasive Computing. Pervasive 2011. Lecture Notes in Computer Science* Vol. 6696 (eds Lyons K., Hightower J. & Huang E. M.) 133–151 (Springer, Berlin & Heidelberg, 2011).
30. Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z. & González, M. C. Unravelling daily human mobility motifs. *J. R. Soc. Interface* **10**, 20130246 (2013).
31. Bagrow, J. P. & Lin, Y.-R. Mesoscopic structure and social aspects of human mobility. *PLoS ONE* **7**, e37676 (2012).
32. Ranjan, G., Zang, H., Zhang, Z.-L. & Bolot, J. Are call detail records biased for sampling human mobility? *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **16**, 33–44 (2012).
33. Zang, H. & Bolot, J. Anonymization of location data does not work: a large-scale measurement study. In *Proc. 17th Annual International Conference on Mobile Computing and Networking* 145–156 (ACM, 2011).
34. Kossinets, G. & Watts, D. J. Empirical analysis of an evolving social network. *Science* **311**, 88–90 (2006).
35. Kossinets, G. & Watts, D. J. Origins of homophily in an evolving social network 1. *Am. J. Sociol.* **115**, 405–450 (2009).
36. Romero, D. M., Meeder, B., Barash, V. & Kleinberg, J. Maintaining ties on social media sites: the competing effects of balance, exchange, and betweenness. In *Proc. 5th International AAAI Conference on Weblogs and Social Media* (AAAI, 2011).
37. Martin, J. L. & Yeung, K.-T. Persistence of close personal ties over a 12-year period. *Soc. Networks* **28**, 331–362 (2006).
38. Miritello, G., Lara, R., Cebrian, M. & Moro, E. Limited communication capacity unveils strategies for human interaction. *Sci. Rep.* **3**, 1950 (2013).
39. Saramäki, J. et al. Persistence of social signatures in human communication. *Proc. Natl Acad. Sci. USA* **111**, 942–947 (2014).
40. Burt, R. S. Decay functions. *Soc. Networks* **22**, 1–28 (2000).
41. Arnaboldi, V., Conti, M., Passarella, A. & Dunbar, R. Dynamics of personal social relationships in online social networks: a study on twitter. In *Proc. 1st ACM Conference on Online Social Networks* 15–26 (ACM, 2013).
42. Isaacman, S. et al. Human mobility modeling at metropolitan scales. In *Proc. 10th International Conference on Mobile Systems, Applications, and Services* 239–252 (ACM, 2012).
43. Lee, K., Hong, S., Kim, S. J., Rhee, I. & Chong, S. SLAW: a new mobility model for human walks. In *Proc. IEEE INFOCOM 2009* 855–863 (IEEE, 2009).
44. Kim, M., Kotz, D. & Kim, S. Extracting a mobility model from real user traces. In *Proc. 25th IEEE International Conference on Computer Communications* 1–13 (IEEE, 2006).
45. Jia, T., Jiang, B., Carling, K., Bolin, M. & Ban, Y. An empirical study on human mobility and its agent-based modeling. *J. Stat. Mech. Theory Exp.* **2012**, P11024 (2012).
46. 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).
47. Pappalardo, L., Rinzivillo, S. & Simini, F. Human mobility modelling: exploration and preferential return meet the gravity model. *Procedia Comput. Sci.* **83**, 934–939 (2016).
48. Stopczynski, A. et al. Measuring large-scale social networks with high resolution. *PLoS ONE* **9**, e95978 (2014).
49. Kiukkonen, N., Blom, J., Dousse, O., Gatica-Perez, D. & Laurila, J. Towards rich mobile phone datasets: Lausanne Data Collection Campaign. In *Proc. ACM International Conference on Pervasive Services* (ICPS, 2010).
50. Laurila, J. K. et al. The mobile data challenge: big data for mobile computing research. In *Proc. Workshop on the Nokia Mobile Data Challenge, in Conjunction with the 10th International Conference on Pervasive Computing* EPFL-CONF-192489 (2012).
51. Eagle, N. & Pentland, A. S. Reality mining: sensing complex social systems. *Pers. Ubiquit. Comput.* **10**, 255–268 (2006).
52. Eagle, N., Pentland, A. S. & Lazer, D. Inferring friendship network structure by using mobile phone data. *Proc. Natl. Acad. Sci. USA* **106**, 15274–15278 (2009).
53. Çolak, S., Alexander, L. P., Alvim, B. G., Mehndiretta, S. R. & González, M. C. Analyzing cell phone location data for urban travel: current methods, limitations and opportunities. In *Proc. Transportation Research Board 94th Annual Meeting* 15-5279 (2015).
54. Lenormand, M. et al. Influence of sociodemographic characteristics on human mobility. Preprint at https://arxiv.org/abs/1411.7895 (2014).
55. Heaps, H. S. *Information Retrieval: Computational and Theoretical Aspects* (Academic Press, Orlando, 1978).
56. Horton, F. E. & Reynolds, D. R. Effects of urban spatial structure on individual behavior. *Econ. Geogr.* **47**, 36–48 (1971).
57. Mazey, M. E. The effect of a physio-political barrier upon urban activity space. *Ohio J. Sci.* **81,** 212–217 2981).
58. Yuan, Y. & Raubal, M. Analyzing the distribution of human activity space from mobile phone usage: an individual and urban-oriented study. *Int. J. Geogr. Inf. Sci.* **30**, 1594–1621 (2016).
59. Sherman, J. E., Spencer, J., Preisser, J. S., Gesler, W. M. & Arcury, T. A. A suite of methods for representing activity space in a healthcare accessibility study. *Int. J. Health Geogr.* **4**, 24 (2005).

60. Zhou, C., Bhatnagar, N., Shekhar, S. & Terveen, L. Mining personally important places from GPS tracks. In *Proc. 2007 IEEE 23rd International Conference on Data Engineering Workshop* 517–526 (IEEE, 2007).
61. Barbosa, H., de Lima-Neto, F. B., Evsukoff, A. & Menezes, R. The effect of recency to human mobility. *EPJ Data Sci.* **4**, 21 (2015).
62. Szell, M., Sinatra, R., Petri, G., Thurner, S. & Latora, V. Understanding mobility in a social petri dish. *Sci. Rep.* **2**, 457 (2012).
63. Axhausen, K. W. Activity spaces, biographies, social networks and their welfare gains and externalities: some hypotheses and empirical results. *Mobilities* **2**, 15–36 (2007).
64. Costa, P. T. & McCrae, R. R. Four ways five factors are basic. *Pers. Individ. Diff.* **13**, 653–665 (1992).
65. Kalish, Y. & Robins, G. Psychological predispositions and network structure: the relationship between individual predispositions, structural holes and network closure. *Soc. Networks* **28**, 56–84 (2006).
66. Pollet, T. V., Roberts, S. G. & Dunbar, R. I. Extraverts have larger social network layers: but do not feel emotionally closer to individuals at any layer. *J. Individ. Diff.* **32**, 161–169 (2011).
67. Eagle, N. *The Reality Mining Data* (Massachusetts Institute of Technology, 2010).
68. Sapiezynski, P., Gatej, R., Mislove, A. & Lehmann, S. Opportunities and challenges in crowdsourced wardriving. In *Proc. 2015 Internet Measurement Conference* 267–273 (ACM, 2015).
69. John, O. P. & Srivastava, S. in *Handbook of Personality: Theory and Research* 2nd edn (eds Pervin, L. & John, O. P.) 102–138 (Guilford, New York, NY, 1999).
70. Cuttone, A., Lehmann, S. & Larsen, J. E. Inferring human mobility from sparse low accuracy mobile sensing data. In *Proc. 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* 995–1004 (ACM, 2014).
71. Laurila, J. K. et al. From big smartphone data to worldwide research: the Mobile Data Challenge. *Pervasive Mob. Comput.* **9**, 752–771 (2013).

Acknowledgements

This work was partially supported by the Villum Foundation (High Resolution Networks project, for which S.L. is the principal investigator), a UCPH-2016 grant (Social Fabric project, for which S.L. is a co-principal investigator) and the Danish Council for Independent Research (Microdynamics of Influence in Social Systems project, for which S.L. is the principal investigator; grant ID 4184-00556). Portions of the research in this paper used the MDC Database made available by the Idiap Research Institute, Switzerland and owned by Nokia. V.S. was supported by Sony Mobile Communications. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. V.S. thanks H. Jonsson for invaluable technical assistance.

Author contributions

L.A., S.L. and A.B. designed the research. L.A., P.S. and V.S. pre-processed the data. L.A. performed the data analysis. L.A., S.L. and A.B. analysed the results and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-018-0364-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.L. or A.B.