

## MIT Open Access Articles

*Unravelling daily human mobility motifs*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Schneider, C. M., V. Belik, T. Couronne, Z. Smoreda, and M. C. Gonzalez. "Unravelling Daily Human Mobility Motifs." *Journal of The Royal Society Interface* 10, no. 84 (April 24, 2013): 20130246–20130246.

**As Published:** <http://dx.doi.org/10.1098/rsif.2013.0246>

**Publisher:** Royal Society, The

**Persistent URL:** <http://hdl.handle.net/1721.1/89420>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike



# Unraveling Daily Human Mobility Motifs

Christian M. Schneider,<sup>1</sup> Vitaly Belik,<sup>1,2</sup> Thomas Couronné,<sup>3</sup> Zbigniew Smoreda,<sup>3</sup> and Marta C. González<sup>1</sup>

<sup>1</sup>*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology,  
77 Massachusetts Avenue, Cambridge, MA 02139, USA*

<sup>2</sup>*Max Planck Institute for Dynamics and Self-Organization, Am Faßberg 17, 37077 Göttingen, Germany*

<sup>3</sup>*Sociology and Economics of Networks and Services department,  
Orange Labs, 38 rue du Général Leclerc, 92794 Issy les Moulineaux, France*

(Dated: March 17, 2013)

Human mobility is differentiated by time scales. While the mechanism for long-time scales has been studied, the underlying mechanism on the daily scale is still unrevealed. Here, we uncover the mechanism responsible for the daily mobility patterns by analyzing the temporal and spatial trajectories of thousands of persons as individual networks. By using the concept of motifs from network theory only 17 unique networks are present in daily mobility, following simple rules. These networks, called here motifs, are sufficient to capture up to 90% of the population in surveys and mobile phone data sets for different countries. Each individual exhibits a characteristic motif, which seems to be stable over several months. Consequently, daily human mobility can be reproduced by an analytically tractable framework for Markov chains by modeling periods of high frequency trips followed by periods of lower activity as the key ingredient.

Our modern society and the environment are shaped by people's mobility patterns at different scales. The long-time and long-distance trips consist generally of rare and infrequent events like international flights or movements between cities. In contrast short time trips mostly consist of intracity travels like commuting to work or grocery shopping. These trips exhibit high regularity, typically following the daily circadian rhythm. Studies of human mobility at large scales, motivated by understanding the global spreading of epidemics [1–6], have unraveled interesting properties of the underlying mobility patterns.

Nowadays, large scale human mobility patterns are described by three widely accepted indicators: the trip distance distribution  $p(r)$ , the radius of gyration  $r_g(t)$ , and the number of visited locations  $S(t)$  over time [7–9]. The trip distance distribution of the entire population follows a power-law  $p(r) \sim r^{-\beta}$  with  $\beta \approx 1.59$ [7].

Individual trajectories can be extracted from mobile phone data [10–13]. This enables the study of the area an individual visits which is characterized by the radius of gyration  $r_g(t)$  [8]. This individual  $r_g$  can be understood as the characteristic distance an individual travels during a given time-period  $t$ . The distribution of the radius of gyration reveals heterogeneity in the population; most individuals travel within a short radius, but others cover long distances on a regular basis. Thus each individual follows  $p(r)$  within his characteristic distance  $r_g(t)$ . The distribution  $p(r_g)$  within the population yields to the power-law observed in the aggregated trip distance distribution  $p(r)$ .

The frequent return to previously visited locations is captured by the number of visited places over time  $S(t)$ . This value grows sub-linearly as  $S(t) \sim t^\mu$  with  $\mu = 0.6$  capturing individuals tendency for revisiting locations [9]. These three measures contain the basic ingredients to describe the individual trajectories, in which frequent travels occur between a limited number of places, with less

frequent trips to new places outside an individual radius. Such behaviour for large time scales can be reproduced by an exploration and preferential return model with the displacement distribution as an input [9] which can be used to model epidemic spreading on the airline network [14].

However, the current model is designed to capture the long-term mobility behaviour. For example, the number of visited locations  $S(t)$  does not show a robust scaling exponent  $\mu$ , for  $t < 24h$  [9]. Additionally, the radius of gyration stabilizes only after a few months of observation [8]. These indications suggest different underlying mechanisms for modeling mobility at the intercity and the intracity scale.

Current studies at the daily, intracity scale focus on forecasting traffic demand and on predicting human decisions based on optimizing a score function or a utility function. Such modeling approaches assume that each individual human tries to minimize his/her effort depending on socioeconomic characteristics [15]. Therefore, agent-based models have been deployed, usually based on detailed data from travel surveys [16–20]. These models include diverse ingredients, like distances between places to precisely reproduce individual trajectories [21–23].

In the present article we investigate the common underlying mechanisms for daily human mobility patterns by combining the advantages of different large scale data sources. In each dataset we observe ubiquitous daily mobility patterns, which we statistically reproduce with an analytical model. Since the generated patterns of our model are only sensitive to the presence or absence of periods of activity followed by periods of inactivity, it implies that humans' daily trips follow a universal law.

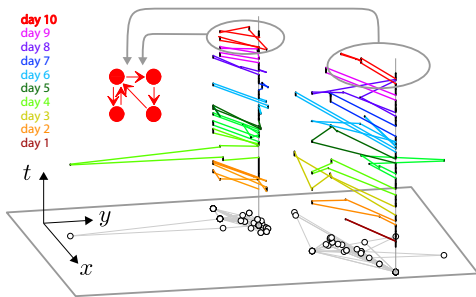


FIG. 1. Decomposition of the mobility profile over 10 days into daily mobility patterns for two anonymous mobile phone users. The home location of the users are highlighted and connected over the entire observation period with a gray line. While the entire mobility profiles (black circles and gray lines in the  $xy$ -plane) are rather diverse, the individual daily profiles (brown to red from bottom to top for different days) share common features. The aggregated networks consists of  $N = 16(22)$  nodes and  $M = 37(43)$  edges with an average degree of  $\langle k \rangle = 2M/N = 4.6(3.9)$ . In contrast the daily average number of nodes is  $\langle N \rangle = 4.4 \pm 1.8(3.9 \pm 1.3)$  and the average number of edges is  $\langle M \rangle = 5.3 \pm 2.8(4.2 \pm 2.2)$ . The left user prefers commuting to one place and visits the other locations during a single tour, while the right user prefers to visit his daily locations during a single tour. On the last day, both users visit not only four locations but also share the same daily profile consisting of two tours with one and two destinations, respectively.

## I. HUMAN MOBILITY PATTERNS

Human mobility is characterized by a sequence of visited locations and the trips among them. As an example we show in Fig. 1 the aggregated mobility profile of two users and their corresponding daily profiles for a ten-day observational period. The time-dependent trajectories for different days are coloured from brown (1st day) to red (10th day) from bottom to top. The black circles and gray lines in the  $xy$ -plane are the projection of the daily trajectories. Both the daily and the aggregated profiles can be described as directed networks, in which nodes represent the visited locations and directed edges stands for trips between them. To classify these networks on a daily basis, we further discard any additional information about the purpose of the activity, the travel time and the activity duration as well as the distances between the visited locations, consequently neither the nodes nor the edges are weighted. Only the trip direction is incorporated by the direction of the edge as highlighted for the last day in Fig. 1.

We first investigate the distribution of the number of different visited locations, which is the size distribution of the daily networks. As shown in Fig. 2 the size distributions  $f(N)$  of the networks is similar for all datasets (see section Perturbation Model for more details). The shape of the observed distributions  $f(N)$  can be approximated

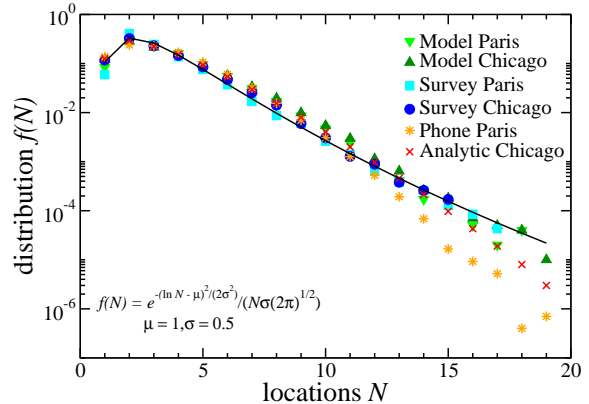


FIG. 2. Daily human mobility patterns seem to follow a universal law. The daily number of visited location can be approximated with a log-normal distribution  $f(N) = \exp(-(\ln N - \mu)^2 / (2\sigma^2)) / (\sqrt{2\pi}N\sigma)$  with  $\mu = 1$  and  $\sigma = 0.5$ . The distributions extracted from activity and travel surveys as well as from mobile phone billing data show similar behaviour. Moreover, the distributions of our perturbation model (see section Model and Fig. 6 for details) generated both analytically and numerically, have the same shape. The broad distribution shows that although most of the people visit less than five locations, a small fraction behave significantly different since people report to visits up to 17 different places within a day in our surveys. Note that due to the mobile phone data limitations the tail of the corresponding distribution is below the other data sets.

by a log-normal distribution

$$f(N) \sim e^{-(\ln N - \mu)^2 / (2\sigma^2)} / (\sigma x \sqrt{2\pi}) \quad (1)$$

with the parameters  $\mu = 1 \pm 0.1$  and  $\sigma = 0.5 \pm 0.1$ . The average number of locations  $\langle N \rangle \approx 3$  is small, hence most people visit only a few locations. In fact 90% of the population visit less than seven locations on a daily basis. All three data sets follow the same distribution despite the difference between the cities and if the dataset is a travel survey or phone data.

To further study the observed daily mobility patterns, the number of different daily networks is investigated. These networks unravel whether people prefer to visit different locations in a single round tour before returning to the starting location, or if they prefer to return to their starting location before visiting another location. In fact, for a given network size  $N$ ,  $N_p$  edge combinations exist:

$$N_p(N) = 2^{N^2 - N}. \quad (2)$$

Since we are interested in networks that picture human daily trips, the number of reasonable networks can be significantly reduced mainly due to two constraints: the need for sleep and the consistency of trips. The need for sleep imposes that the trips start and finish at the

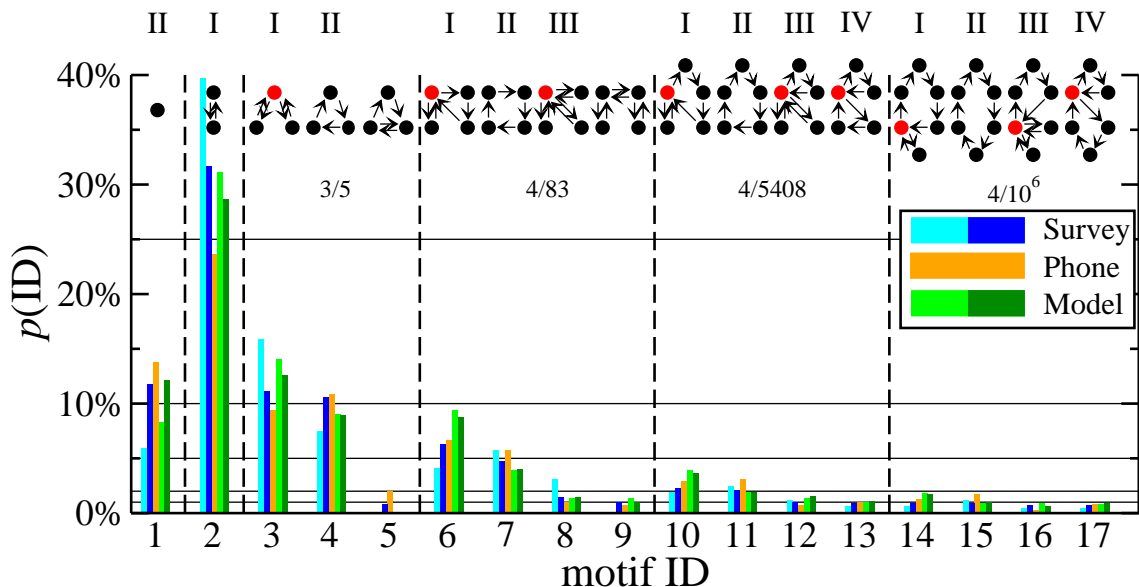


FIG. 3. Possible daily mobility patterns are limited, since up to 90% of the identified daily mobility networks can be described with only 17 different motifs. The probability  $p(\text{ID})$  to find one of these 17 motifs in the surveys (cyan - Paris, blue - Chicago), the phone data (orange - Paris), and the model (light green - Paris, dark green - Chicago) is presented. The motifs are grouped according to their size separated by dashed lines. For each group the fraction of observed over feasible motifs  $N_o/N_f$  is shown and the central nodes are highlighted. Most motifs can be classified by four rules: I) Motifs of size  $N$  consist of a tour with only one stop and another tour with  $N - 2$  stops. II) Motifs of size  $N$  consist of only a single tour with  $N$  stops. III) Motifs of size  $N$  consist of two tours with one stop and another tour with  $N - 3$  stops. IV) Motifs of size  $N$  consist of a tour with two stops and another tour with  $N - 3$  stops. Despite the fact that the number of workers is significantly different in both cities, the rank and the probability to find a specific motif exhibit similar behaviour.

same location, most likely at home. The consistency ensures that each of the  $N$  locations is visited at least once. These two conditions imply that for  $N > 1$  all nodes have at least one ingoing and one outgoing edge. By counting the number of feasible daily networks that fulfill these two constraints, we obtain a large number  $N_f$  increasing rapidly with the number of locations ( $N_f(1) = 1$ ,  $N_f(2) = 1$ ,  $N_f(3) = 5$ ,  $N_f(4) = 83$ ,  $N_f(5) = 5,048$ ,  $N_f(6) = 1,047,008$ ). Nevertheless, up to 90% of the measured trips can be described with only 17 different daily networks for the surveys and the mobile phone data.

We call these 17 daily networks motifs in analogy to motifs in complex networks [24]. Many systems represented as networks consist of various subnetworks, either topological or temporal [25]. If these subnetworks occur more often than in randomized versions of the entire network, these subnetworks are called motifs. Since, randomized versions of the mobility networks are not feasible, we call motifs the daily networks which are found on average more often than 0.5% in the datasets (see ESM for further networks). Consequently, nearly the entire aggregated mobility network of a population can be constructed with these motifs.

In Fig. 3 the motifs obtained from Chicago and Paris surveys, mobile phone data from Paris, and our proposed model are compared. They are ordered by their size and their frequency of occurrence. Although the data sources

cover different cities from different countries the frequencies to observe a specific motif behave similarly. We can suppose that the extracted motifs are general daily mobility characteristics that can be further used to model and simulate urban activity. The most common motif (ID 2) consists of two visited locations and two trips among them, followed by a motif with only a single location (ID 1). The next likely motifs are three locations with four trips, all starting and ending at the same location (ID 3), or with one round trip (ID 4). Interestingly, in none of the datasets a motif with size  $N$  and more than  $N + 2$  trips is observed.

All motifs have at most one 'central' location, defined as a node with more than two directed edges, except the motifs with ID 5 and ID 9. This central node is the origin for a tour  $T(x)$ , a trip visiting  $x$  other locations before returning to the origin with  $x < N$ . The presence of a unique central node ensures that the edges of the motifs belong to exactly one tour. Hence, multiple trips along the same directed edge are suppressed and the entire motifs are composed of a single Eulerian cycle: it is possible to visit all edges exactly once and this path ends at the starting node.

The motifs can be classified by four rules:

- I)  $T(1)$  and  $T(N - 2)$
- II)  $T(N - 1)$

III)  $T(1)$ ,  $T(1)$ , and  $T(N-3)$

IV)  $T(2)$  and  $T(N-3)$

The rule that describes each motif is written on the top of Fig. 3. If a rule leads to a motif with a tour  $T(x)$  visiting a negative number of nodes  $x < 0$ , the motif is forbidden. In contrast, if a rule leads to a tour visiting no nodes, only this tour  $T(0)$  is ignored. For a given number of locations  $N$ , the likelihood of observing a motif is related with the rule number, thus the most likely motif can be described with the first rule. For  $N \leq 6$  the upper limit of daily tours is three, thus the larger the size of the motif the more trips within a tour. Furthermore, we have found that the most common daily networks with more than six locations also follow these rules (see ESM).

Previous results on human predictability [13], as shown in the trajectories in Fig. 1, suggest that each individual have a typical daily motif, thus the observed motifs are similar over several days. To verify this stability, the correlations between motifs of individual users are studied based on phone data, because our surveys provide only information for up to two days. The observed sequence is compared with the sequence of an average user based on the distribution from Fig. 3. In Fig. 4 the correlations are shown:

$$C_{ij} = \begin{cases} \frac{N(i)N(j)}{N_r(i)N_r(j)} - 1 & \text{for } \frac{N(i)N(j)}{N_r(i)N_r(j)} > 1 \\ 1 - \frac{N_r(i)N_r(j)}{N(i)N(j)} & \text{for } \frac{N(i)N(j)}{N_r(i)N_r(j)} \leq 1 \end{cases} \quad (3)$$

with the observed  $N(i)$  and average  $N_r(i)$  number of motifs with ID  $i$ . First, the highest correlation of each motif is the self-correlation  $C_{ii}$  which is usually 10 to 30 times more likely than expected by selecting individual motifs according to the observed distribution. Second, the likelihood to find a motif with similar number of visited places with small variations ( $\pm 2$  locations) behave like the average, but for higher differences the probability is significantly suppressed. Additionally, active users  $N > 4$  seem to be active during the entire observational period, since they have significant higher probability to visit any motif with  $N > 4$ . Interestingly, within the blocks of motifs with size four, five and six some correlations are suppressed or enhanced. We observe that the correlations are enhanced, if both motifs follow the same rule with different number of visited locations  $N$ , for example visiting all locations within one tour. In contrast, the correlations are suppressed, if the motifs are less similar; i.e. if the number of tours differs by more than one. In fact, this is observed for motifs created according to rule II) and III).

In general, motifs may not be unique, because a person may repeat a tour several times within a day. However, the repetition of tours is uncommon, thus an edge corresponds to exactly one trip. In the survey data, the observed motifs without multiple trips are sufficient to reproduce over 95% of the travel behaviour correctly and we observe that tours  $T(x)$  with  $x > 1$  are performed

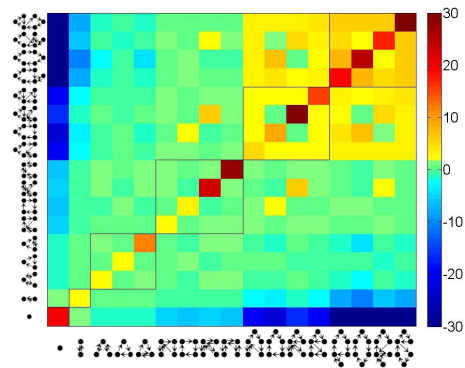


FIG. 4. Daily human mobility patterns are stable over several months. The values, calculated by equation (3), show how more or less likely a motif is found during the observation period of 6 months under the condition that the individual has a given motif on another day. Positive values (yellow to red colours) indicate that these motifs are more likely than expected and negative values (cyan to blue colours) that these motifs are suppressed. The probability to find the same daily motif during another day is significantly larger compared to the randomized dataset. Additionally, active users, which visit more than four locations per day, seem to be active over time, while inactive users remain inactive. The emerging patterns of transitions between active motifs could be explained by the similarity of motifs. While transitions between motifs of group II are preferred, transitions between the groups II and III are suppressed, since the number of tours is most different. As a guide to the eye, motifs with the same number of locations are marked with boxes.

only once during a day (for details see ESM).

These observations imply that each person has a characteristic daily motif although the visited locations can change. Thus a user has a personal number of preferred places on a daily basis, which are most likely visited in a specific sequence given by its characteristic motif.

## II. PERTURBATION-BASED MODEL

It is surprising that nearly the entire population can be described with a few unique daily motifs. To understand this observation, we study the time spent at certain locations as well as the time between the starting time of an activity and the next activity of the same kind.

From both surveys the frequency of staying at a place for a particular time period is extracted for three groups of activities, home, work, and other, as shown in Fig. 5a. The time spent for working and staying at home is relatively flat distributed with some characteristic durations of 3.5h and 8.6h at work, and 14h at home, respectively. In contrast the probability of an activity at another place decreases with its duration. This staying time distribution has no characteristic duration, suggesting that the location changes are not distributed evenly over time, but in groups interspersed with periods of inactivity. To

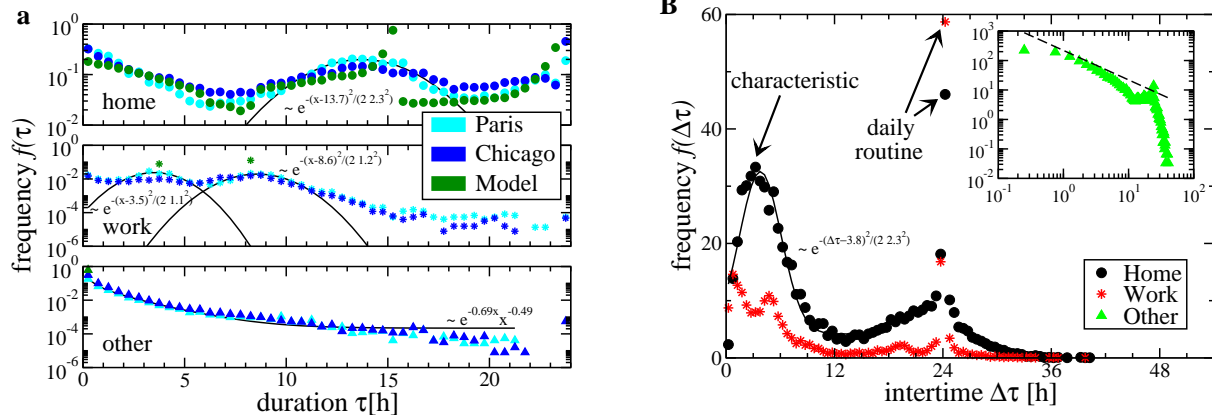


FIG. 5. Fundamental differences between home/work and other locations. a) The duration spend at either home or work is relatively flat distributed with peaks around characteristic time spans of 14h at home as well as 3.5h and 8.6h at work. In contrast the time spend at other activities is broadly distributed. For a guide to the eye, Gaussian distributions are fitted around the characteristic durations for home/work locations and a power-law with an exponential cutoff is fitted for other locations. Our model captures these main characteristics. b) The frequency of observing an inter-event time  $\tau$  between the beginning of two similar activities, if another location has been visited in between. For the home and work location daily routines dominate the distribution with additional characteristic times. In contrast the distribution for other locations exhibits a broad distribution dominated by short inter-event times with a suppressed daily routine. For a guide to the eye, the characteristic inter-event time between home location is approximated by a Gaussian distribution and in the inset a power-law with exponent  $-1$  is included.

support this observation, we study the time between two similar activities, shown in Fig. 5b. While the time based on home and work are governed by the daily routines, the time between other locations follows a broad distribution. Such short inter-event time dynamics has been reported in specific human activities such as web browsing, printing patterns, email, and phone communication [26–37], but it has not been incorporated in models of human mobility. Inspired by these observations, we developed a perturbation-based model, to reproduce not only the observed daily motifs, but also their frequency of occurrence.

In the following the model for a non-working (NW) agent is explained and additional, minor features for working (W) agents are described in the ESM. Accounting for the difference of home and other locations, the model assumes a fixed activity at home and any number of flexible activities elsewhere (shopping, recreation, etc.). Agents prefer staying at home and perform other activities as a kind of perturbation only, thus they return home after finishing a flexible activity, if they have no other flexible activity scheduled. On the other hand, when people are already perturbed, it is more likely that they perform another flexible activity afterwards (e.g. after having dinner in the city, visiting a nearby bar).

In the model the day is divided into  $K = 48$  30-minute intervals. The actual number of discrete time slots is insignificant as long as it is larger than the maximal number of visited locations  $K > 20$ . For each of these time slots, the agent receives a task with the corresponding time-

dependent probability  $p_{NW}(t)$ , and assigns it to the next free time slot. Initially all time slots are free, beside a 9h sleeping period during night. Since most tasks occur and are executed during daytime, we use the simple assumption that the probability to receive a task is related to the circadian daily rhythm. This rhythm is approximated by the normalized phone activity  $p(t)$  of the entire population as shown in Fig. 6b:  $\int_{0h}^{24h} p(t)dt = \gamma_{NW}$  with the parameter  $\gamma_{NW}$  (see ESM). The most important ingredient for modeling the observed motifs from surveys and phone data is the assumption that after receiving a task  $p_{NW}(t) = p(t)$ , the probability to get another task  $p_{NW}(t+1) = \alpha p(t+1)$  for the next time slot is significantly higher,  $\alpha > 1$  as shown in Fig. 6c. For simplicity, we increase the probability by one order of magnitude  $\alpha = 10$ . This ensures that the inter-event time distribution of flexible activities is dominated by short times as observed in Fig. 5b and generates the daily tours. In Fig. 6a-d an example of modeling a non-working agent is shown. The peaks in  $p_{NW}(t)$  in Fig. 6c correspond to activities outside home.

Note that the model has no assumptions about the locations of the individual tasks, their number, or the number of trips. Only the average number of different visited locations is controlled by the parameter  $\gamma_{NW}$  and the fraction of working and non-working agents is preset. However, this is sufficient to reproduce the overall behaviour of the data as shown in Figs. 2, 3, and 5. Additionally, the model also reproduces the fraction of trips between home, work, and other locations with an absolute error

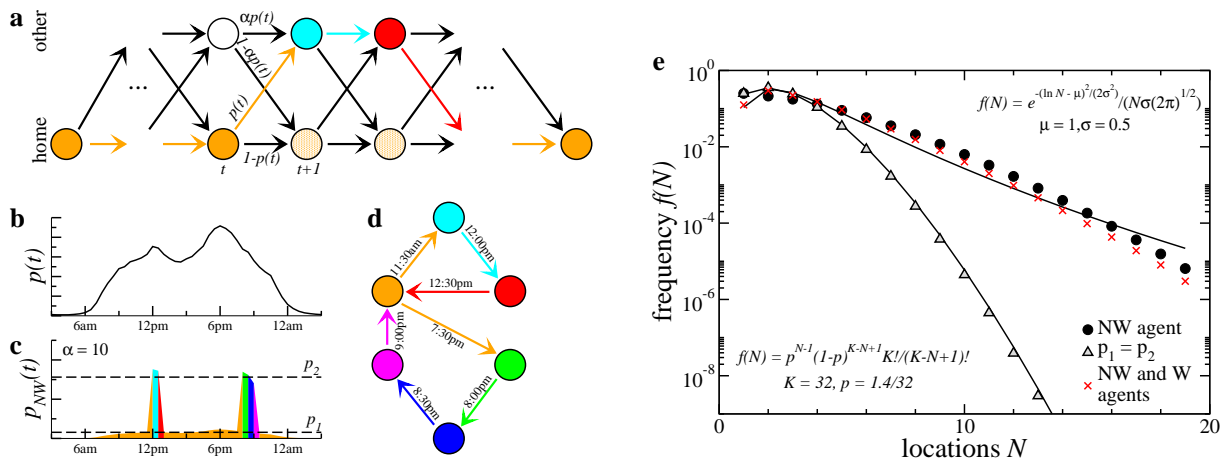


FIG. 6. The introduced model is illustrated for a non-working agent. a) The possible trajectories of an agent are shown. The agent starts the day at home and finishes it at home. At a given time  $t$ , depending on the actual location of the agent, the probability to be at home at time  $t + 1$  is either  $1 - p(t)$  or  $1 - \alpha p(t)$  with a parameter  $\alpha$  and with probability  $p(t)$  and  $\alpha p(t)$  the agent travel to another location. The filled circles and the coloured path is an exemplary trajectory. The time-dependent probabilities can be related to the circadian rhythm of activity, shown in b). c) The location dependent probabilities for the exemplary agent with  $\alpha = 10$  are shown. The time-dependent probabilities can be approximated by only two values,  $p_1$  and  $p_2$  for being at home and being at another location. With this approximation, the model can be solved analytically. d) The exemplary trajectory is converted into the corresponding motif with six locations and seven trips among them. e) The daily number of visited location obtained from the analytical model under three different conditions is shown. While the removal of workers does not change the tail of the log-normal distribution,  $\alpha = 1$  leads to a binomial distribution, thus periods of activities are key for the observed behaviour. Note that the absolute difference between analytical and numerical model is less than 0.01.

of at most 2% (see ESM).

The model can be treated analytically by mapping it on a coin flipping or independent non-identical Bernoulli trials problem, with the reasonable assumption of only two different probabilities  $p_1 = \langle p(t) \rangle$  and  $p_2 = 10p_1$  instead of a time-dependent variable (see Fig. 6c). A person, having  $K$  free time slots, flips a coin to change the location in the next slot. A success  $H$  leads to stay at home or return home, while failures  $T$  lead to the exploration of new locations. The coin flipping occurs with different probabilities dependent on the current state:

$$\begin{aligned} H &\xrightarrow{p_1} T \\ H &\xrightarrow{1-p_1} H \\ T &\xrightarrow{p_2} T \\ T &\xrightarrow{1-p_2} H. \end{aligned}$$

By applying the modified finite Markov chain embedding technique [38] for independent non-identical Bernoulli trials, the probability for the number  $N$  of locations visited during a day or equivalently the number of successes  $P(N)$  after  $K$  Bernoulli trials can be written as

$$P(N) = \xi_0 \left( \prod_{t=1}^K \Lambda_t \right) U'(C_N), \quad (4)$$

with  $\xi_0$  being an initial condition vector in the state space of the corresponding Markov chain,  $\Lambda_t$  is the transition probability matrix, and  $U'(C_N)$  is a transposed vector

corresponding to the subspace with  $N$  successes (for details see ESM). As one can see in Fig. 2 this simple coin flipping model can reproduce the empirical findings very well.

To confirm that the assumption  $\alpha \ll 1$  is the key to get the broad distribution of the number of daily visited locations, we show in Fig. 6e the analytical results for three different models: one with two kind of agents, one with only non-workers, and one with only one probability  $\alpha = 1$ . While the presence of two kinds of agents have a minor impact on the overall motifs and their size distribution, the removal of the perturbation ( $p_2 = p_1$ ) changes the results from an approximately log-normal size distribution, to a binomial size distribution. Moreover, not only the motif distribution changes, but different motifs which are not present in the surveys, mostly star-like ones, emerge. Therefore, the 'perturbed' behaviour  $p_2 = 10p_1$  is the crucial ingredient to reproduce daily mobility.

### III. FINAL REMARKS

Advances in transforming large data into meaningful information are essential to improve our understanding of socio-technical systems. In our study we contribute to this end by analyzing networks of daily trips obtained from individual's surveys and anonymized mobile phone data. We found that both travel surveys and phone traces

from two different cities reveal the same set of ubiquitous networks, that we called motifs. We can suppose that these motifs are general human mobility characteristics that can be further used to model and simulate urban activity. Besides, we found that perturbed states with periods of high activity follows by periods of low activity, is the indispensable ingredient to correctly reproduce those motifs. We remark that due to the limited observation period of at most two days in our survey, the question whether a heavy tail occurs in the inter-event time distribution in Fig. 5b remains open.

Our model successfully reproduces the frequency of visiting different locations and the occurrence rate of the motifs, but it is designed for a single day and therefore it does not incorporate the correlations of motifs between different days. The model captures main characteristics of the duration spent at home by assuming fixed duration for the other activities. The model's inter-event time distributions share some common features with the data, but due to the duration differences as well as the restriction to a single day, it can not accurately reproduce the observed distributions (see ESM).

The future avenues for related research are diverse. Understanding daily routines promises a better assessment of planning and control, which is the core interest of urban and epidemiological applications. Our findings reduce the dimensionality of choices in agent-based modeling helping to enhance current urban simulators [39, 40]. In epidemic spreading, usually only up to three locations, daily visited by a host, are considered in modeling contagious dynamics [41–44]. Thus, our presented insights can straightforwardly extend mobility in current epidemiological models.

#### IV. MATERIALS AND METHODS

To identify motifs we use three different datasets: a survey and mobile phone billing data from Paris and a survey from Chicago [45]. In the surveys 23,764 and 23,429 weekdays of people were selected in such a way that the data is representative for the entire population of Chicago and Paris, respectively. In the Chicago survey each participant answered a questionnaire with his/her activity information for one or two entire day, containing the following information: weekday, duration, location, reason, and mode of trip. With this information it is possible to reproduce the entire daily activity patterns of the anonymous individuals. The Paris survey has the same information, but instead of geographical locations only the trip length are provided. Since weekday and weekend behaviour can be rather different, we focus in this study only on weekdays.

From phone billing data of millions of mobile phone users

the extraction of relevant information needs preprocessing. The phone company provides information about the incoming and outgoing calls and short-message-services (SMS). Thus we have locations of the operating towers, time of the events and user identification numbers (UID). With this information, we reconstruct daily mobility networks of the users during a six month period. The main challenge is converting call information into the corresponding mobility profile of a user. Therefore, only the 39,820 most active users are investigated according the following scheme (the rules are visualized in ESM Fig. 1 and ESM Fig. 2):

- 1) The day, starting at 3am, is divided into 48, 30-minutes slots for each of the 154 days.
- 2) To remove towers, which are only used during travel, all towers which are less frequently visit than a certain threshold are ignored; in this study less than 0.5% during the entire observational period.
- 3) To eliminate signal transitions between neighboring towers, these towers are merged for one day, if more than three back and forward transitions between them are recorded during a single day.
- 4) To remove towers used during travel on daily basis, records are taken into account only if the next records have the same tower location.
- 5) To identify an activity location, only the most frequently observed location during each time slot is assigned as an activity location for this time slot.
- 6) A day is discarded, if less than a certain number (in this case eight) of time slots exhibit location information. A too small number would favor smaller motifs, while a too large threshold would exclude too many individuals. The results are stable for different threshold values.
- 7) To overcome the small number of night calls the location, which is visited most frequently during all nights between 12pm and 6am of a single user, is assigned as the user's home location; in our survey this assumption correctly identifies over 98% of the home locations for a single day. User starts and finishes its day at home, if the user has no other information in the corresponding night time slots at 3am and 3.5am.
- 8) Based on the activity locations for each time slot, the motifs shown in Fig. 3 are constructed for weekdays only. We have published c++ code of our proposed model, the algorithms how to identify motifs and simulated data to test all algorithms on our website [46].

#### V. ACKNOWLEDGMENT

VB kindly acknowledges the financial support by the Volkswagen Foundation. This work was funded by New England UTC Year 23 grant, awards from NEC Corporation Fund, the Solomon Buchsbaum Research Fund.

[1] Anderson, R. M. & May, R. M. 1992 *Infectious Diseases in Humans*, Oxford University Press.

[2] Lloyd, A. L. & May, R. M. 2001 How viruses spread among computers and people. *Science* **292**, 1316. (DOI

- 10.1126/science.1061076)
- [3] Hufnagel, L., Brockmann, D. & Geisel, T. 1992 Forecast and control of epidemics in a globalized world *Proc. Natl. Acad. Sci. USA* **101**, 15124. (DOI 10.1073/pnas.0308344101)
  - [4] Colizza, V., Pastor-Satorras, R. & Vespignani, A. 2007 Reaction-diffusion processes and metapopulation models in heterogeneous networks *Nat. Phys.* **3**, 276. (DOI 10.1038/nphys560)
  - [5] Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J. & Vespignani, A. 2009 Multiscale mobility networks and the spatial spreading of infectious diseases *Proc. Natl. Acad. Sci. USA* **106(51)**, 21484-21489. (DOI 10.1073/pnas.0906910106)
  - [6] Belik, V., Geisel, T. & Brockmann, D. 2011 Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases. *Physical Review X* **1**, 011001. (DOI 10.1103/PhysRevX.1.011001)
  - [7] Brockmann, D., Hufnagel, L. & Geisel, T. 2006 The scaling laws of human travel *Nature* **439**, 462-465. (DOI 10.1038/nature04292)
  - [8] González, M. C., Hidalgo, C. A. & Barabási, A.-L. 2008 Understanding individual human mobility patterns. *Nature* **453(7196)**, 779-782. (DOI 10.1038/nature06958)
  - [9] Song, C., Koren, T., Wang, P. & Barabási, A.-L. 2010 Modelling the scaling properties of human mobility. *Nature Physics* **6**, 818823. (DOI 10.1038/nphys1760)
  - [10] Eagle, N. & Pentland, A. 2006 Reality mining: sensing complex social systems. *Personal and ubiquitous computing* **10**, 255-268. (DOI 10.1007/s00779-005-0046-3)
  - [11] Onnela, J., Saramaki, J., Hyvonen, J., Szabo, G., Lazer, D., Kaski, K., Kertész, J. & Barabasi, A.-L. 2007 Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. USA* **104**, 7332-7336. (DOI 10.1073/pnas.0610245104)
  - [12] Eagle, N., Pentland, A. & Lazer, D. 2009 Inferring friendship network structure by using mobile phone data. *Proc. Natl. Acad. Sci. USA* **106**, 1527415278. (DOI 10.1073/pnas.0900282106)
  - [13] Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. 2010 Limits of predictability in human mobility. *Science* **327**, 1018-1021. (DOI 10.1126/science.1177170)
  - [14] Nicolaides, C., Cueto-Felgueroso, L., González, M. C. & Juanes, R. A. 2012 Metric of Influential Spreading during Contagion Dynamics through the Air Transportation Network *PLoS ONE* **7(7)**, e40961. (DOI 10.1371/journal.pone.0040961)
  - [15] Zipf, G. K. 1949 *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley, Cambridge, MA.
  - [16] Ben-Akiva, M. E. & Bowman, J. L. 1998 Integration of an activity-based model system and a residential location model *Urban Studies* **35(7)**, 1231-1253. (DOI 10.1080/0042098984529)
  - [17] Kitamura, R., Chen, C., Pendyala, R. M. & Narayanan, R. 2000 Microsimulation of activity-travel patterns for travel demand forecasting *Transportation* **27(1)**, 2551. (DOI 10.1023/A:1005259324588)
  - [18] Keuleers, B., Wets, G., Arentze, T. & Timmermans, H. 2001 Association rules in identification of spatial-temporal patterns in multiday activity diary data *Transportation Research Record* **1752**, 3237. (DOI 10.3141/1752-05)
  - [19] Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfuser, G. & Haupt, T. 2002 Observing the rhythms of daily life: A six-week travel diary *Transportation* **29(2)**, 95-124. (DOI 10.1023/A:1014247822322)
  - [20] Schlich, R. & Axhausen, K. W. 2003 Habitual travel behaviour: Evidence from a six-week travel diary *Transportation* **30(1)**, 1336. (DOI 10.1023/A:1021230507071)
  - [21] Beckman, R. J., Baggerly, K. A. & McKay, M. D. 1996 Creating synthetic baseline populations *Transportation Research Part A Policy and Practice* **30(6)**, 415429. (DOI 10.1016/0965-8564(96)00004-3)
  - [22] Arentze, T., Hofman, F., van Mourik, H. & Timmermans, H. 2000 ALBATROSS: Multiagent, rule-based model of activity pattern decisions *Transportation Research Record* **1706**, 136144. (DOI 10.3141/1706-16)
  - [23] Charypar, D. & Nagel, K. 2005 Generating complete all-day activity plans with genetic algorithms. *Transportation*, **32(4)**, 369397. (DOI 10.1007/s11116-004-8287-y)
  - [24] Alon, U. 2007 Network motifs: theory and experimental approaches *Nature Reviews Genetics* **8**, 450-461. (DOI 10.1038/nrg2102)
  - [25] Kovanen, L., Karsai, M., Kaski, K., Kertész, J. & Saramäki 2011 Temporal motifs in time-dependent networks *J. Stat. Mech.* P111005 (DOI 10.1088/1742-5468/2011/11/P11005)
  - [26] Eckmann, J., Moses, E. & Sergi, D. 2004 Entropy of dialogues creates coherent structures in e-mail traffic *Proc. Natl. Acad. Sci. USA* **101**, 14333-14337. (DOI 10.1073/pnas.0405728101)
  - [27] Barabási, A.-L. 2005 The origin of bursts and heavy tails in human dynamics *Nature* **435**, 207-211. (DOI 10.1038/nature03459)
  - [28] Harder, U. & Paczuski, M. 2006 Correlated dynamics in human printing behavior *Physica A* **361(1)**, 329-336. (DOI 10.1016/j.physa.2005.06.079)
  - [29] Candia, J., González, M. C., Wang, P., Schoenhar, T., Madey, G. & Barabási, A.-L. 2008 Uncovering individual and collective human dynamics from mobile phone records *J. Phys. A: Math. Theor.* **41(22)**, 224015. (DOI 10.1088/1751-8113/41/22/224015)
  - [30] Malmgren, R. D., Stouffer, D. B., Motter, A. E., Amaral & L. A. N. 2008 A Poissonian explanation for heavy tails in e-mail communication *Proc. Natl. Acad. Sci. USA* **105**, 18153-18158. (DOI 10.1073/pnas.0800332105)
  - [31] Iribarren, J. L. & Moro, E. 2009 Impact of Human Activity Patterns on the Dynamics of Information Diffusion *Phys. Rev. Lett.* **103**, 038702. (DOI 10.1103/PhysRevLett.103.038702)
  - [32] Miller, K. J., Sorensen, L. B., Ojemann, J. G. & den Nijs, M. 2009 Power-Law Scaling in the Brain Surface Electric Potential *PLoS computational biology* **5(12)**, e1000609. (DOI 10.1371/journal.pcbi.1000609)
  - [33] Rybski, D., Buldyrev, S. V., Havlin, S., Liljeros, F. & Makse, H. A. 2009 Scaling laws of human interaction activity *Proc. Natl. Acad. Sci. USA* **106**, 12640. (DOI 10.1073/pnas.0902667106)
  - [34] Ratkiewicz, J., Fortunato, S., Flammini, A., Menczer, F. & Vespignani, A. 2010 Characterizing and Modeling the Dynamics of Online Popularity *Phys. Rev. Lett.* **105**, 158701. (DOI 10.1103/PhysRevLett.105.158701)
  - [35] Hennig, H., Fleischmann, R., Fredebohm, A., Hagmayer, Y., Nagler, J., Witt, A., Theis, F. & Geisel, T. 2011 The nature and perception of fluctuations in human musical rhythms *PLoS One* **6(10)**, e26457. (DOI 10.1371/jour-

- nal.pone.0026457)
- [36] Jo, H.-H., Karsai, M., Kertész, J. & Kaski, K. 2012 Circadian pattern and burstiness in mobile phone communication *New J. Phys.* **14**, 013055. (DOI 10.1088/1367-2630/14/1/013055)
- [37] Karsai, M., Kaski, K., Barabási, A.-L. & Kertész, J. 2012 Universal features of correlated bursty behaviour *Scientific Reports* **2**, 397. (DOI 10.1038/srep00397)
- [38] Fu, J. C. & Koutras, M. V. 1994 Distribution Theory of Runs: A Markov Chain Approach *Journal of the American Statistical Association* **89**, 1050-1058. (DOI 10.1080/01621459.1994.10476841)
- [39] <http://www.matsim.org/>
- [40] <http://code.google.com/p/transims/>
- [41] Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C. & Burke, D. S. 2006 Strategies for mitigating an influenza pandemic *Nature* **442**, 448-452. (DOI 10.1038/nature04795)
- [42] Viboud, C., Bjornstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A. & Grenfell, B. T. 2006 Synchrony, waves, and spatial hierarchies in the spread of influenza *Science* **312**, 447-451. (DOI 10.1126/science.1125237)
- [43] Ciofi degli Atti, M. L., Merler, S., Rizzo, C., Ajelli, M., Massari, M., Manfredi, P., Furlanello, C., Tomba, G. S. & Iannelli, M. 2008 Mitigation Measures for Pandemic Influenza in Italy: An Individual Based Model Considering Different Scenarios *PLoS ONE* **3(3)**, e1790. (DOI 10.1371/journal.pone.0001790)
- [44] Ajelli, M. et al. 2010 Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models *BMC Infectious Diseases* **10**, 190. (DOI 10.1186/1471-2334-10-190)
- [45] <http://www.cmap.illinois.gov/travel-tracker-survey>
- [46] <http://humnet.scripts.mit.edu/wordpress/downloads/>