Modelling individual routines and spatio-temporal trajectories in human mobility

Luca Pappalardo and Filippo Simini

Abstract Human mobility modelling is of fundamental importance in a wide range of applications, such as the developing of protocols for mobile ad hoc networks or for what-if analysis and simulation in urban ecosystems. Current generative models generally fail in accurately reproducing the individuals' recurrent daily schedules and at the same time in accounting for the possibility that individuals may break the routine and modify their habits during periods of unpredictability of variable duration. In this article we present DITRAS (DIary-based TRAjectory Simulator), a framework to simulate the spatio-temporal patterns of human mobility in a realistic way. DITRAS operates in two steps: the generation of a mobility diary and the translation of the mobility diary into a mobility trajectory. The mobility diary is constructed by a Markov model which captures the tendency of individuals to follow or break their routine. The mobility trajectory is produced by a model based on the concept of preferential exploration and preferential return. We compare DITRAS with real mobility data and synthetic data produced by other spatio-temporal mobility models and show that it reproduces the statistical properties of real trajectories in an accurate way.

Luca Pappalardo
ISTI, CNR, Italy
Department of Computer Science, University of Pisa, Italy
E-mail: lpappalardo@di.unipi.it, luca.pappalardo@isti.cnr.it

Filippo Simini
Department of Engineering Mathematics, University of Bristol, UK
E-mail: f.simini@bristol.ac.uk

1 Introduction

Understanding the complex mechanisms governing human mobility is of fundamental importance in different contexts, from public health [8, 29] to official statistics [31, 40], urban planning [55, 29, 10] and transportation engineering [20]. In particular, human mobility modelling has attracted a lot of interest in recent years for two main reasons. On one side, it is crucial in performance analysis of networking protocols such as mobile ad hoc networks, where network users' displacements are exploited to route and deliver the messages [9, 23, 17]. On the other side human mobility modelling is crucial for urban simulation and what-if analysis [33, 24], e.g., simulating changes in urban mobility after the construction of a new infrastructure or when traumatic events occur like epidemic diffusion, terrorist attacks or international events. In both scenarios the developing of generative models that reproduce human mobility patterns in an accurate way is fundamental to design more efficient and suitable protocols, as well as to design smarter and more sustainable infrastructures, economies, services and cities.

Clearly, the first step in human mobility modelling is to understand how people move. The availability of big mobility data, such as massive traces from GPS devices and mobile phone networks, offers nowadays the possibility to observe human movements at large scales and in great detail [36, 15]. Many studies relied on this opportunity to provide a series of novel insights on the quantitative spatio-temporal patterns characterizing human mobility. These studies observe that human mobility is characterized by a stunning heterogeneity of human travel patterns, i.e., a heavy tail distribution in both trip distances [5, 15, 36] and the characteristic distance traveled by individuals, the so-called radius of gyration [15, 36, 39]. Moreover human mobility is characterized by a high degree of predictability [50], a strong tendency to spend most of the time in just a few locations [15, 36, 50, 51], and a propensity to visit specific locations at specific times [30, 21, 43].

Building upon the above findings, many generative human mobility models have been proposed which try to reproduce the characteristic properties of human mobility trajectories [23]. The goal of generative models of human mobility is to create a population of synthetic individuals whose mobility patterns are statistically indistinguishable from those of real individuals. Typically each generative model focuses on just a few properties of human mobility. A class of models aims to realistically represent spatial properties: they are mainly concerned with reproducing the trip distance distribution [5, 50] or the visitation frequency to a set of preferred locations [50, 39, 3]. Another class of models focus on the accurate representation of the time-varying behavior of individuals, relying on detailed schedules of human activities [21, 43]. However, the major challenge for generative models lies in the creation of realistic temporal patterns, in which various temporal statistics observed empirically are simultaneously reproduced, including the number and sequence of visited locations together with the time and duration of the visits. In particular, the biggest hurdle consists in the simultaneous description of an individual's routine and

sporadic mobility patterns. Currently there is no model able to reproduce the individuals' recurrent or quasi-periodic daily schedules, and at the same time to allow for the possibility that individuals may break the routine and modify their habits during periods of unpredictability of variable duration.

In this work we present DITRAS (DIary-based TRAjectory Simulator), a framework to simulate the spatio-temporal patterns of human mobility. The key idea of DITRAS is to separate the temporal characteristics of human mobility from its spatial characteristics. In order to do that, DITRAS operates in two steps. First, it generates a mobility diary which captures the temporal patterns of human mobility by specifying the arrival time and the time spent in each location visited by the individual. We propose a diary generator where the mobility diary is generated through a Markov model that is constructed by a data-driven procedure based on real mobility data. The Markov model captures the propensity of individuals to follow quasi-periodic daily schedules as well as to break the routine and modify their mobility habits. Second, DITRAS transforms the mobility diary into a mobility trajectory by using proper mechanisms for the exploration of locations on the mobility space, so capturing the spatial patterns of human movements. The trajectory generator we propose embeds mechanisms to explore new locations and return to already visited locations, where the exploration phase takes into account both the distance between locations and their relevance on the mobility space, though taking into account the underlying urban structure and the distribution of population density. We instantiate DITRAS using the proposed diary and trajectory generators and compare it with nation-wide mobile phone data and synthetic trajectories produced by other mobility models on a set of nine different standard mobility measures. We show that DITRAS simulates the main spatio-temporal properties of human mobility in a realistic manner, typically reproducing the mobility patterns of real individuals better than the other considered models. Moreover, we show that the distribution of standard mobility measures can be accurately reproduced only by modelling both the spatial and the temporal aspects of human mobility. In other words, the spatial mechanisms and the temporal mechanisms have to be modeled together by proper diary and trajectory generators in order to reproduce the observed human mobility patterns in an accurate way. The model we propose, able to capture both the spatial and the temporal dimensions of human mobility, is hence a useful tool to develop more reliable protocols for ad hoc networks as well as to perform realistic simulation and what-if scenarios in urban contexts.

This paper is organized as follows. Section 2 revises the relevant literature on human mobility modelling. In Section 3 we present the structure of the DITRAS framework. Section 4 describes the first step of DITRAS, the generation of the mobility diary, and in Section 4.1 we propose a diary generator based on Markov models. Section 5 describes the second step of DITRAS, the generation of the mobility trajectory, and in Section 5.1 we propose a trajectory generator called d-EPR. Section 6 shows the comparison between an instantiation of DITRAS with the proposed diary and trajectory generators with real trajectory

data and the trajectories produced by other mobility models. In Section 6.3 we discuss the obtained results and, finally, Section 7 concludes the paper.

2 Related Work

All the main studies in human mobility document a stunning heterogeneity of human travel patterns that coexists with a high degree of predictability: individuals exhibit a broad spectrum of mobility ranges while repeating daily schedules dictated by routine [15, 51, 11]. Brockmann et al. [5] study the scaling laws of human mobility by observing the circulation of bank notes in United States, finding that travel distances of bank notes follow a power-law behavior. González et al. [15] analyze a nation-wide large-scale mobile phone dataset and find a large heterogeneity in human mobility ranges: (i) travel distances of individuals follow a power-law behavior, confirming the results by Brockmann et al.; (ii) the radius of gyration of individuals, i.e., their characteristic traveled distance, follows a power-law behavior with an exponential cutoff. Song et al. [50] observe on mobile phone data that individuals are characterized by a power-law behavior in waiting times, i.e. the time between a displacement and the next displacement by an individual. Pappalardo et al. [36] find the same mobility patterns as González et al. [15] and Song et al. [50] on a dataset storing the GPS traces of 150,000 private vehicles traveling during one month in Tuscany, Italy. Song et al. [51] study the entropy of individuals' movements and find a high predictability in human mobility, with a distribution of users' predictability peaked at approximately 93% and having a lower cutoff at 80%. Pappalardo et al. [39] analyze mobile phone data and GPS tracks from private vehicles and discover that individuals split into two profiles, returners and explorers, with distinct mobility and geographical patterns. Several studies focus on the prediction of the kind of activity associated to individuals' trips on the only basis of the observed displacements [30, 21, 43], and to discover geographic borders according to recurrent trips of private vehicles [44, 53], or to predict the formation of social ties [7, 54]

How to combine the discovered patterns to create a mathematical model that reproduces the salient aspects of human mobility is an open task. This task is particularly challenging because mobility models should be as simple, scalable and flexible as possible, since they are generally purposed to large-scale simulation and what-if analysis [23]. In literature many individual human mobility models have been proposed so far.

Some models try to reproduce the heterogeneity of individual human mobility and simulate how individuals visits old and new locations. ORBIT [14] is an example of such models. It splits into two phases: (i) at the beginning of the simulation it generates a predefined set of locations on a bi-dimensional space; (ii) then every synthetic individual selects a subset of these locations and moves between them according to a Markov chain. In the Markov chain every state represents a specific location in the scenario and proper probability of transitions guarantee realistic distribution of location frequencies. The

SLAW (Self-similar Least-Action Walk) model [27, 26] produces mobility traces having specific statistical features observed on human mobility data, namely power-law waiting times and travel distances with heavy-tail distribution. In a first step SLAW generates a set of locations on a bi-dimensional space so that the distance among them features a heavy-tailed distribution. Then, a synthetic individual starts a trip by randomly choosing a location as starting point and making movement decisions based on the LATP (Least-Action Trip Planning) algorithm. In LATP every location has a probability to be chosen as next location that decreases with the power-law of the distance to the synthetic individual's current location. SLAW is used in several studies of networking and human mobility modelling and is the base for other models of human mobility, such as SMOOTH [34], MSLAW [46] and TP [49, 48].

The SWIM (Small World In Motion) model [25] is based on the concept of location preference. First, each synthetic individual is assigned to a *home* location, which is chosen uniformly at random on a bi-dimensional space. Then the synthetic individual selects a destination for the next move depending of the weight of each location, which grows with the popularity of the location and decreases with the distance from the home location. The popularity of a location depends on a collective preference calculated as the number of other people encountered the last time the synthetic individual visited the location.

In contrast with many human mobility models, the Exploration and Preferential Return (EPR) model [50] does not fix in advance the number of visited locations on a bi-dimensional space but let them emerge spontaneously. The model exploits two basic mechanisms that together describe human mobility: exploration and preferential return. Exploration is a random walk process with a truncated power-law jump size distribution [50]. Preferential return reproduces the propensity of humans to return to the locations they visited frequently before [15]. A synthetic individual in the model selects between these two mechanisms: with a given probability the synthetic individual returns to one of the previously visited places, with the preference for a location proportional to the frequency of the individual's previous visits. With complementary probability the synthetic individual moves to a new location, whose distance from the current one is chosen from the truncated power-law distribution of travel distances as measured on empirical data [15]. The probability to explore decreases as the number of visited locations increases and, as a result, the model has a warmup period of greedy exploration, while in the long run individuals mainly move around a set of previously visited places. Recently the EPR model has been improved in different directions, such as by adding information about the recency of location visits during the preferential return step [3], or adding a preferential exploration step to account for the collective preference for locations and the returners and explorers dichotomy, as in the d-EPR model [39, 37]. It is worth noting that although the models described above are able to reproduce accurately the heterogeneity of mobility patterns, none of them can reproduce realistic temporal patterns of human movements.

Recent research on human mobility show that individuals are characterized by a high regularity and the tendency to come back to the same few locations

over and over at specific times [15,36]. Temporal models focus on these temporal patterns and try to reproduce accurately human daily activities, schedules and regularities.

Zheng et al. [57] use data from a national survey in the US to extract realistic distribution of address type, activity type, visiting time and population heterogeneity in terms of occupation. They first describe streets and avenues on a bi-dimensional space as horizontal and vertical lines with random length, and then use the Dijkstra's algorithm to find the shortest path between two activities taking into account different speed limits assigned to each street. The WDM (Working Day Movement) model [12] is a scalable approach able to distinguish between inter-building and intra-building movements. It consists indeed of several different submodels to describe mobility in home, office, evening and different transportation means. For example a home model reproduces a sojourn in a particular point of a home location while an office model reproduces a star-like trajectory pattern around the desk of an individual at specific coordinates inside an office building. Although Zheng et al.'s and WDM models provide an extremely thorough representations of human movements in particular scenarios, they suffer two main drawbacks: (i) they represent specific scenarios and their applicability to other scenarios is not guaranteed; (ii) they are too complex for analytical tractability; (iii) they generally fail in capturing some global mobility patterns observed in individual human mobility, e.g., the distribution of radius of gyration. A recent study [32] proposes methods to identify and predict departures from routine in individual mobility using information-theoretic metrics, such as the instantaneous entropy, and developing a Bayesian framework that explicitly models the tendency of individuals to break from routine.

Position of our work. From the literature it clearly emerges that existing mobility models are not able to accurately capture at the same time the heterogeneity of human travel patterns and the temporal regularity of human movements. On the one hand exploration models accurately reproduce the heterogeneity of human mobility but do not account for regularities in human temporal patterns. On the other hand temporal models accurately reproduce human mobility schedules paying the price in complexity, but fail in capturing some important global mobility patterns observed in human mobility. In this paper we try to fill this gap and propose DITRAS, a scalable mobility model that creates synthetic individual trajectories able to capture both the heterogeneity of human mobility and the regularity of human movements. The modelling framework to generate synthetic mobility trajectories proposed in this work can be integrated in transportation forecast models to infer trip demand. Our approach has some similarity with activity-based models [4], as they both aim to estimate trip demand by reproducing realistic individual temporal patterns, however there are important differences between the two approaches. In fact, while activity-based models' goal is to produce detailed agendas filled with activities performed by the synthetic individuals and are calibrated on surveys with a limited number of participants, our framework

produces abstract mobility diaries containing the time and duration of the visits in the various locations without explicitly specifying the type of activity performed there, and is calibrated on a large population of mobile phone users.

A recent paper to appear in PNAS introduces TimeGeo, a modelling framework to generate a population of synthetic agents with realistic spatio-temporal trajectories. Similarly to the modelling framework presented here, TimeGeo combines a Markov model to generate temporal patterns with the correct periodicity and duration of visits, with a modified version of the EPR model to reproduce spatial patterns with the characteristic number of visits and distribution of distances. Albeit having similar aims, there are important differences between our modelling approach and TimeGeo's. In fact, while TimeGeo proposes a parsimonious model which is based on few tunable parameters and is to some extent analytically tractable, the approach proposed in this paper is markedly data driven, with a greater level of complexity which ensures the necessary flexibility to reproduce realistic temporal patterns.

3 Generating mechanism of DITRAS

DITRAS is a modelling framework to simulate the spatio-temporal patterns of human mobility in a realistic way. The key idea of DITRAS is to separate the temporal characteristics of human mobility from its spatial characteristics. For this reason, DITRAS consists of two main phases (Fig. 1): first, it generates a mobility diary which captures the temporal patterns of human mobility; second it transforms the mobility diary into a sampled mobility trajectory which captures the spatial patterns of human movements. In this section we define the main concepts which constitute the mechanism of DITRAS.

The output of a DITRAS simulation is a sampled mobility trajectory for a synthetic individual. A mobility trajectory describes the movement of an object as a sequence of time-stamped locations. The location is described by two coordinates, usually a latitude-longitude pair or ordinary Cartesian coordinates, as formally stated by the following definition:

Definition 1 (Mobility trajectory) A mobility trajectory is a sequence of triples $T = \langle (x_1, y_1, t_1), \ldots, (x_n, y_n, t_n) \rangle$, where t_i $(i = 1, \ldots, n)$ is a timestamp, $\forall_{1 \leq i < n} \, t_i < t_{i+1}$ and x_i, y_i are coordinates on a bi-dimensional space.

For modelling purposes it is convenient to define a sampled mobility trajectory, $S^{(t)}$, which can be obtained by sampling the mobility trajectory at regular time intervals of length t seconds:

Definition 2 (Sampled mobility trajectory) A sampled mobility trajectory is a sequence $S^{(t)} = \langle l_1, \ldots, l_N \rangle$, where l_i $(i = 1, \ldots, N)$ is the geographic location where the individual spent the majority of time during time slot i, i.e., between $(i - 1)t$ and ti seconds from the first observation. N is the total number of time slots considered. A location l_i is described by coordinates on a bi-dimensional space.

To generate a sampled mobility trajectory DITRAS exploits two probabilistic models: a diary generator and a trajectory generator (see Fig. 1). In this paper we propose as diary generator $MD^{(t)}$, a Markov model responsible for reproducing realistic temporal mobility patterns, such as the distribution of the number of trips per day and the tendency of individuals to change location at specific hours of the day [15, 21]. Essentially, $MD^{(t)}$ captures the tendency of individuals to follow or break a mobility temporal routine at specific times. As trajectory generator we propose the density-Exploration and Preferential Return model or d-EPR [39, 37], a generative model able to reproduce realistic spatial mobility patterns, such as the heavy-tail distributions of trip distances [5, 15, 36] and radii of gyration [15, 36, 39], as well as the characteristic visitation patterns, such as the uneven distribution of time spent in the various locations [50, 36]. The d-EPR embeds a mechanism to choose a location to visit on a bi-dimensional space given the current location, the spatial distances between locations and the relevance of each location.

Figure 1 provides an outline of DITRAS. DITRAS is composed of two main steps. During the first step, the diary generator builds a mobility diary D of N time slots, each of duration t. The operation of this step is described in detail in Section 4. During the second step, DITRAS uses the trajectory generator and a given spatial tessellation L to transform the mobility diary into a sampled mobility trajectory. We describe in detail the second step of DITRAS in Section 5. Note that the two-step process described above is a general framework common to many generative models of human mobility, which are often composed by two sequential parts, the first generating temporal patterns and the second determining the spatial trajectory. However, in some models the division between the temporal and the spatial mechanisms is present but not explicitly acknowledged.

In Section 6 we will instantiate DITRAS by using $MD^{(t)}$ and d-EPR and compare it with other generative models obtained combining diary generators (first step) with trajectory generators (second step).

4 Step 1: Generation of mobility diary

A diary generator G produces a mobility diary, $D^{(t)}$, containing the sequence of trips made by a synthetic individual during a time period divided in time slots of t seconds. For example, $G^{(3600)}$ and $G^{(60)}$ produce mobility diaries with temporal resolutions of one hour and one minute, respectively. In Section 4.1 we illustrate how to compute a diary generator, $MD^{(t)}$, using real mobility trajectory data such as mobile phone data.

To separate the temporal patterns from the spatial ones, we define the abstract mobility trajectory, $A^{(t)}$, which contains the time ordered list of the "abstract locations" visited by a synthetic individual during a period divided in time slots of t seconds. An abstract location uniquely identifies a place where the individual is stationary, like home or the workplace, but it does not contain any information on the specific geographic position of the location
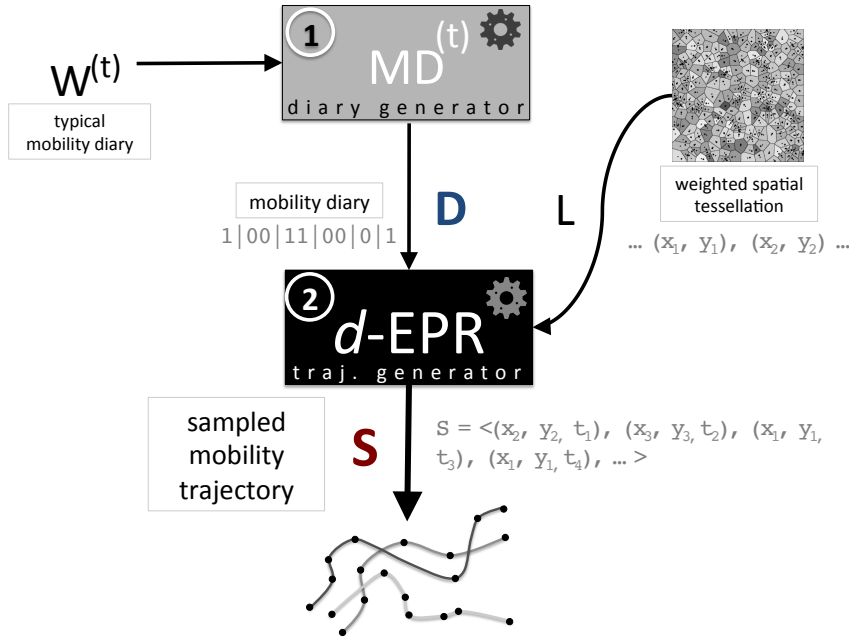
Fig. 1 Outline of the DITRAS framework. DITRAS combines two probabilistic models: a diary generator (e.g., $MD^{(t)}$) and trajectory generator (e.g., d-EPR). The diary generator uses a typical diary $W^{(t)}$ to produce a mobility diary D. The mobility diary D is the input of the trajectory generator together with a weighted spatial tessellation of the territory L. From D and L the trajectory generator produces a sampled mobility trajectory S.

(i.e., its coordinates). The abstract mobility trajectory is thus equivalent to the sampled mobility trajectory where the geographic locations, l_k, are substituted by placeholders, a_k, called abstract locations:

Definition 3 (Abstract mobility trajectory) An abstract mobility trajectory is a sequence $A^{(t)} = \langle a_1, \ldots, a_N \rangle$, where a_i $(i = 1, \ldots, N)$ is the abstract location where the individual spent the majority of time during time slot i, i.e., between $(i-1)t$ and it seconds from the first observation.

The mobility diary, $D^{(t)}$, is generated with respect to a typical mobility diary, $W^{(t)}$, which represents the individual's routine. $W^{(t)}$ is a sequence of time slots of duration t seconds and specifies the typical and most likely abstract location the individual visits in every time slot. Here we consider the simplest choice of typical mobility diary, in which the most likely location where a synthetic individual can be found at any time is her home location.

Definition 4 (Typical mobility diary) A typical mobility diary is a sequence $W^{(t)} = \langle w_1, \ldots, w_N \rangle$ where $w_k = w \quad \forall k = 1, \ldots, N$ denotes the home location of the synthetic individual. N is the total number of time slots considered.

The mobility diary, $D^{(t)}$, specifies whether an individual's abstract mobility trajectory, $A^{(t)}$, follows her typical mobility diary, $W^{(t)}$, or not. In particular, for every time slot i, $D^{(t)}(i)$ can assume two values:

- $D^{(t)}(i) = 1$ if $A^{(t)}(i) = W^{(t)}(i)$, meaning that the individual visits the abstract location $W^{(t)}(i)$ following her routine, i.e., she is at home;
- $D^{(t)}(i) = 0$ if $A^{(t)}(i) \neq W^{(t)}(i)$, meaning that the individual visits a location other than the abstract location $W^{(t)}(i)$ being out of her routine.

Definition 5 (Mobility Diary) A mobility diary is a sequence $D^{(t)}$ of time slots of duration t seconds generated by the regular language $\mathbb{L} = (1^+|(0^+|)^*)^*$, where 1 at time slot i indicates that the individual visits the abstract location in her typical diary at time i, $W^{(t)}(i)$, and 0 indicates a visit to a location different from the abstract location $W^{(t)}(i)$. The symbol "|" indicates a transition or trip between two different abstract locations.

An example of mobility diary generated by language $\mathbb{L}$ is $D^{(t)} = \langle 11|00|0|1 \rangle$. The first two entries indicate that $A^{(t)}(1) = W^{(t)}(1)$ and $A^{(t)}(2) = W^{(t)}(2)$, i.e., the individual follows her routine and she is at home. Next, the third, fourth and fifth entries indicate that $A^{(t)}(3) \neq W^{(t)}(3)$, $A^{(t)}(4) \neq W^{(t)}(4)$ and $A^{(t)}(5) \neq W^{(t)}(5)$, i.e., the individual breaks the routine and visits a non-typical location for two consecutive time slots, then she visits a different non-typical location for one time slot. Finally, the last time slot indicates that $A^{(t)}(6) = W^{(t)}(6)$, the individual follows the routine and returns home. We assume that the travel time between any two locations is of negligible duration.

4.1 Construction of MD$^{(t)}$

In this section we propose diary generator MD$^{(t)}$ and illustrate how to compute it from the abstract mobility trajectories of a set of real individuals. We use a Markov model to describe the probability that an individual follows her routine and visits a typical location at the usual time, or she breaks the routine and visits another location. First, we translate mobility trajectory data of real individuals into abstract mobility trajectories (Section 4.1.1). Second, we use the obtained abstract trajectory data to compute the transition probabilities of the Markov model MD$^{(t)}$ (Section 4.1.2).

4.1.1 Mobility trajectory data

The construction of MD$^{(t)}$ is based on mobility trajectory data of real individuals. We assume that raw mobility trajectory data describing the movements of a set of individuals are in the form $\langle (u_1, x_1, y_1, t_1), \ldots, (u_n, x_n, y_n, t_n) \rangle$ where u_i indicates the individual who visits location (x_i, y_i) at time t_i, $\forall_{1 \leq i < n} \, t_i < t_{i+1}$.

Mobility trajectory data can be obtained from various sources (e.g., mobile phones, GPS devices, geosocial networks) and describe the movements of individuals on a territory [52]. Since the purpose of MD$^{(t)}$ is to capture the

temporal patterns regardless the geographic position of locations, we translate raw mobility trajectory data into abstract mobility trajectories (see definition in Section 3).

Starting from the raw trajectory data, we assign an abstract location to every time slot in an individual's abstract mobility trajectory $A^{(t)}$ according to the following method. If the individual visits just one location during time slot i, we assign that location to i. If the individual visits multiple locations during a time slot i, we choose the most frequent location in i, i.e., the location where the individual spend most of the time during the time slot. If there are multiple locations with the same visitation frequency in time slot i, we choose the location with the highest overall frequency. If there is no information in abstract trajectory data about the location visited in time slot i (e.g., no calls during the time slot in the case of mobile phone data), we assume no movement by the individual and choose the location assigned to time slot $i - 1$.

To clarify the method let us consider the following example. A mobile phone user has the following hourly time series of calls: $[A, A, \bullet, \bullet, B, (C, C, B, B)]$, where A, B, C are placeholders for different cell phone towers (i.e., abstract locations). Here the symbol $\bullet$ indicates that there is no information in the data about the location visited during the 1-hour time slot, while all the locations in round brackets are visited during the same time slot. Using the method described above, the abstract mobility trajectory of the individual becomes $A^{(3600)} = \langle A, A, A, A, B, B \rangle$ because: (i) the two $\bullet$ symbols in the third and fourth time slots are substituted by location A assuming no movement with respect to the second time slot; (ii) the location assigned to the last time slot is B since C and B have the same visitation frequency in (C, C, B, B) but $f(B) > f(A)$, i.e., B has the highest overall visitation frequency.

It is worth noting that the choice of the duration of the time slot, t, is crucial and depends on the specific kind of mobility trajectory data used. GPS data from private vehicles, for example, generally provide accurate information about the location of the vehicle every few seconds. In this scenario, a time slot duration of a minute can be a reasonable choice. In contrast when dealing with mobile phone data a time slot duration of an hour or half an hour is a more reliable choice, since the majority of individuals have a low call frequency during the day [39].

4.1.2 Markov model transition probabilities

Let $A_u = \langle a_0^{(u)}, \ldots, a_{n-1}^{(u)} \rangle$ and $W_u = \langle w_0^{(u)}, \ldots, w_{n-1}^{(u)} \rangle$ be the abstract mobility trajectory and the typical mobility diary of individual $u \in U$, where U is the set of all individuals in the data – we omit the superscript (t) for clarity. Elements $a_h^{(u)} \in A_u$ and $w_h^{(u)} \in W_u$ denote the abstract and the typical locations visited by individual u at time slot h with $h = 0, \ldots, n - 1$.

A state in the Markov model MD is a tuple of two elements $s = (h, R)$. The state's first element, h, is the time slot of the time series denoted by an integer between 0 and $n - 1$. The state's second element, R, is a boolean variable that is 1 (True) if at time slot h the individual is in her typical

location, $w_h^{(u)}$, and 0 (False) otherwise – just like in the mobility diary. In total there are $n \times 2 = 2n$ possible states in the model. The transition matrix, MD, is a $2n \times 2n$ stochastic matrix whose element $\mathrm{MD}_{ss'}$ corresponds to the conditional probability of a transition from state s to state s', $\mathrm{MD}_{ss'} \equiv p(s'|s)$. The normalization condition imposes that the sum over all elements of any row s is equal to 1, $\sum_{s'} \mathrm{MD}_{ss'} = 1, \forall s$. We consider two types of transitions, $s \to s'$, depending on whether in state s the individual is in typical location or not:

- if the individual is in the typical location at time slot h, i.e., $s = (h, 1)$, then she can either go to the next typical location at time slot $h + 1$, $s = (h, 1) \to s' = (h + 1, 1)$, or go to a non-typical location and stay there for τ time slots, $s = (h, 1) \to s' = (h + \tau, 0)$;
- if instead the individual is not in the typical location at time slot h, i.e., $s = (h, 0)$, then she can either go to the typical location at time slot $h + 1$, $s = (h, 0) \to s' = (h + 1, 1)$, or go to a different non-typical location and stay there for τ time slots, $s = (h, 0) \to s' = (h + \tau, 0)$.

The formulae to compute the empirical frequencies for the four types of transitions are shown in Table 1. In the table, $\delta_x^u(a) = \delta(a_x^{(u)}, w_x^{(u)})$, $\hat{\delta}_x^u(a) = \delta(a_x^{(u)}, a_{x+1}^{(u)})$, where $\delta(i, j) = 1$ if $i = j$ and 0 otherwise, is the Kronecker delta. By convention, the product $\prod_{i=1}^{\tau-1} \ldots$ is equal to 1 if $\tau = 1$.

Transition, $s \to s'$	Frequency, $\mathrm{MD}_{ss'}$
$(h, 1) \to (h + 1, 1)$	$\dfrac{\sum_{u \in U} \sum_{a \in A_u} \delta_h^u(a) \delta_{h+1}^u(a)}{\sum_{u \in U} \sum_{a \in A_u} \delta_h^u(a)}$
$(h, 1) \to (h + \tau, 0)$	$\dfrac{\sum_{u \in U} \sum_{a \in A_u} \delta_h^u(a)[1-\delta_{h+1}^u(a)] \prod_{i=1}^{\tau-1} \hat{\delta}_{h+i}^u(a)[1-\hat{\delta}_{h+\tau}^u(a)]}{\sum_{u \in U} \sum_{a \in A_u} \delta_h^u(a)}$
$(h, 0) \to (h + 1, 1)$	$\dfrac{\sum_{u \in U} \sum_{a \in A_u} [1-\delta_h^u(a)]\delta_{h+1}^u(a)}{\sum_{u \in U} \sum_{a \in A_u} [1-\delta_h^u(a)]}$
$(h, 0) \to (h + \tau, 0)$	$\dfrac{\sum_{d \in D} [1-\delta_h^u(a)][1-\delta_{h+1}^u(a)][1-\hat{\delta}_h^u(a)] \prod_{i=1}^{\tau-1} \hat{\delta}_{h+i}^u(a)[1-\hat{\delta}_{h+\tau}^u(a)]}{\sum_{u \in U} \sum_{a \in A_u} [1-\delta_h^u(a)]}$

Table 1: Formulae to compute the transition probabilities of the Markov chain MD.

5 Step 2: Generation of sampled mobility trajectory

Starting from the mobility diary $D^{(t)}$, the sampled mobility trajectory $S^{(t)}$ is generated to describe the movement of a synthetic individual between a set of discrete locations called weighted spatial tessellation. A weighted spatial tessellation is a partition of a bi-dimensional space into locations each having a weight corresponding to its relevance.

Definition 6 (Weighted spatial tessellation) A weighted spatial tessellation is a set of tuples $L = \{(l_1, r_1), \ldots, (l_m, r_m)\}$, where $r_j \in \mathbb{N}$ $(j = 1, \ldots, m)$

is the relevance of a location and the l_j are a set of non-overlapping polygons that cover the bi-dimensional space where individuals can move. The location of each polygon is identified by the geographic coordinates of its centroid, (x_j, y_j).

The weighted spatial tessellation indicates the possible physical locations on a finite bi-dimensional space a synthetic individual can visit during the simulation. The relevance of a location measures its popularity among real individuals: locations of high relevance are the ones most frequently visited by the individuals [39, 37]. The relevance is introduced to generate synthetic trajectories that take into account the underlying urban structure. An example of weighted spatial tessellation is the one defined by a set of mobile phone towers, where the relevance of a tower can be estimated as the number of calls performed by mobile phone users during a period of observation, and the polygons correspond to the regions obtained from the Voronoi partition induced by the mobile phone towers.

First, DITRAS assigns to every abstract location in the typical mobility diary $W^{(t)}$ a physical location on the weighted spatial tessellation L, creating $W_m^{(t)}$, a typical mobility diary where each location has a specific geographic position. Next, DITRAS scans $D^{(t)}$ to assign a physical location to every entry. For every entry $D^{(t)}(i) \in D^{(t)}$ we have two possible scenarios:

- $D^{(t)}(i) = 1$, the entry indicates a visit to a typical location, i.e., the abstract location in $W^{(t)}(i)$. In this scenario the synthetic individual visits location $l = W_m^{(t)}(i)$ which is added to the sampled trajectory at time slot i, i.e. $S^{(t)}(i) = W_m^{(t)}(i)$;

- $D^{(t)}(i) = 0$, the entry indicates a visit to a non-typical location. In this second scenario DITRAS calls the trajectory generator to choose a location l to visit, where $l \neq W_m^{(t)}(i)$. The chosen location l is added to the sampled mobility trajectory k times, where k is the number of consecutive 0 characters before the next separator character '|' appears in $D^{(t)}$, i.e., the total number of time slots spent in location l.

Example of trajectory generation. To clarify how the second step of DITRAS works let us consider the following example. A synthetic individual is assigned a mobility diary $D^{(t)} = \langle 1|00|1 \rangle$ and the chosen typical diary is $W^{(t)} = \langle w, w, w, w \rangle$, where w denotes the individual's home. To generate a synthetic sampled mobility trajectory S, DITRAS operates as follows. First, DITRAS assigns a physical location to the individual's home w, generating $W_m^{(t)} = \langle (x_1, y_1), (x_1, y_1), (x_1, y_1), (x_1, y_1) \rangle$. Next, DITRAS starts from the first entry $D^{(t)}(1)$. Since $D^{(t)}(1) = 1$ the synthetic individual is at home. Therefore, tuple $(x_1, y_1, 1)$ is added to trajectory S. Next, DITRAS processes the second entry $D^{(t)}(2)$, sees a separator and then proceeds to entry $D^{(t)}(3)$. Since $D^{(t)}(3) = 0$, the synthetic individual is not at home in the third time slot. Hence, DITRAS calls a trajectory generator (e.g., d-EPR) which chooses to

visit physical location (x_2, y_2). DITRAS hence adds the tuples $(x_2, y_2, 2)$ and $(x_2, y_2, 3)$ to trajectory S, since there two 0 characters until the next separator in $D^{(t)}$. The last entry $D^{(t)}(6) = 1$ indicates that the synthetic individual returns home in the fourth time slot. So, DITRAS adds tuple $(x_1, y_1, 4)$ to trajectory S. At the end of the execution, the sampled mobility trajectory generated by DITRAS is $S = \langle (x_1, y_1, 1), (x_2, y_2, 2), (x_2, y_2, 3), (x_1, y_1, 4) \rangle$.

5.1 The d-EPR model

As trajectory generator we propose the d-EPR individual mobility model [39, 37] that assigns a location on the bi-dimensional space to an entry in mobility diary $D^{(t)}$. The d-EPR (density-Exploration and Preferential Return) is based on the evidence that an individual is more likely to visit relevant locations than non-relevant locations [39, 37]. For this reason d-EPR incorporates two competing mechanisms, one driven by an individual force (preferential return) and the other driven by a collective force (preferential exploration). The intuition underlying the model can be easily understood: when an individual returns, she is attracted to previously visited places with a force that depends on the relevance of such places at an individual level. In contrast, when an individual explores she is attracted to new places with a force that depends on the relevance of such places at a collective level. In the preferential exploration phase a synthetic individual selects a new location to visit depending on both its distance from the current location, as well as its relevance measured as the collective location's relevance in the bi-dimensional space. In the model, hence, the synthetic individual follows a personal preference when returning and a collective preference when exploring. The d-EPR uses the gravity model [58, 22, 28] to assign the probability of a trip between any two locations in L, which automatically constrains individuals within a territory's boundaries. The usage of the gravity model is justified by the accuracy of the gravity model to estimate origin-destination matrices even at the country level [13, 56, 47, 1, 28].

The d-EPR model assigns a location on the bi-dimensional space defined by a spatial tessellation L for an entry in mobility diary $D^{(t)}$. The d-EPR takes in input two variables: (i) the current sampled mobility trajectory of the synthetic individual $S = \langle (x_1, y_1, t_1), \dots, (x_n, y_n, t_n) \rangle$; (ii) a probability matrix P indicating, for every pair of locations $i, j \in L, i \neq j$ the probability of moving from i to j. Every probability p_{ij} is computed as:

$$p_{ij} = \frac{1}{Z} \frac{r_i r_j}{d_{ij}^2}, \qquad (1)$$

where $r_{i(j)}$ is the relevance of location $i(j)$ as specified in the weighted spatial tessellation L, d_{ij} is the geographic distance between i and j, and $Z = \sum_{i,j \neq i} p_{ij}$ is a normalization constant. The matrix P is computed before the execution of the DITRAS model by using the spatial tessellation L.

With probability $p_{new} = \rho N^{-\gamma}$ where N is the number of distinct locations in S and $\rho = 0.6$, $\gamma = 0.21$ are constants [39, 37, 50], the individual chooses to

explore a new location, otherwise she returns to a previously visited location. If the individual explores and is in location i, the new location $j \neq i$ is selected according to the probability $p_{ij} \in P$. If the individual returns to a previously visited location, it is chosen with probability proportional to the number of her previous visits to that location. The d-EPR model hence returns the chosen location j.

6 Results

In this section we show the results of simulation experiments where we instantiate DITRAS by using $\mathrm{MD}^{(t)}$ as diary generator and d-EPR as trajectory generator (we refer to this model as d-$\mathrm{EPR_{MD}}$). We use this model to generate the sampled mobility trajectories of 10,000 synthetic individuals and compare the resulting sampled mobility trajectories with:

– the trajectories of 10,000 mobile phone users whose mobility is tracked during 11 weeks in a European country;

– the sampled mobility trajectories produced by other 8 spatio-temporal mobility models created combining different diary and trajectory generators through the DITRAS framework.

In Section 6.1 we describe the mobile phone data we use in our experiments to describe the mobility of real individuals and the pre-processing operations we carry out on the data. In Section 6.2 we provide a comparison on a set of spatio-temporal mobility patterns of d-$\mathrm{EPR_{MD}}$'s trajectories, mobile phone data's trajectories and the trajectories produced by the other models. All the simulations are performed by using the same weighted spatial tessellation induced by the mobile phone towers and a time slot duration $t = 3600\mathrm{s} = 1\mathrm{h}$.

6.1 CDR data

We have access to a set of Call Detail Records (CDRs) gathered by a European carrier for billing and operational purposes. The dataset records all the calls made during 11 weeks by ≈ 1 million anonymized mobile phone users. CDRs collect geographical, temporal and interaction information on mobile phone use and show an enormous potential to empirically investigate the structure and dynamics of human mobility on a society wide scale [42, 18, 15, 21, 6, 39, 40]. Each time an individual makes a call the mobile phone operator registers the connection between the caller and the callee, the duration of the call and the coordinates of the phone tower communicating with the reserved phone, allowing to reconstruct the user's approximate position. Table 2 illustrates an example of the structure of CDRs.

(a)

timestamp	tower	caller	callee
2007/09/10 23:34	36	4F80460	4F80331
2007/10/10 01:12	36	2B01359	9H80125
2007/10/10 01:43	38	2B19935	6W1199
⋮	⋮	⋮	⋮

(b)

tower	latitude	longitude
36	49.54	3.64
37	48.28	1.258
38	48.22	-1.52
⋮	⋮	⋮

Table 2 Example of Call Detail Records (CDRs). Every time a user makes a call, a record is created with timestamp, the phone tower serving the call, the caller identifier and the callee identifier (a). For each tower, the latitude and longitude coordinates are available to map the tower on the territory (b).

CDRs have been extensively used in literature to study different aspects of human mobility, due to several advantages: they provide a means of sampling user locations at large population scales; they can be retrieved for different countries and geographic scales given their worldwide diffusion; they provide an objective concept of location, i.e., the phone tower. Nevertheless, CDR data suffer different types of bias [41, 19], such as: (i) the position of an individual is known at the granularity level of phone towers; (ii) the position of an individual is known only when she makes a phone call; (iii) phone calls are sparse in time, i.e., the time between consecutive calls follows a heavy tail distribution [15, 2]. In other words, since individuals are inactive most of their time, CDRs allow to reconstruct only a subset of an individual's mobility. Several works in literature study the bias in CDRs by comparing the mobility patterns observed on CDRs to the same patterns observed on GPS data [36, 39, 35, 38] or handover data (data capturing the location of mobile phone users recorded every hour or so) [15]. The studies agree that the bias in CDRs does not affect significantly the study of human mobility patterns.

Data preprocessing. In order to cope with sparsity in time of CDRs and focus on individuals with reliable call statistics, we carry out some preprocessing steps. Firstly, for each individual u we discard all the locations with a visitation frequency $f = n_i/N \leq 0.005$, where n_i is the number of calls performed by u in location i and N the total number of calls performed by u during the period of observation [45, 39]. This condition checks whether the location is relevant with respect to the specific call volume of the individual. Since it is meaningless to analyze the mobility of individuals who do not move, all the individuals with only one location after the previous filter are discarded. We select only active

individuals with a call frequency threshold of $f = N/(h*d) \geq 0.5$ calls per hour, where N is the total number of calls made by u, $h = 24$ is the hours in a day and $d = 77$ the days in our period of observation. Starting from ≈ 1 millions users, the filtering results in $50,000$ active mobile phone users.

Weighted Spatial Tessellation. The weighted spatial tessellation L we use in the experiments is defined by the mobile phone towers in the CDR data. The relevance of a phone tower is estimated as the total number of calls served by that tower by the 50,000 active mobile phone users during the 11 weeks. Every location's position on the space is identified by the latitude and longitude coordinates of a phone tower.

6.2 Simulated mobility trajectories

We use the DITRAS framework to build nine models which use different combinations for the diary generator and the trajectory generator. In particular, we consider three diary generators – MD, RD and WT – and three trajectory generators – d-EPR, SWIM and LATP. For every model we simulate the mobility of 10,000 synthetic individuals for a period of $N = 1,848$ hours (11 weeks). Table 3 shows the ability of every model in reproducing a set of characteristic statistical distribution derived from the CDR data quantified by the Root Mean Square Error, $RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$ where $\hat{y}_i$ indicates a point of the synthetic distribution, y_i the corresponding point in the empirical distribution and n the number of observations. Here we use the notation TG_{DG} to specify that trajectory generator TG is used in combination with diary generator DG. For example, d-EPR_{MD} indicates the model using diary generator MD in combination with trajectory generator d-EPR. Notation $\text{TG}_{\{\text{DG}_1,...,\text{DG}_k\}}$ indicates the set of models $\{\text{TG}_{\text{DG}_1}, \ldots, \text{TG}_{\text{DG}_k}\}$.

Diary and trajectory generators. In the Random Diary (RD) generator a synthetic individual is in perpetuum motion: in every time slot of the simulation she chooses a new location to visit. We use RD to highlight the difference between the diary generator we propose, MD (Section 4.1), and the temporal patterns of a non-realistic diary generator. In the Waiting Time (WT) diary generator a synthetic individual chooses a waiting time Δt between a trip and the next one from the empirical distribution $P(\Delta t) \sim \Delta t^{-1-\beta} \exp^{-\Delta t/\tau}$, with $\beta = 0.8$ and $\tau = 17$ hours as measured on CDR data [50]. WT is the temporal mechanism usually used in combination with mobility models like EPR [50] and SWIM [25]. It reproduces in a realistic way the distribution of the time between two consecutive trips [50,36] but does not model the circadian rhythm and the tendency of individuals to be in certain places and specific times.

The trajectory generator SWIM [25] is a modelling approach based on location preference. The model initially assigns to each synthetic individual a home location L_h chosen randomly from the spatial tessellation. The synthetic

		Δr trip distance	r_g radius of gyration	S^{unc} mobility entropy	T trips per hour	D trips per day	Δt time of stays	V visits per location	N locations per user	$f(L)$ location frequency
MD	*d*-EPR	**0.0001**	**0.0026**	0.9643	0.0061	0.0659	0.0014	**2.6e^{-5}**	0.0218	0.0122
	SWIM	0.0005	0.0064	3.6069	0.0062	0.0683	0.0029	5.6e^{-5}	-	0.0669
	LATP	0.0001	0.0061	3.2236	0.0062	0.0684	0.0027	6.3e^{-5}	-	0.0625
RD	*d*-EPR	0.0004	**0.0027**	**1.1745**	0.0232	**0.2098**	0.0024	4.1e^{-5}	**0.0235**	**0.0521**
	SWIM	0.0041	0.0072	10.1573	0.0232	-	0.0033	7.2e^{-5}	-	0.0947
	LATP	**0.0002**	0.0065	3.9747	0.0232	-	0.0033	4.6e^{-5}	-	0.0874
WT	*d*-EPR	0.0003	0.0024	**1.1666**	0.0232	**0.1790**	0.0023	4.0e^{-5}	**0.0224**	**0.0502**
	SWIM	0.0033	0.0072	10.1534	0.0232	0.2036	0.0033	**1.9e^{-5}**	-	0.0943
	LATP	0.0001	0.0065	4.0291	0.0232	0.2037	0.0033	7.2e^{-5}	-	0.0866

Table 3 Error of fit between the CDR data and the synthetic data. Every row i in the table corresponds to a model and every column j to a mobility measure. A cell (i,j) indicates the RMSE (Root Mean Square Error) of a synthetic distribution with respect to the CDR distribution. The best RMSE values for a given mobility measure are in blue. The best values for a combination of mobility measure and diary generator are in bold. The symbol - indicates that the the synthetic distribution is not comparable with the CDR distribution.

individual then selects a destination for the next moves depending on the weight of each location [25]:

$$w(L)_{\text{swim}} = \alpha * d(L_h, L) + (1 - \alpha) * r(L), \quad \alpha = 0.75 \qquad (2)$$

which grows with the relevance $r(L)$ of the location and decreases with the distance from the home $d(L_h, L) = \frac{1}{(1+distance(L_h,L))^2}$ [25]. SWIM tries to model both the preference for short trips and the preference for relevant locations, though it does not model the preferential return mechanism.

The trajectory generator LATP (Least Action Trip Planning) [27,26] is a trip planning algorithm used as exploration mechanism in several mobility models, such as SLAW [27,26], SMOOTH [34], MSLAW [46] and TP [49,48]. In LATP a synthetic individual selects the next location to visit according to a weight function [27,26]:

$$w(L)_{\text{latp}} = \frac{1}{distance(c, L)^{1.5}}. \qquad (3)$$

LATP only models the preference for short distances and does not consider the relevance of a location nor model the preferential return mechanism.

We compare the synthetic mobility trajectories of the nine models with CDR trajectories on the distributions of several measures capturing salient characteristics of human mobility. Table 3 displays the mobility measures we consider, which are: trip distance Δr [15,36], radius of gyration r_g [15,36,39], mobility entropy S^{unc} [51,11,40], location frequency $f(L)$ [50,16,36], visits per location V [37], locations per user N [37], trips per hour T [15,36], time of stays Δt [50,16] and trips per day D.

Trip distance. The distance of a trip Δr is the geographical distance between the trip's origin and destination locations. We compute the trip distances for every individual and then plot the distribution $P(\Delta r)$ of trip distances in Figure 2a-c. Figure 2a compares the distribution of trip distance of CDR data with the distributions produced by d-EPR$_{MD}$, SWIM$_{MD}$ and LATP$_{MD}$. We observe that d-EPR$_{MD}$ and LATP$_{MD}$ are able to reproduce the distribution of $P(\Delta r)$ although slightly overestimating long-distance trips. In contrast SWIM$_{MD}$ cannot reproduce the shape of the empirical distribution resulting in a RMSE(SWIM$_{MD}$) higher than the other two models (see Table 3). The shape of the synthetic distributions do not vary significantly by changing the diary generator (Figure 2, b-c). In other words, the choice of the diary generator does not affect the ability of the model to capture the distribution $P(\Delta r)$. This is also evident from Table 3 where the RMSEs in the first column vary a little by changing the diary generator. Model LATP$_{MD}$ produces the best fit with CDR data, as we note both in Figure 2c and Table 3. This suggests that modelling preferential return and location preference is not crucial to reproduce $P(\Delta r)$, since it is just regulated by the preference for short-distance trips. Although SWIM embeds a preference for short-distance trips (Equation 2) the distance is chosen with respect to the home location L_h leading to an underestimation of short-distance trips (Figure 2a-c).

Radius of gyration. The radius of gyration r_g is the characteristic distance traveled by an individual during the 11 weeks [15, 36, 39]. In detail, r_g characterizes the spatial spread of the locations visited by an individual u from the trajectories' center of mass (i.e., the weighted mean point of the locations visited by an individual), defined as:

$$r_g = \sqrt{\sum_{i \in L^{(u)}} p_i (l_i - l_{cm})^2}, \tag{4}$$

where l_i and l_{cm} are the vectors of coordinates of location i and center of mass, respectively [15, 36, 39], $L^{(u)} \subseteq L$ is the set of locations visited by individual u, $p_i \equiv n_i/|L^{(u)}|$ is the individual's visitation frequency of location l_i, equal to the number of visits to l_i divided by the total number of visits to all locations. In Figure 2a we observe that d-EPR$_{MD}$ is the only model capable of reproducing the shape of $P(r_g)$, though overestimating the presence of large radii (see Figure 2d). RMSE(d-EPR$_{MD}$) for r_g is indeed lower than RMSE(SWIM$_{MD}$) and RMSE(LATP$_{MD}$) as shown in Table 3. SWIM$_{MD}$ and LATP$_{MD}$ cannot reproduce the shape of $P(r_g)$ because r_g also depends on the preferential return mechanism [50, 39] which is not modeled in SWIM and LATP. In a previous work [37] we also show that $P(r_g)$ depends on the preferential exploration mechanism of d-EPR since a version of d-EPR without preferential exploration – the s-EPR model – is not able to reproduce the shape of $P(r_g)$. We also observe that while d-EPR$_{\{MD, RD, WT\}}$ produce similar distributions of r_g, SWIM and LATP produce different distributions of r_g with different choices of the diary generator (Figure 2e, f).

Mobility entropy. The mobility entropy S^{unc} of an individual u is defined as the Shannon entropy of her visited locations [51,11,40]:

$$S^{unc}(u) = \frac{\sum_{i \in L^{(u)}} p_i \log(p_i)}{\log |L^{(u)}|}, \tag{5}$$

where p_i is the probability that individual u visits location i during the 11 weeks and $\log |L^{(u)}|$ is a normalization factor. The mobility entropy of an individual quantifies the possibility to predict individual's future whereabouts. Individuals having a very regular movement pattern possess a mobility entropy close to zero and their whereabouts are rather predictable. Conversely, individuals with a high mobility entropy are less predictable.

We observe that the average $\bar{S}^{unc}$ produced by d-EPR$_{\mathrm{MD}}$ data equals the average $\bar{S}^{unc} = 0.61$ in CDR data, although d-EPR$_{\mathrm{MD}}$ underestimates the variance of distribution $P(S^{unc})$ (Figure 2g). In contrast, SWIM$_{\mathrm{MD}}$ and LATP$_{\mathrm{MD}}$ largely overestimate $\bar{S}^{unc}$ and underestimate the variance of $P(S^{unc})$, resulting in a RMSE much higher than RMSE(d-EPR$_{\mathrm{MD}}$), as shown in Table 3. This is because SWIM and LATP do not model the preferential return mechanism, which increases the predictability of individuals since they tend to come back to already visited locations. $P(S^{unc})$ is not robust to the choice of diary generator: diary generator RD and WT make the models to largely overestimate $\bar{S}^{unc}$ (Figures 2h, i). In particular SWIM$_{\{\mathrm{RD, WT}\}}$ and LATP$_{\{\mathrm{RD, WT}\}}$ produce distributions with $\bar{S}^{unc} \approx 1$, indicating that the typical synthetic individual is much more unpredictable than a typical individual in CDR data. This is also evident from Table 3 where RMSE(SWIM$_{\{\mathrm{RD, WT}\}}$) and RMSE(LATP$_{\{\mathrm{RD, WT}\}}$) are around three times RMSE(SWIM$_{\mathrm{MD}}$) and RMSE(LATP$_{\mathrm{MD}}$) respectively. Hence, distribution $P(S^{unc})$ highly depends on both the choice of the trajectory generator and the choice of the diary generator.

Location frequency. Another important characteristic of an individual's mobility is the probability of visiting a location given the location's rank. The rank of a location depends on the number of times the individual visits the locations over the 11 weeks. For instance, rank 1 represents the most visited location (generally home place); rank 2 the second most visited location (e.g., work place) and so on. We compute the frequency of each of these ranked locations for every individual and plot the distribution of frequencies $f(L_i)$ in Figure 3a-c. We observe that d-EPR$_{\mathrm{MD}}$ reproduces the shape of $f(L_i)$ (with RMSE=0.0122) better than SWIM$_{\mathrm{MD}}$ and LATP$_{\mathrm{MD}}$ (which have RMSE=0.0670 and RMSE=0.0250 respectively). If we change the diary generator in the model, d-EPR$_{\{\mathrm{RD, WT}\}}$ underestimate the frequency of the top-ranked location and slightly overestimate the the frequency of the less visited locations with respect to CDR data (Figure 3b, c). A reason for this discrepancy is that RD and WT do not take into account the circadian rhythm of individuals, hence underestimating the number of returns to the most frequent location (usually the home place). In SWIM$_{\mathrm{MD}}$ and LATP$_{\mathrm{MD}}$,
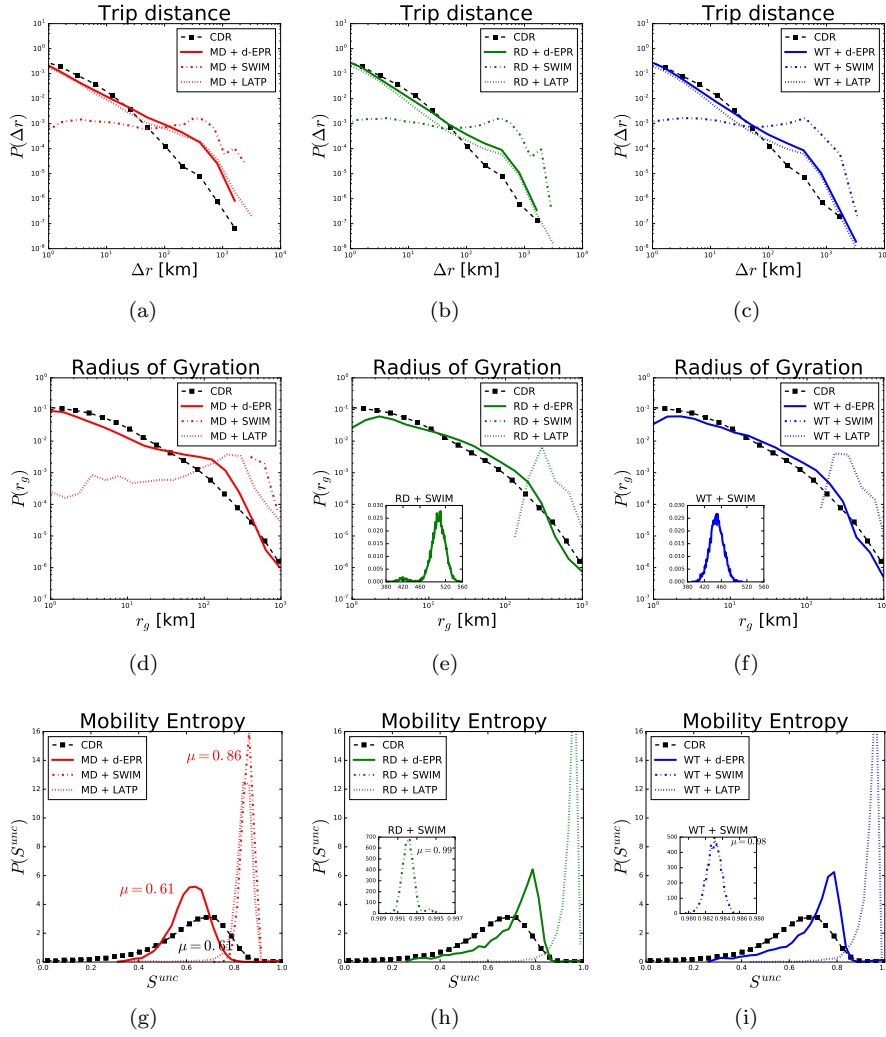
Fig. 2 Distributions of human mobility patterns. The figure compares the mobility models and CDR data on three mobility measures. Plots in (a), (b) and (c) show the distribution of trip distances $P(\Delta r)$ for CDR data (black squares) and data produced by three trajectory generators (d-EPR, SWIM and LATP) in combination with the MD generator (a), the RD generator (b) and the WT generator (c). Plots in (d), (e) and (f) show the distribution of radius of gyration r_g, while plots in (g), (h) and (i) show the distribution of mobility entropy S^{unc}.

the absence of a preferential return mechanism produce a more uniform distribution of location frequencies (Figure 3b-c), which is further exacerbated for $\text{SWIM}_{\{RD,\ WT\}}$ and $\text{LATP}_{\{RD,\ WT\}}$. Location frequency $f(L_i)$ is another case where the choice of the diary generator and the choice of the trajectory

generator are both crucial to reproduce the shape of the distribution in an accurate way.

Visits per location. A useful measure to understand how a set of individuals exploit the mobility space is the number V of overall visits per location, i.e., the total number of visits by all the individuals in every location during the 11 weeks. For every dataset, we compute the number of visits for every location of the weighted spatial tessellation and plot the distribution $P(V)$ in Figure 4d-f. As for CDR data, d-EPR$_{\mathrm{MD}}$ produces a $P(V)$ which follows a heavy tail distribution: the majority of locations have just one visit while a minority of locations have up to several thousands visits during the 11 weeks. The value of V of a location depends on two factors: (i) its relevance in the weighted spatial tessellation; (ii) its position in the weighted spatial tessellation. The higher the relevance of a location in the weighted spatial tessellation, the higher is the probability for the location to be visited in the exploration mechanisms of d-EPR and SWIM. Indeed, from Figure 4e-f we observe that d-EPR and SWIM are the models which better fit $P(V)$. In contrast LATP does not take into account the relevance of a location during the exploration being unable to capture the shape of $P(V)$.

Locations per user. The number N of distinct locations visited by an individual during the 11 weeks describes the degree of exploration of an individual, i.e., how the single individuals exploit the mobility space. In Figure 3g we observe that the MD models do not capture the shape of $P(N)$ in CDR data: the average number of distinct locations $\bar{N}$ according to d-EPR$_{\mathrm{MD}}$ is about twice $\bar{N}$ in CDR data, while SWIM$_{\mathrm{MD}}$ and LATP$_{\mathrm{MD}}$ produce distributions whose $\bar{N}$ is more than ten times $\bar{N}$ in CDR data. By changing diary generator (Figure 3h, i) the difference with CDR data becomes even larger: d-EPR$_{\{\mathrm{RD,\ WT}\}}$ produce a much broader variance of $P(N)$, SWIM$_{\{\mathrm{RD,\ WT}\}}$ and LATP$_{\{\mathrm{RD,\ WT}\}}$ predict a number of distinct visited locations very far from CDR data. These results suggest that the considered models overestimate the degree of exploration of individuals. In the case of d-EPR$_{\mathrm{MD}}$ the overestimation may depend on the distribution of time of stays, as the distribution of time stays $P(\Delta t)$ produced by d-EPR$_{\mathrm{MD}}$ overestimates the number of short stay times, leading to a larger total number of visited locations (Figure 4g).

Trips per hour. Human movements follow the circadian rhythm, i.e., they are prevalently stationary during the night and move preferably at specific times of the day [15,36]. To verify whether the considered models are able to capture this characteristic of human mobility, we compute the number of trips T made by the individuals at every hour of the 11 weeks. Figure 4a-c show how T distribute across the 24 hours of the day. We observe that, regardless the trajectory generator used, diary generator MD produces a distribution of trips per hour very similar to CDR data (Figure 4a). The Markov model proposed in Section 4.1 is hence able to create mobility diaries which reproduce the circadian rhythm of individuals in an accurate way. In contrast, diary
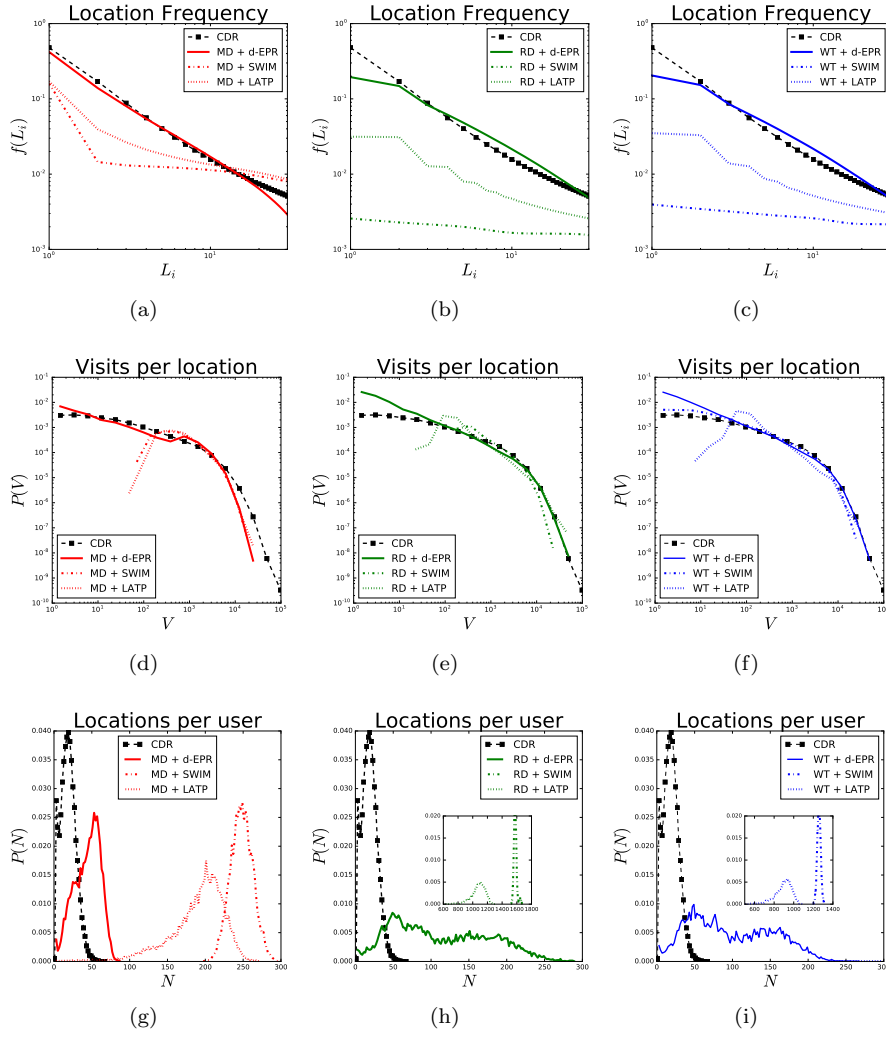
Fig. 3 Distributions of human mobility patterns. The figure compares the distribution of mobility measures produced by mobility models built with different combinations of diary generator and trajectory generator. Plots in (a), (b) and (c) show the distribution of location frequency $f(L)$ for d-EPR, SWIM and LATP used in combination with MD, RD and WT respectively. Plots in (d), (e) and (f) show the distribution of the number V of visits per location and plots in (g), (h), (i) show the distribution of the number N of distinct visited locations per user.

generators RD and WT are not able to capture this distribution, regardless the trajectory generator used (Figure 4b, c). This is because: (i) in RD individuals are always in motion; (ii) WT takes into account the waiting times but not the preference of individuals to move at specific times of the day.

Trips per day. The number of trips per day D indicates the tendency of individuals to travel in their every-day life. For every dataset, we compute the number of trips per day made by each individual during the 11 weeks and plot the distribution $P(D)$ in Figure 4d-f. We observe that d-EPR$_{MD}$, SWIM$_{MD}$ and LATP$_{MD}$ are able to capture the shape of $P(D)$ but overestimate the variance of the distribution (Figure 4d). The other diary generators, RD and WT, are not able to reproduce the CDR distribution since the average number $\bar{D}$ of trips per day is much higher than CDR data (Figure 4e-f). Again, this is because in RD individuals are always in motion and because WT does not take into account the circadian rhythm of individuals.

Time of stays. The distribution of stay times Δt is another important temporal features observed in human mobility. Stay time is the amount of time an individual spends at a particular location. In our experiments we compute the stay time as the number of hours every individual spends in her visited locations and plot the distribution $P(\Delta t)$ in Figure 3g-i. We observe that d-EPR$_{\{MD, RD, WT\}}$ capture the shape of the distribution while the other models do not, though overestimating the presence of short time stays.

6.3 Discussion of results

Two main results emerge from our experiments. First, model d-EPR$_{MD}$ produces sampled mobility trajectories having in general the best fit to CDR data (i.e., having the lowest RMSE for most of the measures), as evident in Table 3. Diary generator MD, indeed, simulates in a realistic way temporal human mobility patterns such as the distribution of location frequency (Figure 3a) and the distribution of trips per hour (Figure 4a). This is mainly because MD reproduces the circadian rhythm of individuals, while RD and WT do not. Moreover, trajectory generator d-EPR embeds two mobility mechanisms: preferential return and preferential exploration. The preferential return mechanism – absent in SWIM and LATP – allows for a realistic simulation of, for example, the distribution of radius of gyration (Figure 2d) and the distribution of stay times (Figure 4g). The preferential exploration mechanism, which is modeled by both d-EPR and SWIM but it is absent in LATP, allows for a realistic description of the territory exploitation by individuals, in terms of the distribution of the number of visits per location (Figure 3d).

Second interesting result is that the temporal and the spatial mechanisms have different roles in shaping the distribution of standard mobility measures. Some measures, such as trip distance (Figure 2a-c), radius of gyration (Figure 2d-f), visits per location (Figure 3d-f) and time of stays (Figure 2g-i) mainly depend on the choice of the trajectory generator, i.e., on the spatial mechanism of the model. Indeed, by changing the underlying diary generator the shape of these distribution and the RMSE w.r.t. CDR data do not change in a significant way. Other measures, such as trips per hour (Figure 4a-c) and trips per day (Figure 4d-f) mainly depend on the choice of the diary generator, i.e.,
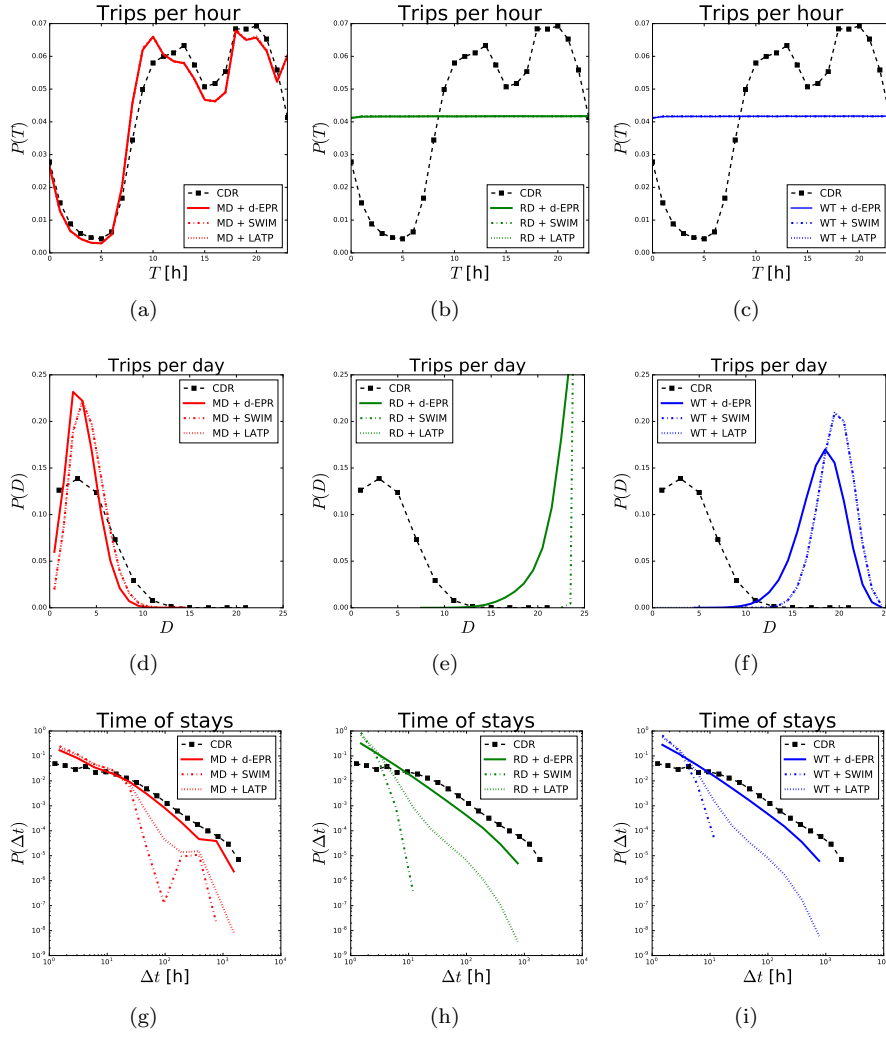
Fig. 4 Distributions of human mobility patterns. The figure compares the distribution of mobility measures produced by mobility models built with different combinations of diary generator and trajectory generator. Plots in (a), (b) and (c) show the distribution of the number T of trips per hour of the day for d-EPR, SWIM and LATP used in combination with MD, RD and WT respectively. Plots in (d), (e) and (f) show the distribution of the number D of trips per day, plots in (g), (h), (i) show the distribution of time of stays Δt.

on the temporal mechanism of the model. Conversely, both the spatial and the temporal mechanism are determinant in reproducing the distribution of some other measures like mobility entropy (Figure 2g-i) and locations per user (Figure 3g-i). Moreover the right combination of diary and trajectory generator, d-EPR$_{MD}$, leads to more accurate fits w.r.t. CDR data for the majority of measures (Table 3). Human mobility patterns depend on both where people

go and when people move: our results show that to reproduce them in an accurate way we need proper choices for the spatial and the temporal generative models to use in the Ditras framework.

7 Conclusion and future works

In this paper we propose Ditras, a framework to simulate the spatio-temporal patterns of individual human mobility. The model consists of two basic steps: (i) the generation of mobility diary, where we build a mobility diary for a synthetic individual; (ii) and the generation of mobility trajectory, where we transform the mobility diary into a sampled mobility trajectory. We implement the model to simulate the mobility of 10,000 synthetic individuals visiting the locations on a large European country. The generated sampled mobility trajectories are compared with CDR data and the trajectories produced by other 8 models, each using a different combination of diary generator and trajectory generator in the Ditras framework. Among the considered models, d-EPR$_{MD}$ produces the best fit w.r.t. CDR data. We also observe that different combinations of diary and trajectory generators show different abilities to reproduce the distribution of standard mobility measures. This result highlights the importance of considering both the spatial and temporal dimensions in human mobility modelling.

The instantiation of Ditras we propose, d-EPR$_{MD}$, can be further improved in several directions. First, in this work the construction of the Markov model MD$^{(t)}$ is based on the simplest possible typical diary $W^{(t)}$, where the most likely location where a synthetic individual can be found at any time is her home location. More complex typical diaries can be used specifying, for example, the typical times where an individual can be found at work, school, friends' home and so on. Such a composition of $W^{(t)}$ can be constructed by using surveys or generative models describing the daily schedule of human activities [43,21,30] as a way to enrich an individual's trajectory with information about the type of activity associated to a location. Second, in d-EPR the preference for short-distance trips is embedded in the preferential exploration phase only. A preference for short-distance trips can be introduced during the preferential return mechanisms as well, in order to eliminate the overestimation of long-distance trips and long-distance radii observed in Figures 2a and 2d. Third, current generative mobility models typically do not embed information about the time of travel between a location to another. As a consequence, they can produce sequences of trips with non-realistic velocities. A way to overcome this issue is to embed a speed constraint during the preferential return or the preferential exploration mechanisms. Moreover, another interesting improvement can be to map the sampled mobility trajectories to a road network specifying specific road routes with specific velocities. This mapping would be of great help, for example, in what-if analysis where we want to study how human mobility changes with the construction of a new

infrastructure in an urban context. We leave all these improvements of DITRAS as future works.

References

1. Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J.J., Vespignani, A.: Multiscale mobility networks and the spatial spreading of infectious diseases. Proceedings of the National Academy of Sciences **106**(51), 21,484–21,489 (2009). DOI 10.1073/pnas.0906910106
2. Barabási, A.L.: The origin of bursts and heavy tails in human dynamics. Nature **435**(7039), 207–211 (2005). DOI 10.1038/nature03459
3. Barbosa, H., de Lima-Neto, F.B., Evsukoff, A., Menezes, R.: The effect of recency to human mobility. EPJ Data Science **4**(1), 1–14 (2015). DOI 10.1140/epjds/ s13688-015-0059-8. URL http://dx.doi.org/10.1140/epjds/s13688-015-0059-8
4. Bellemans, T., Kochan, B., Janssens, D., Wets, G., Arentze, T., Timmermans, H.: Implementation framework and development trajectory of feathers activity-based simulation platform. Transportation Research Record: Journal of the Transportation Research Board (2175), 111–119 (2010)
5. Brockmann, D., Hufnagel, L., Geisel, T.: The scaling laws of human travel. Nature **439**(7075), 462–465 (2006). URL http://dx.doi.org/10.1038/nature04292
6. Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., Ratti, C.: Real-time urban monitoring using cell phones: A case study in rome. IEEE Transactions on Intelligent Transportation Systems **12**(1), 141–151 (2011). DOI 10.1109/TITS.2010.2074196
7. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'11, pp. 1082–1090. ACM (2011)
8. 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**(1), 1–16 (2007). DOI 10.1371/journal.pmed.0040013
9. Conti, M., Giordano, S., May, M., Passarella, A.: From opportunistic networks to opportunistic computing. Comm. Mag. **48**(9), 126–139 (2010). DOI 10.1109/MCOM.2010. 5560597. URL http://dx.doi.org/10.1109/MCOM.2010.5560597
10. De Nadai, M., Staiano, J., Larcher, R., Sebe, N., Quercia, D., Lepri, B.: The death and life of great italian cities: A mobile phone data perspective. In: Proceedings of the 25th International Conference on World Wide Web, WWW '16, pp. 413–423. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2016). DOI 10.1145/2872427.2883084. URL http://dx.doi.org/10.1145/2872427.2883084
11. Eagle, N., Pentland, A.S.: Eigenbehaviors: identifying structure in routine. Behavioral Ecology and Sociobiology **63**(7), 1057–1066 (2009). DOI 10.1007/s00265-009-0830-6
12. Ekman, F., Keränen, A., Karvo, J., Ott, J.: Working day movement model. In: Proceedings of the 1st ACM SIGMOBILE Workshop on Mobility Models, MobilityModels '08, pp. 33–40. ACM, New York, NY, USA (2008). DOI 10.1145/1374688.1374695. URL http://doi.acm.org/10.1145/1374688.1374695
13. Erlander, S., Stewart, N.F.: The Gravity model in transportation analysis : theory and extensions. Topics in transportation. VSP, Utrecht, The Netherlands (1990). URL http://opac.inria.fr/record=b1117869
14. Ghosh, J., Philip, S.J., Qiao, C.: Sociological orbit aware location approximation and routing in manet. In: 2nd International Conference on Broadband Networks, 2005., pp. 641–650 Vol. 1 (2005). DOI 10.1109/ICBN.2005.1589669
15. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. Nature **453**(7196), 779–782 (2008). DOI 10.1038/nature06958

16. Hasan, S., Schneider, C., Ukkusuri, S., González, M.: Spatiotemporal Patterns of Urban Human Mobility. Journal of Statistical Physics **151**(1-2), 304–318 (2013). DOI 10.1007/ s10955-012-0645-0. URL `http://dx.doi.org/10.1007/s10955-012-0645-0`

17. Hess, A., Hummel, K.A., Gansterer, W.N., Haring, G.: Data-driven human mobility modeling: A survey and engineering guidance for mobile networking. ACM Comput. Surv. **48**(3), 38:1–38:39 (2015). DOI 10.1145/2840722. URL `http://doi.acm.org/10. 1145/2840722`

18. Hidalgo, C.A., Rodriguez-Sickert, C.: The dynamics of a mobile phone network. Physica A: Statistical Mechanics and its Applications **387**(12), 3017–3024 (2008). DOI http://dx.doi.org/10.1016/j.physa.2008.01.073. URL `http://www.sciencedirect.com/ science/article/pii/S0378437108000976`

19. Iovan, C., Olteanu-Raimond, A.M., Couronné, T., Smoreda, Z.: Moving and calling: Mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies. In: Springer (ed.) 16th International Conference on Geographic Information Science (AGILE'13), pp. 247–265 (2013). DOI 10.1007/978-3-319-00615-4 14. URL `http://dx.doi.org/10.1007/978-3-319-00615-4_14`

20. Janssens, D.: Data Science and Simulation in Transportation Research, 1st edn. IGI Global, Hershey, PA, USA (2013)

21. Jiang, S., Jr, J.F., González, M.: Clustering daily patterns of human activities in the city. Data Mining and Knowledge Discovery **25**(3), 478–510 (2012). DOI 10.1007/ s10618-012-0264-z

22. Jung, W.S., Wang, F., Stanley, H.E.: Gravity model in the korean highway. EPL (Europhysics Letters) **81**(4), 48,005 (2008). URL `http://stacks.iop.org/0295-5075/81/ i=4/a=48005`

23. Karamshuk, D., Boldrini, C., Conti, M., Passarella, A.: Human mobility models for opportunistic networks. IEEE Communications Magazine **49**(12), 157–165 (2011). DOI 10.1109/MCOM.2011.6094021. URL `http://ieeexplore.ieee.org/xpls/abs_ all.jsp?arnumber=6042290&tag=1`

24. Kopp, C., Kochan, B., May, M., Pappalardo, L., Rinzivillo, S., Schulz, D., Simini, F.: Evaluation of spatio–temporal microsimulation systems. In: L.K. D. Janssens A. Yasar (ed.) Data on Science and Simulation in Transportation Research. IGI Global (2014)

25. Kosta, S., Mei, A., Stefa, J.: Small world in motion (SWIM): Modeling communities in ad-hoc mobile networking. In: 2010 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), pp. 1–9. IEEE (2010). DOI 10.1109/secon.2010.5508278. URL `http://dx.doi.org/10.1109/ secon.2010.5508278`

26. Lee, K., Hong, S., Kim, S.J., Rhee, I., Chong, S.: Slaw: A new mobility model for human walks. In: INFOCOM 2009, IEEE, pp. 855–863 (2009). DOI 10.1109/INFCOM.2009. 5061995

27. Lee, K., Hong, S., Kim, S.J., Rhee, I., Chong, S.: Slaw: Self-similar least-action human walk. IEEE/ACM Trans. Netw. **20**(2), 515–529 (2012). DOI 10.1109/TNET.2011. 2172984. URL `http://dx.doi.org/10.1109/TNET.2011.2172984`

28. Lenormand, M., Bassolas, A., Ramasco, J.J.: Systematic comparison of trip distribution laws and models. Journal of Transport Geography **51**, 158 – 169 (2016). DOI 10. 1016/j.jtrangeo.2015.12.008. URL `http://www.sciencedirect.com/science/article/ pii/S0966692315002422`

29. Lenormand, M., Gonçalves, B., Tugores, A., Ramasco, J.J.: Human diffusion and city influence. Journal of The Royal Society Interface **12**(109) (2015). DOI 10.1098/rsif. 2015.0473

30. Liao, L., Patterson, D.J., Fox, D., Kautz, H.: Learning and inferring transportation routines. Artif. Intell. **171**(5-6), 311–331 (2007). DOI 10.1016/j.artint.2007.01.006

31. Marchetti, S., Giusti, C., Pratesi, M., Salvati, N., Giannotti, F., Pedreschi, D., Rinzivillo, S., Pappalardo, L., Gabrielli, L.: Small area model-based estimators using big data sources. Journal of Official Statistics **31**(2), 263–281 (2015). DOI 10.1515/jos-2015-0017

32. McInerney, J., Stein, S., Rogers, A., Jennings, N.R.: Breaking the habit: Measuring and predicting departures from routine in individual human mobility. Pervasive and Mobile Computing **9**(6), 808–822 (2013)

33. Meloni, S., Perra, N., Arenas, A., Gómez, S., Moreno, Y., Vespignani, A.: Modeling human mobility responses to the large-scale spreading of infectious diseases. Scientific Reports **1**(62) (2011). DOI 10.1038/srep00062. URL `http://dx.doi.org/10.1038/srep00062`

34. Munjal, A., Camp, T., Navidi, W.C.: Smooth: A simple way to model human mobility. In: Proceedings of the 14th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, MSWiM '11, pp. 351–360. ACM, New York, NY, USA (2011). DOI 10.1145/2068897.2068957. URL `http://doi.acm.org/10.1145/2068897.2068957`

35. Pappalardo, L., Rinzivillo, S., Pedreschi, D., Giannotti, F.: Validating general human mobility patterns on gps data. In: Proceedings of the 21th Italian Symposium on Advanced Database Systems, (SEBD2013) (2013)

36. Pappalardo, L., Rinzivillo, S., Qu, Z., Pedreschi, D., Giannotti, F.: Understanding the patterns of car travel. The European Physical Journal Special Topics **215**(1), 61–73 (2013). DOI 10.1140/epjst\%252fe2013-01715-5. URL `http://dx.doi.org/10.1140/epjst%252fe2013-01715-5`

37. Pappalardo, L., Rinzivillo, S., Simini, F.: Human mobility modelling: Exploration and preferential return meet the gravity model. Procedia Computer Science **83**, 934 – 939 (2016). DOI http://dx.doi.org/10.1016/j.procs.2016.04.188. URL `http://www.sciencedirect.com/science/article/pii/S1877050916302216`. The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016) / The 6th International Conference on Sustainable Energy Information Technology (SEIT-2016) / Affiliated Workshops

38. Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F.: Comparing general mobility and mobility by car. In: Proceedings of the 2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence, BRICS-CCI-CBIC '13, pp. 665–668. IEEE Computer Society, Washington, DC, USA (2013). DOI 10.1109/BRICS-CCI-CBIC.2013.116. URL `http://dx.doi.org/10.1109/BRICS-CCI-CBIC.2013.116`

39. Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., Barabasi, A.L.: Returners and explorers dichotomy in human mobility. Nat Commun **6** (2015). DOI 10.1038/ncomms9166. URL `http://dx.doi.org/10.1038/ncomms9166`

40. Pappalardo, L., Vanhoof, M., Gabrielli, L., Smoreda, Z., Pedreschi, D., Giannotti, F.: An analytical framework to nowcast well-being using mobile phone data. International Journal of Data Science and Analytics pp. 1–18 (2016). DOI 10.1007/s41060-016-0013-2. URL `http://dx.doi.org/10.1007/s41060-016-0013-2`

41. Ranjan, G., Zang, H., Zhang, Z.L., Bolot, J.: Are call detail records biased for sampling human mobility? SIGMOBILE Mob. Comput. Commun. Rev. **16**(3), 33–44 (2012). DOI 10.1145/2412096.2412101. URL `http://doi.acm.org/10.1145/2412096.2412101`

42. Reades, J., Calabrese, F., Sevtsuk, A., Ratti, C.: Cellular census: Explorations in urban data collection. IEEE Pervasive Computing **6**(3), 30–38 (2007). DOI 10.1109/MPRV.2007.53

43. Rinzivillo, S., Gabrielli, L., Nanni, M., Pappalardo, L., Pedreschi, D., Giannotti, F.: The purpose of motion: Learning activities from individual mobility networks. In: Proceedings of the 2014 International Conference on Data Science and Advanced Analytics, DSAA'14, pp. 312–318 (2014). DOI 10.1109/DSAA.2014.7058090

44. Rinzivillo, S., Mainardi, S., Pezzoni, F., Coscia, M., Pedreschi, D., Giannotti, F.: Discovering the geographical borders of human mobility. Künstliche Intelligenz **26**(3), 253–260 (2012). DOI 10.1007/s13218-012-0181-8

45. Schneider, C.M., Belik, V., Couronné, T., Smoreda, Z., González, M.C.: Unravelling daily human mobility motifs. Journal of The Royal Society Interface **10**(84) (2013). DOI 10.1098/rsif.2013.0246. URL `http://rsif.royalsocietypublishing.org/content/10/84/20130246`

46. Schwamborn, M., Aschenbruck, N.: Introducing geographic restrictions to the slaw human mobility model. In: 2013 IEEE 21st International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems, pp. 264–272 (2013). DOI 10.1109/MASCOTS.2013.34

47. Simini, F., González, M.C., Maritan, A., Barabási, A.L.: A universal model for mobility and migration patterns. Nature **484**, 96–100 (2012). DOI 10.1038/nature10856

48. Solmaz, G., Akbaş, M.I., Turgut, D.: Modeling visitor movement in theme parks. In: Local Computer Networks (LCN), 2012 IEEE 37th Conference on, pp. 36–43 (2012). DOI 10.1109/LCN.2012.6423650
49. Solmaz, G., Akbaş, M.I., Turgut, D.: A mobility model of theme park visitors. IEEE Transactions on Mobile Computing **14**(12), 2406–2418 (2015). DOI 10.1109/TMC.2015. 2400454
50. Song, C., Koren, T., Wang, P., Barabási, A.L.: Modelling the scaling properties of human mobility. Nature Physics **6**(10), 818–823 (2010). DOI 10.1038/nphys1760. URL `http://dx.doi.org/10.1038/nphys1760`
51. Song, C., Qu, Z., Blumm, N., Barabási, A.L.: Limits of predictability in human mobility. Science **327**(5968), 1018–1021 (2010). DOI 10.1126/science.1177170
52. Spinsanti, L., Berlingerio, M., Pappalardo, L.: Mobility and geo-social networks. In: Mobility Data: Modeling, Management, and Understanding, pp. 315–333. Cambridge press (2013)
53. Thiemann, C., Theis, F., Grady, D., Brune, R., Brockmann, D.: The structure of borders in a small world. PloS one **5**(11), e15,422 (2010)
54. Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabási, A.L.: Human mobility, social ties, and link prediction. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11, pp. 1100–1108. ACM, New York, NY, USA (2011). DOI 10.1145/2020408.2020581
55. Wang, P., Hunter, T., Bayen, A.M., Schechtner, K., González, M.C.: Understanding road usage patterns in urban areas. Scientific Reports **2**(1001) (2012). DOI 10.1038/ srep01001
56. Wilson, A.G.: The use of entropy maximising models, in the theory of trip distribution, mode split and route split. Journal of Transport Economics and Policy pp. 108–126 (1969). DOI 10.2307/20052128
57. Zheng, Q., Hong, X., Liu, J., Cordes, D., Huang, W.: Agenda driven mobility modelling. IJAHUC **5**(1), 22–36 (2010). DOI http://dx.doi.org/10.1504/IJAHUC.2010.03
58. Zipf, G.K.: The p1p2/d hypothesis: On the intercity movement of persons. American Sociological Review **11**(6), 677–686 (1946)