Assessing reliable human mobility patterns from higher-order memory in mobile communications

Joan T. Matamalas*, Manlio De Domenico*, and Alex Arenas
Departament d'Enginyeria Informàtica i Matemàtiques,
Universitat Rovira i Virgili, 43007 Tarragona, Spain

Understanding how people move within a geographic area, e.g. a city, a country or the whole world, is fundamental in several applications, from predicting the spatio-temporal evolution of an epidemics to inferring migration patterns. Mobile phone records provide an excellent proxy of human mobility, showing that movements exhibit a high level of memory. However, the precise role of memory in widely adopted proxies of mobility, as mobile phone records, is unknown. Here we use 560 millions of call detail records from Senegal to show that standard Markovian approaches, including higher-order ones, fail in capturing real mobility patterns and introduce spurious movements never observed in reality. We introduce an adaptive memory-driven approach to overcome such issues. At variance with Markovian models, it is able to realistically model conditional waiting times, i.e. the probability to stay in a specific area depending on individual's historical movements. Our results demonstrate that in standard mobility models the individuals tend to diffuse faster than what observed in reality, whereas the predictions of the adaptive memory approach significantly agree with observations. We show that, as a consequence, the incidence and the geographic spread of a disease could be inadequately estimated when standard approaches are used, with crucial implications on resources deployment and policy making during an epidemic outbreak.

Keywords: Human Mobility, Markovian Model, Epidemics Spreading, Complex Networks, Diffusion

I. INTRODUCTION

People move following complex dynamical patterns at different geographical scales, e.g. among areas of the same city, among cities and regions of the same country or among different countries. Such patterns have been recently revealed by using human mobility proxies [1–5] and, intriguingly, some specific patterns tend to repeat more than others, with evidences [6, 7] of memory of meaningful locations playing a fundamental role in our understanding of human mobility. In fact, human dynamics might significantly affect how epidemics spread [2, 6, 8–10] or how people migrate from one country to another [4].

The collaboration between researchers and mobile operators recently opened new promising directions to gather information about human movements, country demographics and health, faster and cheaper than before [1, 10–18]. In fact, mobile phones heterogeneously penetrated both rural and urban communities, regardless of richness, age or gender, providing evidences that mobile technologies can be used to obtain real-time information about individual's location and social activity, in order to build realistic demographics and socio-economics maps of a whole country [19]. Mobile data have been successfully used in a wide variety of applications, e.g., to estimate population densities and their evolution at national scales [13], to confirm social theories of behavioral adaptation [20] and to capture anomalous behavioral patterns associated to religious, catastrophic or massive social events [21]. Even more recently, the public availability of mobile phone data sets further revolutionized the field, e.g., by allowing ubiquitous sensing to map poverty, to monitor social segregation and to optimize information campaigns to reduce epidemics spreading [14, 22], to cite just some of them [18].

Although some limitations, mobile phone data still provide one the most powerful tools for sensing complex social systems and represent a valuable proxy for studies where human mobility plays a crucial role [1–4, 6, 8–10, 15, 22–24]. Milestone works in this direction have shown that human trajectories exhibit more temporal and spatial regularity than previously thought. Individuals tend to return to a few highly frequented locations and to follow simple reproducible patterns [1, 5], allowing a higher accuracy in predicting their movements [3] and significantly affecting the spreading of transmittable diseases [6]. However, the increasing interest for using mobile phone data in applications should be accompanied by a wise usage of the information they carry on. In fact, an inadequate model accompanied by incomplete data and scarce knowledge of other fundamental factors influencing the model itself, might lead, for instance, to a wrong estimation of the incidence of an epidemics and its evolution [25].

Here, we used high-quality mobile phone data, consisting of more than 560 millions of call detail records, to show that standard approaches might significantly overestimate mobility transitions between distinct geographical areas, making difficult to build a realistic model of human mobility. To overcome this issue, we developed an adaptive memory-driven model based on empirical observations that better captures existing correlations in human dynamics, showing that it is more suitable than classical memoryless or higher-order models to understand how individuals move and, for instance, might spread a disease.

II. MATERIALS AND METHODS

A. Markovian model of human mobility

Let us consider a physical mobility network composed by nodes, representing geographic areas, connected by weighted edges, representing the fraction of individual movements among them. Usually, the weights are inferred from geolocated activities of individuals, e.g. the consecutive airports where a plane departs and lands or, as in this work, the cell towers where a person makes consecutive calls.

A standard approach to deal with mobility models of dynamics [3, 4, 6, 10, 15, 16, 26] is to consider each node as a state of a Markov process, obtaining the flux between any pair of nodes from consecutive calls, and to build a mobility matrix F_{ij} encoding the probability that an individual in node i will move to node j $(i, j = 1, 2, ..., n)$. Here, we use a similar approach to build the mobility matrix for each individual $\ell = 1, 2, ..., \mathcal{L}$ separately and we then average over the whole set of mobility matrices, to obtain the transition probability of an individual, on average:

$$F_{ij} = \frac{\sum\limits_{\ell=1}^{\mathcal{L}} f_{ij}^{(\ell)}}{\sum\limits_{\ell=1}^{\mathcal{L}} \sum\limits_{k=1}^{n} f_{ik}^{(\ell)}}, \tag{1}$$

where $f_{ij}^{(\ell)}$ is the number of times the individual ℓ makes at least one call in node j after making at least one call in node i.

We did not impose a specific time window to calculate transitions, to avoid introducing biases and undesired effects due to the choice of the temporal range and it is worth remarking that other normalizations can be considered depending on data and metadata availability [15]. Where not otherwise specified, we considered the mobility matrix obtained from the whole period of observation. This model is known as "first-order" (or 1-memory) because the present state is the only information required to choose the next state. Although very useful, this has the fundamental disadvantage that it does not account for mobility memory. In fact, it is very likely that an individual moves to a neighboring area (by means of a car or public transportation) to work and after a few hours he or she will go back to the original position. This effect has been shown to be relevant, for instance, at country level, where individuals fly from one city to another and often go back to their origin instead of moving towards a different city [7]. This memory is an intrinsic property of human mobility and must be taken into account for a realistic modeling of people movements between different geographic areas. When memory is taken into account, each physical node (e.g., $i \in \mathcal{A}$) is replaced by the corresponding state-nodes (e.g., $\mathrm{i} \lhd \mathrm{j} \in \tilde{\mathcal{A}}$ if memory is of order 2) encoding the information that an individual is in node i when he or she comes from j. While $\mathbf{F}$ encodes information about the network of n physical nodes, we need to introduce a new matrix $\mathbf{H}$ to encode information about the network of n^2 state-nodes, accounting for the allowed binary combinations (e.g. $\mathrm{k} \lhd \mathrm{j}$, $j, k = 1, 2, ..., n$) between physical nodes. Similarly, higher-order memory can be taken into account by building appropriate matrices.

We use different mobility matrices to build different mobility models. Let $N_i(t)$ indicate the population of the physical node $i \in \mathcal{A}$ at time t, then the n mobility equations describing how the flux of people diffuses through the network are given by

$$N_i(t+1) = \sum_{j=1}^{n} F_{ji} N_j(t). \tag{2}$$

In the case of $\tau-$memory, we indicate by $\tilde{N}_\alpha(t)$ the population of the state-node $\alpha \in \tilde{A}$ at time t and the n^τ mobility equations required to describe the same process are given by

$$\tilde{N}_\alpha(t+1) = \sum_{\rho=1}^{n^\tau} H_{\rho\alpha} \tilde{N}_\rho(t). \tag{3}$$

The population in each physical node at time t is given by the sum of the population in the corresponding state-nodes. It is worth remarking that, in general, the matrix $\mathbf{H}$ can be a function of time as well and the equations would keep their structural form.

B. Adaptive memory model of human mobility

However, spatial human mobility is quite complex and (higher-order) Markovian dynamics might not be suitable to model peculiar patterns such as returning visits and conditional waiting times, i.e. the probability to stay in a location depending on the origin of the travel.

We will discuss better this point in the following. Let us consider, for instance, the call sequence *BBBBCCCSSS* made by an individual traveling between three American cities: Chicago, Boston and San Antonio. The main drawbacks of Markovian models – of order lower than three – become evident in a scenario like this one, because the number of consecutive calls in the same city exceeds the memory of the model and the spatial information about previously visited locations is lost. Clearly, in presence of more complicated patterns, increasing the order of the model will not solve the issue and some information will be inevitably lost. Alternatively, we could aggregate consecutive calls in the same place to a single identifier, e.g. the previous sequence would be reduced to *BCS*. In this case, a Markovian model would preserve the spatial information and correctly identify the transitions between the three cities, at the price of losing information about how many calls have been made in each place.

In absence of detailed temporal information about calling activity, the number of consecutive calls in a specific location can be used as a proxy: higher the number of calls larger the waiting time. The temporal information about the amount of time spent in each location is critical for many dynamical processes like spreading or congestion. We assert that this time, like the next visited location, is conditioned by previous movements of the individuals. To illustrate this, we use the example shown in Fig. 1, where people from three different places (nodes blue, green and orange) go to the same destination (node red), stay some time in there, and come back to the origin of their trip. The self-loops in the central (red) node represent the time spent there, the color encoding individuals coming from different origins and the size encoding the amount of time spent. For instance, individuals coming from the blue node wait more than individuals coming from the green node. This type of dependence is what we call conditional waiting time.

To better appreciate this fact, let us consider holiday trips. Individuals making expensive intercontinental trips tend to spend more time visiting the destination than individuals making cheaper trips, achieving a good trade-off between the travel cost and the time spent. Another emblematic case is urban mobility. For instance, the red node might be an expensive commercial area, the green node a wealthy neighborhood and the blue node a less wealthy area. In this scenario, that should be considered only for illustrative purposes, individuals coming from the less wealthy area are more likely to be qualified workers in the commercial one, with long and frequent visits. Conversely, individuals from the wealthy are more likely to make unfrequent and shorter visits for shopping, for instance.

The importance of accounting for conditional waiting times will be evident later, when we will consider the spreading of an epidemics in a country.

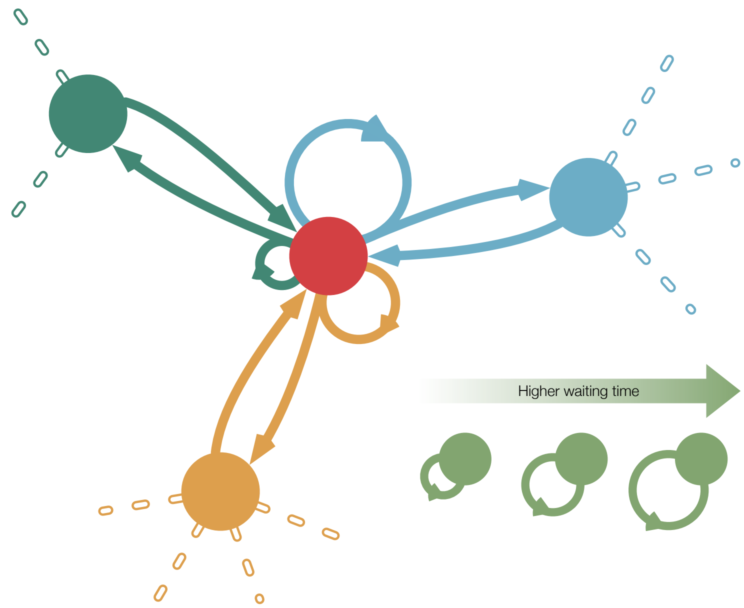


FIG. 1. **Conditional waiting times.** An example of human mobility between four different places. Individuals from green, blue and orange nodes move to the red central node and, after some time, go back to their previous location. The amount of time spent in the red node by individuals coming from the other nodes depends on their previous location, and it is represented by self-loops of different size.

Here, we propose a mobility model, that we name *adaptive memory*, able to account for conditional waiting times. At first order, the method is equivalent to a classical first-order Markovian model, whereas significant differences emerge for increasing memory with respect to standard approaches. For instance, at second order, the 2-memory mobility matrix is built between all possible pairs of nodes (2-states), as in a standard second-order Markovian model. However, instead of considering transitions between areas in the sequence of calls, as a second-order Markovian model does, transitions in the sequence of distinct geographical areas are considered. This point is crucial, and we better clarify it with the example shown in Fig. 2, where the differences between adaptive memory and Markovian models, in terms of probability assigned to different mobility patterns, are reported.

The importance of such differences is reflected in the ability of each model to predict successive individual movements. In fact, the presence of spurious or under-represented patterns might significantly affect the results, as shown in Fig. 3. In this example, two sequences of phone calls generated by two different users moving between three cities – B, C and S – are considered. Markovian models generate spurious patterns that are never observed in the data, issue not affecting the adaptive memory model by construction. Morover, our approach predicts the next movement with

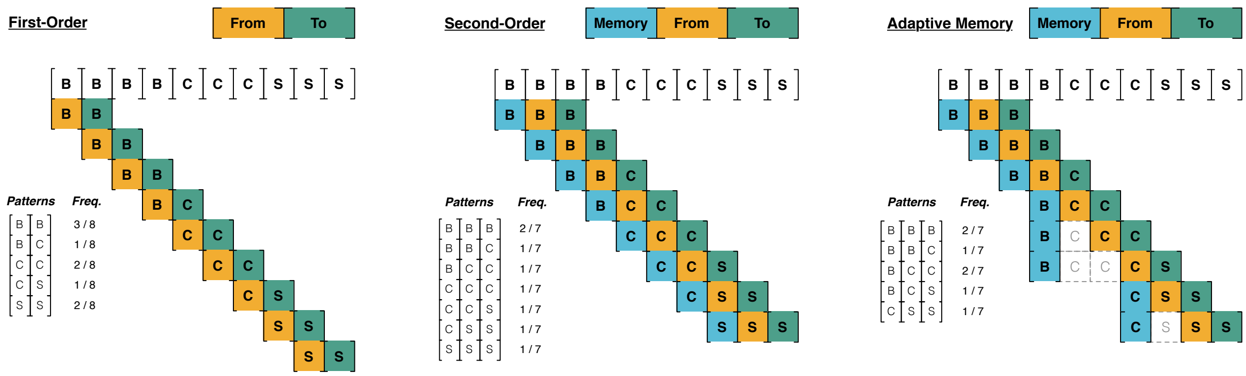


FIG. 2. **Comparing different mobility models**. Mobility models built from a representative sequence of mobile phone calls (BBBBCCCSSS) made, for instance, by an individual during travels between three American cities, namely Chicago (C), San Antonio (S) and Boston (B). Let us focus on the pattern $S \leftarrow C \leftarrow B$, that is the real sequence of movements in the geographical space. The first-order model predicts a probability of $\frac{1}{64}$, the second-order model a probability of $\frac{1}{49}$, whereas the adaptive 2-memory estimates a probability of $\frac{1}{7}$, closer to observation.

more accuracy than Markovian ones, because it correctly takes into account conditional waiting times.

The difference between the adaptive memory and Markovian models becomes more evident when the corresponding transition matrices are compared. There is no difference at the first order, thus we will focus on the comparison between τ-order Markovian and adaptive $\tau-$memory models, in the following.

In both models, the number of possible transitions between state nodes is the same and equals $n^{2\cdot(\tau-1)}$, where n is the number of physical nodes. For instance, in second-order models, there are $n^2 \times n^2$ transition matrices with n^3 possible transitions between state nodes, as shown in Fig. 4. However, the way how each model stores repeating calls in the same physical node is very different. While adaptive memory stores this information into the n^τ diagonal elements of the matrix, encoding the conditional waiting times discussed in the previous section, Markovian models redistribute this information among off-diagonal entries, because they do not allow this type of self-loops by construction.

More specifically, the information is redistributed among transitions between state nodes of the same physical node. The entries of off-diagonal blocks – corresponding to transitions between state nodes of different physical nodes – are the same in both models. Therefore, while the stationary probability of finding a random walker in a physical node is not different in the two models, it is different at the level of state nodes and, as we will see later, this significantly affects diffusion

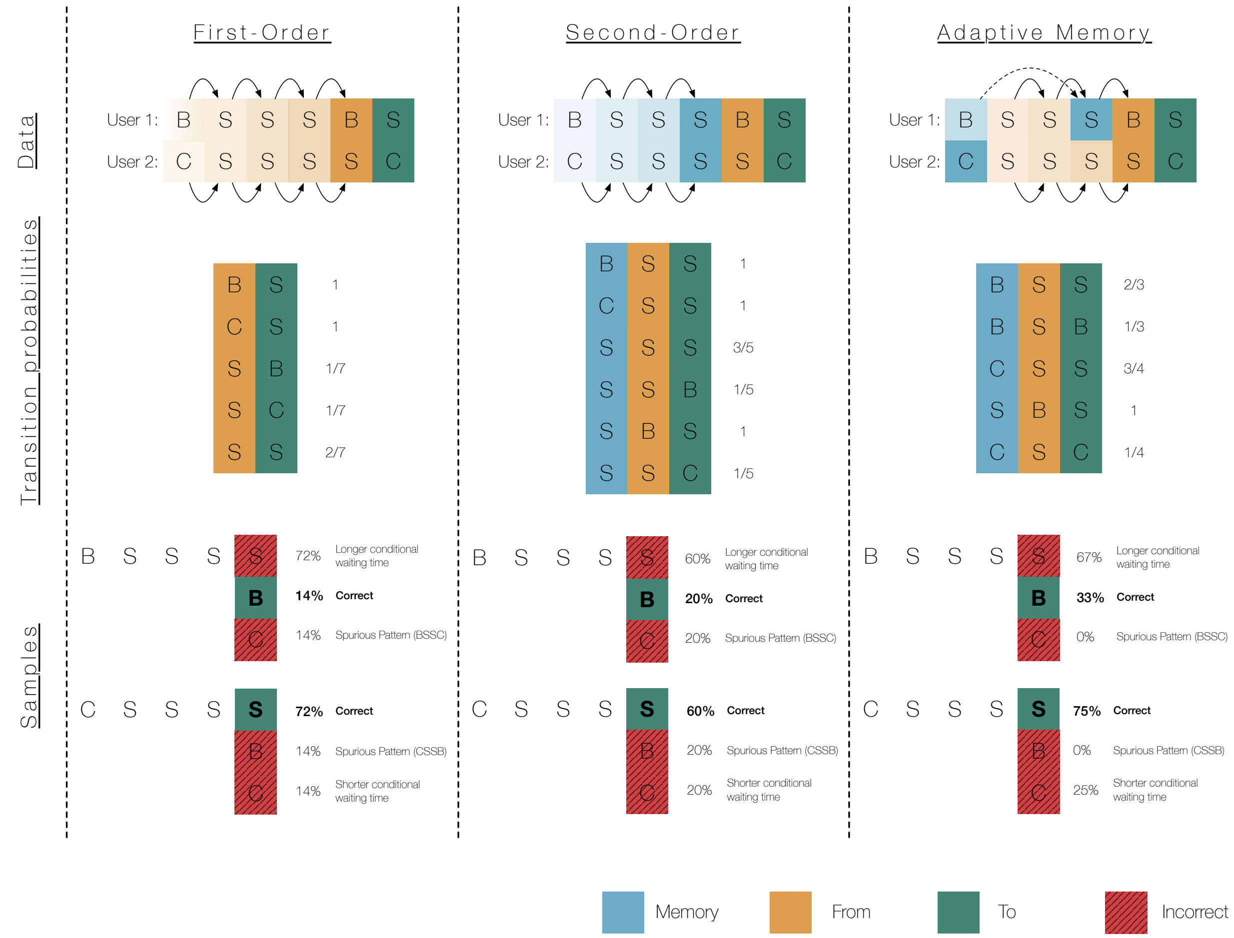


FIG. 3. **Predicting individual mobility.** Using the sequence of calls made by two different users (1 and 2) – starting from two different locations (B and C) and visiting a new location S – we build first-order and second-order Markov models, as well as the adaptive memory one. We use each mobility model to generate the possible mobility sequences. Given that there are two empirical starting points, we originated the sampled sequences in B and C, respectively. In the figure, for each sample, we report the fraction of times it is reproducing observation ("Correct"), it is a non-observed mobility pattern ("Spurious Pattern") and it is underestimating or overestimating waiting times ("Longer/Shorter Conditional Waiting Time").

processes such as epidemics spreading.

C. Overview of the dataset

In the next section, we will quantify the impact of adaptive memory on human mobility modeling by using data sets provided by the Data for Development Challenge 2014 [27] and some supplementary data sets provided by partners of the challenge. Mobile phone data consist of communications among 1666 towers distributed across Senegal. We exploit this information to map

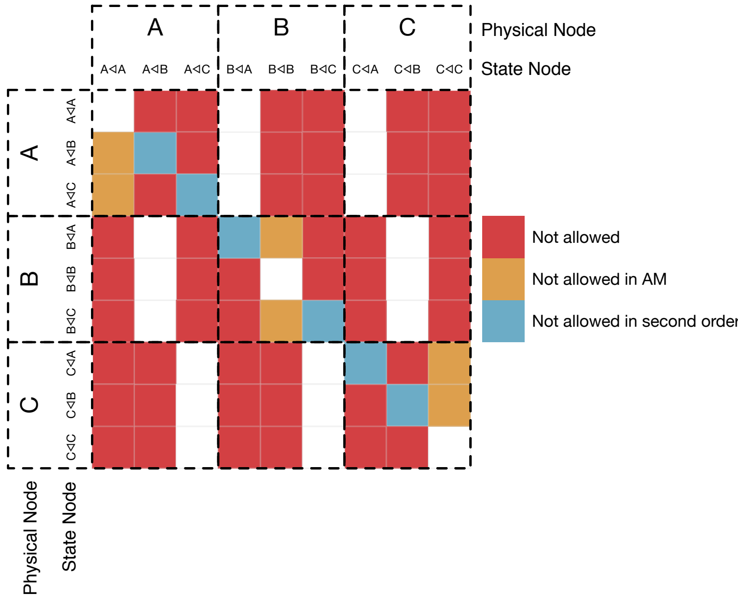


FIG. 4. **Mobility Matrix** Second-order transition matrix for three physical nodes A, B and C. The state nodes are represented with the notation $x \lhd y$ meaning that walkers in this node have traveled from node y to node x. The cells in red are not used by either second-order markovian model or adaptive memory model. The cells in bluish are used only in adaptive memory model, while, the cells in orange are used only in the second-order markovian model. The cells in white are used by both models.

communication patterns between different areas of the country (i.e. the arrondissements). Another subset consists of 560 millions call records of about 150,000 users along one year at at the spatial resolution of arrondissements. We use this information to map individuals' movements among different arrondissements. Demographics information has been obtained from the Senegal data portal [28], an official resource. It is worth noting that information has been manually checked against inconsistencies and data about population for the arrondissements of Bambilor, Thies Sud, Thies Nord, Ndiob and Ngothie were not available. We reconstructed the missing information by combining mobile phone activity and available demographics data (Fig. 5). Such arrondissements did not exist at the time when the population census was obtained, because they were part of larger administrative areas. Information is available for older arrondissements, therefore we devise a procedure to infer the population in the new areas by using phone calls as a proxy to population density.

We have used the data to also infer more realistic contact rates to be used in viral spreading simulations. The contacts among individuals are generally quite difficult to track at country level. Their rate varies depending on several social and demographical factors such as age, gender, location, urban development, *etc.* [29, 30]. Nevertheless, there are evidences from European and African countries that, on average, the number of daily physical contacts among individuals range from 11 to 22 [29, 30]. There are no available data about contact rate in each arrondissement of Senegal, therefore we need to infer this information from available sources. We first estimate the population density for each region, an administrative level coarser than arrondissement, using available data about number of inhabitants and area. As a plausible range of contact rates, we consider 10 and 25. Under the assumption that the contact rate is proportional to the population density, we assign a value to each region that ranges between 10 and 25, with extremal values assigned to the regions with lowest and highest population density, respectively. Therefore, we assign the same contact rate to all arrondissements pertaining to the same region. We obtain a contact rate between 10 and 11 for all regions, except Dakar which has the highest population density.

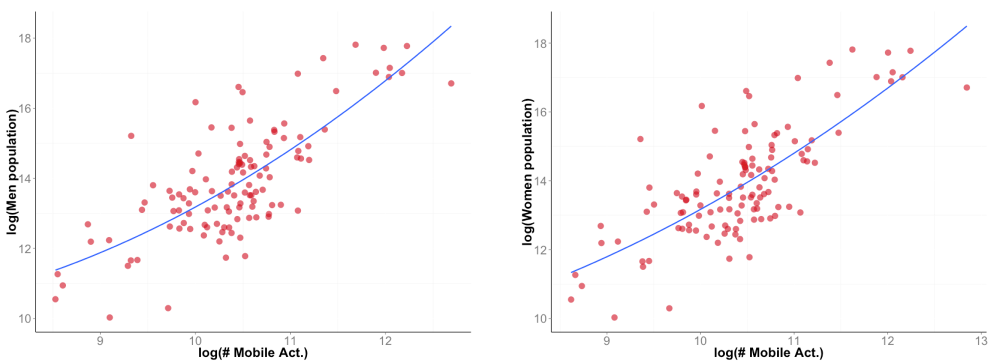
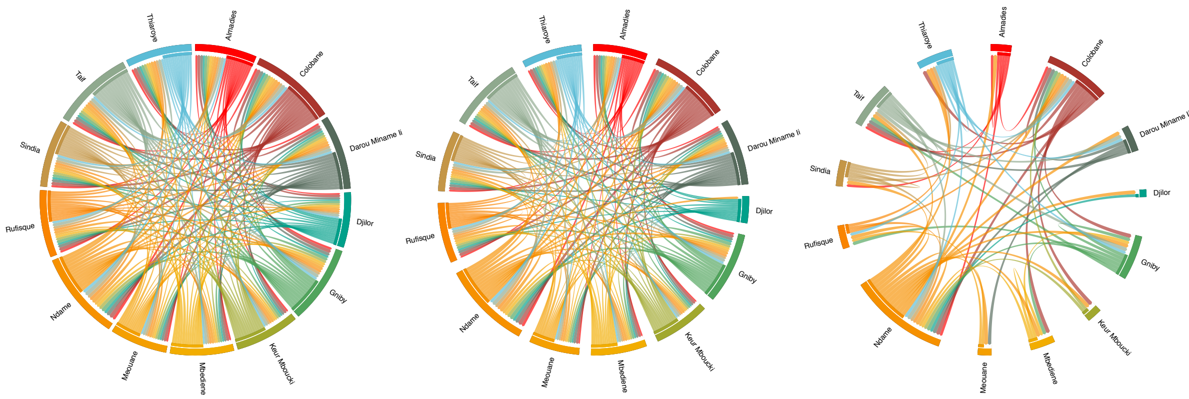


FIG. 5. **Inferring men and women populations.** Second-order polynomial model (solid line) fitting the log-log relationships between the observed mobile phone data and demographics data (points). Men (A) and women (B) population were fitted separately, thanks to data availability, and have been used to infer the populations in the arrondissements of Bambilor, Thies Sud, Thies Nord, Ndiob and Ngothie.

III. RESULTS

A. Understanding human mobility flow

We show in Fig. 6 the significant differences in modeling the mobility flow using first-order (FO), second-order (SO) and adaptive memory (AM) models. Markovian models provide very similar transition patterns, whereas adaptive memory provides very different results. The adaptive memory model exhibits significantly less returning transitions than Markovian models, but – on average – with much higher probability of observing them. In fact, 47.4% of patterns captured by the first-order approach and 43.4% captured using second-order are spurious because they are not observed in reality. Remarkably, the probability that an individual comes back to her origin is on average six times higher using adaptive memory models than using first-order, and five times higher using second-order. In the Supplementary Material, we show the result of the same analysis for the gravity model [31, 32] and the more recent radiation model [4, 33], two widely adopted approaches to model human mobility.



(a) First-Order (b) Second-Order (c) Adaptive Memory

FIG. 6. **Mobility flow among a sub-set of Senegal's arrondissements**. For simplicity, we illustrate the effects of each model by considering a subset of 13 arrondissements and patterns that goes through one specific arrondissement (Kael, in this example) after departing from their origin and before reaching their destination. The figure shows the mobility modeled by means of first-order (A), second-order (B) and adaptive 2-memory (C), putting in evidence the different mobility patterns between Markovian models and adaptive memory. For instance, the adaptive memory module captures returning patterns (i.e. movements like X $\rightarrow$ Kael $\rightarrow$ X) better than the first-order model. See Supplementary Material for results obtained from gravity and radiation models.

To compare the accuracy of both models against the mobility behavior observed in data, we use the coverage, defined as the fraction of nodes visited by an individual within a given amount of time. We calculate the coverage for each individual in the data, over a period of one months, and then we average over all arrondissements to obtain a measure at country level. For the same period of time, we generate three transition matrices $\mathbf{F}$, $\mathbf{H}$ and $\mathbf{A}$ encoding the mobility dynamics for first-order, second-order and adaptive memory models, respectively. To better replicate the calling behavior of the individuals in the data set, we extract information about the distribution of time between calls and we use this information in our simulations (see Supplementary Material).

In Fig. 7A and B we show that people diffuse in the country too fast using Markovian models,

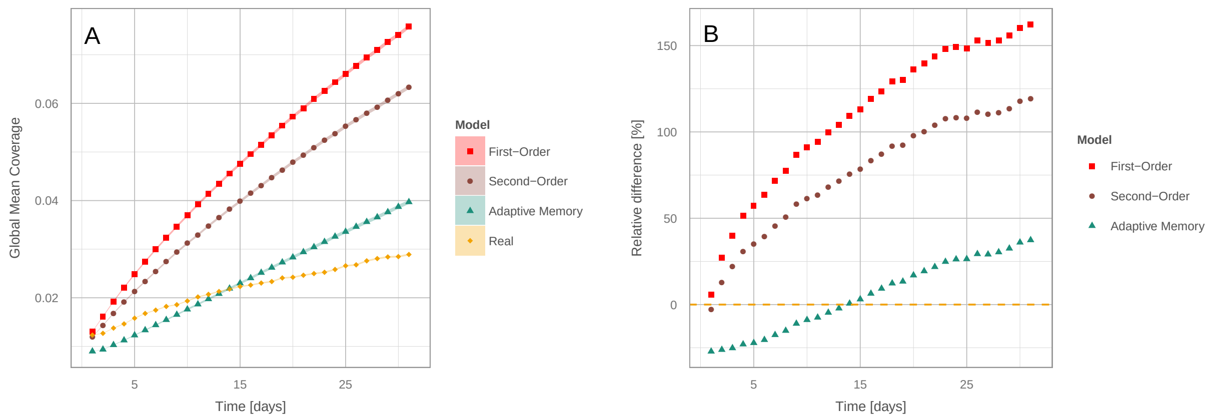


FIG. 7. **Observed human mobility and theoretical predictions**. (A) Temporal evolution of the global mean coverage calculated from real data and from simulations using first-order (FO), second-order (SO) and adaptive memory (AM) models. (B) Relative difference between the coverage observed in real human mobility and the one obtained from simulations. See Supplementary Material for results obtained from gravity and radiation models.

whereas significantly slower diffusion is found with adaptive memory, in agreement with empirical observation. In the Supplementary Material, we show the result of the same analysis for both the gravity and the radiation models. We observe that the gravity model is not suitable to reproduce the observation, whereas the radiation model provides results comparable with the adaptive memory model proposed in this study.

These results have deep implications, for instance, in short-term or long-term predictions of epidemic spreading or national infrastructure planning.

B. Impact of human mobility models on the spreading of epidemics

Here, we focus on epidemic spreading. How infectious individuals move among different locations has a strong influence in how diseases diffuse in a population. We considered each arrondissement as a meta-population where any individual can interact with a limited number of other individuals. We use a SEIR compartmental model [34] to characterize the epidemics evolution within each arrondissement and mobility models to simulate people traveling in the country.

The discrete time step of the following models is $\Delta t \approx 1$ hour, approximately the observed median between two successive calls from the same individual. The parameters are demographical and epidemiological. Demographics parameters include the birth $B = \tilde{B}\Delta t$ and death $\delta = \tilde{\delta}\Delta t$ probability, whereas epidemiological parameters correspond to the latent period τ_E of the infection, from which the probability $\epsilon = \Delta t/\tau_E$ to pass to the infectious state is calculated, and the infectious period τ_I, from which the probability $\gamma = \Delta t/\tau_I$ to recover from or die because of the infection is calculated. The last parameter is the effective transmission probability

$$\beta_i(t) = 1 - \left(1 - \tilde{\beta}\Delta t \frac{I_i(t)}{N_i(t)}\right)^{c_i \Delta t}, \tag{4}$$

an arrondissement-dependent parameter that depends on the average number of contacts per unit of time c_i experienced by an individual in node i, the fraction of infected individuals in that node and the transmission risk $\tilde{\beta}\Delta t$ in case of contact with an infectious individual. In fact, the definition of $\beta_i(t)$ induces a type-II reaction-diffusion dynamics [9] accounting for the fact that each individual does not interact with *all* the other individuals in the meta-population, but only with a limited sample. If the number of infected agents is small (i.e. $I_i(t) \approx 0$) the Taylor expansion of $\beta_i(t)$ truncated at the first order gives the classical factor $\tilde{\beta}\Delta t c_i \Delta t \frac{I_i(t)}{N_i(t)}$ [34]. It follows that the equations describing the average spreading of a disease according to a SEIR model coupled to first-order mobility are given by

$$\begin{aligned}
S_i(t+1) &= \sum_{j=1}^{n} F_{ji}\left[(1-\delta-\beta_j(t))S_j(t) + BN_j(t)\right] \\
E_i(t+1) &= \sum_{j=1}^{n} F_{ji}\left[(1-\epsilon-\delta)E_j(t) + \beta_j(t)S_j(t)\right] \\
I_i(t+1) &= \sum_{j=1}^{n} F_{ji}\left[(1-\gamma-\delta)I_j(t) + \epsilon E_j(t)\right] \\
R_i(t+1) &= \sum_{j=1}^{n} F_{ji}\left[(1-\delta)R_j(t) + \gamma I_j(t)\right]
\end{aligned} \tag{5}$$

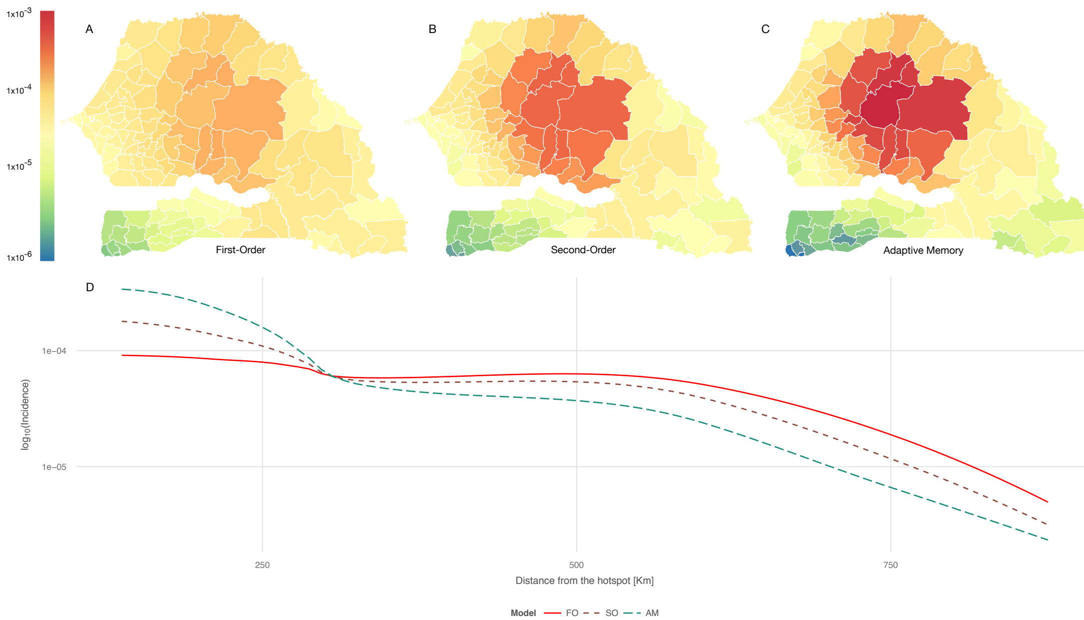


FIG. 8. **Spreading of an influenza-like outbreak in Senegal**. We show the incidence of an influenza-like virus over Senegal arrondissements a week after the infection onset, using first-order (A), second-order (B) and adaptive 2-memory (C) mobility models. The infection started in Barkedji (center of Senegal), where three individuals are initially infected. A SEIR compartmental dynamics with parameters $\beta = 0.05$, $\epsilon = 0.2$, $\gamma = 0.5$ is used to simulate the spreading of the disease within each arrondissement. We found that the number of arrondissements with infected individuals is higher using Markovian dynamics. Conversely, the adaptive memory favors a higher concentration of infected individuals in the arrondissements around the initial location of the infection. In fact, the location of the onset of the epidemic can be better identified using adaptive memory rather than Markovian models. (D) Relation between the incidence in a region and the distance from the hotspot of the infection using the three models. Adaptive memory models spread the incidence on regions closer to the hotspot and this effect is even more evident when higher memory is used.

whereas the coupling to the second-order model is given by

$$\tilde{S}_\alpha(t+1) = \sum_{\psi=1}^{n^2} H_{\psi\alpha}\left[(1-\delta-\tilde{\beta}_\psi^{(\alpha)}(t))\tilde{S}_\psi(t) + B\tilde{N}_\psi(t)\right]$$
$$\tilde{E}_\alpha(t+1) = \sum_{\psi=1}^{n^2} H_{\psi\alpha}\left[(1-\epsilon-\delta)\tilde{E}_\psi(t) + \tilde{\beta}_\psi^{(\alpha)}(t)\tilde{S}_\psi(t)\right]$$
$$\tilde{I}_\alpha(t+1) = \sum_{\psi=1}^{n^2} H_{\psi\alpha}\left[(1-\gamma-\delta)\tilde{I}_\psi(t) + \epsilon\tilde{E}_\psi(t)\right]$$
$$\tilde{R}_\alpha(t+1) = \sum_{\psi=1}^{n^2} H_{\psi\alpha}\left[(1-\delta)\tilde{R}_\psi(t) + \gamma\tilde{I}_\psi(t)\right]$$
$$\tilde{\beta}_\psi^{(\alpha)}(t) = 1 - \left(1 - \tilde{\beta}\Delta t \frac{\sum_{\rho=\lfloor\frac{\alpha}{n}\rfloor n+1}^{\lfloor\frac{\alpha}{n}\rfloor n+n} \tilde{I}_\rho(t)}{\sum_{\rho=\lfloor\frac{\alpha}{n}\rfloor n+1}^{\lfloor\frac{\alpha}{n}\rfloor n+n} \tilde{N}_\rho(t)}\right)^{c_i\Delta t} \tag{6}$$

where $N(t) = \sum_{\psi=1}^{n^\tau} \tilde{N}_\psi(t)$ is the total population in the country at time t, $\lfloor\cdot\rfloor$ indicates the floor function and is used to identify the sub-set of state-nodes corresponding to the same physical node the population $\tilde{S}_\alpha$ belongs to. The equations for the adaptive memory model are the same, except that the transition matrix $\mathbf{A}$ is used instead of $\mathbf{H}$.

We initiate the simulation by infecting five individuals in Barkedji, at the center of Senegal. The differences between the diffusion of the infective process using each mobility model are quite visible in Fig. 8. The spreading is faster for Markovian models, with some arrondissement populated by more infected individuals than adaptive memory. The incidence, i.e. the fraction of infected individuals in an arrondissement, follows different spatial patterns in the three models (see Fig. 8A–C), with a higher incidence observed in the origin of the infection that decreases as we move far from there. This effect is significantly stronger using adaptive memory because it tends to concentrate more infectious individuals close to the origin (see Fig. 8D).

DISCUSSION

Modeling how people move among different locations is crucial for several applications. Given the scarcity of information about individuals' movements, often human mobility proxies such as call detail records, GPS, *etc*, are used instead. Here, we have shown that dynamical models built from human mobility proxies can be significantly wrong, underestimating (or overestimating) real mobility patterns or predicting spurious movements that are not observed in reality. We have proposed a general solution to this issue, by introducing an adaptive memory modeling of human mobility that better captures observed human dynamics and dramatically reduces spurious patterns with respect to memoryless or higher-order Markovian models. However, it is worth remarking that this approach, as all other methods in the literature, is based on the assumption that an individual makes a call in each place he or she visits. In fact, this is not always true and care must be taken when interpreting the results. Fortunately, an appropriate choice of the spatial granularity, for instance at administrative levels corresponding to cities or larger areas, reduces this unavoidable effect. We have validated our model on a data set consisting of 560 millions of call detail records from Senegal. We have found that individuals tend to diffuse faster with standard mobility models than what observed in reality, whereas the adaptive memory approach reconciles empirical observations and theoretical expectations. Our findings have, for instance, a deep impact on predicting how diseases spread in a country. While standard approaches tending to overestimate the geographical incidence of the infection, the more realistic modeling obtained by means of adaptive memory can improve the inference of the hotspot of the infection, helping to design better countermeasures, e.g. more effectives quarantine zones, improved resources deployment or targeted information campaigns.

COMPETING INTERESTS

The authors declare that they have no competing interests.

DATA ACCESSIBILITY

The data used in this manuscript were made publicly available during the D4D Senegal Challenge organized by Orange. More information about data can be found here: `http://www.d4d.orange.com/en/presentation/data`

AUTHOR'S CONTRIBUTIONS

JTM and MDD contributed equally to this work. JTM and MDD developed the theoretical model and carried out the statistical analyses; MDD and AA conceived of the study, designed the study and coordinated the study. All authors wrote the manuscript and gave final approval for publication.

FUNDING

M.D.D. acknowledges financial support from the Spanish program Juan de la Cierva (IJCI-2014-20225). JTM was supported by Generalitat de Catalunya (FI-DGR 2015). A.A. acknowledges financial support from ICREA Academia and James S. McDonnell Foundation and Spanish MINECO FIS2015-71582.

---

[1] Gonzalez MC, Hidalgo CA, Barabasi AL. Understanding individual human mobility patterns. Nature. 2008;453(7196):779–782.

[2] Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. PNAS. 2009;106(51):21484–21489.

[3] Song C, Qu Z, Blumm N, Barabási AL. Limits of predictability in human mobility. Science. 2010;327(5968):1018–1021.

[4] Simini F, González MC, Maritan A, Barabási AL. A universal model for mobility and migration patterns. Nature. 2012;484(7392):96–100.

[5] Lima A, Stanojevic R, Papagiannaki D, Rodriguez P, González MC. Understanding individual routing behaviour. Journal of The Royal Society Interface. 2016;13(116).

[6] Balcan D, Vespignani A. Phase transitions in contagion processes mediated by recurrent mobility patterns. Nature Phys. 2011;7(7):581–586.

[7] Rosvall M, Esquivel AV, Lancichinetti A, West JD, Lambiotte R. Memory in network flows and its effects on spreading dynamics and community detection. Nature Communications. 2014;5:4630.

[8] Ferguson NM, Cummings DA, Cauchemez S, Fraser C, Riley S, Meeyai A, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature. 2005;437(7056):209–214.

[9] Colizza V, Pastor-Satorras R, Vespignani A. Reaction–diffusion processes and metapopulation models in heterogeneous networks. Nature Phys. 2007;3(4):276–282.

[10] Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, et al. Quantifying the impact of human mobility on malaria. Science. 2012;338(6104):267–270.

[11] Eagle N, Pentland A. Reality mining: sensing complex social systems. Personal and ubiquitous computing. 2006;10(4):255–268.

[12] Eagle N, Pentland AS, Lazer D. Inferring friendship network structure by using mobile phone data. PNAS. 2009;106(36):15274–15278.

[13] Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE, et al. Dynamic population mapping using mobile phone data. PNAS. 2014;111(45):15888–15893.

[14] Amini A, Kung K, Kang C, Sobolevsky S, Ratti C. The impact of social segregation on human mobility in developing and industrialized regions. EPJ Data Science. 2014;3(1):6.

[15] Tizzoni M, Bajardi P, Decuyper A, Kon Kam King G, Schneider C, et al . On the Use of Human Mobility Proxies for Modeling Epidemics. PLoS Comput Biol. 2014;10(7):e1003716.

[16] Wesolowski A, Buckee C, Bengtsson L, Wetter E, Lu X, Tatem A. Commentary: Containing the Ebola outbreak–the potential and challenge of mobile network data. PLoS Curr Outbr. 2014;.

[17] Calabrese F, Ferrari L, Blondel VD. Urban Sensing Using Mobile Phone Network Data: A Survey of Research. ACM Computing Surveys (CSUR). 2014;47(2):25.

[18] Blondel VD, Decuyper A, Krings G. A survey of results on mobile phone datasets analysis. To appear in EPJ Data Science. 2015;.

[19] Wesolowski A, Eagle N, Noor AM, Snow RW, Buckee CO. Heterogeneous mobile phone ownership and usage patterns in Kenya. PLoS ONE. 2012;7(4):e35319.

[20] Eagle N, de Montjoye Y, Bettencourt LM. Community computing: Comparisons between rural and urban societies using mobile phone data. In: Computational Science and Engineering, 2009. CSE'09. International Conference on. vol. 4. IEEE; 2009. p. 144–150.

[21] Dobra A, Williams NE, Eagle N. Spatiotemporal Detection of Unusual Human Population Behavior Using Mobile Phone Data. PLoS ONE. 2014;p. DOI:10.1371/journal.pone.0120449.

[22] Lima A, De Domenico M, Pejovic V, Musolesi M. Disease Containment Strategies based on Mobility and Information Dissemination. Scientific Reports. 2015;5:10650.

[23] Colizza V, Vespignani A. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. J Theor Bio. 2008;251(3):450–467.

[24] Tatem AJ, Adamo S, Bharti N, Burgert CR, Castro M, Dorelien A, et al. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. Popul Health Metr. 2012;10(8).

[25] Pandey A, Atkins KE, Medlock J, Wenzel N, Townsend JP, Childs JE, et al. Strategies for containing Ebola in west Africa. Science. 2014;346(6212):991–995.

[26] Wesolowski A, Buckee CO, Pindolia DK, Eagle N, Smith DL, Garcia AJ, et al. The use of census migration data to approximate human movement patterns across temporal scales. PLoS ONE. 2013;8(1):e52971.

[27] de Montjoye YA, Smoreda Z, Trinquart R, Ziemlicki C, Blondel VD. D4D-Senegal: The Second Mobile Phone Data for Development Challenge. arXiv:14074885. 2014;.

[28] ;. http://donnees.ansd.sn/en/.

[29] Rohani P, Zhong X, King AA. Contact network structure explains the changing epidemiology of pertussis. Science. 2010;330(6006):982–985.

[30] Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. PloS one. 2014;9(8):e104786.

[31] Zipf GK. The P 1 P 2/D hypothesis: on the intercity movement of persons. American sociological review. 1946;11(6):677–686.

[32] Barthélemy M. Spatial networks. Physics Reports. 2011;499(1):1–101.

[33] Ren Y, Ercsey-Ravasz M, Wang P, González M, Toroczkai Z. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. Nature communications. 2013;5:5347–5347.

[34] Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton University Press; 2008.