The 5th International Workshop on Agent-based Mobility, Traffic and Transportation Models, Methodologies and Applications (ABMTRANS)

Human mobility modelling: exploration and preferential return meet the gravity model

Luca Pappalardo[a,b,*], Salvatore Rinzivillo[b], Filippo Simini[c]

[a]*Department of Computer Science, University of Pisa, 56127 Pisa, Italy*
[b]*Institute of Information Science and Technologies (ISTI), National Research Council (CNR), 56124 Pisa, Italy*
[c]*Department of Engineering Mathematics, University of Bristol, Merchant Venturers Building, Woodland Road, BS8 1UB Bristol, UK*

Abstract

Modeling the properties of individual human mobility is a challenging task that has received increasing attention in the last decade. Since mobility is a complex system, when modeling individual human mobility one should take into account that human movements at a collective level influence, and are influenced by, human movement at an individual level. In this paper we propose the d-EPR model, which exploits collective information and the gravity model to drive the movements of an individual and the exploration of new places on the mobility space. We implement our model to simulate the mobility of thousands synthetic individuals, and compare the synthetic movements with real trajectories of mobile phone users and synthetic trajectories produced by a prominent individual mobility model. We show that the distributions of global mobility measures computed on the trajectories produced by the d-EPR model are much closer to empirical data, highlighting the importance of considering collective information when simulating individual human mobility.

1. Introduction

The analysis of the patterns of human mobility has received increasing attention in the last decade, given the availability of massive digital traces of human movements and its importance in domains such as urban planning, sustainability, transportation engineering, public health, and economic forecasting. Particular interest has been put on modeling the properties of individual human mobility, with the purpose of reproducing the movements of an individual in a realistic manner[1]. For example in the prominent Exploration and Preferential Return (EPR) model an individual can choose either to return to previously visited locations (preferential return) or to explore new locations at a given distance from the current location (exploration), according to well-known distributions of standard mobility measures such as the waiting time and the jump length[2]. In the EPR model and its recent improvements[3,4] no

* Corresponding author
   E-mail address: lpappalardo@di.unipi.it

collective information about the movements of other moving individuals is taken into account when deciding the new location an individual explores. Notwithstanding, human mobility is a complex system and the movements of an individual influence, and are influenced by, the collective mobility behavior of other individuals on the mobility space. Omitting this information can produce simulations that are unable to capture accurately human mobility patterns at a global level, e.g., the distribution of the radius of gyration, the distribution of location relevance, or the distribution of population density on the space[5,6]. Here, we advocate that the movements of individuals on a space are also driven by a *preferential exploration* force, which depends on the *collective* relevance of locations on the mobility space.

In this paper, we propose the *d*-EPR model, which improves the EPR model by using collective information and the gravity model to drive the movements of a synthetic individual. In particular, the model exploits information about the relevance of locations on the space: when an individual explores a new location, she is attracted to new places with a force that depends on the relevance of such places at a collective level (preferential exploration). We implement the *d*-EPR model to simulate the mobility of 50,000 synthetic individuals and compare the synthetic movements with real trajectories of mobile phone users and synthetic trajectories produced by a spatial version of the EPR model where individuals are constrained to move in a confined geographical space. We observe that the distributions of global mobility measures computed on the trajectories produced by the *d*-EPR model are much closer to empirical data than those produced by the EPR model. Our results highlight the importance of considering collective information when simulating individual mobility, enforcing the intuition that individual movements are strongly influenced by the collective mobility behavior of other people. In other words, individuals express *individual* preference when returning the previously visited places and *collective* preference when exploring new places on the mobility space.

The paper is organized as follows. Section 2 introduces the EPR model which is the base of our model. Section 3 describes in detail the *d*-EPR model and introduces the algorithm to reproduce it. In Section 4 we compare the results of our model with real trajectories of mobile phone users and the synthetic trajectories produced by the EPR model. Finally, Section 5 concludes the paper and discuss some possible extensions and improvements of the proposed model.

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[5,7]. How to combine such ingredients to create a realistic model which captures the salient aspects of individual human mobility is a challenging task. Many individual human mobility models have been proposed so far, the majority of which do not use spatio-temporal realism about population densities thus producing unrealistic mobility patterns[1]. The model proposed by Isaacman et al.[8], for example, exploits several distributions sampled from mobile phone data or census data to simulate the movements of individuals between a predefined number of locations on a given territory. Although this model produces realistic population density distributions, it is not able to produce realistic distributions of standard mobility measures, such as the radius of gyration.

Among the many proposed models, the Exploration and Preferential Return (EPR) model is one of the most used ones, especially because it does not fix in advance the number of visited locations but let them emerge spontaneously[2]. 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[2]. Preferential return reproduces the propensity of humans to return to the locations they visited frequently before[5]. An agent in the model selects between these two mechanisms: with a given probability the individual returns to one of the *S* previously visited places, with the preference for a location proportional to the frequency of the individual's previous visits. With complementary probability the individual moves to a new location, whose distance from the current one is chosen from the truncated power-law distribution of displacements as measured on empirical data[5]. The probability to explore decreases as the number of visited locations *S* 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 information about moving from home or other places[4]. It is worth noting that in the EPR model both exploration and preferential return mechanisms depend on individual forces. During a preferential return the individual returns to one of her previously visited locations, during an exploration the individual explores a new location randomly chosen at a given distance: none of the two mechanisms take into

account the relevance of locations on the space or its population density. In this paper, we advocate the need of considering such information during the exploration phase, relying on the intuition that individuals move preferably to dense places, where the variety and the number of locations available are large. For this reason we propose the *d*-EPR model, which improves the EPR model in two directions: first, it works on a finite mobility space using a predefined tessellation of the space into locations; second, it considers the relevance of locations on the space when choosing a new location to explore, hence defining a preferential exploration step.

3. The *d*-EPR model

The *d*-EPR model 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, an individual selects a new location to visit depending on both its distance from the current position, as well as its relevance measured as the total number of visits of all users. In the model, hence, the synthetic individual follows a personal preference when returning and a collective preference when exploring new locations. We use the gravity model[9,10] to assign the probability of a trip between any two locations, 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[11,12,13,14].

Algorithm 1 describes the *d*-EPR model in detail. The model takes in input several variables: (i) a list L of tuples each representing a location on the space; (ii) an integer $MaxTime$, the length (in hours) of the time period during which the individual moves on the space; (iii) the parameters of the waiting time distribution β and τ; (iv) the parameters for defining the probability of returning ρ and γ. Every tuple in L contains information about the geographical coordinates of the location and its relevance.

Given the input list, the algorithm computes for every pair of locations i, j the probability of moving from i to j (Algorithm 1, line 2). Every probability is computed as $p_{ij} = \frac{1}{N} \frac{d_i d_j}{r_{ij}^2}$, where $d_{i(j)}$ is the relevance of location $i(j)$, r_{ij} is the geographic distance between i and j, and $N = \sum_{i,j \neq i} p_{ij}$ is a normalisation constant (see Algorithm 1, function `computeProbabilityMatrix`). Starting from a location chosen randomly according to its relevance (Algorithm 1, line 3), until $time < MaxTime$ the algorithm iterates four basic steps: (i) waiting time choice, (ii) action selection, (iii) movement, (iv) variable updates.

In the waiting time choice step, the model extracts a waiting time Δt from the distribution $P(\Delta t) \sim \Delta t^{-1-\beta} \exp^{(-\Delta t/\tau)}$ (Algorithm 1, line 7)[2]. In the action selection phase, with probability $P_{new} = \rho S^{-\gamma}$ where S is the number of distinct locations previously visited[2], the individual chooses to explore a new location (Algorithm 1, line 10), otherwise she returns to a previously visited location (Algorithm 1, line 16). If the individual explores and is in location i, the new location $j \neq i$ is selected according to the precomputed probability p_{ij} (Algorithm 1, function `PreferentialExploration`) and the number of distinct locations visited, S, is increased by one. 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 (Algorithm 1, function `preferentialReturn`). After the movement step, the time elapsed (Algorithm 1, line 20) and the current location (line 21) are updated. When the maximum time expires ($time \geq MaxTime$), the algorithm terminates and returns in output the sequence V of locations visited by the individual.

For a comparison with the EPR model we design the *s*-EPR model, a spatial version of the original EPR model where individuals are constrained to move in a confined geographical space. The *s*-EPR model differs from the original EPR model in the exploration phase: when an individual explores a new location a distance Δr is extracted from the distribution $P(\Delta r) = \Delta r^{-(1+\alpha)}$, and an angle θ between 0 and 2π is extracted with uniform probability; if the location at distance Δr and angle θ from the current location is not in space's boundaries a new distance and a new angle are extracted until this condition is satisfied. It is worth highlighting an important difference between the *s*-EPR model and the *d*-EPR model. In the former, the exploration phase depends on the individual, i.e., when exploring the individual does not take into account the location relevance on the mobility space. In contrast, in the *d*-EPR the individual does take into account location relevance and is more likely to explore relevant locations.

---

input : *MaxTime*: the period of time the individual moves on the space

   L: a list of tuples $[t_1, t_2, \ldots, t_n]$ where $t_i = (x_i, y_i, d_i)$ describes a location

   $\beta, \tau, \rho, \gamma$: parameters of distributions

output: *V*: the sequence of locations visited by the synthetic individual

1   $S = 1$, *time* $= 0$ `// S is the number of visited locations`

2   $M = \text{computeProbabilityMatrix}(L)$ `// computes for every pair i, j the probability of moving from i to j`

3   $i = \text{weightedRandom}(L)$ `// choose randomly a location according to its relevance`

4   $v_i = (x_i, y_i, 1)$

5   $V.\text{append}(v_i)$

6   **while** *time* $\leq$ *MaxTime* **do**

7     $\Delta t = \text{getWaitingTime}()$ `// Extract a waiting time from the distribution` $P(\Delta t) \sim \Delta t^{-1-\beta} \exp^{(-\Delta t/\tau)}$

8     $P_{new} = \text{getReturnProbability}()$ `// Choose a probability to return or to explore` $P_{new} = \rho S^{-\gamma}$

9   **if** $P_{new} \leq \rho S^{-\gamma}$ **then**

10       $j = \text{PreferentialExploration}(i, M)$ `// Explore a new location`

11       $v_j = (x_j, y_j, 1)$

12       $V.\text{append}(v_j)$

13       $S = S + 1$

14   **end**

15   **else**

16       $j = \text{PreferentialReturn}()$ `// Return to a previously visited location`

17       $v_j = (x_j, y_j, count_j + 1)$

18       $V.\text{update}(v_j)$

19   **end**

20   *time* $=$ *time* $+ \Delta t$

21     $i = j$

22 **end**

1 **Function** computeProbabilityMatrix(*L*)

2   **foreach** $t_i \in L$ **do**

3   **foreach** $t_j \in L, j \neq i$ **do**

4       $p_{ij} = \frac{d_i * d_j}{dist(i,j)^2}$ `// compute probability according to locations' density and gravity model`

5       $M[i, j] = p_{ij}$

6   **end**

7   **end**

8   $N = \sum_{i, j \neq i} M[i, j]$ `// N is a normalization factor to ensure` $p_{ij} \in [0, 1]$

9   **foreach** $t_i \in L$ **do**

10   **foreach** $t_j \in L, j \neq i$ **do**

11       $M[i, j] = M[i, j]/N$

12   **end**

13   **end**

15   **return** M

1 **Function** PreferentialExploration(*i*)

2   $j = \text{weightedRandom}(M[i])$ `// choose randomly a location j according to its probability in list M[i]`

4   **return** j

1 **Function** PreferentialReturn()

2   $j = \text{weightedRandom}(V)$ `// choose randomly a location j according to` $count_i$ `in list V`

4   **return**

Algorithm 1: The algorithm describing how the *d*-EPR model works.

4. Model validation

We implement the *d*-EPR model to simulate the mobility of 50,000 synthetic individuals. Each individual moves for a period of three months (2,160 hours) between a set *L* of locations consisting in GSM towers dislocated on a European country. We estimate the relevance of each location in *L* as the number of calls from that location made during three months by 50,000 anonymized mobile phone users. We set the input parameters to $\beta = 0.8$, $\tau = 17$ hours, $\rho = 0.6$ and $\gamma = 0.21$, which are the parameters' values for the waiting time distribution and the probability of returning estimated by Song et al. on GSM data[2].

We compare the results of the *d*-EPR model with two other mobility datasets. The first one is an anonymized GSM dataset collected by a European carrier for billing and operational purposes[5,2,6]. The dataset consists of Call Detail Records (CDR) describing each phone call performed by 50,000 users in a period of three months. Each call is characterized by timestamp, caller and callee identifiers, duration of the call and the geographical coordinates of the tower serving the call. The time ordered list of towers from which a user made her calls forms a trajectory, capturing her movements during the period of observation. The other dataset consists of the mobility trajectories produced by 50,000 synthetic individuals obtained by running the *s*-EPR model[6], where agents are constrained within a country boundary (the same country as GSM data and *d*-EPR model)*. We set the exponent for the distribution of distance lengths to $\alpha = 0.55$, as estimated by Gonzalez et al. on GSM data[5].
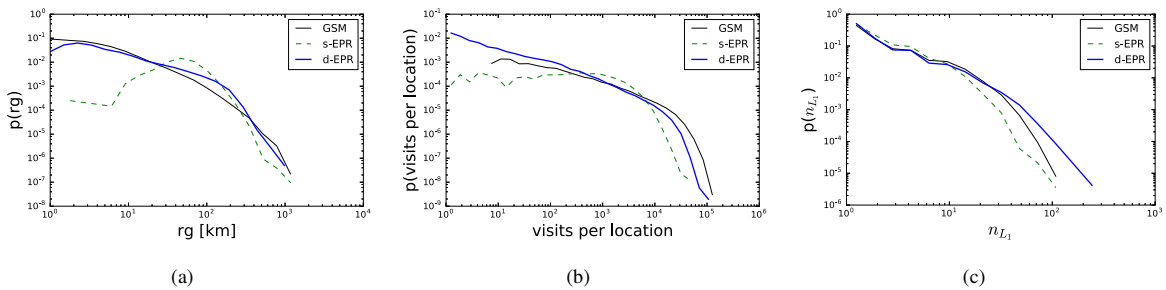

Fig. 1. **A comparison of *d*-EPR model, *s*-EPR model and empirical GSM data**. *(a)* The distribution of the radius of gyration r_g of individuals computed on the three datasets. We observe that r_g for *d*-EPR and GSM data (blue and black solid curves) are similar and well approximated by a power-law with exponential cut-off, while for *s*-EPR model (dashed curve) we observe a peaked distribution. *(b)* The distribution of overall number of visits per location for the three datasets. Also in this case the distribution for the *s*-EPR model differs from the other two distributions. *(c)* The distribution of n_{L_1}, the number of individuals for which a location is the most frequent location. We observe that n_{L_1} for *d*-EPR is more similar to GSM data than the distribution for *s*-EPR.

Figure 1 compares the three datasets on: (i) the distribution of radius of gyration r_g, a measure of the characteristic distance traveled by a given individual during the period of observation defined as $r_g = \sqrt{\frac{1}{N}\sum_{i\in L} d_i(r_i - r_{cm})^2}$ where *N* is the total number of visits to any location by the individual, *L* is the set of locations visited, d_i is the relevance of location *i*, r_i are the coordinates of location *i*, r_{cm} the coordinates of the center of mass of the individual[5,15]; (ii) the distribution of overall visits per location, i.e., the total number of visits by all the individuals in that location during the period of observation; (iii) the distribution of n_{L_1} per location, where n_{L_1} is the number of individuals for which that location is the most frequent location L_1, i.e. the phone tower where the user performs the highest number of calls during the period of observation. We observe that the distribution of the radius of gyration for GSM data and *d*-EPR data are similar (a power-law distribution with exponential cutoff), while for *s*-EPR data it is a peaked distribution (Figure 1(a), green dashed curve). Similarly, the distribution of the number of visits per location of *s*-EPR data differs from the other two distributions, which are similar to each other (Figure 1(b)). In Figure 1(c) we plot the distributions of n_{L_1}, the number of individuals for whom a given location is the most frequent locations (L_1), an estimate of the number of individuals living in a given location. We observe that all the three distributions are heavy-tailed, reflecting an uneven distribution of population density on the space. However the curves for GSM data and *d*-EPR data are more

---

*The original EPR model works in an infinite mobility space, we implement the *s*-EPR model (which works on a finite mobility space) to make the results of the model comparable with the GSM and the *d*-EPR datasets.

similar to each other than the curve for *s*-EPR data, which starts to differ from the others for low values of $n_{L_1} \approx 10$ (Figure 1(c)). These results show that the *s*-EPR model fails in capturing some global human mobility patterns, and that we can overcome this shortcoming by implementing a preferential exploration phase.

5. Conclusion

In this paper we proposed the *d*-EPR, a generative model to simulate individual human mobility. In contrast with the EPR, our model exploits collective information about location relevance and implements a preferential exploration phase, producing results that are much in better agreement with empirical data. Our results show that the patterns of individual mobility are driven by two competing forces: an individual force during the preferential return phase, and a collective force during the exploration phase where the movements of an individual are influenced by the relevance of locations on the mobility space. In the approach we proposed, the distribution of visitation relevance of locations is given as input variable to the *d*-EPR model. Although such a distribution can be easily computed from mobile phone data or census data, information about location relevance on a space are not always available. As future work, we plan to turn the individual model into a *collective* model, making the relevance of locations to emerge naturally during the running of the model: in the preferential exploration phase the probability of an individual to visit a new location will be proportional to the number of visits to that location made by other synthetic agents moving at the same time on the space. It will be interesting to investigate whether the empirical distribution of visitation relevance emerges spontaneously from the collective model.

Acknowledgements

References

1. D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella, "Human mobility models for opportunistic networks," *Communications Magazine, IEEE*, vol. 49, pp. 157–165, December 2011.
2. C. Song, T. Koren, P. Wang, and A.-L. Barabási, "Modelling the scaling properties of human mobility," *Nature Physics*, vol. 6, pp. 818–823, Sept. 2010.
3. H. Barbosa, F. B. de Lima-Neto, A. Evsukoff, and R. Menezes, "The effect of recency to human mobility," *EPJ Data Science*, vol. 4, no. 21, 2015.
4. Y. Yang, S. Jiang, D. Veneziano, S. Athavale, and M. C. Gonzalez, "Timegeo: a spatiotemporal framework for modeling urban mobility without surveys," *PNAS*, 2015.
5. M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, June 2008.
6. L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti, and A.-L. Barabasi, "Returners and explorers dichotomy in human mobility," *Nature Communications*, vol. 6, no. 8166, 2015.
7. C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
8. S. Isaacman, R. Becker, R. Cáceres, M. Martonosi, J. Rowland, A. Varshavsky, and W. Willinger, "Human mobility modeling at metropolitan scales," in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, MobiSys '12, (New York, NY, USA), pp. 239–252, ACM, 2012.
9. G. K. Zipf, "The p1p2/d hypothesis: On the intercity movement of persons," *American Sociological Review*, vol. 11, no. 6, pp. 677–686, 1946.
10. W. S. Jung, F. Wang, and H. E. Stanley, "Gravity model in the korean highway," *EPL (Europhysics Letters)*, vol. 81, p. 48005, 2008.
11. S. Erlander and N. F. Stewart, *The gravity model in transportation analysis: theory and extensions*. Vsp, 1990.
12. A. G. Wilson, "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.
13. F. Simini, M. C. González, A. Maritan, and A. L. Barabási, "A universal model for mobility and migration patterns," *Nature*, vol. 484, p. 96, 2012.
14. D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, p. 21484, 2009.
15. L. Pappalardo, S. Rinzivillo, Z. Qu, D. Pedreschi, and F. Giannotti, "Understanding the patterns of car travel," *The European Physical Journal Special Topics*, vol. 215, no. 1, pp. 61–73, 2013.