STS-EPR: Modelling individual mobility considering the spatial, temporal, and social dimensions together

Giuliano Cornacchia and Luca Pappalardo

April 29, 2021

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2021) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

The 11th International Conference on Ambient Systems, Networks and Technologies (ANT)
March 23 - 26, 2021, Warsaw, Poland

STS-EPR: Modelling individual mobility considering the spatial, temporal, and social dimensions together

Giuliano Cornacchia[a,b], Luca Pappalardo[b]

[a] *Department of Computer Science, University of Pisa, Pisa, Italy*
[b] *Institute of Information Science and Technologies (ISTI), National Research Council of Italy (CNR), Pisa, Italy*

Abstract

Modelling human mobility is crucial in several areas, from urban planning to epidemic modeling, traffic forecasting, and what-if analysis. On the one hand, existing models focus on the spatial and temporal dimensions of mobility only, while the social aspect is often neglected. On other hand, those models that capture some social aspects of human mobility have trivial and unrealistic spatial and temporal mechanisms. In this paper, we propose STS-EPR, a modeling framework that embeds mechanisms to capture the spatial, temporal and social aspects together. Our experiments show that STS-EPR outperforms existing spatial-temporal or social models on a set of standard mobility metrics. STS-EPR, which is open-source and tested on open data, is a step towards the design of a mechanistic models that can capture all the aspects of human mobility in a comprehensive way.

Keywords: human mobility; generative models; synthetic trajectories; social network; data science; mechanistic models

1. Introduction

Mobility data are crucial in different contexts, such as computational epidemiology, traffic forecasting, urban planning, what-if analysis, and the design of protocols for ad hoc and opportunistic networks [3, 6, 8, 13, 17, 19]. Unfortunately, privacy implications restrict sharing mobility datasets because they contain sensitive information about the individuals whose movements are described [10, 9, 15, 16]. A way to overcome this issue is to design generative mobility models [8, 3], i.e., algorithms that generate synthetic trajectories that reproduce human mobility patterns.

Most individual models focus on capturing the spatial patterns, such as the power-law distribution in jump lengths and characteristic distances [4, 6, 11] and the tendency to return to locations visited before [6, 11]. For example, the

* Corresponding authors: Giuliano Cornacchia, Luca Pappalardo
E-mail address: giuliano.cornacchia@phd.unipi.it, luca.pappalardo@isti.cnr.it

1877-0509

Exploration and Preferential Return model (EPR) [18] is based on two competing mechanisms: *exploration*, a random walk process with truncated power-law jump length distribution; and *preferential return*, which reproduces humans' propensity to return to previously visited locations. Among the mechanistic models that improve EPR by adding increasingly sophisticated mechanisms [12, 2, 1], TimeGeo [7] and DITRAS [13] embed a temporal mechanism that captures the circadian propensity to travel. The social dimension of human mobility is often neglected in EPR-like models, although about 10-30% of human movements has social purposes [5]. As an exception, GeoSim [20] use an individual preference and social influence mechanism, but its spatial and temporal realism is limited.

In this paper, we propose STS-EPR (Spatial, Temporal, and Social EPR), a model that combines: *(i)* a mechanism that takes into account the spatial distance between locations and their collective relevance [17, 14]; *(ii)* a temporal mechanism to capture the individuals' tendency to follow a circadian rhythm [13]; *(iii)* a mechanism that models the social dimension of human mobility [20]; *(iv)* an action-correction mechanism that deals with borderline cases during the simulation. We conduct experiments using check-ins of thousands of users in three cities worldwide and show that STS-EPR's trajectories are realistic with respect to several spatial, temporal, and social aspects of mobility.[1]

2. STS-EPR model

We define a mobility trajectory as a sequence $T = \langle (r_1, t_1), \dots, (r_n, t_n) \rangle$ where t_i is a timestamp $(\forall i \in [1, n); t_i < t_{i+1})$ and $r_i = (x_i, y_i)$ where x_i and y_i are coordinates on a bi-dimensional space [8]. We assume that agents move on a spatial tessellation L, representing a bi-dimensional space's tiling, resulting in a non-overlapped set of locations. Every location has a weight corresponding to its collective relevance and as a representative point the tile's centroid expressed as a pair of coordinates: $L = \langle (r_1, w_1), \dots, (r_n, w_n) \rangle$ where w_j is the weight of the tile j and r_j its representative point. We represent an agent's a visitation pattern as a location vector lv_a of $|L|$ locations. The vector's *j-th* element, $lv_a[j]$, contains the number of times a visited r_j. We assume that an agent's network of contacts G influences their movements. $G = (V, E)$ is a graph where V is the set of agents and E the social ties between them.

STS-EPR takes as input the number N of synthetic agents, the spatial tessellation L, the undirected graph G, a mobility diary generator MDG, and the time interval of the simulation. The model outputs N synthetic trajectories. STS-EPR consists of four phases: initialization, action selection, location selection, and action-correction (Figure 1). After the initialization phase, the agents execute the action selection, the location selection, and the action-correction phases until a stopping criterion is satisfied (e.g., the number of hours to simulate is reached).

Initialization. Each edge's weight in G indicates the mobility similarity of the linked agents (the cosine similarity of their location vectors). The weights are initialized to 0 and updated during the simulation. The model assigns to each agent a mobility diary produced by MDG, a Markov model that captures the individuals' probability to follow or break their routine at specific times of the day [13]. A mobility diary MD for an agent a is defined as $\mathrm{MD}_a = \langle (ab_0, t_1), (ab_1, t_2), \dots (ab_j, t_{j+1}), (ab_0, t_{j+2}), (ab_1, t_{j+3}) \dots) \rangle$, where ab is an abstract location, ab_0 denotes a's starting location, t_i is a timestamp. The probability $p(r_i)$ for an agent of being assigned to a starting physical location $r_i \in L$ is $\propto w_i$, where w_i is the location's relevance. Each agent moves according to its mobility diary's entries at the time specified. If the current abstract location is ab_0, the agent visits the starting location; otherwise, ab_i is converted into a physical one through the following steps.

Action selection. When moving, an agent can select between two competing mechanisms: exploration and preferential return. Exploration models the decreasing tendency to explore new locations over time [17]. Preferential return reproduces individuals' propensity to return to locations they explored before [17]. An agent explores a new location with probability $P_{exp} = \rho S^{-\gamma}$, or returns to a previously visited one with a complementary probability $P_{ret} = 1 - \rho S^{-\gamma}$, where S is the agent's number of unique visited locations and $\rho = 0.6$, $\gamma = 0.21$ are constants [17]. At that point, the agent determines whether or not the location's choice will be affected by the other agents, selecting between the individual and the social influence mechanisms. With a probability $\alpha = 0.2$, the agent's social contacts influence its movement [20]. With a complementary probability of $1 - \alpha$, the agent's choice is not influenced by the other agents.

Location selection. At this point, the agent decides which location will be the destination of its next displacement. The sets of locations an agent a can visit or return are $exp_a = \{i \mid lv_a[i] = 0\}$ and $ret_a = \{i \mid lv_a[i] > 0 \wedge i \neq s_a\}$,

[1] The code to reproduce our model and the experiments is available at: https://github.com/kdd-lab/2019_Cornacchia

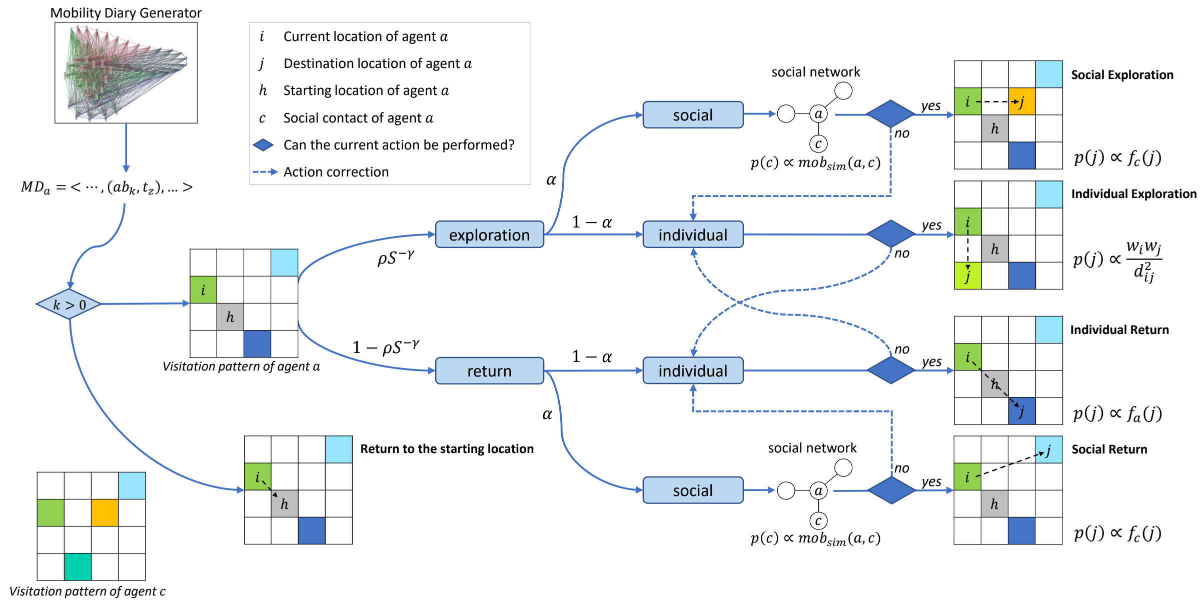

Fig. 1. A schematic description of STS-EPR. First, when an individual moves according to the entry in its mobility diary, if the abstract location is ab_0 the individual returns to its starting location, otherwise it decides whether to explore a new location or return to a previously visited one. At that point, the agent determines whether or not its social contacts affect its choice for the location to visit next. If the selected action cannot be performed, it is corrected with an executable one (dashed arrows indicate action corrections).

respectively, where s_a denotes the index of the starting location of agent a. The visitation frequency of a to a location r_i is: $f_a(r_i) = \frac{lv_a[i]}{\sum_{j=1}^{|L|} lv_a[j]}$.

- **Individual Exploration:** a chooses a new location to explore from exp_a. Individuals are more likely to move at small rather than long distances but also take into account the location's collective relevance [13]. We use the gravity law to couple distance and relevance [12]. If a is currently at location r_j, during the Individual Exploration action selects an unvisited location r_i, with $i \in exp_a$, with probability $p(r_i) \propto \frac{w_i w_j}{d_{ij}^2}$, where d_{ij} is the distance between locations r_i and r_j with relevances w_i, w_j.
- **Social Exploration:** a selects an agent c among its social contacts in G. The probability $p(c)$ for c to be selected is proportional to the mobility-similarity between them: $p(c) \propto mob_{sim}(a, c)$. After the contact c is chosen, the candidate location to explore is an unvisited location for a that was visited by c, i.e., the location is selected from set $A = exp_a \cap ret_c$; the probability $p(r_i)$ for a location r_i, with $i \in A$, to be selected is proportional to the visitation pattern of c, namely $p(r_i) \propto f_c(r_i)$.
- **Individual Return:** a picks the return location from the set ret_a with a probability proportional to its visitation pattern. The probability for a location r_i with $i \in ret_a$ to be chosen is: $p(r_i) \propto f_a(r_i)$.
- **Social Return:** c is selected as in the Social Exploration action from the set $A = ret_a \cap ret_c$; the probability $p(r_i)$ for a location r_i to be selected is proportional to the visitation pattern of the agent c, namely $p(r_i) \propto f_c(r_i)$.

Action correction. The set of possible locations an agent can reach is limited. For example, it may happen that the agent visited all locations at least once and there are no locations to explore. To comply with these constraints, we include an action-correction phase, executed if the location selection phase does not allow movements in any location.

- **No location in social choices**: If an agent a decides to move with the influence of a social contact c but $ret_a \cap ret_c = \varnothing$ or $exp_a \cap ret_c = \varnothing$ (no locations visited by both c and a or no locations visited by c and unvisited by a), we execute an individual action preserving a's choice to explore or return.

- **No new location to explore**: When an agent a decides to explore but it visited all the locations at least once ($exp_a = \varnothing$), we force the agent to make an Individual Return.
- **No return location**: If an agent a, currently at location r_i, decides to perform an Individual Return, and r_i is the only location visited so far (besides the starting location), it cannot return to any location ($ret_a = \varnothing$). We force a to make an Individual Exploration.

3. Results

We simulate the mobility of 1001, 622, and 4396 agents in New York City, London, and Tokyo, respectively, using STS-EPR, DITRAS [13], and GeoSim [20]. For each city, we compare the synthetic trajectories with real ones extracted from Foursquare checkins [21]: New York City has 1001 users, 37,4891 checkins and 1755 edges; London has 622 users, 14,895 checkins and 1185 edges; Tokyo has 4396 users, 231,471 checkins and 18,183 edges. We evaluate the models' realism with respect to the distribution of:

- Jump Length Δr, the distance between two consecutive locations visited by an individual [6, 11];
- Radius of Gyration r_g, the typical distance traveled by an individual during the period of observation [6, 14];
- Visits per Location V_l, the relevance of a location described as its attractiveness at a collective level;
- Location Frequency $f(r_i)$, the probability of visiting a location r_i [6];
- Waiting Time Δt, the elapsed time between two consecutive visited locations;
- Entropy E_{unc}, the predictability of the movements of an individual u [18];
- Activity per Hour $t(h)$, the number of movements made by the individuals at every hour of the day [13, 7];
- Mobility Similarity mob_{sim}, the cosine-similarity of two individuals' location vectors [20, 5, 21];

We quantify the similarity between each measure's distributions for real and synthetic trajectories using the Kullback–Leibler divergence (KL). Table 1 reports our results.

Regarding the spatial measures, Δr and r_g, STS-EPR is more realistic than GeoSim, which cannot reproduce neither the Δr distribution (Figures 2(a), 2(e) and 2(i)) nor the r_g distribution. STS-EPR is also slightly more realistic than DITRAS concerning Δr, except for New York City (Table 1). DITRAS is the best model concerning r_g, with the exception of Tokyo, where DITRAS and STS-EPR achieve similar scores. GeoSim generates trajectories with the most realistic distribution of $f(r_i)$ but fails in reproducing the V_l distribution (Figures 2(b), 2(f) and 2(j)). STS-EPR is more realistic than DITRAS on $f(r_i)$, and for V_l the KL scores of STS-EPR are 58.62% (New York City), 27.73% (London), and 79.47% (Tokyo) better than DITRAS.

The temporal measures, Δt and $t(h)$ (Figures 2(c), 2(g) and 2(k)), are better reproduced by STS-EPR and DITRAS, since they use the same temporal mechanism (MDG). The small fluctuations of the scores obtained from these two models are caused only by the pseudo-random nature of each execution. Although GeoSim can reproduce the Δt distribution, it fails in reproducing $t(h)$ because its trajectories that do not follow the circadian rhythm.

Regarding the distribution of the social measure mob_{sim}, STS-EPR reproduces it better than GeoSim, especially for Tokyo (KL=0.01, Table 1), presumably because the social graph of users in Tokyo is the largest in terms of nodes and edges, giving a more realistic representation of the individuals' sociality. STS-EPR can capture the distribution's trend better than GeoSim; the latter can reproduce correctly the distribution only for values ≤ 0.25 (Figures 2(d), 2(h) and 2(l)). None of the models can replicate the distribution of E_{unc}, though the best one is this sense is STS-EPR. The results for New York City, London, and Tokyo (Table 1) are consistent, suggesting that STS-EPR do not depend on the specific characteristics of the geographic area.

4. Discussion

STS-EPR models human mobility considering the spatial, temporal and social dimensions. Our results show that STS-EPR can generate realistic trajectories for all the three dimensions. It is worth noting that the inclusion of the social dimension in STS-EPR help improve the realism with respect to the spatial and temporal measures. This result highlights the importance of sociality: though often neglected in generative mobility models, it is essential to model properly individual human mobility. Modeling the social dimension together with the spatial and temporal ones is

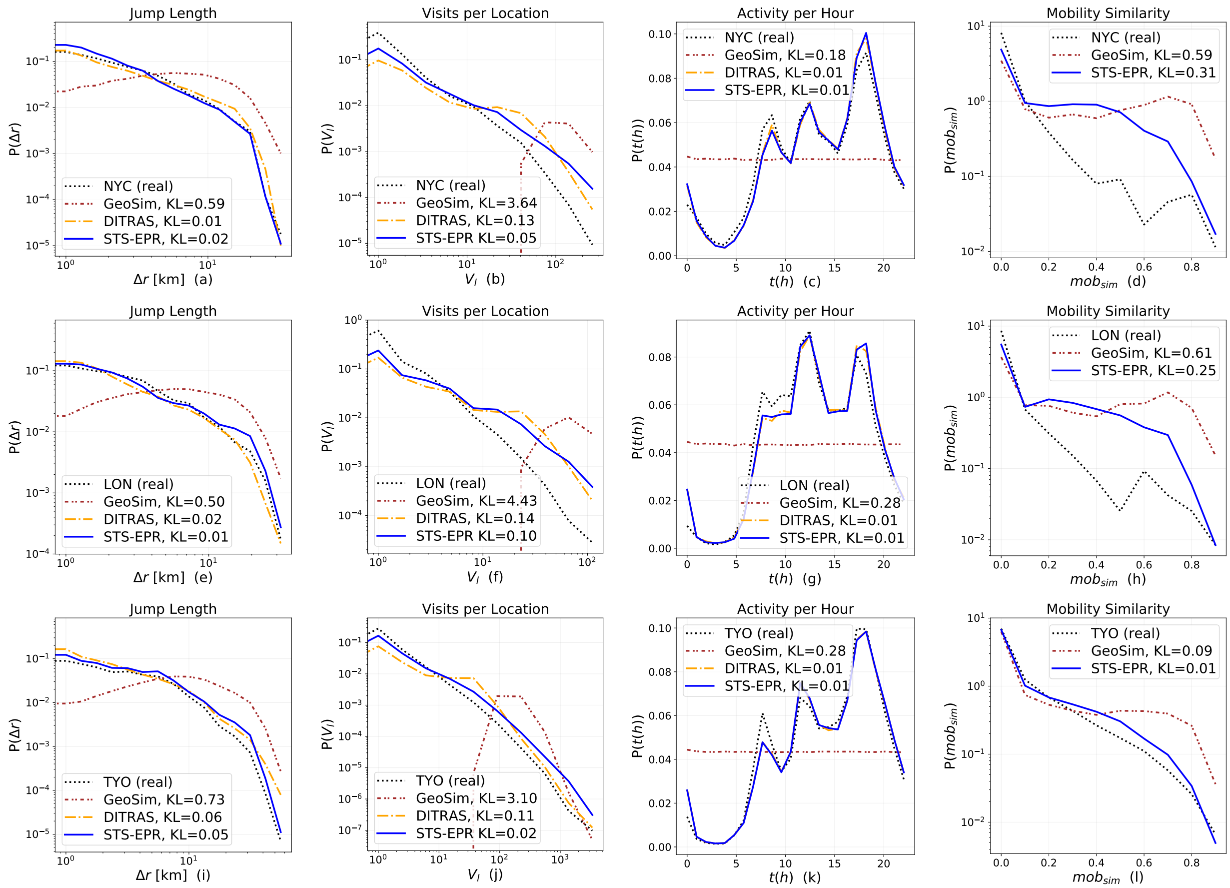

Fig. 2. Distributions of four human mobility measures (columns) referred to different cities (rows) computed for real and generated trajectories.

	Model	Δr	r_g	$f(r_i)$	Vl	Δ_t	$t(h)$	E_{unc}	mob_{sim}
London	GeoSim	0.5036 ±0.0075	4.9381 ±0.0932	**0.0016 ±0.0001**	4.427 ±0.0069	0.1962 ±0.0043	0.281 ±0.0003	8.5182 ±0.0003	0.6097 ±0.0079
	DITRAS	0.0221 ±0.0022	**0.1813 ±0.0239**	0.1094 ±0.0	0.1428 ±0.006	0.166 ±0.0031	0.0119 ±0.0004	3.8816 ±0.1897	- -
	STS-EPR	**0.0108 ±0.0016**	0.4609 ±0.233	0.0097 ±0.0003	**0.1032 ±0.0126**	**0.1626 ±0.0035**	**0.0116 ±0.001**	**2.6749 ±0.1169**	**0.2543 ±0.01**
Tokyo	GeoSim	0.7257 ±0.002	4.8165 ±0.0042	**0.0002 ±0.0**	3.0957 ±0.0148	0.2354 ±0.0003	0.2837 ±0.0006	7.1242 ±0.0593	0.0931 ±0.0017
	DITRAS	0.0628 ±0.0025	**0.2417 ±0.0171**	0.1409 ±0.0	0.1101 ±0.0048	0.2007 ±0.003	**0.0074 ±0.0001**	5.0034 ±0.2708	- -
	STS-EPR	**0.0485 ±0.0013**	0.2504 ±0.0746	0.0108 ±0.0002	**0.0226 ±0.0019**	**0.2001 ±0.0024**	0.0076 ±0.0001	**4.8717 ±0.2247**	**0.014 ±0.0009**
New York City	GeoSim	0.5947 ±0.0062	5.3913 ±0.0051	**0.0071 ±0.0004**	3.6418 ±0.0069	0.1973 ±0.0004	0.18 ±0.0005	8.0483 ±0.0579	0.5879 ±0.0149
	DITRAS	**0.0091 ±0.0006**	**0.2987 ±0.0359**	0.193 ±0.0026	0.1281 ±0.0044	**0.1665 ±0.0032**	**0.0066 ±0.0003**	**4.8881 ±0.0248**	- -
	STS-EPR	0.0188 ±0.0015	0.3886 ±0.0284	0.0318 ±0.0008	**0.0531 ±0.004**	0.1705 ±0.0047	0.0071 ±0.0005	5.028 ±1.1511	**0.3066 ±0.0044**

Table 1. Results London, Tokyo and New York City. For each measure we show the KL divergence and its standard deviation. Best results in bold.

crucial in many disciplines. For example, in computational epidemiology both movements and social contacts may be exploited to model the spreading process. Future work regards the evaluation of STS-EPR's realism on a wider range of features. In the meantime, our model is a step towards the design of a mechanistic models that can capture all the aspects of human mobility in a comprehensive way.

Acknowledgements. Luca Pappalardo has been supported by project H2020 SoBigData++ grant agreement 871042.

References

[1] Alessandretti, L., Sapiezynski, P., Lehmann, S., Baronchelli, A., 2018. Evidence for a conserved quantity in human mobility. Nature Human Behaviour 2. doi:10.1038/s41562-018-0364-x.

[2] Barbosa, H., de Lima-Neto, F.B., Evsukoff, A., Menezes, R., 2015. The effect of recency to human mobility. EPJ Data Science 4, 21. URL: https://doi.org/10.1140/epjds/s13688-015-0059-8, doi:10.1140/epjds/s13688-015-0059-8.

[3] Barbosa-Filho, H., Barthelemy, M., Ghoshal, G., James, C., Lenormand, M., Louail, T., Menezes, R., Ramasco, J.J., Simini, F., Tomasini, M., 2018. Human mobility: Models and applications. Physics Reports 734, 1–74.

[4] Brockmann, D., Hufnagel, L., Geisel, T., 2006. The scaling laws of human travel. Nature 439, 462–5. doi:10.1038/nature04292.

[5] Cho, E., Myers, S., Leskovec, J., 2011. Friendship and mobility: User movement in location-based social networks, pp. 1082–1090. doi:10.1145/2020408.2020579.

[6] Gonzalez, M.C., Hidalgo, C., Barabasi, A.L., 2008. Understanding individual human mobility patterns. Nature 453, 779–82. doi:10.1038/nature06958.

[7] Jiang, S., Yang, Y., Gupta, S., Veneziano, D., Athavale, S., Gonzalez, M.C., 2016. The timegeo modeling framework for urban mobility without travel surveys. Proceedings of the National Academy of Sciences 113, 201524261. doi:10.1073/pnas.1524261113.

[8] Luca, M., Barlacchi, G., Lepri, B., Pappalardo, L., 2020. Deep learning for human mobility: a survey on data and models. arXiv:2012.02825.

[9] Montjoye, Y.A., Gambs, S., Blondel, V., Canright, G., Cordes, N., Deletaille, S., Engø-Monsen, K., García-Herranz, M., Kendall, J., Kerry, C., Krings, G., Letouzé, E., Luengo-Oroz, M., Oliver, N., Rocher, L., Rutherford, A., Smoreda, Z., Steele, J., Wetter, E., Bengtsson, L., 2018. On the privacy-conscientious use of mobile phone data. Scientific Data 5, 180286. doi:10.1038/sdata.2018.286.

[10] Montjoye, Y.A., Hidalgo, C., Verleysen, M., Blondel, V., 2013. Unique in the crowd: The privacy bounds of human mobility. Scientific reports 3, 1376. doi:10.1038/srep01376.

[11] Pappalardo, L., Rinzivillo, S., Qu, Z., Pedreschi, D., Giannotti, F., 2013. Understanding the patterns of car travel. The European Physical Journal Special Topics 215, 61–73. URL: https://doi.org/10.1140/epjst/e2013-01715-5, doi:10.1140/epjst/e2013-01715-5.

[12] Pappalardo, L., Rinzivillo, S., Simini, F., 2016. Human mobility modelling: Exploration and preferential return meet the gravity model. Procedia Computer Science 83. doi:10.1016/j.procs.2016.04.188.

[13] Pappalardo, L., Simini, F., 2017. Data-driven generation of spatio-temporal routines in human mobility. Data Mining and Knowledge Discovery 32. doi:10.1007/s10618-017-0548-4.

[14] Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., Barabasi, A.L., 2015. Returners and explorers dichotomy in human mobility. Nature Communications 6. doi:10.1038/ncomms9166.

[15] Pellungrini, R., Pappalardo, L., Pratesi, F., Monreale, A., 2017. A data mining approach to assess privacy risk in human mobility data. ACM Trans. Intell. Syst. Technol. 9. URL: https://doi.org/10.1145/3106774, doi:10.1145/3106774.

[16] Pellungrini, R., Pappalardo, L., Simini, F., Monreale, A., 2020. Modeling adversarial behavior against mobility data privacy. IEEE Transactions on Intelligent Transportation Systems , 1–14.

[17] Song, C., Koren, T., Wang, P., Barabasi, A.L., 2010a. Modelling the scaling properties of human mobility. Nature Physics 6. doi:10.1038/nphys1760.

[18] Song, C., Qu, Z., Blumm, N., Barabasi, A.L., 2010b. Limits of predictability in human mobility. Science (New York, N.Y.) 327, 1018–21. doi:10.1126/science.1177170.

[19] Tomasini, M., Mahmood, B., Zambonelli, F., Brayner, A., Menezes, R., 2017. On the effect of human mobility to the design of metropolitan mobile opportunistic networks of sensors. Pervasive and Mobile Computing 38, 215 – 232. URL: http://www.sciencedirect.com/science/article/pii/S1574119216301195, doi:https://doi.org/10.1016/j.pmcj.2016.12.007.

[20] Toole, J., Herrera-Yague, C., Schneider, C., Gonzalez, M.C., 2015. Coupling human mobility and social ties. Journal of the Royal Society, Interface / the Royal Society 12. doi:10.1098/rsif.2014.1128.

[21] Yang, D., Qu, B., Yang, J., Cudre-Mauroux, P., 2019. Revisiting user mobility and social relationships in lbsns: A hypergraph embedding approach, pp. 2147–2157. doi:10.1145/3308558.3313635.