

Route Selection and Trajectory Filling in Mobility Traces

Augusto C.S.A. Domingues¹,
Antonio A.F. Loureiro¹ (Advisor), Fabrício A. Silva² (Co-advisor)

¹Computer Science Department - Federal University of Minas Gerais
Belo Horizonte, MG - Brazil

²Federal Univeristy of Viçosa - *campus* Florestal
Florestal, MG - Brazil

{augusto.souza,loureiro}@dcc.ufmg.br, fabricio.asilva@ufv.br

Abstract. *The feasibility of vehicular networks is directly related to the understanding of mobility patterns, which is a necessary knowledge for the elaboration and application of novel algorithms and technologies for such networks. However, there are few traces that represent urban mobility with details and precision. The main objective of this dissertation is to characterize the mobility of vehicles in urban environments and to use this information to propose an algorithm to generate and enrich data in this environment.*

1. Introduction

Mobile Ad Hoc Networks (MANETs) are networks composed of entities that can move freely in a space while communicating wirelessly. Therefore, the link between devices is directly related to their mobility, i.e., they will not communicate if the distance between them is greater than the communication range. The lack of knowledge about the movement of such entities can impose barriers regarding the correct functioning of the network, reducing the end-user satisfaction [Cotta et al. 2017, Silva et al. 2015]. Additionally, certain types of mobile networks are more prone to the effects of mobility, such as Vehicular Ad Hoc Networks (VANETs), in which each vehicle represents a node. In such networks, the topology is highly dynamic, and the speed in which they move is higher when compared to the natural human movement, affecting the occurrence and detection of contacts. Thus, it is important to analyze the mobility behavior in this network to address this problem. Such analysis is useful to improve VANETs regarding routing algorithms, simulations, validation of protocols and the development of applications and services.

In the literature, most studies in the field of VANETs use simulation to evaluate their proposals. Thus, the better we understand the behavior of vehicles, the better we can simulate them. For this, there are mobility traces, which are datasets related to the movement of entities in a space. These traces are frequently applied in the simulation of vehicular networks and in the behavior analysis of entities. Thus, the quality of this data reflects directly in the results. For example, one of the most common problem found in these datasets is the sparsity of their samples, which generates gaps in the trajectory of a node, making it difficult to understand their actions and nullifying the detection of possible interactions with others. Thus, it is fundamental that the choice of mobility traces applied in a study take these factors into consideration. However, there are few traces available in the literature and most of them are not complete. The generation of

new real or synthetic traces and the refinement of existing ones offer to researchers novel opportunities.

This work analyzes the mobility of vehicles in terms of temporal, spatial and contextual factors. The objective of this analysis is the understanding of the mobility behaviors of users, and how we can benefit from them. With this information in hand, we propose an algorithm to fill sparse trajectories of vehicles in traces where none or few is known about the ground-truth trajectory. We then show how the filling of such trajectories allows the enrichment of existing datasets, consequently improving the quality of simulations and experiments.

1.1. Background

In Chapter 2 of the dissertation, we present a review of the literature in terms of the usage of mobility traces and the issues faced when working with them. For the former, we highlight many studies ranging from analysis about how the mobility affects the connectivity of the network in large-scale urban environments [Alipour et al. 2018, Cotta et al. 2017, Xia et al. 2017, Hou et al. 2016], to the effects of mobility to the consumption of services in the network [Lu et al. 2018, Çatay and Keskin 2017], the mobility prediction of users [Sadri et al. 2017, Qiao et al. 2017b], to the general analysis of human mobility for various purposes [Wang et al. 2018a, Wang et al. 2018b, Garcia et al. 2018, Yao 2018, Diniz et al. 2017, Wang et al. 2017, Lu et al. 2017, Qiao et al. 2017a, Xia et al. 2017].

For the quality of mobility traces, we searched for works that address issues related to faulty mobility data, such as trajectory completion and enrichment, that justifies the outcomes of this dissertation. Given the importance of mobility in networks and the number of different analysis that can be made from this information, it is important that the data used to perform the analysis be complete and trustworthy. We analyze the techniques used to fill gaps, such as interpolation, and describe studies that look at how you can improve the results of analysis when you improve the quality of the data used [Silva et al. 2015]. Finally, from these works, it is clear that the available solutions in the literature are limited to solve the problems found, thus opening new research opportunities to solve this problem.

2. Main Contributions

The main contributions of this dissertation are (i) the detailed characterization of vehicular mobility in a large city regarding factors such as space and time through a metric that measures the willingness of a driver to follow the shortest path in a trip, and (ii) an algorithm to generate trajectories in highly sparse mobility traces, such as when only the origin and destination points of a trip are known, using information extracted from these users and that considers the road mesh to generate the new points in the trajectories.

2.1. Mobility Characterization

In Chapter 3 of the dissertation, we present a characterization of vehicular mobility towards route selection. The understanding of route preferences and the effects of time and space to these preferences presents numerous applications in mobile networks, such as the selection of candidates to transmit a message in opportunistic networks, traffic management and balancing, and so on. Few studies have considered the characterization of

routes regarding their optimality, i.e., if they tend to follow the best path, which can be defined as the shortest one in distance or even the quickest in time.

To study the mobility patterns in a vehicular environment, we use a real vehicular trace, the San Francisco Taxi Trace [Piorowski et al. 2009], containing the mobility of taxicabs in San Francisco, USA, collected in 2008. The dataset contains information about the vehicles' location collected through an embedded GPS sensor device installed. There is a specific motivation for selecting this trace. Taxi drivers tend to select their trajectories according to the best route, being less driven by particular reasons. Therefore, their behavior tend to be less biased. Additionally, there is contextual information about the trips, i.e., when the taxi picks up a passenger at a defined point and delivers them to a defined point (the drop off point). Having well-defined origin and destination points is an essential factor to achieve a precise characterization of the data.

In order to enrich the information about the context in which the taxi trips happened, we pulled information from external sources to complete the trace, obtaining data from two sources: first, from Open Street Maps, we gathered information about the road mesh of San Francisco, in order to match the points in the trace to the real roads in which they happened. Additionally, the road mesh data also contains information about the road length, connections, road type and speed limits. Second, we obtained information from Google maps to match the points to their locations in the city (the neighborhoods), and their context, i.e., their main activities, such as residential housing, entertainment and nightlife, tourism, and business housing. This allows us to relate the effects of traffic and time to the specific locations and their contexts in the city.

Next, we performed the mobility characterization, which is divided into two parts: in the first one, we looked at how we can detect traffic changes from the data through mobility metrics, and define which metric is the most appropriate for this purpose. Given that, we define a new metric, called path optimality, which represents the rate of trips from a point A to a point B that followed the best route between both points, where the best route is defined as the shortest in distance. We showed how this proposed metric surpasses the existing others in relation to the results found. Additionally, the metric works well even with low-grained traces. From this analysis, we come to the result that approximately 70% of the trips in the trace tend to follow the best path. This result by itself is already an interesting contribution, which can be used in the modelling of drivers' behavior in the construction of mobility models and synthetic mobility traces.

Given that, we used the path optimality metric to perform the mobility characterization given spatial and temporal factors. Considering the temporal dimension, Figure 1 shows the rate of trips that followed the shortest path grouped by day and time. The Figure shows that the traffic peak hours affects the path selection by the drivers, i.e. drivers tend to take more detours during rush hours. Adding to that, Figure 2 shows the path optimality rate grouped by the context of the location in which the trips departed and arrived. From it, we can highlight the influence that the place of origin of a trip has on the occurrence of detours during the trajectory (that can be caused by numerous reasons, such as road infrastructure). Finally, Figure 3 shows an example of the joint analysis of both spatial and temporal dimensions, with the path optimality rate for trips arriving at business locations grouped by time. It shows that, for certain regions with strongly defined periodicity (e.g. offices), there is an even higher contrast of changes during the day

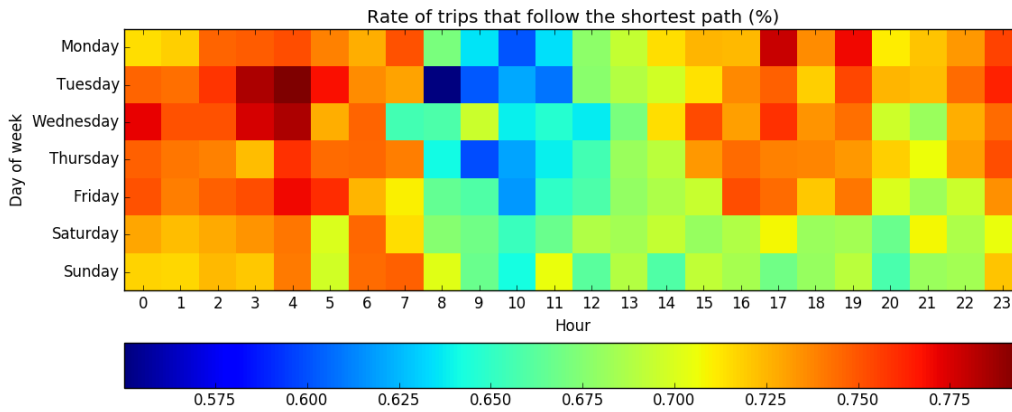


Figure 1. Rate of trips that followed the shortest path grouped by day and time. The figure shows clearly that the traffic peak hours affects the taking of detours by the drivers

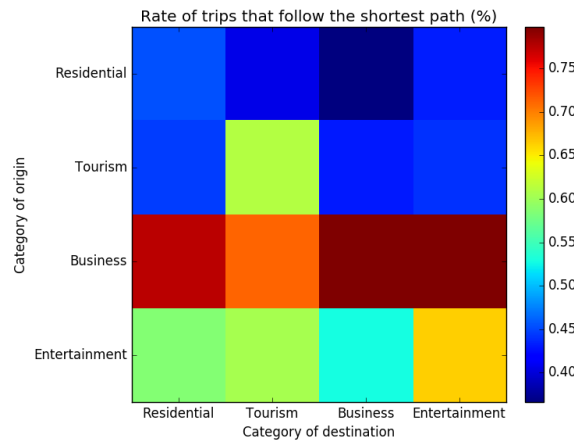


Figure 2. Rate of trips that followed the shortest path grouped by the context of the location in which they departed and arrived. The category of origin tends to affect the trips similarly, independent of the category of destination

and week. The information presented in the figures cited above can be used in message dissemination strategies and load-balancing algorithms, as well as in traffic optimization and the development of mobility models for vehicular networks.

2.2. Trajectory Filling Algorithm

In Chapter 4 of the dissertation, we use the characterization results to propose an algorithm to fill trajectories with large gaps, e.g., when only the origin and destination points are known. This algorithm is based on the premise that every driver tends to follow the shortest path between two points, which we saw on the characterization that is a correct assumption. In the cases in which the driver does not follow the shortest path, we model the detour taken from the analysis as well. Since we used the enriched data from Open Street Maps to fill the trajectories, the generated points are already calibrated and matched to the map, eliminating the need of any of these steps in the future.

To validate the algorithm, we generated an origin-destination trace from the San

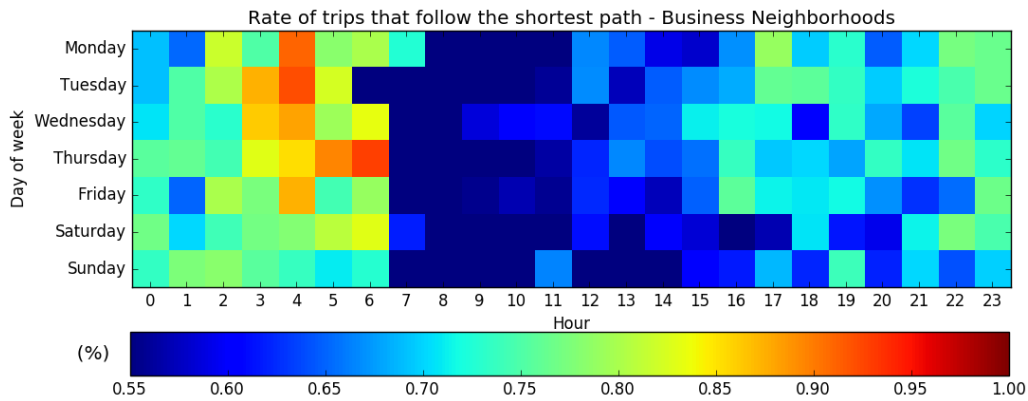


Figure 3. Rate of trips that followed the shortest path grouped by day and time, for trips that arrived in business locations. Certain regions with strongly defined periodicity show a higher contrast of changes

San Francisco Taxi Trace, containing 500 random trips. This trace contains only the first and last points in each trip. From that, we used our algorithm to fill the trajectories between both points, considering the knowledge about shortest paths and detours. Additionally, we used a baseline algorithm that considers that every trajectory follows the shortest path with no occurrence of detours, to fill the trajectories as well. We then compared our algorithm’s generated trajectories and the baseline’s generated trajectories with the ground-truth ones.

Considering all the 500 trips, our algorithm obtained a mean error of 21.3% of the trajectory length, while the baseline presented a mean error of 24.8% of the trajectory length. On the other hand, if we consider the difference in length only for the trips which had a detour (which are ignored by the baseline), the baseline solution obtained a mean error of 38.2%, while our algorithm had a mean error of 14.4%, indicating that our trajectories were closer to the real ones, and therefore represent them better. Figure 4 shows, for each one of the 500 trips, the trajectory length for the ground truth trajectories (in red), for the baseline (in blue), and for our algorithm (in green). Despite some very long trips in which the difference in the trajectories’ lengths is greater, our algorithm is capable of reproducing the trajectories and detours with a higher degree of precision than the baseline. The results showed that indeed our algorithm mimicks well the behavior of drivers, being able to reproduce the trajectories with a low rate of error between the generate trajectories and the original ones.

This algorithm is an important contribution to the research community since it allows the enrichment of existing mobility traces that suffer of spatial data sparsity, i.e. traces with only few points in the trajectory, or even with only the origin and destination points. The main benefit is that by enriching a real trace, we end up with a hybrid set of real traces enriched with formally defined mobility data, thus the final result is more realistic than strategies that simulate the occurrence of trips as well (purely formal methods).

3. Publications

As a result of this dissertation, we published one paper in an International Periodic, two papers in renowned International Conferences, and three papers in National Conferences.

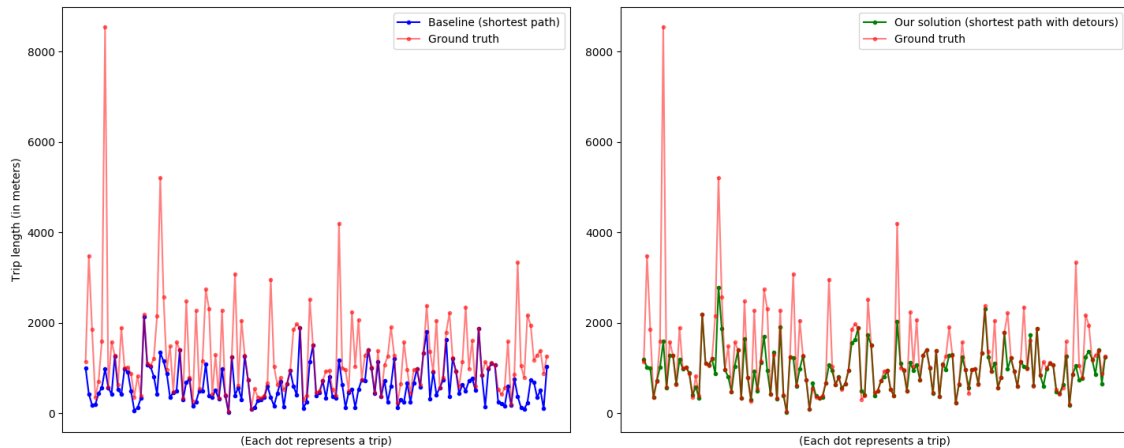


Figure 4. Trajectory length comparison between the ground truth trajectories, the ones generated by our algorithm, and the ones generated by the baseline algorithm

The following publications are related to this dissertation and contributed to its development:

- [Journal B1, 2018] Fabrício A. Silva; **Augusto C.S.A. Domingues**; Antonio A.F. Loureiro, "Discovering mobile application usage patterns from a large-scale dataset". *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 12(5):59
- [Conference A2, 2018] **Augusto C.S.A. Domingues**; Fabrício A. Silva; Antonio A.F. Loureiro, "Space and Time Matter: An Analysis About Route Selection in Mobility Traces". *IEEE International Symposium on Computers and Communications*
- [Conference B1, 2018] Fabrício R. de Souza; **Augusto C.S.A. Domingues**; Pedro O.S. Vaz de Melo; Antonio A.F. Loureiro, "MOCHA: A Tool for Mobility Characterization". *ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*
- [Conference B2, 2018] **Augusto C.S.A. Domingues**; Fabrício A. Silva; Thais R.M.B. Silva; Antonio A.F. Loureiro, "Perfil de Uso de Aplicativos Móveis: Caracterização e Aplicações". *Brazilian Symposium on Computer Networks and Distributed Systems*
- [Conference B2, 2018] Fabrício R. de Souza; **Augusto C.S.A. Domingues**; Pedro O.S. Vaz de Melo; Antonio A.F. Loureiro, "MOCHA: Um framework para caracterização e comparação de traces de mobilidade". *Brazilian Symposium on Computer Networks and Distributed Systems*
- [Conference B2, 2017] **Augusto C.S.A. Domingues**; Luís G.A. Diniz; Fabrício A. Silva; Thais R.M.B. Silva, "Understanding Mobile Application Usage During the 2014 FIFA World Cup". *Brazilian Symposium on Multimedia and the Web*

It is important to highlight the high classification level of the papers published in journal and international conferences (Qualis A2 and B1). Regarding the papers published in national conferences, it is also important to highlight that two of them were published in the Brazilian Symposium of Computer Networks and Distributed Systems, the main national conference in the area.

4. Final Remarks

This paper sums up the contributions of the dissertation in [Domingues 2018]. In it, we explored the field of mobility in Vehicular Ad Hoc Networks (VANETs), with the purpose of improving existing data and creating new realistic data that can be used in many different purposes towards the development of such networks. Given that, the objective was to perform a characterization of the vehicles' mobility and use this information to propose an algorithm to fill trajectories and generate realistic synthetic data. The state-of-the-art of trajectory mining and filling does not consider the rate of trips that follow the shortest path between origin and destination points as an alternative to analyzing the traffic levels in an urban environment. However, we showed that this is an interesting approach to traffic analysis, since its implementation is not complex and its results surpass the ones found in other metrics in the literature.

In order to fill this gap in the literature, we proposed a new metric called path optimality, that represents the rate of trips that followed the shortest path between two points. Additionally, for the trips that did not follow the shortest path, we analyzed the increase percentage (in relation to the shortest path length) that the detour taken provided. Given this information, we looked at how they are affected (positively or negatively) by contextual factors, such as the time of the day, day of week and region. Finally, we presented an algorithm to fill entire trajectory gaps (i.e. when only the origin and destination points are known) based on the analysis performed. The algorithm considers the road mesh to perform the filling, eliminating the need of performing a map matching procedure in the future, and is able to outperform the other existing solution in the literature, providing more realistic trajectories that better represent the behavior of drivers.

References

- Alipour, B., Tonetto, L., Ding, A., Ketabi, R., Ott, J., and Helmy, A. (2018). Analyzing mobility-traffic correlations in large wlan traces: Flutes vs. cellos. *arXiv preprint arXiv:1801.02705*.
- Çatay, B. and Keskin, M. (2017). The impact of quick charging stations on the route planning of electric vehicles. In *IEEE Symposium on Computers and Communications, 2017*, pages 152–157. IEEE.
- Cotta, L., de Melo, P. O. V., and Loureiro, A. A. (2017). Understanding the role of mobility in real mobile ad-hoc networks connectivity. In *IEEE Symposium on Computers and Communications, 2017*, pages 1098–1103. IEEE.
- Diniz, G. R., Cunha, F. D., and Loureiro, A. A. (2017). On the characterization of vehicular mobility. In *Modeling, Analysis and Simulation of Wireless and Mobile Systems, 2017*, pages 23–29. ACM.
- Domingues, A. C. (2018). Route selection and trajectory filling in mobility traces. Master's thesis, Federal University of Minas Gerais.
- Garcia, J. C., Avendaño, A., and Vaca, C. (2018). Where to go in brooklyn: Nyc mobility patterns from taxi rides. In *World Conference on Information Systems and Technologies*, pages 203–212. Springer.

- Hou, X., Li, Y., Jin, D., Wu, D. O., and Chen, S. (2016). Modeling the impact of mobility on the connectivity of vehicular networks in large-scale urban environments. *IEEE Transactions on Vehicular Technology*, 65(4):2753–2758.
- Lu, M., Lai, C., Ye, T., Liang, J., and Yuan, X. (2017). Visual analysis of multiple route choices based on general gps trajectories. *IEEE Transactions on Big Data*, 3(2):234–247.
- Lu, Z., Feng, Y., Zhou, W., Li, X., and Cao, Q. (2018). Inferring correlation between user mobility and app usage in massive coarse-grained data traces. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):153.
- Piorkowski, M., Sarafijanovic-Djukic, N., and Grossglauser, M. (2009). Crawdad dataset epfl-mobility (v. 2009-02-24). Downloaded from <https://crawdad.org/epfl/mobility/20090224>.
- Qiao, L., Shi, Y., and Chen, S. (2017a). An empirical study on the temporal structural characteristics of vanets on a taxi gps dataset. *IEEE Access*, 5:722–731.
- Qiao, Y., Si, Z., Zhang, Y., Abdesslem, F. B., Zhang, X., and Yang, J. (2017b). A hybrid markov-based model for human mobility prediction. *Neurocomputing*.
- Sadri, A., Salim, F. D., and Ren, Y. (2017). Full trajectory prediction: what will you do the rest of the day? In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 189–192. ACM.
- Silva, F. A., Celes, C., Boukerche, A., Ruiz, L. B., and Loureiro, A. A. (2015). Filling the gaps of vehicular mobility traces. In *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 47–54. ACM.
- Wang, D., Liu, Q., Xiao, Z., Chen, J., Huang, Y., and Chen, W. (2017). Understanding travel behavior of private cars via trajectory big data analysis in urban environments. In *Proceedings of the 15th International Conference on Dependable, Autonomic and Secure Computing*, pages 917–924. IEEE.
- Wang, P., Liu, G., Fu, Y., Zhou, Y., and Li, J. (2018a). Spotting trip purposes from taxi trajectories: A general probabilistic model. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 9(3):29.
- Wang, Y., Qin, K., Chen, Y., and Zhao, P. (2018b). Detecting anomalous trajectories and behavior patterns using hierarchical clustering from taxi gps data. *ISPRS International Journal of Geo-Information*, 7(1):25.
- Xia, F., Wang, J., Kong, X., Wang, Z., Li, J., and Liu, C. (2017). Exploring human mobility patterns in urban scenarios: A trajectory data perspective. *IEEE Communications Magazine*.
- Yao, Z. (2018). Exploiting human mobility patterns for point-of-interest recommendation. In *WSDM, 2018*, pages 757–758. ACM.