HAL

archives-ouvertes.fr

Understanding Interactions in Vehicular Networks Through Taxi Mobility

Felipe Domingos da Cunha, Davidysson Abreu Alvarenga, Aline Carneiro
Viana, Raquel A. F. Mini, Antonio A. F. Loureiro

▶ To cite this version:

HAL Id: hal-01199675
https://hal.inria.fr/hal-01199675

Submitted on 24 Sep 2015

Understanding Interactions in Vehicular Networks Through Taxi Mobility *

Felipe D. Cunha
DCC - UFMG
fdcunha@dcc.ufmg.br

Davidysson A. Alvarenga
DCC - PUC Minas
davidysson@ufmg.br

Aline C. Viana
INRIA - Saclay
aline.viana@inria.fr

Raquel A. F. Mini
DCC - PUC Minas
raquelmini@pucminas.br

Antonio A. F. Loureiro
DCC - UFMG
loureiro@dcc.ufmg.br

ABSTRACT

Vehicular Networks (VANETs) are an emerging network that enables the communication among the vehicles, in order to promote a safe and efficient traffic, avoiding crashes and preventing hazards. These vehicles transit on the streets and highways, and during their trajectories, they can communicate with each other or with another network, through interactions among them and road side units. Aiming to better understand these interactions, in this work, we characterize the vehicular mobility through a detailed analysis of dataset traces, which portray the mobility of a group of taxis in a great city. We perform the analysis using statistical techniques, graph theory and network analysis, extracting properties and behaviors from the mobility traces. The results reveal the existence of regularity and common interests in the studied traces.

Categories and Subject Descriptors

C.2.0 [**Computer-Communications Networks**]: General—*Data communications*; C.2.1 [**Computer-Communications Networks**]: Network Architecture and Design—*Wireless communication*

Keywords

VANETs Interactions Social Behaviors Graph Analysis

1. INTRODUCTION

Vehicular Networks (VANETs) are an emerging network that has attracted the interest of the research community. This is a type of ad hoc network, formed by vehicles with processing and wireless communication abilities, traveling on streets or highways. Normally, the vehicles can establish a direct communication or by the use of a roadside unit. Differently from other networks, in VANETs, the nodes are vehicles such as cars, buses, taxis, and trucks. Each type of vehicle presents a different behavior and a mobility pattern. In the meantime, due to the high vehicle mobility, in VANETs, the topology changes frequently [2, 4].

In these VANETs scenarios, the communication is highly influenced by the vehicles' mobility patterns. The vehicles move and stop according to interactions that they perform among the drivers and the limits of speed and directions imposed by the regulatory signs. However, the interactions among day period, the driver routines and vehicles can also influence on this mobility. Observing the traffic during weekdays, we can note rush hours with the traffic slow in some streets and traffic jams. On contrary, in early mornings and weekends the traffic is quieter, with low vehicles' density. This traffic variation in different days and hours reinforces the characteristic of dynamic topology of VANETs, making the communication a challenging task in some scenarios.

When we consider the taxis, the interactions among people, vehicles and places have a great influence in the mobility. In these vehicles, the passenger defines the trip destination. During the weekends, they decide to go to destinations such as malls, theaters, churches and camping are chosen for leisure and entertainment. On weekdays, people tend to repeat their choices, at the same time and to the same destination, such as school, work, offices, university, restaurant and coffee shops.

During their trajectories, taxis interact with other vehicles, go through the same streets, and are subjected to the same traffic conditions. These features suggest the investigation of the taxi mobility, in order to understand the interactions that they present during their trajectories. In addition, to improve the connectivity and services in VANETs, it would be interesting to understand how interactions among vehicles happen.

In order to better understand the daily traffic evolution, the routines and the driver's behavior, in the literature, there are many studies that analyze the mobility of vehicular networks [5, 8, 9, 11]. However, these are studies based on analytical mobility models or realistic traces, and the analysis takes into account short periods. In this work, we focus on the investigation of the following question: *How do taxi vehicles perform interactions in a vehicular network?* Thus, we present statistical and social network analyzes of two real traces that describe the mobility of a taxi group on a great city, extracting properties and behaviors from the taxi mobility, in order to verify how the taxi interactions are performed over the day.

The rest of this paper is organized as follows. In Section 2, we present the related work. In Section 3, we present the main features of the analyzed dataset. In Section 4, we define the interactions, the techniques and metrics used to analyze the traces. In Section 5, we present the quantitative results. Finally, in Section 6, we present the final conclusions and future perspectives.

2. RELATED WORK

When we look at the literature, it is possible to find studies that have the focus on the analysis of physical encounters of nodes on the network, aiming to understand the evolution of the network topology and the connectivity among the nodes. In vehicular networks, due to their peculiarities, it is interesting to perform traffic mobility analysis to better design protocols and services. In this way, Fiore and Härri [5] present a depth analysis of the topological properties of vehicular networks, where they explored the social metrics in different mobility models to investigate the temporal evolution of the network topology. Results show that it is possible to take advantage of the vehicular mobility to improve the performance of the network protocols. However, that work is based only on analytical mobility models.

In the same context, Loulloudes et al. [9] and Pallis et al. [11] discuss how social metrics can be employed to improve the performance of the routing protocols in urban vehicular networks. They perform the analysis in realistic and real traces, and their results show the importance of considering topology aspects to design new routing protocols. Nevertheless, their analysis is based on a period of only two hours of traces, not the entire dataset. Also, Liu et al. [8] discuss several universal laws of social networks, presenting an analysis of two dataset traces. However, they only consider a reduced number of vehicles and three social metrics, i.e., node degree, distance and cluster coefficient.

Uppoor and Fiore [14] present a macro and microscopic evaluation of the Cologne trace. The macroscopic analysis involves the evolution of vehicular density and the dynamics of large-scale flows of vehicles through the metropolitan region; and the microscopic analysis considers the distribution of vehicles in the area and the encounters among them. They present the evolution of the traffic over the area and time, highlighting the importance of designing protocols and services to vehicular networks considering the dynamism of the road traffic. Melo et al. [3] present a strategy to analyze users' interactions in dynamic networks. They define a way to classify the user interactions among random and social relationships. Similarly, they evaluated mobility traces describing different users. However, they use a long slot time, which groups all encounters that happen during one day. In our work, we will consider the whole trace duration, figuring out how the interactions happen among the taxis and the environment.

3. VEHICULAR DATASETS ANALYZED

Traces are a special type of dataset that contains information about the trajectory of vehicles, with the purpose of tracking the vehicles' mobility. Such traces are attractive since they exhibit the real behavior of vehicles in a certain scenario, and they can be either real or realistic (synthetic) traces. Real traces contain information collected from a group of vehicles with the use of some localization system device (e.g., GPS); and realistic or synthetic traces are created by the junction of maps of a particular locality, traffic information of this locality, and a mobility' simulator. In the following, we describe the main features of the two traces used in our evaluation. We chose these traces because they are real traces, available to download on the Internet with the duration of 30 days.

- **San Francisco:** this real trace contains mobility traces of taxi cabs in San Francisco, USA. Each taxi has a GPS receiver, and the trace contains GPS coordinates, at each minute, of 500 taxis, collected over 30 days in the San Francisco area, from May 17 to June 10 of 2008 [12].

- **Rome:** this real trace contains mobility traces of taxi cabs in Rome, Italy. Each taxi has a GPS receiver, and the trace contains GPS coordinates of approximately 320 taxis that work in the center of Rome, collected over 30 days in the Rome area, from February 1 to March 2 of 2014 [1]. Each taxi driver has a tablet that periodically (7 seconds) retrieves the GPS position and sends it to a central server.

4. INTERACTIONS

Aiming to better understand the interactions in vehicular networks, we consider as an interaction the definition presented in [13], which defines it "as an occasion when two or more people or things communicate with or react to each other". Taking into account this definition, we nominated a group of entities, which interact with the vehicles during their routines such as people, signals and roads. In this work, our focus is to analyze the interactions between the taxi vehicles and these entities. In order to perform this analysis, we use temporal graphs along with metrics of complex networks and statistical techniques. In the following sections, we define the modeling used to create the temporal graphs, and the interactions that we evaluate.

4.1 Temporal Graphs

Temporal graphs are created considering the unit disc model, which allows us to make the inference of the encounter between vehicles, i.e., when two vehicles are within communication range of each other. In our evaluation, we consider two communication ranges, 100 m and 150 m, according to the protocol IEEE 802.11p. Thus, we vary the range aiming to capture the effects of range size in the topology and metric behaviors.

After that, we map the vehicles' mobility and encounters described in the trace into a temporal graph. Figure 1 shows the number of components and the edge density of the graph, when we vary the time slot duration. As we can see, a very short time slot leads to a fragmented network, presenting many nodes without connections and a low-density graph. On the other hand, a long time slot generates a small number of components and a higher density graph, which portrays a homogeneous behavior in the metrics. Thus, we divide the whole trace into discrete time slots of duration $t = 15\ min$, aiming to better capture the traffic changes. We generate each temporal graph $G(t)$ using the Growing Time Window technique [6], which puts together all encounters which happened during the same time slot.

The temporal graph at time t is an undirected graph, and can be formally defined as a graph $G(t) = (V, E)$, where V represents the set with all vehicles v_i and E represents the set of edges e_{ij}. In $G(t)$, an edge $e_{ij}(t)$ exists between vehicles v_i and v_j during time t, with $i \neq j$. The evaluation of each metric is applied to each graph, taking into account each temporal graph $G(t)$. Thus, we intend to analyze the encounters that happen in each period t. For each trace, 96 graphs are generated describing the vehicles' encounters per

(a) Rome. (b) San Francisco.
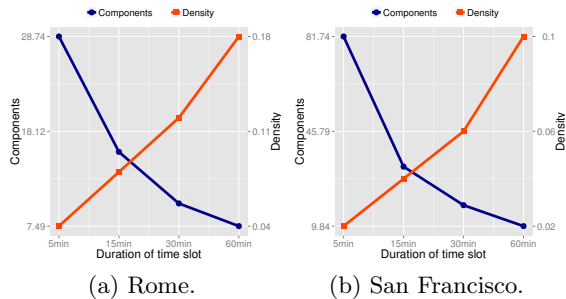
Figure 1: Number of components and edge density varying the time slot window for 150 meters of communication range.

day $G(t) = \{G(t_1), G(t_2), G(t_3), \ldots, G(t_{96})\}$, and the values of the metrics represent the average for each metric in each time slot analyzed. Also, we compare the results with random graphs G^R generated with the same degree distribution $G^R(t) = \{G^R(t_1), G^R(t_2), G^R(t_3), \ldots, G^R(t_{96})\}$ [7].

4.2 Vehicles' Interactions

The interactions between the vehicles can happen when two or more vehicles are inside the communication range of one another. At this moment, they can exchange data between them, and providing services. However, other entities can cooperate to happen an interaction and even they can affect directly to time and duration of this interaction. Thus, according to the entities, bellow we define the interactions and the metrics that represent them.

Vehicles and Calendar: this interaction represents the traffic density variation over time and days. At daybreak, the traffic is quieter and during the day, it presents a peak of high traffic density. In addition, rush hours and special events present a higher traffic density comparing with weekends and holidays. We evaluate this interaction in order to identify the time and days where we find a high traffic density in the city, aiming to promote services and connectivity to the vehicles aware of these variations. We compute and analyze this interaction considering the statistical analysis of the traffic variation over the days.

Vehicles, People and Calendar: this interaction causes variations in the encounter frequency of the vehicles during the day. People who have similar routines normally present a higher encounter frequency, because they cross the same roads at the same hour, which portrays a higher interaction level between these people. Moreover, the weekday can also influence in this interaction. On weekdays, the traffic is higher and the encounter is a lot more frequent than on the weekends. To measure this interaction type, we use the node degree and the edge persistence.

The node degree determines the number of distinct encounters that a vehicle has during a period. This metric can be influenced by the route and the period of the day. If a vehicle goes through a region with a higher traffic density or in a rush hour, its degree tends to be high. On other hand, at regions far from the downtown or on hours with low traffic, its degree can be low or null. The edge persistence denotes the number of times two nodes encounter at the same time-slot window. Taking into account the trace duration, in this work, the edge persistence portrays the number of

times the edge repeats over the days at the same time. With this metric, we can discover the vehicles that present some regularity among them. Thus, vehicles with higher edge persistence present similar behaviors and routines compared with their neighbors [3].

Vehicles, People, and Places: For these interactions, we propose some questions: *what is the time of day with the highest number of vehicles in transit?*, *what is the place most visited in the city?* or *what type of place people prefer to visit?*. Considering these questions and some other features of these places, we can better understand the mobility pattern of vehicles in the city and the interaction pattern between those entities (i.e., frequency, time, duration, or the number of people who visit each place). These important aspects help us to define people's mobility patterns in the city. To analyze this interaction, we use graphs and compute the metrics distance, cluster coefficient and topology overlap.

The metric distance means the length of a path between a pair of nodes in terms of the number of hops. This distance can represent the existence of a common interest among the drivers. If the distance between the vehicles is long, it implies that they are physically separated, and probably do not visit nearby places. Otherwise, if the distance is small, they tend to visit close places or have similar routes.

The cluster coefficient measures the tendency of nodes to cluster on the graph. In other words, we can measure the density of connections among the neighbors in the graphs. This metric can also represent vehicles with similar interests on a network, probably vehicles with similar behaviors and routines. Besides, higher values can represent areas with more traffic jams, increasing the probability of encounters. In the same direction, the topology overlap portrays the similarity of encounters among nodes. Thus, vehicles with similar behaviors and routines will portray a high value to topological overlap. A vehicle with a high node degree and a low topological overlap can indicate a hub. Probably, this vehicle connects to different vehicles and can prosecute an important role in the communication.

5. NUMERICAL RESULTS

In this section, we present the analysis of interactions considering the evaluation of the two traces: San Francisco and Rome. We show the results for two communication ranges, 100 m and 150 m, compared with the results of the random graphs G^R for each analyzed metric. We apply some statistical techniques to measure some behaviors highlighting the interactions. In the next section, we present the discussion and results grouped by the interaction mentioned in the previous section.

5.1 Vehicles and Calendar Interactions

Considering the interactions between vehicles and the calendar, we can analyze the traffic variation during the day and week. We can identify the peak hours in each city, the days with low traffic and possible traffic variations in holidays. Figure 2 presents the weekly and daily evolution of the traffic density for the cities Rome and San Francisco respectively. When we analyze the entire week for the two traces (Figs. 2-(a) and (c)), we can notice that both traces present constant behavior during the weekdays. However, for Rome trace, we observe a great decrease during the weekend. In contrast, for San Francisco, just presenting a little decrease at Sundays. Taking into account the daily traffic

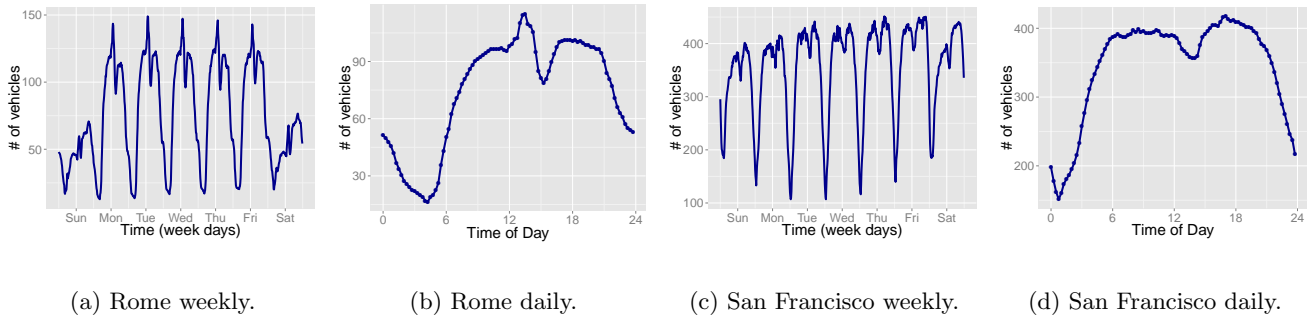(a) Rome weekly.　　(b) Rome daily.　　(c) San Francisco weekly.　　(d) San Francisco daily.

Figure 2: Traffic evolution in Rome and San Francisco.

evolution (Figs. 2-(b) and (d)), we can note that the Rome trace presents a decrease in the daybreak, due to the trips reduction in this time. On the other hand, the San Francisco trace presents a different behavior: it portrays the routine of the city on the traffic, presenting a constant behavior during the whole day. At daybreaks the traffic reduces following the decrease in the number of trips.

5.2 Vehicles, People and Calendar Interactions

The node degree represents the number of interactions a vehicle has during a time slot. Figure 3 shows the daily mean degree evolution for the two traces, considering the communication range variation. For both traces, we can note the same behavior on the curves. In Rome, Figures 3-(a) and (c), we note a decrease in the degree during daybreak and, at 6 am, the value starts to increase, following the beginning of the day. This happens due to the fact that during the daybreak the number of trips surfers a reduction. In San Francisco, Figures 3-(b) and (d), we observe an increase in the degree at 6 am, following the beginning of the day. Moreover, during the lunch time (from 12 pm to 2 pm), the value of degree presents a high increase. We attribute this behavior because, in this period, people tend to go out to the same places (e.g., bars, coffee places and restaurants), and the traffic is intense in some regions.

Considering the degree distribution for the nodes at the transition times (0 am, 6 am, 12 pm and 6 pm), in the Figure 4, we present the CDF distribution for a typical Monday. In the Rome trace (Figure 4-(a) and (c)), we can observe that the major degree value appears at 12 pm, and at this time few taxis present high degree. At 6 am, the taxis present low degree value in accordance to the presence of low traffic density. Also, at 12 pm and 6 pm, the probability to find a vehicle with a high degree is low. Otherwise, in San Francisco trace (Figure 4-(b) and (d)), considering the number of taxis in transit, we can find many vehicles with a low degree, close to 30, and few vehicles presenting high degree. Also, at 12 pm, we found the largest degree value, close to 70, which indicates that the probability to encounter is higher.

We also compute the edge persistence in both traces. We consider the edge frequency, in each time slot during all days. Besides the random generation process, we also compute the edge persistence for random graphs. Figure 5 shows the results for the two traces. For both, we can note that the communication range variation does not imply in behavior changes at the curves, just increases the values. Taking into account the two traces, we can note that the Rome trace

(a) Rome – 100 m　　(b) San Francisco – 100 m
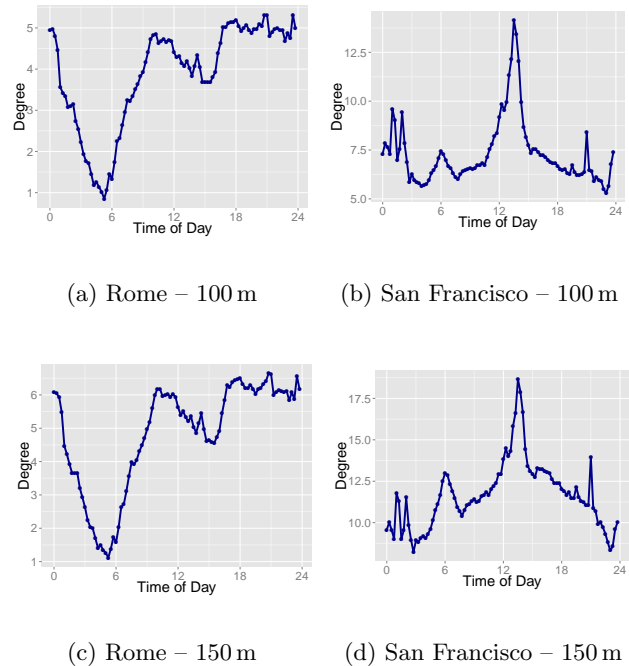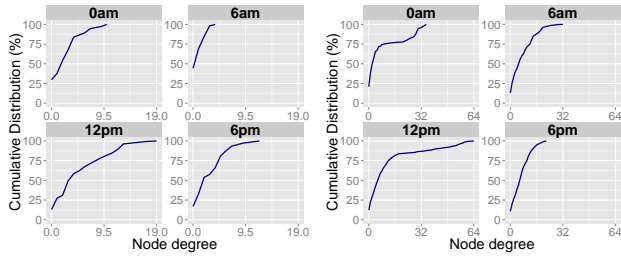

(c) Rome – 150 m　　(d) San Francisco – 150 m

Figure 3: Degree daily evolution for Rome and San Francisco's traces, varying the communication range.

presents a major edge persistence than the San Francisco trace, i.e., in Rome, the edges are repeated on average 30% over the day, compared with San Francisco is 10% of repetitiveness. Also, in San Francisco trace, Figures 5-(b) and (d) show that the greatest value of edge persistence is 30$, which is observed in lunch time where we have more trips with common destinations. This happens because people can go to the same place, but not necessarily they take the same taxi.

Analyzing the encounter duration and frequency for the first Monday at Figure 6, we can see the results for both traces. We note that Rome (Figure 6-(a)) presents a greater encounter duration at daybreak, and during the day, taxis perform a lot of interactions with short durations. On the contrary, in San Francisco (Figure 6-(b)) the interactions are more frequent at lunch time, presenting longer durations. We can also observe a higher number of interactions in San

(a) Rome – 100 m   (b) San Francisco – 100 m
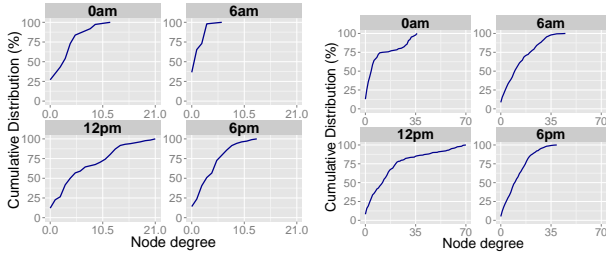
(c) Rome – 150 m   (d) San Francisco – 150 m

Figure 4: CDF for Node Degree in Rome and San Francisco's traces, varying the communication range.

(a) Rome – 100 m   (b) San Francisco - 100 m

(c) Rome – 150 m   (d) San Francisco - 150 m

Figure 5: Daily evolution, edge persistence for Rome and San Francisco's traces, varying the communication range.
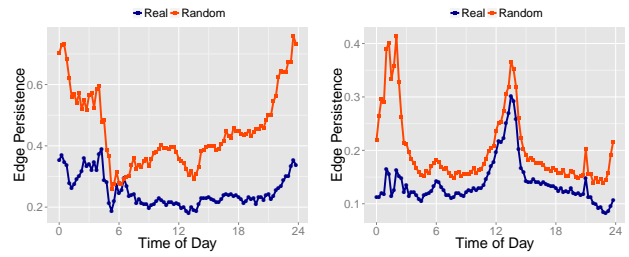
Francisco compared to Rome, because in San Francisco we have more taxi transitings in a small area.

5.3 Vehicles, People, and Places Interactions

Analyzing the interaction among the traffic, people and the city area, Figure 7 presents a heatmap for the two hours (6 am and 6 pm). We chose these hours in order to capture the transitions at the beginning and at the end of the day. Figures 7-(a) and (b) present the results for the Rome trace, and the Figures 7-(c) and (d) present the results for San Francisco trace. Observing the Rome traffic at 6 am, we can see a sparsely traffic at the peripheral areas, and at 6 pm a dense traffic at downtown. Differently from the San Francisco trace, for both hours, we can observe high interactions of vehicles around the Financial District, the main central business district at San Francisco. However, we observe less vehicles transiting at 6 am than at 6 pm
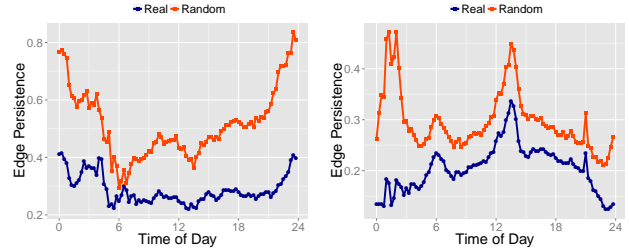
When we analyze the nearby interactions, we compute the distances among the vehicles on the graph, i.e., the number of edges that exist in the path among these nodes. Moreover, when we consider a vehicular network and the communication range as the criteria to add an edge, this metric can also represent how much two vehicles are distant to each other. We can see in Figure 8 the average distance for both traces during the day, considering the communication range variation.

As we can see, Figures 8-(a), (b), (c) and (d) represent the mean distance for random graphs, which have a lower value compared to the real traces. Due to the random criteria to add an edge in a random graph, we believe that the distance decreases by adding short cuts to the graph. Moreover, for the Rome trace, Figures 8-(a) and (c), we note that the distance follows the traffic behavior, decreasing during the

daybreak. In addition, considering the number of vehicles transiting, we can observe a very short distances among them, close to 3 hops.

Regarding the San Francisco trace, Figures 8-(b) and (d) show the mean distance over the day. As we can see, the distance also follows the traffic density, presenting a constant behavior during all day, just with a reduction at daybreak. However, at 12 pm, we can see a little tendency to decrease the number of interactions, since at that time, the node degree increases and more encounters happen. Thus, we can also observe that, to data deliver from the source vehicle to the destination vehicle can spend, in average, more hops during the day compared to night period.

Taking into account the vehicles closely, we compute the clustering coefficient, and Figure 9 presents the results. Re-
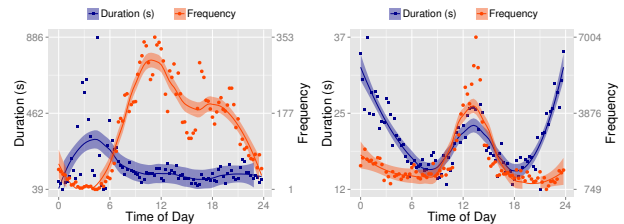
(a) Rome.   (b) San Francisco.

Figure 6: Encounter duration and frequency in Rome and San Francisco's traces for the first Monday analyzed.
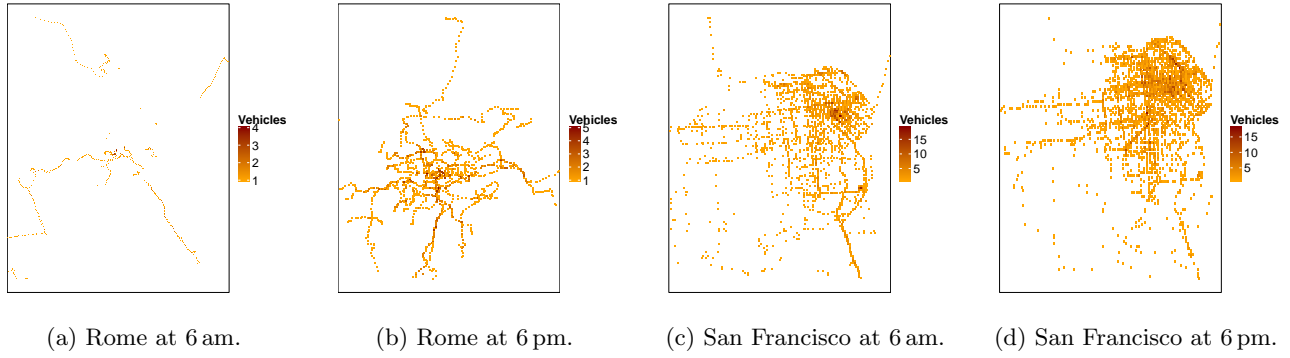
(a) Rome at 6 am. (b) Rome at 6 pm. (c) San Francisco at 6 am. (d) San Francisco at 6 pm.

Figure 7: Traffic concentration in Rome and San Francisco.

(a) Rome – 100 m (b) San Francisco – 100 m

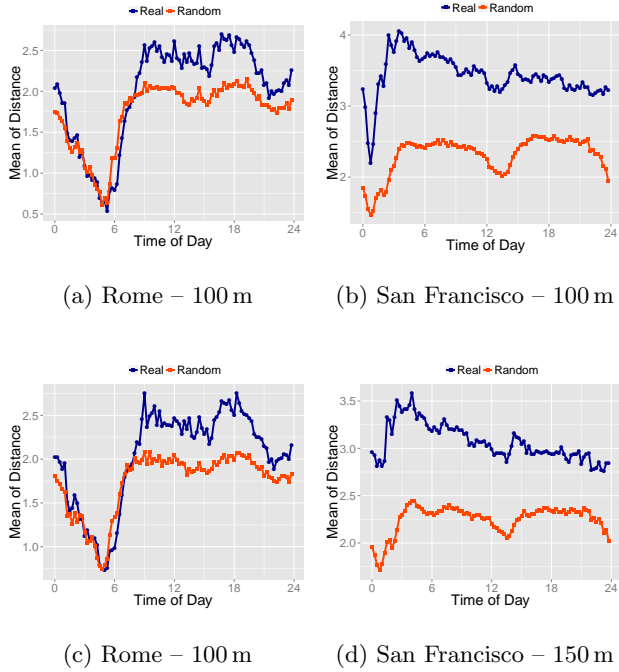(c) Rome – 100 m (d) San Francisco – 150 m

Figure 8: Mean distance daily evolution for Rome and San Francisco's traces, varying the communication range.

marking this metric, we can measure how much node neighbors are connected, forming communities in network. Generally, the communities gather nodes that have same interest. In vehicular networks, when two or more vehicles have the same interest, it means that they cross the same roads, transiting through the same area, or vehicles that interact at the same places.

In Figures 9-(a) and (c), we can see the results to the Rome trace. For both communication ranges, we can note that the random graph presents an expressive difference to the real trace, which indicates that the random generation process breaks the communities from the original graph. Despite the traces portray the taxis' mobility, the vehicles go to the same places, guided by the passengers which present same interests.

In the same context, Figures 9-(b) and (d) show the evolution of the clustering coefficient during the day. First, we can see that at 12 pm, the trace presents the highest value

of clustering coefficient, since this is the time in which a higher number of vehicles share the same destinations. In addition, we can see during the day that despite the trace portrays the taxisâĂŹ behavior, where each trip represents the desire of one person, they have common interests, forming vehicle communities on the graph. Beyond all, the random graph presents a result much less than the real graph, which indicates the presence of structures on the graphs.
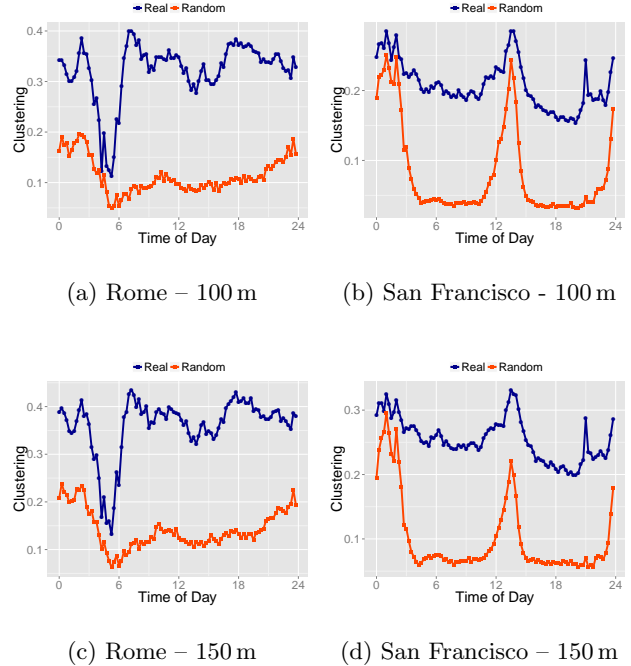

(a) Rome – 100 m (b) San Francisco - 100 m

(c) Rome – 150 m (d) San Francisco – 150 m

Figure 9: Clustering coefficient daily evolution for Rome and San Francisco's traces, varying communication range.

In order to analyze the clustering coefficient distribution among the nodes, in Figure 10, we present the clustering coefficient of the real trace with 100 m of range communication, using box plots graphics. In these graphics, we can see the variability of the values, the median and the outliers for each hour during the day. For the Rome trace, we can note that during daybreak the trace presents a major variation in the values of clustering. This is expected because, at this time, the traffic is sparse and the destination is variable. Also, during the day, mainly in rush times, the clustering value is

more constant, close to 40%. For the San Francisco trace, the values suffer less variation, although we can see the presence of some outliers, i.e., vehicles that present different values from others. Generally, during the high traffic period, the mean value of clustering coefficient is close to 20%.

(a) Rome   (b) San Francisco
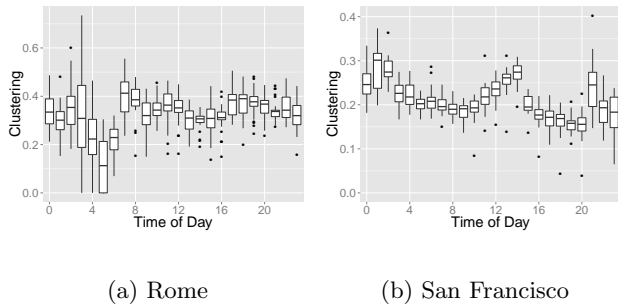
Figure 10: Distribution for Clustering Coefficient for Rome and San Francisco's traces (range 100 m).

Furthermore, to better understand the topology of graphs, in Figure 11, we generate both graphs (temporal and random) of each trace in a typical Friday at 6 pm In each graph, we use the *walktrap community detection algorithm* to identify the communities [10], highlighting the communities. We can observe, comparing the real and random graphs, the presence of communities in the real graph (nodes with the same color). Both graphs present nodes with same interest. Also, due to the fact that San Francisco has more taxis, we can observe a higher number of disconnected communities. Taking into account the presence of communities, we can use this metric to design new services or to reach a specific community, e.g., to perform marketing and publicity or to delivery an alert message that matters only to the group.

Following the clustering coefficient analysis, we also investigate the topological overlap. In Figure 12, we present the values for the topological overlap during the day. In both traces, we observe that the random graphs present a lower value compared with the real traces. This behavior indicates the presence of similar interest, i.e., people go to the same places. Regarding the trace of Rome (Figures 12-(a) and (c)), we see that during the day, the value of the topological overlap follows the traffic density. Also, at the rush times, we have more taxis sharing more neighbors, which can define similar behaviors. Considering the San Francisco trace, (Figures 12-(b) and (d)), we can see that the moment of the day that presents the highest value is also during the lunch time. Whether we have more taxis going to the same destinations, the probability to see higher values of topological overlap is bigger.

Aiming to capture the variation of the topological overlap values and the probability of one vehicle having a large value, Figure 13 presents the cumulative distribution function in the transition times for the first day analyzed. We observe that in each evaluated trace, few vehicles present higher value of topological overlap. Furthermore, the taxis share more vehicles neighbors at the end of the day, i.e., they present the major value of topological overlap. We can see also that the communication range variation does not change the probability and the curves present the same behaviors. Regarding the vehicular network properties, for both traces, we have a group of vehicles sharing neighbors,

which indicate the presence of social proprieties. However, in the dissemination process, it is interesting to choose vehicles with low topological overlap, which guarantees an increase in the delivery ratio, by reducing the number of duplicate messages without compromising the dissemination process.

6. CONCLUSION AND FUTURE WORK

In this paper, we presented an analysis of the interactions in vehicular networks using two vehicular traces (San Francisco and Rome) that describe the mobility of taxis in the metropolitan area. For both traces, we could verify higher interactions of the passenger's routines on the taxis' traffic. Despite being taxis passengers, they define the places to go and people have similar routines, which are portrayed in the interactions between the traffic and the time of day, with the presence of rush times and traffic jams. Also, we verified that they share destinations and interests, forming communities on the graph. Thus, in accordance to the work presented in [3], we can classify the encounters among the taxi vehicles as *acquaintance*: group of vehicles that share many common encounters, but not meeting often.

As future work, we plan to extend this evaluation grouping more traces that describe different mobility behaviors such as taxis, buses, cars, and bikes. We will improve the evaluation using calibrated traces comparing original ones. Moreover, we will use more statistic tests, tools and techniques in order to define mobility models that better portray the vehicles behaviors. Furthermore, we will use these results as an input to services and protocols in vehicular networks, aiming to improve the performance of services and ameliorate the connectivity among vehicles.

7. REFERENCES

[1] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, and A. Rabuffi. CRAWDAD data set roma/taxi (v. 2014-07-17). Downloaded from http://crawdad.org/roma/taxi/, July 2014.

[2] F. Cunha, A. Carneiro Viana, R. A. F. Mini, and A. A.F. Loureiro. Is it possible to find social properties in vehicular networks? In *IEEE Symposium on Comp.and Com.(ISCC '14)*, 2014.

[3] P. O. V. de Melo, A. C. Viana, M. Fiore, K. Jaffrès-Runser, F. L. Mouël, A. A. Loureiro, L. Addepalli, and C. Guangshuo. Recast: Telling apart social and random relationships in dynamic networks. *Performance Evaluation*, 87(0):19 – 36, 2015. Special Issue: Recent Advances in Modeling and Performance Evaluation in Wireless and Mobile Systems.

[4] M. Faezipour, M. Nourani, A. Saeed, and S. Addepalli. Progress and challenges in intelligent vehicle area networks. *Communications ACM*, 55(2):90–100, feb 2012.

[5] M. Fiore and J. Härri. The networking shape of vehicular mobility. In *ACM Int. Symposium on Mobile Ad Hoc Net. and Computing (MobiHoc '08)*, pages 261–272, 2008.

[6] T. Hossmann, T. Spyropoulos, and F. Legendre. Know thy neighbor: Towards optimal mapping of contacts to social graphs for dtn routing. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9, March 2010.
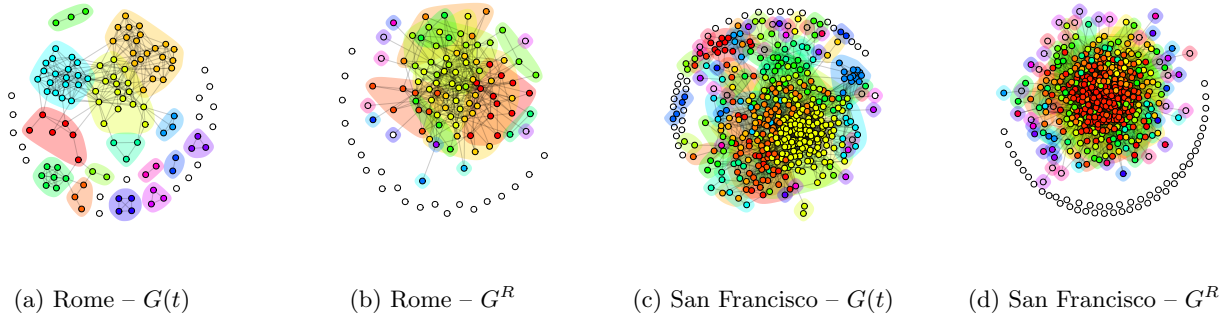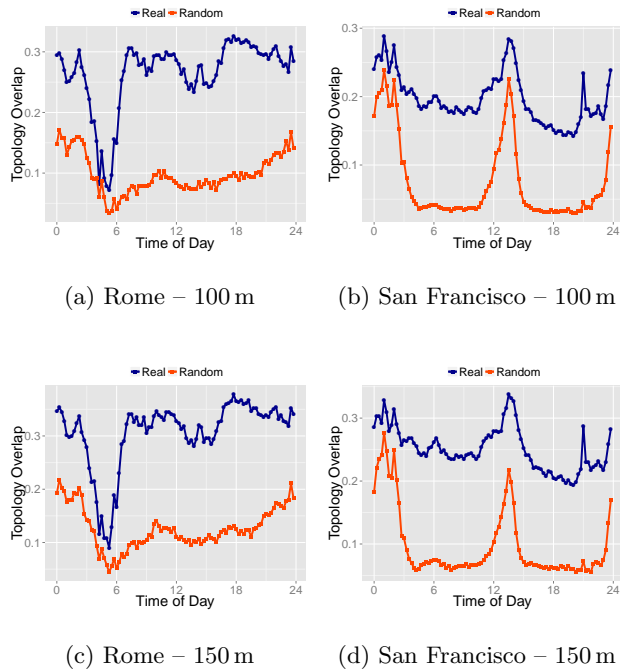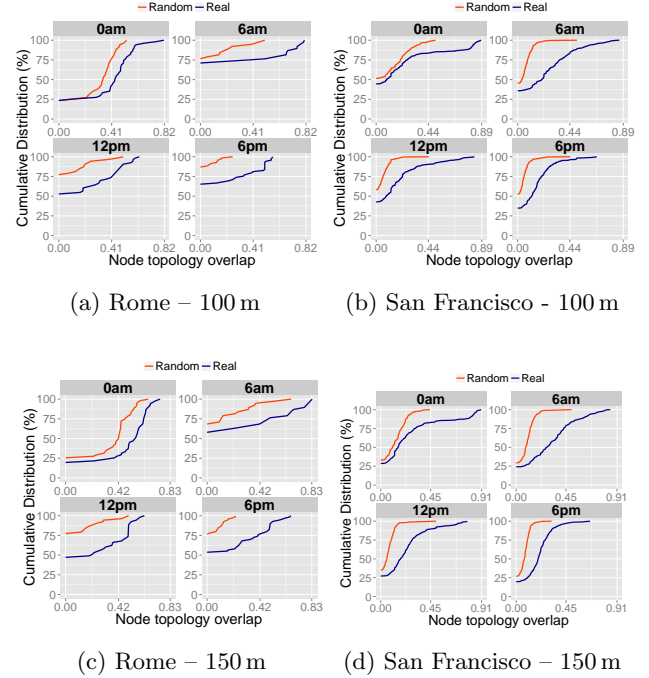
(a) Rome – $G(t)$ (b) Rome – G^R (c) San Francisco – $G(t)$ (d) San Francisco – G^R

Figure 11: Snapshot for $G(t)$ and G^R for Rome and San Francisco's traces at $6pm$.

(a) Rome – $100\,\mathrm{m}$ (b) San Francisco – $100\,\mathrm{m}$

(c) Rome – $150\,\mathrm{m}$ (d) San Francisco – $150\,\mathrm{m}$

Figure 12: Topological Overlap Daily evolution for Rome and San Francisco's traces, varying the communication range.

(a) Rome – $100\,\mathrm{m}$ (b) San Francisco - $100\,\mathrm{m}$

(c) Rome – $150\,\mathrm{m}$ (d) San Francisco – $150\,\mathrm{m}$

Figure 13: Cumulative distribution function for the Topological Overlapping for Rome and San Francisco's traces, varying the communication range.

[7] N. L. Johnson and S. Kotz. *Urn Models and Their Applications: An Approach to Modern Discrete Probability Theory*. Wiley, New York, 1977.

[8] X. Liu, Z. Li, W. Li, S. Lu, X. Wang, and D. Chen. Exploring social properties in vehicular ad hoc networks. In *Proceedings of the Fourth Asia-Pacific Symposium on Internetware*, Internetware '12, pages 24:1–24:7, New York, NY, USA, 2012. ACM.

[9] N. Loulloudes, G. Pallis, and M. D. Dikaiakos. The dynamics of vehicular networks in urban environments. *CoRR*, abs/1007.4106, 2010.

[10] G. Orman and V. Labatut. A comparison of community detection algorithms on artificial networks. In J. Gama, V. Costa, A. Jorge, and P. Brazdil, editors, *Discovery Science*, volume 5808 of *Lecture Notes in Computer Science*, pages 242–256. Springer Berlin Heidelberg, 2009.

[11] G. Pallis, D. Katsaros, M. Dikaiakos, N. Loulloudes, and L. Tassiulas. On the structure and evolution of vehicular networks. In *Modeling, Analysis Simulation of Computer and Telecommunication Systems, 2009. MASCOTS '09. IEEE International Symposium on*, pages 1–10, Sept 2009.

[12] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser. CRAWDAD data set epfl/mobility (v. 2009-02-24). Downloaded from http://crawdad.cs.dartmouth.edu/epfl/mobility, Feb. 2009.

[13] C. U. Press. Cambridge dictionaries online, 2015.

[14] S. Uppoor and M. Fiore. Insights on metropolitan-scale vehicular mobility from a networking perspective. In *Proceedings of the 4th ACM international workshop on Hot topics in planet-scale measurement*, HotPlanet '12, pages 39–44, New York, NY, USA, 2012. ACM.